Journal of Ocean and Coastal Economics

Volume 6 | Issue 1 Article 7

June 2019

Actual Versus Hypothetical WTA Stated Values for Recreational Fishing Licenses: Experimental Evidence of Fish Tales

Scott Steinback
NOAA Fisheries, Northeast Fisheries Science Center

Kristy Wallmo
NOAA Fisheries, Office of Science and Technology

Sabrina Lovell
NOAA Fisheries, Office of Science and Technology

Eric Thunberg
NOAA Fisheries, Northeast Fisheries Science Center

John Foster
NOAA Fisheries, Office of Science and Technology

Recommended Citation

Steinback, Scott; Wallmo, Kristy; Lovell, Sabrina; Thunberg, Eric; and Foster, John (2019) "Actual Versus Hypothetical WTA Stated Values for Recreational Fishing Licenses: Experimental Evidence of Fish Tales," *Journal of Ocean and Coastal Economics*: Vol. 6: Iss. 1, Article 7.
DOI: https://doi.org/10.15351/2373-8456.1111

Actual Versus Hypothetical WTA Stated Values for Recreational Fishing Licenses: Experimental Evidence of Fish Tales

Acknowledgments

The authors would like to thank Fred Serchuk, Jorge' Holzer, and Craig Landry for providing valuable suggestions that improved this manuscript. Sincere thanks are also owed to Rita Curtis and the Massachusetts Division of Marine Fisheries whose unwavering support made this research possible. While NOAA Fisheries supported and funded this project, the views expressed here are those of the authors and do not necessarily reflect the views or policies of NOAA Fisheries.

1. INTRODUCTION

Stated preference values elicited from contingent valuation methods are often viewed with skepticism because these values are derived from hypothetical—rather than actual— transactions (e.g., Hausman 2012). Empirical evidence developed over the past 40 years comparing hypothetical and actual valuations for the same good (public or private) has largely contributed to this skepticism. While exceptions exist, most experimental studies have found systematic disparities between valuations elicited from hypothetical transactions with those obtained in an actual marketplace. This disparity, known as hypothetical bias, has been well documented for both laboratory and field experiments (List and Gallet 2001; Murphy et al. 2005; Penn and Hu 2018) and has spurred an interest in explaining and finding approaches to eliminate the problem.

Another issue in the use of contingent valuation methods is the choice between eliciting estimates of Willingness to Accept (WTA) compensation for foregoing rights to resources, or eliciting Willingness to Pay (WTP) to access resources. A voluminous literature has shown that the choice matters since WTA and WTP have consistently been found to be empirically different. Meta-analyses of this evidence has revealed that WTA values are often much larger than WTP values and that the causes of the disparity are not limited to issues with survey techniques alone (Horowitz and McConnell 2002; Tuncel and Hammitt 2014). A number of alternative explanations for the disparity have been suggested. Two are consistent with standard economic theory: income effects and substitution. Income effects are possible because WTP is strictly limited by budget constraints while WTA is not. If the income elasticity of demand is large enough that income significantly constrains ability to pay, WTA may exceed WTP (Randall and Stoll 1980). The difference between WTA and WTP has also been shown to be related to the absence of substitutes (Hanemann 1991). That is, holding income effects constant, the fewer substitutes available the greater the potential ratio of WTA/WTP because it becomes more difficult to maintain an individual at a constant level of utility through compensation. In the case of zero substitutability it would be impossible to compensate an individual to forgo a resource or give up a good, thereby leading to extreme WTA values. A third explanation offered for the WTA-WTP disparity, outside the framework of conventional economic theory, emphasizes the psychological importance of reference points suggested by Prospect Theory (Kahneman and Tversky 1979). Prospect Theory posits that people define gains and loses based on a reference point, typically the status quo, and losses from this reference point are valued higher than corresponding gains. Thaler (1980) was the first to apply Prospect Theory to the disparity between WTA and WTP and introduced the term Endowment Effect to explain the phenomenon. Individuals asked for their WTA for a good will consider the good part of their endowment while individuals asked for their WTP will not. These conditions can result in WTA compensation values that are higher than WTP values for the same good.

In our study we focus on eliciting WTA values for use in recreational fisheries damage assessments (e.g., to establish levels of compensation for recreational fishermen after a chemical spill) and for benefit-cost analyses of restrictive fisheries policies (e.g., regulations that curtail access). In both cases, angler welfare would be expected to decline so WTA measures are more appropriate because WTA provides a selling price, relevant for valuing a proposed loss, whereas WTP provides a purchase price, relevant for valuing a proposed gain (Bromley 1995; Brown and Gregory 1999; Knetsch 2010; Whittington, Adamowicz, and Lloyd-Smith 2017). Perceived or actual property rights are also theorized to be a necessary criterion for employing a WTA format. If respondents do not perceive any rights over the resources or good being valued then a WTA elicitation format could lead to biased results (Freeman 2003). The respondents in our study purchased a quasi-public good – a saltwater recreational fishing license issued by the state of Massachusetts (MA) for the right to access a public resource.[1] While a fishing license does not infer exclusive property rights to public resources, it clearly assigns de jure access rights to the resources.

1.1 Actual Versus Hypothetical WTA

The first objective of our research was to compare hypothetical WTA values for MA saltwater recreational fishing licenses with WTA values obtained in an actual (simulated) marketplace. An actual market for fishing licenses was simulated by offering anglers real cash in return for giving up their 2012 fishing license. A separate treatment group of anglers was offered hypothetical cash to relinquish their license. Using the actual treatment group as a benchmark for assessing hypothetical bias, parametric and non-parametric evaluations revealed that hypothetical WTA responses overstated values based on actual transactions. These findings generally align with the fairly limited and dated number of WTA studies comparing hypothetical to actual valuations. A literature search revealed only eight studies that elicited both hypothetical and actual WTA values, and all were published prior to 2004 (Table 1). While the majority of the nine WTA observations reported in these studies pointed to an overestimation problem, the evidence was not entirely unanimous, as mean ratios of hypothetical WTA to actual WTA ranged from 0.70 to 28.20. Bishop and Heberlein's (1990) sealed-bid auction experiment of deer hunters has been the only study to report a mean hypothetical/actual ratio less than one (0.70). In contrast, List and Shogren's (2002) random price auction experiment of WTA to surrender holiday gifts yielded a mean ratio of hypothetical to actual WTA of 1.05, but this difference was statistically indistinguishable. Smith and Mansfield (1998) came to the same conclusion in their dichotomous choice experiment which revealed no significant differences between real and hypothetical WTA statements for the opportunity to spend time in a second set of interviews on an undisclosed topic. The mean

[1] Quasi-public goods are those provided by the government, typically for a nominal fee that is not determined by market forces (see Chambers, Chambers, and Whitehead 1998, Carson and Groves 2007). Common examples include recreational fishing and hunting licenses, public parks, historic buildings, and campgrounds located near public lakes. Quasi-public goods have both private and public attributes because it is possible to exclude members of the public by charging a price to access or use the good.

https://cbe.miis.edu/joce/vol6/iss1/7
DOI: 10.15351/2373-8456.1111

hypothetical/actual ratios for the remaining six WTA observations though were all significantly greater than one. These studies evaluated real and hypothetical WTA statements for goose hunting permits (Bishop and Heberlein 1979), to experience a drop of the bitter-tasting sucrose octa-acetate (Coursey, Hovis, and Schulze 1987), alternative densities of trees in a new, public park (Brookshire and Coursey 1987), deer hunting permits (Bishop and Heberlein 1990), and holiday gift giving (List and Shogren 1998). Nape et al. 2003 also compared actual and hypothetical WTA responses for a private good (a Far Side cartoon wall calendar). While the authors do not report mean WTA valuations, their evidence aligns with ours and supports the existence of overstated WTA hypothetical bias.

