COMMODITY DEFINITION AND CONTENT VALIDITY IN STATED PREFERENCE VALUATION: A META-ANALYSIS OF WATER QUALITY WELFARE ESTIMATES

By

Jamie Arnett Hansen

RECOMMENDED:

Dr. Jungho Baek

Dr. Peter Fix

Dr. Joseph Little, Advisory Committee Chair

Dr. Joseph Little, Program Director, M.S. Resource and Applied Economics

APPROVED:

Dr. Mark Herrmann, Dean, School of Management

Dr. John Eichelberger, Dean of the Graduate School

29 April 2013

Date

COMMODITY DEFINITION AND CONTENT VALIDITY IN STATED PREFERENCE VALUATION: A META-ANALYSIS OF WATER QUALITY WELFARE ESTIMATES

A THESIS

Presented to the Faculty

of the University of Alaska Fairbanks

in Partial Fulfillment of the Requirements

for the Degree of

MASTER OF SCIENCE

By

Jamie Arnett Hansen, B.A., M.A.

Fairbanks, Alaska

May 2013

Abstract

This paper applies a meta-analysis to investigate variation in willingness to pay estimates that arise from the use of different commodity descriptions in stated preference valuation surveys. To maintain commodity consistency, the data set for this meta-analysis is composed of willingness to pay estimates from contingent valuation, conjoint analysis, and choice experiment studies valuing water quality change in surface water bodies in the United States. The analysis uses an ordinary least squares regression with a cluster command to correct for potential correlation between observations drawn from the same study. The primary contribution of this study is the identification of systematic variation across stated preference studies resulting from changes in how the environmental commodity is presented and defined. By identifying the directional effect of these differences, this analysis provides insight into interpreting stated preference estimates and guidance for producing well-designed stated preference studies capable of eliminating bias and context effects.

Table of Contents

Page

List of Figures

List of Tables

List of Appendices

Acknowledgements

This thesis would not have been possible without the support of my professors, colleagues, and family. I would like to express my heartfelt thanks to:

My thesis committee, in particular, Dr. Joseph Little and Dr. Jungho Baek who both offered guidance and support for this project as it developed.

My family with special thanks to my sister, Sarah Arnett, for her encouragement and inspiration and to my husband, Steven Hansen, for his unconditional support.

1 Introduction

Stated preference valuation is commonly used to quantify the benefits from environmental resources and services. Initially spurred by the 1989 legal decision from the District of Columbia Court of Appeals, which affirmed the use of contingent valuation (CV) and the inclusion of passive use values in total compensable damages, the use of CV and related stated preference methods have grown markedly. The rise in stated preference valuation studies has led to a corresponding rise in the number of investigations into the validity of stated preference welfare estimates. The 1993 National Oceanic and Atmospheric Administration (NOAA) panel evaluated the most common stated preference method, CV, to determine the validity of CV measures of value (Arrow et al., 1993). The NOAA panel guidelines, which recommend protocols for study design and validity tests, remain a reference for both critics and practitioners of CV. While tests and recommendations on specific aspects of content validity – payment mechanism, elicitation format, and presentation of the environmental commodity – are covered in the NOAA protocols, a focused review of commodity definition, in essence the core of a stated preference study, is largely missing.

In a Resources for the Future report, Boyd and Krupnick (2009) argue the lack of protocols on how environmental commodities are presented and defined opens value estimations up to questions of bias or even relevance, e.g. what exactly did the study value. A river, forest, open space, wetland, or other nonmarket environmental resource is a lumpy commodity comprised of a bundle of ecological inputs, outputs, and processes; the precise definition of an environmental commodity is of paramount importance, regardless of the good's perceived level of complexity. Imprecision in defining the environmental commodities to be valued compelled Boyd and Krupnick (2009) to question the accuracy of nonmarket valuation estimates. In a similar line of inquiry, Johnston, Schultz, Segerson, Besedin, and Ramachandran (2012) review commodity definition in stated preference valuation methods and find them lacking in terms of quantifiable and ecologically relevant indicators. Both studies raise grave concerns about the content validity of these estimation practices, questioning the presumed comparability of estimates and their use in benefit transfer applications.

This paper investigates the variation in willingness to pay (WTP) estimates, arising from changes in the commodity descriptions found in stated preference valuation studies. Drawing on Boyd and Krupnick's (2009) critique and Johnston et al.'s (2012) study-specific findings, this analysis contributes to the literature by identifying the systematic variation across stated preference studies that results from changes in how the environmental commodity is presented and defined. By identifying the directional effect of these differences, this analysis provides insight into interpreting stated preference estimates and guidance for producing well-designed stated preference studies capable of eliminating bias and context effects.

This study employs the meta-analysis method, which is uniquely capable of identifying variations in WTP across nonmarket valuation studies. The data for this meta-analysis are drawn from stated preference studies conducted between 1989 and 2005 that estimate the mean household willingness to pay for water quality improvements of surface water in the United States with distinct consumptive uses[1], such as recreational and agricultural activities. This meta-analysis uses commodity definition components identified by Boyd and Krupnick (2009) and Johnston et al. (2012) and employs the recommended practices outlined by Nelson and Kennedy (2009) to set up and test the degree to which commodity definition characteristics are determinants of willingness to pay. The meta-analysis regression employs ordinary least squares for both preliminary and final analysis with an additional cluster command used in the final analysis to account for potential correlation between observations collected from the same study. A series of diagnostic tests and specification comparisons is undertaken including tests for heteroskedasticity, outliers, natural log transformation of the dependent variable, and related diagnostics on the goodness of fit to select the preferred model and regression method.