Table 1. **WTA Studies**

Study	Year	Type of good	Ratio of mean hypothetical WTA to mean actual WTA
Bishop and Heberlein	1979	Quasi-public	1.60
Coursey, Hovis, and Schulze	1987	Private	2.00
Brookshire and Coursey[a]	1987	Public	25.79
			28.20
Bishop and Heberlein	1990	Quasi-public	0.70
			2.74
Smith and Mansfield	1998	Private	1.07
List and Shogren	1998	Private	1.42
List and Shogren	2002	Private	1.05
Nape et al.	2003	Private	N/A

[a] The authors report more than one value.

1.2 Accounting for Uncertainty to Mitigate Hypothetical Bias

Despite WTA often being the more appropriate measure to assess welfare losses, hypothetical WTA studies have comparatively seldom been performed because of a longstanding view that they result in a greater degree of bias than equivalent hypothetical WTP formats (Mitchell and Carson 1989; Arrow et al. 1993; OMB 2003). The reality though is that studies using the WTP format have generally suffered from a substantial overestimation bias as well. Meta-analyses show that the statistical evidence does not support the argument that WTP studies are less prone to hypothetical bias than WTA studies (e.g., Little, Broadbent, and Berrens 2012; Penn and Hu 2018).

In the more recent literature, the reluctance to pose hypothetical WTA questions to respondents when assessing welfare losses is diminishing as researchers have increasingly recognized that substituting a WTP value for a desired WTA value will tend to undervalue environmental assets (e.g., Bush et al. 2013; Kaczan, Swallow, and Adamowicz 2013). To the extent that WTA measures of value will provide a more accurate reflection of welfare losses in many cases, an issue to be resolved is how to correct for overstated WTA hypothetical bias. Towards this end, the second objective of our study was to evaluate the effectiveness of applying an ex-post certainty adjustment technique for removing overstated hypothetical WTA bias. The certainty adjustment approach is based on a follow-up certainty question with two degrees of certainty and has been applied in a

number of dichotomous choice WTP settings. We apply the approach for the first time to hypothetical WTA responses.

The certainty adjustment technique was originally developed based on the hypothesis that overstated hypothetical WTP bias might arise, in part, from respondents' uncertainty about whether they would actually pay under a real rather than a dichotomous choice hypothetical WTP setting (Blumenschein et al. 1998). Proponents of this approach suggest that overstated hypothetical WTP bias can be reduced or eliminated by recoding a *yes* response to *no* if the respondent expresses a high degree of uncertainty in a follow-up certainty question. Applying this practical approach to our study was particularly appealing because respondents were asked about a good that they had likely never thought of in monetary terms.

Two ex-post certainty adjustment techniques have primarily been used in the WTP literature and each has its own advantages. The first collects information about the degree of certainty of the hypothetical *yes* responses on a 10-point rating scale, usually between 1 and 10, with 1 labelled "very uncertain" or "very unsure" and 10 labelled "very certain" or "very sure." The advantage of the 10-point certainty scale is that it provides a flexible approach for respondents that state they would pay the hypothetical amount to indicate how sure they are that they would actually pay. *Yes* respondents whose level of certainty falls below some specified threshold are then recoded to *no*. It is presumed that respondents who indicate *yes* but are fairly uncertain would likely reject the WTP question if it was real. Studies using this approach have generally found that the recoded hypothetical responses correspond to actual payments. The disadvantage of the approach is determining the appropriate certainty cutoff point for recoding a *yes* response to *no*. Champ et al. (1997) who evaluated real and hypothetical donations for road removal in Grand Canyon National Park, and Blumenschein et al. (2001) who compared real and hypothetical purchase decisions for a pharmacist provided asthma management program, found that it was necessary to recode all certainty values below 10 to *no* before mean hypothetical WTP was equivalent to actual WTP. In contrast, in a study of donations for wind power in Wisconsin (Champ and Bishop 2001) and in a public goods study by Norwood (2005), a cut-off of 8 was required to equalize hypothetical and actual WTP. While Ethier et al. (2000) and Poe et al. (2002) who both evaluated stated preference responses for green-pricing electricity programs, and Morrison and Brown (2009) in a study of donations to a Red Cross children's breakfast fund, found that a cut-off of 7 produced hypothetical and actual WTP values that were statistically similar. Champ, Moore, and Bishop (2009) suggest that the (sample) mean level of certainty may be an appropriate cutoff point, but it is apparent that more progress in this area of research is needed.

The second ex-post adjustment approach is more appealing because it avoids the need for determining the appropriate certainty cutoff point by committing to two distinct categories. In a series of field experiments by Blumenschein and colleagues (1998, 2001, 2008) for a private good (e.g., sunglasses) and two health-related quasi-public goods (the opportunity to participate in asthma and diabetes management programs), respondents were first asked a dichotomous choice WTP contingent valuation question and then asked a follow-up certainty question divided into *probably sure* and *definitely sure* responses. Only *yes* responses from those that replied *definitely*

sure were treated as *yes* responses. *Probably sure yes* responses were treated the same as *no* responses. When recoded responses were compared to real purchases obtained in separate treatment groups, they found that the recoding approach removed the inflated WTP hypothetical bias.

The bottom line is that these two certainty adjustment approaches have consistently been shown to mitigate or eliminate overstated hypothetical WTP bias. However, while the WTP literature on using respondent uncertainty to mitigate hypothetical bias continues to grow, the literature has yet to extend either technique to WTA responses. In this study we modify the narrative approach developed by Blumenschein and colleagues to address overstated hypothetical WTA responses. A modification of the recoding technique was necessary because while response uncertainty leads to inflated *yes* responses in WTP settings, response uncertainty leads to inflated *no* responses in WTA settings. That is, respondents in our study who refused the hypothetical offer to relinquish their fishing license (i.e., a *no* response) but were uncertain would likely have accepted if given a real—rather than a hypothetical—opportunity to respond. Therefore, in contrast to previous studies that treated *probably sure yes* responses the same as *no* responses in an attempt to remove overstated hypothetical WTP bias, we treated *probably sure no* responses the same as *yes* responses in an effort to eliminate overstated hypothetical WTA bias. While the modified recoding approach resulted in more consistent estimates of hypothetical and actual WTA in our study, the evidence suggests it was not effective in completely eliminating the overstated hypothetical bias.

The remainder of this paper is organized as follows. Section 2 explains the experimental design. Section 3 describes the sample data and Section 4 contains the empirical analysis of the sample data. The results are then discussed in Section 5 and the paper ends with some concluding comments in Section 6.

2. EXPERIMENTAL DESIGN

Individuals who are 16 years and older and wish to recreationally fish in MA marine waters are required by the MA Division of Marine Fisheries (MA DMF) to obtain a recreational fishing license (except for disabled persons and anglers fishing on for-hire charter or party boats). The license costs $10 (free to those aged over 60), is not transferrable, and is valid for the entire calendar year in which it is purchased. Fishing licenses can be purchased on-line or from select retail stores in MA at any time during the year beginning in December of the previous year. In 2012, more than 154,000 fishing licenses were issued to MA recreational saltwater anglers.

In our 2012 survey, two separate sample groups of licensed anglers were drawn at random. The first comprised 700 anglers who received hypothetical cash offers in return for relinquishing their 2012 MA fishing license (herein referred to as the “HWTA sample group”). For the second sample group, we simulated an actual market for fishing licenses by offering 500 anglers cash in

return for relinquishing their 2012 MA fishing license (AWTA sample group). Anglers in both the HWTA and AWTA sample groups were informed that acceptance of the offer would mean giving up their license and thus their right to fish recreationally in MA marine waters for the remainder of 2012.

Potential survey participants were drawn randomly each month from February through May 2012. In each month, a random draw was conducted of 300 anglers who had obtained a license during the previous month and the individuals selected were subsequently assigned to one of the two sampling groups: 175 were assigned to the HWTA group, and 125 to the AWTA group. A total of 1,200 questionnaires (300 individuals per month x 4 months) were mailed to MA saltwater anglers who had obtained a license during the first four months of 2012. The MA saltwater recreational fishing season generally begins during late May-June each year, so to elicit consistent valuation information across licensed anglers it was important to obtain responses before the fishing season began.

A modified Dillman Tailored Design (Dillman, 2007) was used to administer the survey. Four mailings were conducted (copies of all mailing forms associated with this article can be obtained from the first author). The first mailing consisted of a letter notifying each selected individual that they had been chosen for the survey, explained the importance of their participation, and that the purpose of the study was to measure the economic value of recreational fishing in MA marine waters to license holders like themselves. The second mailing was sent out 3-5 business days later and contained the survey questionnaire along with a stamped return envelope. For the AWTA sample group, this mailing also contained a personalized check that varied in value from $15 to $500. The cover letter for this mailing reiterated that the purpose of the study was to determine the monetary value anglers like themselves place on being able to go saltwater recreational fishing in MA waters. The third mailing was a postcard that thanked anglers who had already responded, and served as a reminder to those who had not yet responded to complete and return their questionnaires. The postcards were mailed five business days after the second mailing. The fourth mailing was sent to anglers who had not responded within three weeks of the third mailing. This final mailing included a replacement questionnaire and explained the importance of responding to the survey as the data collection aspect of the survey would soon be terminating. As well as the valuation questions, the surveys inquired about the number of days fished in 2011, the expected number of days fished in 2012, boat ownership, and demographics.

The first sample group (HWTA) received hypothetical cash offers in return for relinquishing their 2012 fishing license. These offers ranged randomly in log-linear amounts from $15 to $500, and the dichotomous choice question was worded as follows:

> "Imagine that a check payable to you for $[ ] was included along with this survey as a cash offer for your 2012 Massachusetts Recreational Saltwater Fishing Permit.[2] Knowing that the fishing permit is required for anyone that fishes recreationally in Massachusetts marine waters during 2012, would you be willing to send us your permit, and in doing so, <u>give up your right to fish recreationally in Massachusetts marine waters for the rest of this year in return for $[ ]</u>?"

[2] In this study the words "permit" and "license" are interchangeable.

Short cheap-talk script was provided before the respondent indicated *yes* or *no* that encouraged anglers to try and answer the hypothetical question as if it were an actual cash offer.

> "Although the question is hypothetical, please consider the offer carefully and try to answer the question as if this was an actual cash offer to you for your 2012 Massachusetts Recreational Saltwater Fishing Permit, and then check the indicated box below. [ ] Yes [ ] No"

Following their *yes* or *no* response, a certainty question asked the respondents to indicate if they were *probably sure* or *definitely sure* that they would [or would not, if answered no] give up their right to saltwater fish in MA waters during 2012 in return for $[ ]?

Individuals in the AWTA sample received actual cash offers (in the form of a personalized check), in amounts varying between $15 and $500 to relinquish their 2012 license. The offer letter described the importance of the study, explained how the angler was selected for participation, and provided the following instructions:

> "The enclosed check in the amount of $____ is yours to cash if you fill out and return the enclosed one-page questionnaire and your 2012 Massachusetts saltwater fishing permit to us in the return envelope provided. If you accept our offer by cashing your check and mailing us the questionnaire and your 2012 fishing permit, your permit will be invalidated by the Massachusetts Division of Marine Fisheries. You will then be unable to obtain another one for the remainder of 2012. This means that you will be giving up your right to fish recreationally in Massachusetts marine waters for the rest of this year in return for $[ ]."

> "If you decide to decline our cash offer and keep your 2012 Massachusetts saltwater fishing permit (so that you can fish in Massachusetts marine waters during the rest of 2012), we ask that you please complete and return the one-page questionnaire along with the check. Please consider this offer carefully. Your check will expire on _____ so we look forward to hearing from you soon."

The personalized checks expired six weeks from the mail date and, given the simulated marketplace developed for this sample group no cheap-talk or certainty questions were included on the questionnaire.

3. SAMPLE DATA

3.1 Response Rates by Treatment Group

A total of 465 questionnaires were returned from the HWTA group (66% of 700 potential respondents). Of these, 11 were excluded from the analysis; nine were removed because the HWTA question was not answered and two were excluded due to a possible publicity bias. Shortly after the first monthly wave of questionnaires was mailed, a story about the study was published in the Boston Globe newspaper on 2 March 2012. Within a week, six other newspapers in MA printed their own stories about the study. This created the possibility that readers of the stories, who had no intention of going saltwater recreational fishing in 2012, would purchase a $10 license

hoping to be mailed one of the future surveys that contained a $500 check (i.e., pay $10 for the chance to win $500). To minimize this possible publicity bias, we excluded two questionnaires received after March 2 from respondents who indicated they did not fish in 2011 and did not expect to fish in 2012. We felt that these two respondents were the ones most likely influenced by the newspaper stories.[3] Table 2 shows the response rates by treatment group and offer amount, after removing invalid responses.

Table 2. **Response Rates by Offer Amount and Treatment**

	Useable/Mailed Surveys (Response Rate)	
Offer Amount ($'s)	**HWTA**	**AWTA**
15	57/89 (.64)	39/64 (.61)
25	55/80 (.69)	30/59 (.51)
40	45/69 (.65)	26/52 (.50)
55	36/56 (.64)	22/40 (.55)
75	34/49 (.69)	21/34 (.62)
100	34/49 (.69)	18/34 (.53)
125	28/49 (.57)	24/33 (.73)
160	29/42 (.69)	14/30 (.47)
200	30/42 (.71)	15/29 (.52)
250	20/35 (.57)	13/24 (.54)
300	21/28 (.75)	13/21 (.62)
350	15/28 (.54)	13/20 (.65)
400	19/28 (.68)	12/20 (.60)
450	12/28 (.43)	11/20 (.55)
500	19/28 (.68)	13/20 (.65)
All	454/700 (.65)	284/500 (.57)

Of the 301 license holders who completed the AWTA questionnaire (60% of 500 potential respondents), 284 either accepted or rejected the offer by returning their check or their license. Seventeen of the 301 respondents were excluded from subsequent analysis; 12 were removed because they kept their license and cashed their check, one was eliminated because the offer was accepted on the questionnaire and the license returned, but the check was never cashed, and four others were excluded because of the possible publicity bias.

The success of the simulated AWTA market depended on the effectiveness of excluding anglers from access if they accepted the real offer. All AWTA anglers were clearly informed that they would be giving up their right to fish in MA waters if their offer checks were cashed. As only 12 AWTA anglers (4% of the 301 AWTA respondents) did not play by the rules, the language in the questionnaire seemed sufficient in overcoming this potential obstacle. Nonetheless, we do not know if any of the anglers who accepted their AWTA offers decided to fish during 2012 without a valid permit. Based on comments received from some survey participants, it was obvious that

[3] If our data cleaning approach was not sufficient at removing publicity-influenced responses, our estimates of WTA could be underestimated. As all of the 15 offer amounts exceeded the cost of the fishing license, any publicity-influenced angler that participated in our study would likely have accepted their real or hypothetical offer. Inclusion of these responses would have raised the acceptance rates leading to underestimates of WTA. On the other hand, the publicity generated from the newspaper stories may have actually had a positive effect on response rates by alleviating erroneous beliefs about the purpose of the study and concerns that the checks were illegitimate.

the monetary offers were considered very seriously. Hence, we believe that accepting the cash offer and then fishing without a permit was minimal or non-existent. In addition, the MA DMF made numerous announcements indicating increased enforcement of the permit requirement during the 2012 fishing season (Paul Diodati, Director MA DMF, in discussion with first author, March 2012).

All 199 of the AWTA non-respondents (i.e., those who did not return the questionnaire) were also eliminated from the analysis. Survey recipients in our study may have thrown the envelope away without opening it, may not have received the questionnaire in time to respond for some reason (e.g., mailed to a vacation house address), doubted the validity of the study and ignored it, or simply never took the time to fill out the questionnaire. If our budget would have been larger, we would have attempted to contact non-respondents to determine the reason for their non-response. We considered treating AWTA non-responders the same as a *no* response (i.e., refused their offer), by making the assumption that all non-responders actually received and read the questionnaire, but chose not to return it because the cash offer was deemed too low. If this were true we would expect the response rates to be related to the offer amounts and increase as the offer amounts increased. A comparison of the AWTA response rates across the 15 offer levels shows no discernable increase (Table 2). We tested this empirically in a logistic regression by regressing response (modeled as 1 if responded to the AWTA questionnaire and 0 if did not respond) on the 15 offer amounts. The regression results indicated that the likelihood of response was not significantly correlated to the offer amount ($\chi^2 = 1.79$; $p = 0.1286$). Therefore, since response was not related in a statistically significant manner to the offer amount, there was no strong basis for treating the AWTA non-responders as if they legitimately refused their offer.