[1] In this paper, consumptive use excludes the use of surface water for municipal or other drinking water supplies.

2 Literature Review

Nonmarket valuation methods identify the instrumental value of environmental commodities using two fundamental tenets of neoclassical welfare economics: economic activity is undertaken to increase the well-being of an individual, and the individual is best able to determine his or her own state of well-being (Freeman, 2003). A related microeconomic assumption is that individuals will be able to identify the characteristics, attributes, or scenarios they prefer from a set of goods or services (Flores N. , 2003). Stated preference (SP) methods ask an individual to state his or her preference rather than observing behavior or using proxy indicators to identify preferences. The goal of SP welfare measures is to identify the Hicksian or compensated demand in which utility is fixed while the prices of market goods and the level of nonmarket goods vary. For a nonmarket good, key information is missing that must be elicited: the amount of variation in the market good's price and amount of variation in the nonmarket good's levels that will return an individual to a given level of utility (whether utility is fixed at status quo levels or changed to a new level) (Flores N. , 2003). Since the nonmarket goods considered here are public, the levels of provision are shared while marginal values vary among individuals. The identification of an individual's marginal value for a given utility level, based on the levels of a particular nonmarket good, is the critical piece of information nonmarket valuation techniques are designed to estimate.

To accomplish this with nonmarket environmental goods, SP methods use surveys to present an environmental good or goods with hypothetical changes in provision levels, such as quality, along with the hypothetical values for these levels (Brouwer, Dekker, Rolfe, & Windle, 2010). The environmental commodities valued cover a broad range of resources, from general biodiversity to the survival of specific rare species, and from wetland restoration to fish catch rates in recreational freshwater bodies. This analysis reviews the literature on stated preference methods as applied to various resources but narrows the focus to water quality studies for the meta-analysis.

2.1 Stated Preference Methods

The most commonly used stated preference methods in economics are easily divided into contingent valuation and conjoint analysis (CJ) and the related choice experiments (CE). Contingent valuation has a long history and remains the most widely used and heavily criticized stated preference method. Though developed and used primarily in the fields of marketing and transportation, conjoint analysis and choice experiments are increasingly used as complements or alternatives to contingent valuation. These methods produce comparable welfare measures – the Hicksian compensating variation or surplus value estimates (Hoyos, 2010). The values estimated capture both use and passive use value for the environmental commodity of interest (Adamowicz, Boxall, Williams, & Louviere, 1998). The methods share the same basic design, which is developing and administering a survey describing the environmental commodity to be valued, eliciting the value at which the respondent would be willing to pay to acquire the commodity, and offering a mechanism through which the respondent would pay for this commodity (Carson & Louviere, 2011; Powe, Garrod, & McMahon, 2005). The phases of survey design in terms of information gathering, design and testing of the instrument, sample selection, data collection, and modeling are quite similar among the SP methods, but the elicitation question and information context are distinct (Boxall, Adamowicz, Swait, Williams, & Louviere, 1996; Powe et al., 2005). CV describes the changes in an environmental commodity as a choice between the status quo and the change, either an improvement or degradation. The choice set presented in CE and CJ is a set of characteristics and features (often described as attributes) that describe the situation in a multi-dimensional manner. Whereas the CV scenario is typically a single decision for all or nothing, the CE and CJ scenario is a choice set ranging from 2 to 16 or more decisions on specific commodity attributes of which price is only one (Boxall et al., 1996). Due to the nature of the elicitation question and choice set for CE and CJ, the approach to researching and constructing an environmental commodity is different from that typically used in CV.

Though not as extensively researched as the difference between the welfare measures of revealed and stated preference methods, the difference between these stated preference methods has some empirical basis. Adamowicz et al. (1998) show the error variances and welfare values of CV and CE are straight alternatives or complements to each other. Limited research conflicting with this conclusion has identified differences in the consideration of substitutes, hypothetical bias, and sensitivity to scope, as well as the varying results from alternative elicitation methods. Boxall et al. (1996) conducted simultaneous CV and CE surveys with the same group of respondents and found the CV estimates were over twenty times higher than the CE estimates. This study tentatively concludes the difference results from the fact that the CE survey explicitly allows respondents to incorporate and examine substitutes, while the CV method does not. This explanation is echoed in later reviews and studies (Powe et al., 2005). Other factors that may lead to different estimates include the lower hypothetical bias in CE and CJ and greater sensitivity to scope (Hoyos, 2010). Finally, the elicitation method is typically different. CE and CJ use a sequential binary or multinomial choice set while CV often relies on a single binary choice. This is discussed in more detail later in this analysis, but McNair, Bennett, and Hensher (2011) found the choice formats used in CJ and CE result in smaller WTP estimates than the single binary choice format used in CV.

The evidence to date points to a clear difference in value estimates among stated preference methods. This meta-analysis controls for this difference without losing the variation in commodity definition. Indeed, the literature has not identified commodity definition as a potential factor in the difference between stated preference estimates, making research in the area rich with possibility.

2.2 Validity of Stated Preference Methods

Three types of validity are considered when evaluating the reliability of a stated preference measure: criterion, construct, and content. Criterion validity assesses the validity of an SP estimate by comparing it against another measure, closer to the theoretical construct being researched – in this case preferences (Brown, 2003).

Construct validity evaluates the measure based on a comparison with other measures expected to provide the same or similar values, such as revealed preference measures, or against other factors theoretically expected to have a direct relationship to the size of the measure, such as income, use, or size of change (Brown, 2003).