3.2 Respondent Characteristics by Treatment Group

Table 3 summarizes the personal characteristics of the survey respondents. The variables presented in this table are those that are included in the regression analysis that will be reported shortly. No statistically significant differences ($p<0.05$) were detected between the sample groups using two-sided t-tests for continuous variables and the Rao-Scott design corrected chi-square test for categorical variables[4], with the exception of household income which was higher in the AWTA group ($\chi^2 = 17.12$; $p = 0.0166$). We control for differences in personal characteristics in the logistic regression models we subsequently develop to examine whether the probability of acceptance differs between the HWTA and AWTA groups.

Table 3. **Personal Characteristics by Treatment Group**

	HWTA	AWTA
Own a boat (%)	35.37	31.05
Expected number of days will fish in 2012	29.26	26.26
Age (years)	55.12	55.67

[4] The Rao-Scott chi-square test is computed from the Pearson chi-square statistic and a design correction is applied to account for the stratified random sampling design of our survey data. See Rao and Scott (1981) for details about the design-adjusted chi-square test.

Annual household income composition ($1,000)		
< 15 (%)	6.32	1.95
15 - 34,999 (%)	12.40	16.24
35 - 49,999 (%)	15.54	13.08
50 - 74,999 (%)	25.15	16.21
75 - 99,999 (%)	13.09	12.40
100 - 149,999 (%)	18.43	24.73
150 – 199,999 (%)	6.18	9.16
> 200,000 (%)	2.90	6.22
Number of people in household	2.70	2.81
Employed fulltime (%)	48.40	53.32
Education composition		
Some high school (%)	6.16	10.09
High school (%)	29.19	22.54
Some college (%)	17.50	16.01
Two-year college degree of technical school (%)	10.32	12.32
Four-year college degree (%)	17.58	19.03
Some graduate work but not a graduate degree (%)	5.82	6.03
Graduate degree	13.44	13.99
Married (%)	75.14	74.12

4. EMPIRICAL ANALYSIS AND RESULTS

Table 4 shows the weighted acceptance rates by dollar amount for the two sample groups.[5] As expected, the proportion accepting the hypothetical and actual cash offers is generally higher at the higher dollar amounts. The acceptance rates in both groups are thus consistent with downward sloping demand curves. However, a comparison of the acceptance rate distributions reveals that the AWTA acceptance rates are higher at all offer amounts above $55. These differences are statistically significant at the 5% level for the $300, $350, and $450 offer amounts and at the 10% level for the $75 offer amount. Overall, the acceptance rate for the AWTA group is almost twice that for the HWTA group (39% versus 23%), and the difference is statistically significant ($p<0.05$). This result suggests that anglers are more likely to relinquish their permits for actual rather than hypothetical dollars.

Table 4. **Weighted Acceptance Rates by Offer Amount**

			Probably Sure No Responses Recoded as Yes Responses
	AWTA	**HWTA**	**HWTA**
Offer Amount ($'s)	**Proportion Accepting**		**Proportion Accepting**
15	.08	.01[b]	.01[b]
25	0	.10	.12

[5] Estimation of population parameters from samples drawn from a stratified random design must be weighted according to strata characteristics and related survey procedures (Cochran, 1977). As the sampling rates in our study varied by month, treatment group, and offer levels, weights were calculated that reflect both the sample and experimental designs of the study. The full weighting procedures are detailed in the Appendix.

DOI: 10.15351/2373-8456.1111

40	.08	.10	.16
55	.13	.26	.30
75	.32	.10[b]	.13
100	.30	.16	.26
125	.41	.25	.38
160	.41	.18	.24
200	.55	.31	.39
250	.34	.21	.28
300	.63	.24[a]	.36
350	.83	.21[a]	.37[a]
400	.52	.39	.45
450	.67	.38[a]	.48
500	.67	.62	.65
All	.39	.23[a]	.32[b]

[a] Probability value of the difference compared to AWTA was rejected at the 5% level using the Rao-Scott design corrected chi-square test.

[b] Probability value of the difference compared to AWTA was rejected at the 10% level using the Rao-Scott design corrected chi-square test.

Applying the modified certainty adjustment technique outlined earlier, we converted 35 *probably sure* HWTA *no* responses to *yes* responses. This reduced the differences in acceptance rates between the HWTA and AWTA groups (Table 4), but the AWTA acceptance rates remained higher than the HWTA rates at all offer amounts above $55 and the overall acceptance rates were still statistically different when the criterion was the 10% level of significance.[6]

While the weighted acceptance rates by dollar amount are informative, they do not take account of potential differences in personal characteristics among the two groups (shown in Table 3). We control for these differences by combining the responses of both sample groups in pooled logistic regression models (using household income and the other personal characteristics as covariates in the models), and constructing a hypothetical dummy variable that has the value of 1 if the angler is in the HWTA group and 0 if the angler is in the AWTA group.

The basic relationship specified in a dichotomous choice logit model is:

$$Prob(Yes) = 1 - \{1 + \exp[\beta_0 - \beta_1(\$Amount) + \beta_2(Z_1) + \ldots \beta_n(Z_n)]\}^{-1} \quad [1]$$

where the probability of a *yes* response (i.e., acceptance) is related to the dichotomous choice cash offer amount ($Amount) and also to the personal characteristic variables (Z) assumed to affect

[6] Per a reviewer's recommendation, we also calculated Turnbull-based lower bound estimates of WTA from the weighted acceptance rates shown in Table 4. Refer to Haab and McConnell (2002) for calculating Turnbull lower bound non-parametric point estimates in a multi-price setting. This approach yielded lower bound estimates of WTA for the unadjusted HWTA sample ($213.68) and the recoded HWTA sample ($188.35) that were both higher and significantly different from the AWTA sample ($138.34) at the 5% level. These findings align with the acceptance rate comparisons and indicate that the hypothetical recoding approach mitigated, but did not completely eliminate the hypothetical bias.

utility (Hanemann 1984). From Eq. [1] Hanemann (1989) provides a formula for a non-negative mean WTA:

$$Mean\ WTA = (1/\beta_1) * (\ln\left(1 + \exp\left(-\left(\beta_0 + \sum(\beta_n(Z_n))\right)\right)\right)) \quad [2]$$

where β_n is the vector of coefficients and Z_n are the sample means of the associated independent variables. Confidence intervals around mean WTA were then calculated by the Krinsky-Robb method using 50 thousand draws (Krinsky and Robb, 1986).

4.1 Regression Models

The results of the WTA logistic regression models, estimated using maximum likelihood techniques, are shown in Table 5. Two WTA logistic regression equations were developed. The first equation considered all *no* responses to be *no* responses in the HWTA group. The second equation recoded all *probably sure no* responses to *yes* responses in the HWTA group.

The number of observations used in the regression models are lower than given in Table 2, due primarily to missing household income values. Imputation procedures were considered to replace the missing income data, but we decided that this was unnecessary because similar regressions that excluded the household income variable gave virtually identical values for the other coefficients. In addition, statistical comparisons of the remaining covariates in the models for those that answered the household income question and those that did not showed no differences at the 5% level. We also tested a logarithmic functional form that yielded similar results to our logistic equations, as did specification of a probit model.