The final form of validity, which will be investigated in this analysis, is content validity. Content validity examines the survey instrument used to obtain the value, seeking to verify if the instrument is able to elicit true WTP. Content validity questions may include "(1) are the items to be valued unambiguously described to respondents?, (2) is the payment vehicle … likely to be accepted as reasonable?, (3) does the sample represent the population?, and (4) is the statistical model appropriate?" (Brown, 2003, p. 104). While content validity covers the entire study, certain areas within content validity are far more heavily researched than others. The literature on CV and to a lesser extent CE and CJ is filled with studies on elicitation mechanisms, choice complexity, payment mechanisms, and statistical models used to estimate the welfare value. The on-going criticism of the validity of stated preference methods is continually refuted with the contention that a well-designed survey is capable of minimizing or eliminating the concerns surrounding the validity of value estimates (Brown, 2003). This is guided by a set of established and evolving protocols for these areas with regard to incentive compatibility and scenario credibility (Arrow et al., 1993; Boyle, 2003; Brown, 2003; Carson & Groves, 2007; Carson & Louviere, 2011; Champ & Welsh, 2007; Hannemann, 1994; Hoyos, 2010; McConnell, 1990; Stevens, DeCoteau, & Willis, 1997; Welsh & Poe, 1998).

2.3 Content Validity & The Information Effect

There are no guidelines or standards for the description and presentation of commodity information in stated preference studies, though the best practice for CE and CJ is far more defined and established than those for CV (Boyle, 2003; Boyd & Krupnick, 2009; Johnston et al., 2012; Hoyos, 2010). The mechanics of the effect information has on welfare measure estimates, in terms of the way a commodity is presented and defined, have not been addressed as intensively as the effect the elicitation method and payment mechanism have on WTP estimates (Boyle, 2003). Indeed, Boyle (2003) contends most researchers recognize the importance of information in a survey, but the accompanying assumption that respondents actually understand the information or what they are being asked to value may be invalid. Johnson et al. (2012) posit this research void in stated preference literature is compounded not only by the untested assumptions identified by Boyle (2003) but also due to the difficulty of quantitatively testing content validity, which relies on "subjective expert appraisal", in a field that highly values quantitative evaluation (Johnston et al., 2012, p. 102). The absence of an overarching research program or even consistent attention does not mean, however, that the area of commodity definition in content validity has been ignored. A number of strands of research have focused on the role information plays in altering the value and the certainty surrounding welfare estimates obtained through stated preference techniques.

The stated preference method is a "process of information transfer" (Boyle, 1989, p. 59) and fundamental to this transfer is the description of the resource change to be valued (Boyle, 2003). The identification of a specific information effect in the stated preference method has been demonstrated by studies addressing how the amount of information provided on an environmental commodity affects the WTP estimates for those commodities. In sum, additional commodity information increases WTP estimates (Bergstrom, Stoll, & Randall, 1990; Blomquist and Whitehead, 1998; Samples, Dixon, & Gowen, 1986). This empirical conclusion was re-affirmed in the 1993 NOAA panel guidelines for CV, specifying the need for an accurate description of the program or policy being valued (Arrow et. al, 1993). Milon and Scrogin (2006) examined the traditional information effect in a survey using a choice experiment elicitation method. As with earlier studies, Milon and Scrogin (2006) found the use of specific, contextual information, in this case structural attributes or endpoints, results in a higher WTP because the environmental change being valued is described using tangible and relevant services and products. More recent research into the information effect has focused on the pre-existing preferences, information, and perceptions respondents bring to a stated preference valuation survey. In these studies, empirical findings show information is a determinant of WTP but the information effect is also interacted with respondents' personal relevance and familiarity with the resource (Ajzen, Brown, and Rosenthal, 1996; Blomquist and Whitehead, 1998).

Other recent investigations into context and information effects have framed the research in terms of preference construction. As mentioned earlier, nonmarket valuation techniques draw validity from a number of economic theories including the neoclassical theory maintaining that individuals prefer certain goods over others, whether exchanged in the market or no, and these preferences are fixed and stable. The fixed nature of preferences has come under increasing scrutiny and doubt from within economics and from interdisciplinary reviews based in psychology and consumer theory (Fischer & Hanley, 2007; Gregory, Lichtenstein, & Slovic, 2006; Schwarz, 1997).

Drawing on behavioral psychology research, Schwarz (1997) and Gregory et al. (2006) argue economists engaged in nonmarket valuation should explicitly accept the construction of preferences during the elicitation process and design the survey to take an active role in value and preference construction. As Gregory et al. (2006) explain, individuals have cognitive beliefs about goods not sold in the market, but these beliefs are by no means quantified or monetized. Nonmarket valuation that attempts to elicit these preferences must serve a tutorial function with care taken to avoid framing or context details that skew or distort value preferences. To minimize these effects, Gregory et al. (2006) recommend presenting a comprehensive picture of all elements contributing to the value of the good or service. Similarly, Schwarz (1997) agues CV practitioners recognize the need for accurate and detailed information with the aim of avoiding respondent decisions made using "context independent information" (p. 71). The concern is that if respondents draw on information from outside the survey, their responses may not answer the questions asked. This context problem increases with respondent knowledge and decreases with the explicitness of the scenario in the survey (Schwarz, 1997). Carson, Flores, and

Mitchell (1999) argue information rather than familiarity is the key to well-informed choices and stable preference revelation.

A similar line of research focuses on the cognitive demands arising from the valuation of complex goods and questions the validity of results from respondents unfamiliar with these complex goods. These results tend to be subject to the distorting effects that extraneous, imprecise, or inaccurate information have on unfamiliar respondents (Brown, 2003; Barkmann, et al., 2008).

Despite the varying conclusions on information and familiarity, all affirm a well-designed survey is the bulwark against unpredictable results and ensures, to the extent possible, respondents answer the questions posed while using the information provided to construct preferences. The recommendations to counteract these concerns are similar across economics, psychology, and ecology: pre-testing of studies, careful design of surveys with care to present relevant descriptions of the environmental commodity, describing ecosystem services in terms of benefits respondents care about, and placing restrictions on the size of the system change being valued (Barkmann, et al., 2008; Christie et al., 2006; de Groot, Wilson, & Boumans, 2002; Nunes & van den Bergh, 2001).

Work on the information effect has focused on the presumption that more information is useful and, potentially, a means to alleviate problems of unfamiliarity among respondents. Little has been investigated or even noted on the specific information that should be included and how commodity information should be broken apart. Indeed, though studies have looked at the importance of complete information, there is limited guidance or investigation into what “complete information” must include and how researchers should go about presenting it.

2.4 Content Validity & Commodity Definition

The literature on stated preference methods has addressed multiple aspects of content validity but few have tackled the issue of importance in this paper: the description and definition of the key component of the survey, the environmental commodity to be valued. Boyd and Krupnick (2009) and Johnston et al. (2012) step into this line of inquiry and refine the basic criticism summarized by Boyle (2003) that information presented in stated preference surveys is not understood by respondents in the way researchers expect and, as a result, the WTP estimates may not be valid. The approach of Boyd and Krupnick's (2009) critique is starkly different from earlier work on content validity either on the information effect or contemporary work on preference formation. Where split sample survey designs were used to test small changes in key words and/or the level of information provided, Boyd and Krupnick (2009) take an analytical approach and ask the basic question Boyle (2003) raised - do respondents know what they are being asked to value? The follow-up question is to which commodity or commodities are the derived values attached? The critique rests on the assertion that, unless the commodity is disaggregated, survey respondents will bring their own "expansive priors" to the valuation process (Boyd & Krupnick, 2009, p. 29). For example, a simple valuation question on improving the water quality of a river by reducing nutrient runoff has the potential to expand into considerations of the habitat along the river, ecosystem in the river, and benefits to people living near the river. The same vagueness may also have the opposite effect. Defining the environmental commodity as a reduction in phosphorous load without clear endpoints, such as a habitat that supports aquatic life and fishing opportunities, may reduce respondents' interest and perceived loss of utility.

To address this deficiency in commodity definition, Boyd and Krupnick (2009) recommend using ecological endpoints, which are a subset of the biophysical outputs from natural resources. This is different from the technical measures of water quality frequently used in stated preference studies, such as chemical oxygen demand or pH, nitrogen, and phosphorous levels. Endpoints are tangible, known outputs directly used by households, enabling the elicitation of a preference on such items. Endpoints also help disaggregate the bundles of services included in any environmental commodity — a key step to precisely defining the services to be valued (Boyd & Krupnick, 2009).

Johnston et al. (2012) also take up these questions, prompted by Boyd and Krupnick's (2009) analysis, by conducting choice experiments with variants of a survey examining the effect of incomplete information on commodity valuation. Johnston et al. (2012) lay out a case for the serious threat to content validity posed by the widespread and consistent ambiguity in commodity definition found in stated preference surveys, arguing this ambiguity directly impacts the measurement of utility. For example, assume a respondent's utility is in the form $U_i(E(X))$, where $E(\cdot)$ is the environmental commodity familiar to the respondent and X is an environmental process that may or may not be known to the respondent but is an input to E. If a respondent is asked to value X based on its impact on E, this asks the respondent to value input X rather than output E, with which a respondent is likely to be familiar, and assumes the respondent knows the ecological production function that determines the way X affects E (Johnston et al., 2012).[2] This is not a safe assumption and leads in Johnston et al.'s (2012) estimation to speculation and potentially results in the respondents valuing different environmental commodities though they are described identically in the survey (Boyd & Krupnick, 2009; Carson, 1998; Schwarz, 1997).[3] The potential effect of ambiguity on commodity definition in the valuation process has not been addressed directly or extensively in CV literature and deserves more consideration (Boyle, 2003).

2.5 Implications of Commodity Definition Critique

Both critiques by Boyd and Krupnick (2009) and Johnston et al. (2012) detail the dangers ambiguity pose to the content validity of stated preference welfare measures. In concluding their review, Boyd and Krupnick (2009) advocate the use of ecological production theory in which ecological systems are broken up into inputs, outputs, and natural processes. Only through unbundling environmental commodities,

---

[2] Boyd & Krupnick (2009) describe this situation with respect to water quality and the water quality ladder, noting the water quality ladder uses inputs paired with outputs in an attempt to translate inputs (quantitative water quality measures) into endpoint-like outputs (fishable or swimmable water).

[3] Boyle (2003) and Carson et al. (1999) describe how such a situation develops. When a physical description of the resource or service flow change resulting from a policy is not available, the survey must frame questions around the policy change. This leaves respondents with little to no information on the resource change or change in service flow and, as a result, the respondents must make two assumptions 1) how the policy change affects resource conditions and 2) how the change in the resource affects the services they receive.

even those as seemingly simple as water quality in a river basin, is it possible to identify the processes and outputs valued and accurately interpret the values obtained. Boyd and Krupnick (2009) do not offer evidence nor do they hypothesize the direction of bias resulting from this commodity definition ambiguity. The review concludes that the more information provided on a commodity and the more flexible the valuation technique is in separating the inputs, outputs, and biophysical processes, the less exposed the estimation is to bias (Boyd & Krupnick, 2009).