Table 5. **WTA Logistic Regression Analysis**

	Pooled AWTA and HWTA			**Pooled AWTA and HWTA with HWTA Probably Sure No Responses Recoded as Yes Responses**		
	Coefficient	**SE**	**Marginal Effect**	**Coefficient**	**SE**	**Marginal Effect**
Intercept	2.221*	0.947		1.826	0.870	
Hypothetical dummy[a]	-1.157*	0.277	-0.206	-0.517*	0.247	-0.109
$Amount[b]	0.006*	0.001	0.001	0.005*	0.001	0.001
Boat ownership[c]	-1.134*	0.305	-0.202	-0.729*	0.265	-0.154
Expected days will fish in 2012[d]	-0.032*	0.007	-0.006	-0.030*	0.007	-0.006
Annual household gross income[e]	-0.225*	0.099	-0.040	-0.193*	0.091	-0.041
Number of people in household[f]	-0.133	0.115	-0.024	-0.069	0.103	-0.015
Employed fulltime[g]	-0.698*	0.286	-0.124	-0.622*	0.276	-0.131
Education Level[h]	0.022	0.073	0.004	0.021	0.068	0.004
Age[i]	-0.022	0.012	-0.004	-0.016	0.011	-0.003
Married[j]	0.113	0.376	0.020	-0.224	0.341	-0.047
Number of obs.	626			626		
Wald chi-square (p-value)	73.421	(0.000)		79.576	(0.000)	
Log-likelihood	-266.865			-305.161		
McFadden pseudo R^2	0.268			0.217		
% correct prediction	80.4			77.4		
WTA values ($'s)						
Mean AWTA (95% confidence limits)	338	(279 – 396)				
Mean HWTA (95% confidence limits)	520	(437 – 604)		436	(362 – 510)	

* Significance at the 5% level or lower.

[a] Hypothetical dummy (categorical variable): 1, if HWTA group, 0, if AWTA group.

[b] $Amount (categorical variable): 15 cash offer levels between $15 and $500.

[c] Boat ownership (categorical variable): 1, if owned a boat, 0, if did not own a boat.

[d] Expected days will fish in 2012 denotes anticipated fishing days during the year.

[e] Annual household gross income (categorical variable): 1=less than $15,000; 2=$15,000-$34,999; 3=$35,000-$49,999; 4=$50,000-$74,999; 5=$75,000-$99,999; 6=$100,000-$149,999; 7=$150,000-$199,000; 8=$200,000 or over.

[f] Number of people in household denotes number of people living in the household.

[g] Employed fulltime (categorical variable): 1, if employed fulltime, 0, if not employed fulltime.

[h] Education level (categorical variable): 1=some high school; 2=high school; 3=some college; 4=two-year college degree; 5=four-year college degree; 6=some graduate work but not a graduate degree; 7=graduate degree.

[i] Age denotes actual age in years.

[j] Married (categorical variable): 1, if married, 0, if not married.

In the first regression model, the hypothetical group dummy variable is negative, statistically significant ($p<0.05$), and has a marginal effect of 20.6 percentage points. This means that the probability of accepting the amount offered in the hypothetical treatment is on average 20.6 percentage points lower than in the actual treatment. If we consider the responses from the AWTA group as a benchmark for assessing hypothetical bias, these results align with the non-parametric comparisons of acceptance rates discussed earlier (Table 4) and imply a strong downward bias in the probability of acceptance for the HWTA group. In the second regression model, where all *probably sure no* responses were considered to be *yes* responses, the hypothetical group dummy coefficient is closer to zero but still statistically significant. Additionally, the marginal effect is reduced but is still almost 11 percentage points. Hence, adjusting the HWTA responses based on the follow-up certainty question reduced the hypothetical bias, but did not eliminate it.

We can also see that the coefficients on the offer amount ($Amount) and on several of the personal characteristic variables (i.e., boat ownership, expected days will fish in 2012, annual household gross income, and employed fulltime) are statistically significant in both regression equations. The signs for the offer and background variables are all consistent with prior expectations. Higher offer amounts are positively related to the probability of a *yes* response, boat ownership lessens the probability that an angler will accept an offer and sell his/her permit, the more days an angler expected to fish in 2012 reduces the probability of acceptance, and anglers with higher annual household gross incomes or are employed fulltime are less likely to accept an offer. None of the remaining personal characteristic control variables included in the models are statistically significant.

In our final (and perhaps most important) comparison of the responses from the AWTA and HWTA groups, we compared the parametric estimates of mean WTA values. As indicated in Table 3, the *actual* mean value of a 2012 MA saltwater recreational fishing permit is $338 and the *hypothetical* mean value is $520. The difference of $182 is statistically different from zero at the 5% level because the confidence intervals do not overlap. When all *probably sure no* responses are recoded as *yes* responses in the HWTA group, the *hypothetical* mean value declines to $436. Although this value is closer to the *actual* mean value and the confidence intervals overlap in the tails, the 95% confidence interval for the difference between the two group means is 3.67 – 192.61. Because this interval does not contain zero, we rejected the null hypothesis that the group means are the same (Poe, Severance-Lossin, and Welsh 1994). Consistent with the preceding evaluations that found statistically significant differences between the *hypothetical* and *actual* responses, these findings indicate a strong upward bias in the *hypothetical* mean WTA value, even after adjusting the hypothetical responses based on a follow-up certainty question.[7]

[7] As further evidence, we compared the combined AWTA and HWTA models with models estimated separately for each of the two groups. Likelihood ratios and mean WTA values were compared and gave results consistently similar to those from the models in Table 3. Given these similarities, we decided to use the combined estimates.

5. DISCUSSION

Our study evaluated whether differences exist between hypothetical and actual dichotomous WTA responses for a quasi-public good – a MA saltwater recreational fishing license. We compared stated preference dichotomous choice WTA responses with those obtained in an actual market. Our parametric and non-parametric evaluations align with the notably limited previous evidence that suggested hypothetical measures of WTA generally overstate values based on actual transactions. In addition, criterion validity tests revealed that the mean hypothetical WTA value for a 2012 MA saltwater recreational fishing license ($520) was 1.54 times higher than the value derived from actual transactions ($338), and this difference was statistically significant. The mean of the nine hypothetical/actual WTA ratios shown in Table 1 is 7.17. However, two of the WTA observations in Table 1 were from one study (Brookshire and Coursey 1987) and are over nine times higher than the next highest reported observation. If those observations are removed, the mean of the remaining ratios is 1.51 – nearly identical to our result.

We also evaluated the effectiveness of applying a very practical ex-post certainty adjustment technique to remove the overstated WTA hypothetical bias. While prior evidence on the use of the adjustment technique in WTP environments has shown that it can be effective at eliminating the difference between actual and hypothetical mean WTP, the approach had never been extended to WTA responses. We find that when applied in our WTA setting, the approach mitigated but did not eliminate the overstated hypothetical bias. The hypothetical mean WTA value declined to $436 after applying the adjustment approach, but this value was still about 1.3 times higher than the mean value based on actual responses and the difference was statistically significant.

Further applications of the ex-post certainty adjustment technique we employed are needed before larger conclusions can be drawn regarding the link between uncertainty and WTA hypothetical bias. A potential issue with our split-sample comparisons is that they do not establish a definitive link between the degree of certainty in hypothetical responses and actual responses. That is, by treating *probably sure no* responses the same as *yes* responses we increased the incidence of *yes* answers. In turn, this had the obvious effect of mitigating the hypothetical bias and by coincidence the proportion of adjusted *yes* responses could have risen high enough that they equaled the proportion of actual *yes* responses. Our results show otherwise, but this issue could be further evaluated through within-sample comparisons of the same individuals who are first asked a hypothetical WTA valuation question followed by a certainty question, and then confronted with an actual sell decision. Experiments of this kind have been conducted in WTP settings where hypothetical WTP responses were found to overstate values based on actual purchases (Blumenschein et al. 1998; Johannesson, Liljas, and Johansson 1998). The degree of certainty in hypothetical WTP responses was found to be a strong predictor of actual purchasing behavior. Within-sample comparisons of both hypothetical and actual WTA certainty responses would be valuable for assessing the validity of the certainty approach used in our study.

While we cannot definitely explain why the certainty adjustment approach failed to remove the WTA hypothetical bias in our study, we theorize that it was related to: (a) a considerably high

degree of unfamiliarity associated with valuing recreational fishing licenses; (b) protest behavior; and (c) the non-incentive compatible survey design.