Johnston et al. (2012) present the results from CE valuation surveys showing higher values placed on direct effects (probability of fish run survival) of the environmental commodity being valued (water habitat restoration) when information on the overall ecological condition and effects of habitat restoration are omitted. This leads Johnston et al. (2012) to hypothesize the difference is due to overvaluation of the attributes offered when respondents are not provided with the option of valuing the overall change in the ecological condition. This offers limited empirical justification to the calls for explicit and comprehensive descriptions of the ecological effects of the policy change.

Using the areas highlighted by Boyd and Krupnick (2009) and Johnston et al. (2012), this analysis identifies the elements of commodity definition and description expected to affect WTP. The aim is to peel away at the general findings on the information effect in earlier work and, in so doing, locate specific survey elements that may affect WTP. Before starting the analysis, a brief review of previous attempts to investigate commodity definition on WTP is undertaken.

2.6 Meta-Analysis & Variation in WTP

An important tool in the investigation of variation in WTP across studies is the meta-analysis, which is increasingly used in environmental economics. A meta-analysis allows for the identification of determinants of WTP by offering a means to test for systematic variation in WTP across primary valuation studies (Brander, Florax, & Vermaat, 2006; Smith & Pattanayak, 2002). In reviewing published meta-analyses on nonmarket valuation studies, no meta-analyses are found testing hypotheses on commodity definition. The characteristics of nonmarket commodities have been raised by reviews of meta-analyses in environmental economics – specifically the incomparability of seemingly similar environmental commodities (Bergstrom & Taylor, 2006; Nelson & Kennedy, 2009; Smith & Pattanayak, 2002). Boyd and Krupnick (2009) mention two meta-analyses that take a different approach to the problem of commodity heterogeneity and inconsistency (Johnston et al., 2005; Van Houtven, Powers, & Pattanayak, 2007). In an attempt to compare water quality changes across stated preference studies, the two meta-analyses mapped a comparable metric of change, the water quality ladder/ water quality index, onto already completed studies (Johnston et al., 2005; Van Houtven et al. (2007). Both meta-analyses limited the studies included according to Smith and Pattanayak's (2002) recommendation to increase homogeneity in the meta-analysis sample, rather than merely controlling for differences in the studies.

Though an impressive attempt to address the serious issue of commodity heterogeneity, the two meta-analyses focus little attention on the commodity information available to respondents during the valuation process. This disconnect with the presentation in the original study is apparent in the results for both analyses. In the results from Van Houtven et al. (2007), the variable of interest, change in water quality index, was only significant when interacted with the recreation variable, which identified the studies that had used recreation endpoints for water quality change in the original studies. Johnston et al. (2005) found an unexplained negative and statistically significant effect on WTP for studies that originally used the water quality ladder. Both studies noted these results but were unable to explain them other than suggesting the use of recreational terms are important (Van Houtven et al., 2007) or surmising the finding may result from systematic bias in mapping water quality measurements onto the original studies (Johnston et al., 2005). This absence again shows an avoidance of the critical consideration of commodity presentation that was found lacking in earlier work on the information effect.

Building on the work of Van Houtven et al. (2007) and Johnston et al. (2005), this meta-analysis uses many of the techniques employed by the two research groups, such as restricting studies and focusing on strict commodity comparability. This meta-analysis differs in two fundamental ways: 1) the analysis is based on the information originally available to respondents to enable smoother interpretation of results and 2) the analysis focuses on commodity definitions rather than efforts to make studies' quality changes comparable post-facto. Indeed, this meta-analysis is the first to test for systematic variation based on the findings and reviews of Boyd and Krupnick (2009) and Johnston et al. (2012). Specifically, this tests for the information effect, if any, of valuation ambiguity and service specificity on WTP estimates. Previous meta-analyses have provided mixed guidance on criteria for which studies to include and the econometric specifications of meta-models (Johnston, Besedin, & Wardwell, 2003; Johnston et al., 2005; Rosenberger & Loomis, 2000b; Van Houtven et al., 2007). This analysis follows the best practices approach outlined by Nelson and Kennedy (2009) in their assessment of meta-analyses conducted in the field of environmental economics: clear criteria for study selection and multiple model specifications supported by diagnostic tests.

3 Conceptual Framework

The conceptual model for this meta-analysis is a version of the variation function used for water quality change in the meta-analysis conducted by Van Houtven et al. (2007, p. 211).[4] The determinants in this model are predicted to affect WTP based on economic theory and empirical findings on study-specific variation.[5]

$$WTP = V\ (Quality,\ Income,\ Use,\ Study) \qquad \text{(Equation 1)}$$

WTP is the mean household willingness to pay for water quality change; Quality is the level of water quality change valued in the study; Income is mean or median income

[4] The variation function in Van Houtven et al. (2007) is $WTP = V(Q^0,Q^1,P,Y;B)$.

[5] This model is a variant of the model created by McConnell (1990) for interpretation of empirical dichotomous choice CV data and expanded by Whitehead (1995) for use on both continuous and discrete choice data on CV studies valuing quality changes. This function and its properties are explained in detail by McConnell (1990) and Whitehead (1995).

for the study group; Use represents socio-demographic factors related to use and proximity to the water resource. The final set of determinants, Study, is for study characteristics, such as elicitation method.

3.1 Determinants of WTP for Water Quality Change

Quality Characteristics (Quality): The commodity to be valued is water quality change. This analysis disaggregates quality effects into the core commodity, water quality change, and other quality-related presentation effects based on the information effects highlighted by Boyd and Krupnick (2009) and Johnston et al. (2012). The quality changes of interest are based on the information presented in the survey tool. A number of potential variables are raised in the literature; the variables used in this analysis are described in detail later. For all these variables, the goal is to identify variations in commodity definition and information design as they appeared to respondents.