Given that the nominal fee ($10) required to purchase an annual MA saltwater fishing permit is determined by the MA state legislature and not through market forces, most anglers probably had never realistically thought about the value of their fishing permit. Unlike normal private goods that people may repeatedly buy and sell in a market, respondents to our survey could not adequately draw upon previous experiences to guide their acceptance/rejection decision. Saltwater recreational fishing is an inherently unique activity with no perfect substitutes so it is unlikely that many of the respondents had well-defined monetary preferences for their fishing permit. We believe this unfamiliarity was at least partially responsible for the disparities between hypothetical and actual responses. Although the AWTA treatment group was subject to the same unfamiliarity, respondents in the HWTA group had less incentive to really reflect on the matter as no actual money was involved. This hypothesis dates back to at least Freeman (1979) and Bishop, Heberlein, and Kealy (1983). For the HWTA group, we believe that the lack of familiarity combined with the hypothetical nature of the survey caused some anglers to seek conservatively high requirements for compensation. That is, some anglers may have recognized that they would be willing to accept even lower amounts, but because they did not thoroughly explore their preferences, due to the artificiality of the survey, they rejected the offer to be on the safe side.[8]

Our certainty adjustment results lend credence to this hypothesis. The magnitude of the hypothetical bias declined after treating HWTA *probably sure no* responses the same as *yes* responses. Nevertheless, while the certainty adjustment technique reduced the hypothetical bias, it did not eliminate it. We suspect that even some of the HWTA *definitely sure no* respondents would also have accepted compensation if given a real—rather than a hypothetical—opportunity to respond. If so, this may explain, at least in part, why our certainty adjustment approach failed to eliminate the bias.

Another possible cause, first suggested by Hammack and Brown (1974), is that when participants are offered hypothetical compensation for goods which evoke strong moral sentiments or have high intrinsic value, some participants may respond with indignation by rejecting the offer without considering their actual monetary preferences or seek absurdly high compensation requirements.[9] Saltwater recreational fishing in Massachusetts is considered an inalienable right by some anglers, as proclaimed in the newspaper stories written about the study, so this type of "protest" behavior cannot be ruled out. Protest responses may have occurred in both the HWTA and AWTA markets because the only difference between the two markets was the exchange of hypothetical versus real dollars. The AWTA market involved actual money though, and a real market provides more powerful incentives to explore monetary preferences than a hypothetical market. Therefore, we believe that fewer AWTA participants ultimately interpreted their response

[8] This unfamiliarity hypothesis did not translate into longer questionnaire return rates for the hypothetical groups though. There were no statistical differences in the questionnaire return times among the two sample groups.

[9] In addition to Hammack and Brown (1974), further elaboration of this moral statement hypothesis is provided in Kahneman and Knetsch (1992), Kahneman, Ritov, and Schkade (1999), Svedsäter (2003), and Bischoff (2008).

as a moral statement.[10] If so, this would also have contributed to the failure of the certainty adjustment approach to remove the overstated hypothetical bias in our study.

The last possible reason we propose is associated with the non-binding format of the hypothetical dichotomous choice mechanism. In the more recent literature, *consequentiality* and *incentive compatibility* are increasingly being recognized as perhaps the most essential conditions for truthful reporting in stated preference surveys (Carson and Groves 2007; Kling, Phaneuf, and Zhao 2012; Kim, Kling, and Zhao 2015). Carson and Groves (2007) argue that the necessary conditions for truthful reporting require (1) that the respondent believes that the study results may influence something that the respondent cares about (i.e., consequentiality), and (2) having an incentive compatible elicitation mechanism that discourages strategic responses. To our knowledge these conditions have never been extended to a WTA environment, but there is no reason to believe that the same theoretical protocols are not applicable.

We speculate that our dichotomous choice WTA design was perceived as consequential mainly because recipients were notified that they were selected from a small sample of MA saltwater license holders for participation in an important government sponsored study commissioned to measure the monetary value that anglers place on being able to go saltwater recreational fishing in MA – an activity all license holders must care about to varying degrees. However, the HWTA elicitation mechanism did not have the same incentive compatibility properties as the AWTA mechanism because anglers were able to keep their license even if they accepted the hypothetical offers (whereas AWTA anglers were required to relinquish their license to accept an actual cash offer). In an attempt to provide equivalent incentives and discourage strategic behavior by the hypothetical recipients, cheap talk script was added that asked anglers to consider the hypothetical question carefully and to respond as if an actual cash offer had been tendered. While the evidence indicates that cheap talk varies in its effectiveness (Morrison and Brown 2009), the addition of this script in our study may not have been sufficient to discourage some HWTA recipients from responding strategically.

The obvious implication of the non-binding hypothetical format was that it provided strategic incentives for rejecting hypothetical offers that would otherwise have been acceptable if the compensation was real. This is the optimal strategy for anglers that inferred a response of this type would inflate the "true" value of their fishing licenses in our study and ultimately raise the study's estimate of the economic value of saltwater recreational fishing in MA. Anglers responding in this manner would naturally have indicated that they were *definitely sure* that they would not give up their right to saltwater fish in MA waters during 2012 in return for the hypothetical offer. As *definitely sure* responses were taken at "face-value," this type of strategic misrepresentation may also have led to the failure of the certainty adjustment approach to remove the overstated hypothetical bias in our study. In hindsight, we may have been able to better assess the extent of potential strategic responses by including a follow-up question that asked respondents to state their

[10] In all likelihood, the majority of AWTA "protest" responses were excluded from the study. The logical reaction for the AWTA anglers affected by this "moral statement" issue would be to simply refuse to participate in the survey. As previously mentioned, all non-respondents were dropped from the analysis.

reasons for accepting [rejecting] the offer amount. Debriefing questions that explore the nature of respondents' decision-making process have proved to be informative for identifying strategic behavior from stated preference WTP dichotomous choice responses (see Vossler, Doyon, and Rondeau 2012; Moore et al. 2018).

6. CONCLUSION

The results from this research provide the first evidence about the effectiveness of an ex-post certainty adjustment technique in a WTA environment. It would, of course, be unwarranted to draw overarching conclusions about the effectiveness of the certainty adjustment approach we used based on a single WTA experiment, but our findings are worthy of additional exploration. The issue to be resolved is whether our WTA result – that the approach mitigated but did not eliminate the overstated hypothetical bias – is generalizable across alternative goods and experimental designs.

As noted earlier, stricter access restrictions would deprive saltwater anglers of their de jure rights to fishery resources. In such cases, if WTP measures of losses are estimated instead of WTA estimates, the real value of the associated losses will likely be underestimated. To stimulate development of more defensible WTA-based measures of losses for environmental goods and services, there is a need for additional studies that compare hypothetical and actual WTA valuations. This gap in the literature also limits the ability of researchers to explain disparities between hypothetical and actual stated WTA values across studies. While meta-analysis of hypothetical bias has been conducted previously (List and Gallet 2001; Little and Berrens 2004; Little, Broadbent, and Berrens 2012; Penn and Hu 2018), the lack of WTA studies contained in these meta-analyses (about 5% of the total) make it truly difficult to capture the influence of WTA on hypothetical bias. In fact, Murphy et al. (2005) exclude WTA observations in their meta-analysis of stated preference studies due to concerns of insufficient WTA observations. We hope that this issue can be resolved through additional comparison studies of hypothetical and actual WTA valuation responses.

Depending upon the context of the stated preference application though, eliciting values in a non-hypothetical manner may be cost prohibitive or politically infeasible. Indeed, we faced considerable challenges obtaining funding for our study due to the added cost of the real simulated marketplace and the questions it raised about government agencies ultimately paying anglers not to fish. A purely hypothetical WTA study would not have raised the same level of consternation. To overcome these obstacles it may be possible to conduct controlled experiments for a surrogate good that could be used to measure and adjust for the hypothetical bias at hand (Harrison 2006). In cases such as ours, however, where there are no close substitutes for saltwater recreational fishing licenses, there may be no other practical way than to address the issue head on.

REFERENCES

Arrow, Kenneth, Robert Solow, Paul Portney, Edward E. Leamer, Roy Radner, and Howard Schuman. 1993. Report of the NOAA Panel on Contingent Valuation. *Federal Register*, 58 (10): 4602-4614.

Bischoff, Ivo. 2008. "Endowment Effect Theory, Prediction Bias and Publicly Provided Goods: An Experimental Study." *Environmental & Resource Economics* 39 (3): 283-296.

Bishop, Richard C., and Thomas A. Heberlein. 1979. "Measuring Values of Extramarket Goods: Are Indirect Measured Biased?" *American Journal of Agricultural Economics* 61 (5): 926-930.

Bishop, Richard C., Thomas A. Heberlein, and Mary Jo Kealy. 1983. "Contingent Valuation of Environmental Assets: Comparisons with a Simulated Market." *Natural Resources Journal* 23 (3): 619-633.

Bishop, Richard C., Thomas A. Heberlein. 1990. The Contingent Valuation Method, in R.L. Johnson & G.V. Johnson (Eds.) Economic Valuation of Natural Resources: Issues, Theory, and Applications , Boulder, CO: Westview Press.

Blumenschein, Karen, Magnus Johannesson, Glenn C. Blomquist, Bengt Liljas, and Richard M. O'Connor. 1998. "Experimental Results on Expressed Certainty and Hypothetical Bias in Contingent Valuation." *Southern Economic Journal* 65 (1): 169-177.

Blumenschein, Karen, Magnus Johannesson, Krista K. Yokoyama, and Patricia R. Freeman. 2001. "Hypothetical Versus Real Willingness to Pay in the Health Care Sector: Results from a Field Experiment." *Journal of Health Economics* 20 (3): 441-457.

Blumenschein, Karen, Glenn C. Blomquist, B., Magnus Johannesson, Nancy Horn, and Patricia R. Freeman. 2008. "Eliciting Willingness to Pay without Bias: Evidence from a Field Experiment." *The Economic Journal* 118 (525): 114-137.

Bromley, Daniel W. 1995. "Property Rights and Natural Resource Damage Assessments." *Ecological Economics* 14: 129-135.

Brookshire, David S., and Don L. Coursey. 1987. "Measuring the Value of a Public Good: An Empirical Comparison of Elicitation Procedures." *American Economic Review* 77 (4): 554-566.

Brown, Thomas C., and Robin Gregory. 1999. "Why the WTA-WTP Disparity Matters." *Ecological Economics* 28: 323-335.

Bush, Glenn, Nick Hanley, Mirko Moro, and Daniel Rondeau. 2013. "Measuring the Local Costs of Conservation: A Provision Point Mechanism for Eliciting Willingness to Accept Compensation." *Land Economics* 89 (3): 490-513.

Carson, Richard T., and Theodore Groves. 2007. "Incentive and Informational Properties of Preference Questions." *Environmental Resource Economics* 37 (1): 181-210.

Chambers, Catherine. M., Paul. E. Chambers, and John C. Whitehead. 1998. "Contingent Valuation of Quasi-Public Goods: Validity, Reliability, and Application to Valuing a Historic Site." *Public Finance Review* 26 (2): 137-154.

Champ, Patricia A., Richard C. Bishop, Thomas C. Brown, and Daniel W. McCollum. 1997. "Using Donation Mechanisms to Value Nonuse Benefits from Public Goods." *Journal of Environmental Economics and Management* 33: 151-162.

Champ, Patricia A., and Richard C. Bishop. 2001. "Donation Payment Mechanisms and Contingent Valuation: An Empirical Study of Hypothetical Bias." *Environmental and Resource Economics* 19 (4): 383-402.

Champ, Patricia A., Rebecca Moore, and Richard C. Bishop. 2009. "A Comparison of Approaches to Mitigate Hypothetical Bias." *Agricultural and Resource Economics Review* 38 (2): 166-180.

Coursey, Donald L., John L. Hovis, and William D. Schulze. 1987. "The Disparity between Willingness to Accept and Willingness to Pay Measures of Value." *Quarterly Journal of Economics* 102 (3): 679-690.

Dillman, Don A. 2007. Mail and Internet Surveys: The Tailored Design Method, Second Edition. Wiley, New York, NY. 512 p.

Ethier, Robert G., Gregory Poe, William D. Schulze, and Jeremy Clark. 2000. "Comparison of Hypothetical Phone and Mail Contingent Valuation Responses for Green-Pricing Electricity Programs." *Land Economics* 76 (1): 54-67.

Freeman, A. Myrick III. 1979. The Benefits of Environmental Improvement: Theory and Practice. United States: John Hopkins University Press, Baltimore, MD. 277 p.

Freeman, A. Myrick III. 2003. The Measurement of Environmental and Resource Values: Theory and Methods, Second Edition. Resources for the Future, Washington, DC. 491 p.

Fuller, Wayne A. 2009. Sampling Statistics. John Wiley & Sons, Hoboken, NJ. 472 p.

Haab, T.C., and K.E. Mc Connell. 2002. *Valuing Environmental and Natural Resources: The Econometrics of Non-Market Valuation*. Cheltenham, UK: Edward Elgar Publishing.

Hammack, Judd, and Gardner Mallard Brown. 1974. Waterfowl and Wetlands: Toward Bioeconomic Analysis. Johns Hopkins Univ. Press, Baltimore, MD. 95 p.

Hanemann, W. Michael. 1984. "Welfare Evaluations in Contingent Valuation Experiments with Discrete Responses." *American Journal of Agricultural Economics* 66 (3): 332-341.

Hanemann, W. Michael. 1989. “Welfare Evaluations in Contingent Valuation Experiments with Discreet Response Data: Reply.” *American Journal of Agricultural Economics* 71 (4): 1057-1061.

Hanemann, W. Michael. 1991. “Willingness to Pay and Willingness to Accept: How Much Can They Differ?” *The American Economic Review* 81 (3): 635-647.

Harrison, Glenn W. 2006. “Experimental Evidence on Alternative Environmental Valuation Methods.” *Environmental and Resource Economic*s 34 (1): 125-162.

Hausman, Jerry. 2012. “Contingent Valuation: From Dubious to Hopeless.” *Journal of Economic Perspectives* 26 (4): 43-56.

Horowitz, John K., and Kenneth E. McConnell. 2002. “A Review of WTA/WTP Studies.” *Journal of Environmental Economics and Management* 44 (3): 426-447.

Johannesson, Magnus, Bengt Liljas, and Per-Olov Johansson. 1998. “An Experimental Comparison of Dichotomous Choice Contingent Valuation Questions and Real Purchase Decisions.” *Applied Economics* 30 (5): 643-647.

Kaczan, Daniel, Brent M. Swallow, and W.L. Adamowicz. 2013. “Designing a Payments for Ecosystem Services (PES) Program to Reduce Deforestation in Tanzania: An Assessment of Payment Approaches." *Ecological Economics* 95: 20-30.

Kahneman, Daniel, and Amos Tversky. 1979. “Prospect Theory: An Analysis of Decision under Risk.” *Econometricia* 47 (2): 263-292.

Kahneman, Daniel, and Jack L. Knetsch. 1992. “Valuing Public Goods: The Purchase of Moral Satisfaction.” *Journal of Environmental Economics and Management* 22 (1): 57-70.

Kahneman, Daniel, Ilana Ritov, and David Schkade. 1999. “Economic Preferences or Attitude Expressions?: An Analysis of Dollar Responses to Public Issues.” *Journal of Risk and Uncertainty* 19 (1): 203-235.

Kim, Younjun, Catherine L. King, and Jinhua Zhao. 2015. “Understanding Behavioral Explanations of the WTP-WTA Divergence Through a Neoclassical Lens: Implications for Environmental Policy.” *Annual Review of Resource Economics* 7: 169-187.

Kling, Catherine L., Daniel J. Phaneuf, and Jinhua Zhao. 2012. “From Exxon to BP: Has Some Number Become Better than No Number?” *Journal of Economic Perspectives* 26 (4): 3-26.