Income Effects (Income): In the variation function for WTP, a household's income acts as a constraint on WTP just as it does in an indirect utility function for market goods. Since income is predicted to have an impact on WTP, the mean or median income for the sample is included in the model.

Socio-Demographic Factors (Use): Individual preferences for water quality change are determined by the nature of the water quality change and individuals' characteristics that affect preference. This includes proximity or familiarity with the resource as well as use of the resource. Variables identifying sample respondents as being users, owners of property, or living in close proximity to the resource are included in the model.

Study Characteristics (Study): Study characteristics, though not predicted to affect WTP values in economic theory, have been empirically shown to have an effect on the WTP values estimated using stated preference methods. The study characteristics included in the model are the elicitation method; year of the study; quality of the study, often defined as going through peer review and publication; framing of scenario, and the timing of payment. Study characteristics often included in nonmarket meta-analysis but not in this meta-analysis are discussed below.

Location: The location of the study is often included to reflect differences in appreciation, value, proximity, and size of natural resources in different areas of the U.S. The regions are often divided, artificially, according to the USDA regions – Northeast, Mid-Atlantic, Southeast, Southwest, Mid-West, Mountain Pacific, and the West. The environmental commodity considered in this analysis, water quality in surface water with consumptive uses, is found across the United States and used in very similar ways: fishing, recreation, and irrigation. There is no expectation that a significant regional difference would be found. Based on the lack of clear empirical findings to support inclusion and a preliminary analysis, location variables are not included in the model.[6]

Mode of Administration: The 1993 NOAA panel advocated the use of in-person interviews, noting reliable estimates were unlikely to be elicited in mail surveys (Arrow et al., 1993). Champ (2003) reviews survey administration modes as well as studies of those modes and finds the studies are inconclusive on the preferred method or even direction of bias for in-person or telephone administration versus mail survey administration. More recent studies employ mixed methods (phone-mail-phone versus mail-phone) and find no significant difference between the mean willingness to pay (Eisen-Hecht & Kramer, 2002). Reviews of web-based surveys have also found minimal differences between the values and opinions expressed in web-based versus telephone surveys (Berrens, Bohara, Jenkins-Smith, Silva, & Weimer, 2003). Given the lack of empirical evidence for a consistent bias in the value estimates gathered through varying survey administration techniques and the well-documented evidence that survey estimates do vary across modes of administration but not due to mode alone (but because of sample population characteristics and experience of surveyors), the survey administration variable is not included in the meta-analysis model (Boyle, 2003; Champ, 2003; Champ & Welsh, 2007).

[6] The results of a joint test on the location variables' statistical significance are presented in the preliminary analysis.

Payment Mechanism: The payment mechanism is often included in meta-analyses because the expectation is voluntary contributions will be higher than mandatory payments because of a free rider problem, in which there is little incentive to state actual WTP when payment is not required (Brown, 2003). Champ, Flores, Brown, and Chivers (2002) found the WTP for an individual contribution was smaller than the WTP for a referendum tax. Though the results of Champ et al. (2002) support the NOAA panel's concern on potential differences resulting from varied incentives offered by payment mechanisms, the results are also contrary to the expectation noted above. The payment mechanism must be evaluated in terms of incentive compatibility (or ability to elicit true WTP) as well as the mechanism's credibility in the eyes of respondents (Champ et al., 2002). In explaining the results, Champ et al. (2002) note an incentive compatible mechanism may not be credible, depending on the elicitation mechanism or scenario. In short, credibility is likely to affect a respondent's answer. Given the multiple elicitation methods considered in this meta-analysis, the finding of a systematic bias as the result of payment mechanism is not expected and, therefore, the variable is not included in the regression model.

4 Methodology

4.1 Selection and Summary of Data Sources

The focus of this meta-analysis is commodity definition, which requires gathering studies that value highly similar environmental commodities. With this emphasis, the data are drawn from primary stated preference studies estimating total household WTP for water quality changes in U.S. surface water bodies with consumptive uses, such as recreation, fishing, or agriculture. The specific criteria for study selection were: 1) the stated preference valuation method was used;[7] 2) total (use and nonuse) WTP per household was estimated; 3) the primary environmental commodity valued was water quality change for surface water; studies valuing

[7] The three stated preference methods represented are contingent valuation, conjoint analysis, and choice experiments.

changes for groundwater only, drinking water only, water quantity only, recreational uses only such as fish catch rates, or valuing ecosystems changes resulting from large scale clean-up without providing separate values for surface water quality improvements were excluded; 4) the water body is in the U.S.; 5) the data were from unique primary surveys, meaning multiple studies analyzing data gathered from the same survey are not included; and 6) the study was conducted in 1989 or later. The date criterion has two purposes: 1) to update recent meta-analyses on surface water that gathered much of their data from studies conducted in the early 1980s and 2) to reflect the studies likely to be used by the EPA Water Office for benefit transfer.[8]

Literature searches were conducted in the Environmental Valuation Reference Inventory, the Beneficial Use Values Database, ECONLIT, and combined databases with gray literature including reports, damage assessments, and dissertations. The primary studies used in meta-analyses by Van Houtven et al. (2007) and Johnston et al. (2005) were also reviewed and, when the above criteria were met, included.[9]

The resulting metadata include 75 estimates of WTP from 30 unique studies conducted between 1989 and 2005.[10] The studies include 20 peer-reviewed journal articles, four research and academic papers, one damage assessment report, four PhD dissertations, and one book chapter. The number of observations (WTP estimates) exceeds the number of studies because many studies produce multiple WTP values due to identification of subsample WTP and isolation of WTP for different levels of change and commodity attributes. Table 1 summarizes the study characteristics and number of observations from each study used in the metadata.

---

[8] The 1998 effluent guidelines used studies published in 1990 and 1993, indicating that more recent rules and guidelines would be highly unlikely to use data gathered before 1989 (Griffiths et al., 2012).

[9] Robert Johnston provided the dataset used for the 2005 meta-analysis (Johnston et al., 2005). All data included in this meta-analysis was taken directly from the original studies; however, comparing information with data contained in the Johnston dataset was very valuable to the creation of this dataset.

[10] These numbers show the full data set prior to preliminary analysis. After analyses on outliers and elicitation method variables, two studies and seven observations were removed from the initial data set.

Table 1: Studies Analyzed in Meta-Analysis[*]

In Final Dataset	Study Citation	# of Obs	Valuation Method	Elicitation Method	Extent of Change	Quality Change Factors	Other Study Characteristics	Sample Characteristics	Adj. WTP Values[**]
✓	(Abdul-Mohsen, 2005)	1	Contingent Valuation	DC-Single Referendum	Unusable to Full Use	*Description of Change:* Outputs *Endpoints:* Yes *Disaggregated Endpoints:* Yes *Narrow Change Description:* No	*Payment Period*: Lump Sum *Change Frame:* Improvement *Published:* No *Year of Study:* 2003	*Income:*[**] $47, 538 *Sample Population:* General Public in Area *Location:* MidWest	$119
✓	(Azevedo et al., 2001)	5	Contingent Valuation	DC-Single Referendum	Unusable to Usable & Unusable to Full Use	*Description of Change:* Outputs *Endpoints:* Yes *Disaggregated Endpoints:* Yes *Narrow Change Description:* Yes	*Payment Period*: Annual *Change Frame:* Improvement/Prevent Deterioration *Published:* No *Year of Study:* 2000	*Income:* $50, 633 - $56, 962 *Sample Population:* General Public in Area & Users *Location:* Mountain Pacific	$22 - $144

Table 1 continued

In Final Dataset	Study Citation	# of Obs	Valuation Method	Elicitation Method	Extent of Change	Quality Change Factors	Other Study Characteristics	Sample Characteristics	Adj. WTP Values**
✓	(Bishop, Breffle, Lazo, Rowe, & Wytinck, 2000)	5	Choice Experiment	Sequential Binary Choice Referendum	Unusable to Usable & Unusable to Full Use	*Description of Change:* Inputs *Endpoints:* Yes *Disaggregated Endpoints:* Both *Narrow Change Description:* Both	*Payment Period*: Annual *Change Frame:* Improvement *Published:* No *Year of Study: 1999*	*Income:* $56, 283 *Sample Population:* Residents *Location:* MidWest	$31 - $262
✓	(Collins, Benson, Borisova, & D'Souza, 2006)	3	Contingent Valuation	Payment Card Referendum	Unusable to Full Use	*Description of Change:* Inputs *Endpoints:* Yes *Disaggregated Endpoints:* No *Narrow Change Description:* No	*Payment Period*: Annual *Change Frame:* Improvement *Published:* No *Year of Study:* 2005	*Income:* $52, 228 - $61, 795 *Sample Population:* Residents & General Public in Area *Location:* MidAtlantic	$50 - $89

Table 1 continued

In Final Dataset	Study Citation	# of Obs	Valuation Method	Elicitation Method	Extent of Change	Quality Change Factors	Other Study Characteristics	Sample Characteristics	Adj. WTP Values**
✓	(Collins, Rosenberger, & Fletcher, 2005)	3	Choice Experiment	Sequential Binary Choice	Unusable to Full Use	*Description of Change:* Outputs *Endpoints:* Yes *Disaggregated Endpoints:* Yes *Narrow Change Description:* Yes	*Payment Period*: Annual *Change Frame:* Improvement *Published:* Yes *Year of Study:* 2003	*Income:* $50,948 *Sample Population:* Residents & General Public in Area *Location:* MidAtlantic	$51 - $72
✓	(Eisen-Hecht & Kramer, 2002)	2	Contingent Valuation Method	DC-Single Referendum	Usable to Full Use	*Description of Change:* Inputs *Endpoints:* No *Disaggregated Endpoints:* No *Narrow Change Description:* No	*Payment Period*: Annual *Change Frame:* Prevent Deterioration *Published:* Yes *Year of Study:* 1998	*Income:* $74,172 *Sample Population:* Residents *Location:* Southeast	$186 - $259

Table 1 continued

In Final Dataset	Study Citation	# of Obs	Valuation Method	Elicitation Method	Extent of Change	Quality Change Factors	Other Study Characteristics	Sample Characteristics	Adj. WTP Values**
✓	(Farber & Griner, 2000)	3	Conjoint Analysis	Sequential Binary Choice	Unusable to Full Use & Unusable to Usable	*Description of Change:* Outputs *Endpoints:* Yes *Disaggregated Endpoints:* No *Narrow Change Description:* No	*Payment Period*: Annual *Change Frame:* Improvement *Published:* Yes *Year of Study:* 1996	*Income:* $52,083 *Sample Population:* Residents *Location:* MidAtlantic	$37- $105
✓	(Flores & Strong, 2008)	1	Contingent Valuation Method	DC-Multiple Bounded Referendum	Unusable to Full Use	*Description of Change:* Outputs *Endpoints:* Yes *Disaggregated Endpoints:* No *Narrow Change Description:* No	*Payment Period*: Lump Sum *Change Frame:* Improvement *Published:* Yes *Year of Study:* 2004	*Income:* $80,831 *Sample Population:* Residents *Location:* West	$64