Knetsch, Jack L. 2010. “Values of Gains and Losses: Reference States and Choice of Measure.” *Environmental Resource Economics* 46 (2): 179-188.

Krinsky, Itzhak, and A. Leslie Robb. 1986. “On Approximating the Statistical Properties of Elasticities.” *The Review of Economic and Statistics* 68 (4): 715-719.

Lessler, Judith T. and William D. Kalsbeek. 1992. Nonsampling Error in Surveys. John Wiley & Sons, New York, NY. 432 p.

List, John A., and Jason F. Shogren. 1998. "The Deadweight Loss of Christmas: Comment." *American Economic Review* 88 (5): 1350-1355.

List, John A., and Craig A. Gallet. 2001. "What Experimental Protocol Influence Disparities Between Actual and Hypothetical State Values?" *Environmental and Resource Economics* 20 (3): 241-254.

List, John A., and Jason F. Shogren. 2002. "Calibration of Willingness-to-Accept." *Journal of Environmental Economics and Management* 43: 219-233.

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

Little, Joseph, Craig D. Broadbent, and Robert P. Berrens. 2012. "Meta-Analysis of the Probability of Disparity between Actual and Hypothetical Valuation Responses: Extension and Preliminary New Results." Western Economics Forum, *Western Agricultural Economics Association* 11: 1-12.

Mitchell, Robert Cameron, and Richard T. Carson. 1989. "Bishop and Heberlein's Horicon Goose-hunting Experiment." Pp. 195-199 in Using Surveys to Value Public Goods: the Contingent Valuation Method by Robert Cameron Mitchell and Richard T. Carson, Resources for the Future Press, New York, NY. 484 p.

Moore, Chris, Dennis Guignet, Chris Dockins, and Kelly B. Maquire. 2018. "Valuing Ecological Improvements in the Chesapeake Bay and the Importance of Ancillary Benefits." *Journal of Benefit-Cost Analysis* 9(1):1-26.

Morrison, Mark, and Thomas C. Brown. 2009. "Testing the Effectiveness of Certainty Scales, Cheap Talk, and Dissonance-Minimization in Reducing Hypothetical Bias in Contingent Valuation Studies." *Environmental Resource Economics* 44: 307-326.

Murphy, James J., P. Geoffrey Allen, Thomas A. Stevens, and Darryl Weatherhead. 2005. "A Meta-Analysis of Hypothetical Bias in Stated Preference Valuation." *Environmental and Resource Economics* 30 (3): 313-325.

Nape, Steven, Peter Frykblom, Glen W. Harrison, and James C. Lesley. 2003. "Hypothetical Bias and Willingness to Accept." *Economics Letters* 78: 423-430.

Norwood, Bailey F. 2005. "Can Calibration Reconcile Stated and Observed Preferences?" *Journal of Agricultural and Applied Economics* 37 (1): 237-248.

OMB (U.S. Office of Management and Budget), 2003. Circular A-4: Regulatory Analysis. Washington, DC, September 17, 2003.

Poe, Gregory L, Eric K. Severance-Lossin, and Michael P. Welsh. 1994. "Measuring the Difference (X – Y) of Simulated Distributions: A Convolutions Approach." *American Journal of Agricultural Economics* 76: 904-915.

Poe, Gregory L., Jeremy E. Clark, Daniel Rondeau, and William D. Schulze. 2002. "Provision Point Mechanisms and Field Validity Tests of Contingent Valuation." *Environmental and Resource Economics* 23 (1): 105-131.

Randall, Alan, and John R. Stoll. 1980. "Consumer's Surplus in Commodity Space." *The American Economic Review* 70 (3): 449-455.

Rao, J. N. K. and Scott, A. J. 1981. "The Analysis of Categorical Data from Complex Surveys: Chi-Squared Tests for Goodness of Fit and Independence in Two-Way Tables," *Journal of the American Statistical Association*, 76: 221–230.

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

Svedsäter, Henrik. 2003. "Economic Valuation of the Environment: How Citizens Make Sense of Contingent Valuation Questions." *Land Economics* 79 (1): 122-135.

Thaler, Richard. 1980. "Toward a Positive Theory of Consumer Choice." *Journal of Economic Behavior and Organization* 1: 39-60.

Tuncel, Tuba, and James K. Hammitt. 2014. "A New Meta-Analysis on the WTP/WTA disparity." *Journal of Environmental Economics and Management* 68: 175-187.

Vossler, Christian A., Maurice Doyon, and Daniel Rondeau. 2012. "Truth in Consequentiality: Theory and Field Evidence on Discrete Choice Experiments." *American Economic Journal: Microeconomics* 4(4):145-171.

Whittington, Dale, Wiktor Adamowicz, and Patrick Lloyd-Smith. 2017. "Asking Willingness-to-Accept Questions in Stated Preference Surveys: A Review and Research Agenda." *Annual Review of Resource Economics* 9: 317-336.

APPENDIX

Final analysis weights were calculated to reflect both the sample and experimental designs of the study. A stratified random without replacement design was used to select the sample with strata defined by the four 1-month waves (January-April, 2012). Base sample weights were calculated following standard sampling methodology (e.g. Fuller 2009):

$$w_{i,base} = \pi_{hi}^{-1} = \frac{N_h}{n_h}$$

where $w_{i,base}$ was the base sample weight for unit i in stratum h, π_{hi}^{-1} was the inverse of the inclusion probability for unit i, N_h was the total count of sample units in the frame for stratum h, and n_h was the sample size or number of units originally drawn from stratum h using simple random sampling without replacement.

To account for nonresponse, nonresponding units were assumed to be missing completely at random within strata and a poststratification adjustment method (e.g., Lessler and Kalsbeek 1992, p. 183) was employed to adjust the weights of responding units to represent the entire drawn sample:

$$\begin{aligned} w_{i,resp} &= \alpha_h \frac{N_h}{n_h} = \frac{N_h}{n_r} && \text{if } i \text{ is a respondent} \\ &= 0 && \text{if } i \text{ is a nonrespondent} \end{aligned}$$

$$\alpha_h = \frac{\sum_{i=1}^{n_h} \frac{N_h}{n_h}}{\sum_{i=1}^{n_r} \frac{N_h}{n_h}} = \frac{\frac{N_h}{n_h} n_h}{\frac{N_h}{n_h} n_r} = \frac{n_h}{n_r}$$

where $w_{i,resp}$ was the nonresponse adjusted sample weight for unit i in stratum h, α_h was the poststratification adjustment made to the base weight for all responding units in h, and n_r was the number of responding units in h. Applying this adjustment increased the sample weights of the responding units such that the sum of the adjusted sample weights for the responding units, n_r, equaled the sum of the base weights for the full sample, n_h. The respondents now represented both themselves and the nonrespondents.

A second poststratification adjustment was made to account for the experimental design. Sample units were randomly assigned to one of 30 distinct treatments: two valuation groups (HWTA and AWTA) crossed with fifteen different monetary offer levels (15, 25, 40, 55, 75, 100, 125, 160, 200, 250, 300, 350, 400, 450, 500). Final analysis weights were calculated separately by treatment group:

$$w_{i,final} = \alpha_t^{'} w_{i,resp}$$

$$\alpha_t^{'} = \frac{\sum_1^{n_{r1}} w_{i,resp}}{\sum_1^{n_{rt}} w_{i,resp}}$$

where $w_{i,final}$ was the final analysis weight for unit i in treatment group t, $\alpha_t^{'}$ was the poststratification adjustment made to the nonresponse weight for all responding units in t, n_{r1} was the number of responding units in treatment group 1 (HWTA, 15), and n_{rt} was the number of responding units in treatment group t. For this adjustment, the sums of the nonresponse weights for each treatment group were standardized to the sum of the nonresponse weights for the first treatment group (HWTA, 15). Using the first treatment group as the control total was made for convenience and did not have any impact on study results. Giving equal total weight to both treatment groups ensured results would not be biased towards the treatment with larger totals of nonresponse weights, a situation that could occur due to differential nonresponse or initial random assignments among the two treatments.