Table 1 continued

In Final Dataset	Study Citation	# of Obs	Valuation Method	Elicitation Method	Extent of Change	Quality Change Factors	Other Study Characteristics	Sample Characteristics	Adj. WTP Values**
✓	(Heberling, 2000)	3	Choice Experiment	Sequential Binary Choice	Unusable to Full Use	*Description of Change:* Outputs *Endpoints:* Yes *Disaggregated Endpoints:* Yes *Narrow Change Description:* Yes	*Payment Period*: Annual *Change Frame:* Improvement *Published:* No *Year of Study:* 2000	*Income:* $44,301 – $44,667 *Sample Population:* Residents *Location:* MidAtlantic	$316 - $363
✓	(Hite, 2002)	2	Contingent Valuation Method	DC-Single Referendum	Usable to Full Use	*Description of Change:* Inputs *Endpoints:* No *Disaggregated Endpoints:* No *Narrow Change Description:* No	*Payment Period*: Lump Sum *Change Frame:* Prevent Deterioration *Published:* Yes *Year of Study:* 1999	*Income:* $56,805 *Sample Population*: General Public in Area *Location:* Southeast	$61 - $65

Table 1 continued

In Final Dataset	Study Citation	# of Obs	Valuation Method	Elicitation Method	Extent of Change	Quality Change Factors	Other Study Characteristics	Sample Characteristics	Adj. WTP Values**
✓	(Holmes, Bergstrom, Huszar, Kask, & Orr, 2004)	2	Contingent Valuation Method	DC- Double Bounded Referendum	Unusable to Full Use & Unusable to Usable	*Description of Change:* Outputs *Endpoints:* Yes *Disaggregated Endpoints:* Yes *Narrow Change Description:* Yes	*Payment Period*: Annual *Change Frame:* Improvement *Published:* Yes *Year of Study:* 2000	*Income:* $56,962 *Sample Population*: Residents *Location:* Southeast	$7 - $68
✓	(Huang, Haab, & Whitehead, 1997)	1	Contingent Valuation Method	DC- Double Bounded Referendum	Unusable to Full Use	*Description of Change:* Outputs *Endpoints:* Yes *Disaggregated Endpoints:* No *Narrow Change Description:* No	*Payment Period*: Annual *Change Frame:* Improvement *Published:* Yes *Year of Study:* 1995	*Income:* $45,445 *Sample Population*: Residents *Location:* Southeast	$269

Table 1 continued

In Final Dataset	Study Citation	# of Obs	Valuation Method	Elicitation Method	Extent of Change	Quality Change Factors	Other Study Characteristics	Sample Characteristics	Adj. WTP Values**
✓	(Hushak & Bielen, 1999)	2	Contingent Valuation Method	Payment Card	Unusable to Full Use	*Description of Change:* Inputs *Endpoints:* No *Disaggregated Endpoints:* No *Narrow Change Description:* No	*Payment Period*: Annual *Change Frame:* Improvement *Published:* No *Year of Study:* 1998	*Income:* $88,715 *Sample Population*: Users *Location:* MidWest	$29 - $51
✓	(Kaoru, 1993)	1	Contingent Valuation Method	Open Ended	Unusable to Full Use	*Description of Change:* Outputs *Endpoints:* Yes *Disaggregated Endpoints:* No *Narrow Change Description:* Yes	*Payment Period*: Annual *Change Frame:* Improvement *Published:* Yes *Year of Study:* 1989	*Income:* $152,808 *Sample Population*: Residents *Location:* Northeast	$230

Table 1 continued

In Final Dataset	Study Citation	# of Obs	Valuation Method	Elicitation Method	Extent of Change	Quality Change Factors	Other Study Characteristics	Sample Characteristics	Adj. WTP Values**
✓	(Lichtkoppler & Blaine, 1999)	1	Contingent Valuation Method	DC- Double Bounded Referendum	Unusable to Usable	*Description of Change:* Inputs *Endpoints:* No *Disaggregated Endpoints:* No *Narrow Change Description:* No	*Payment Period*: Annual *Change Frame:* Improvement *Published:* Yes *Year of Study:* 1997	*Income:* $43,910 *Sample Population*: General Public in Area *Location:* MidWest	$44
✓	(Lindsey, 1994)	1	Contingent Valuation Method	Payment Card Referendum	Usable to Full Use	*Description of Change:* Inputs *Endpoints:* No *Disaggregated Endpoints:* No *Narrow Change Description:* No	*Payment Period*: Annual *Change Frame:* Improvement *Published:* Yes *Year of Study:* 1989	*Income:* $64,183 *Sample Population*: Residents *Location:* MidAtlantic	$74
✓	(Lipton, 2004)	1	Contingent Valuation Method	Open Ended	Usable to Full Use	*Description of Change:* Outputs *Endpoints:* Yes *Disaggregated Endpoints:* No *Narrow Change Description:* Yes	*Payment Period*: Annual *Change Frame:* Improvement *Published:* Yes *Year of Study:*2001	*Income:* $70,936 *Sample Population*: Users *Location:* MidAtlantic	$67

Table 1 continued

In Final Dataset	Study Citation	# of Obs	Valuation Method	Elicitation Method	Extent of Change	Quality Change Factors	Other Study Characteristics	Sample Characteristics	Adj. WTP Values**
✓	(Loomis, Strange, Fausch, & Covich, 2000)	1	Contingent Valuation Method	DC-Single Referendum	Usable to Full Use	*Description of Change:* Outputs *Endpoints:* Yes *Disaggregated Endpoints:* Yes *Narrow Change Description:* Yes	*Payment Period*: Annual *Change Frame:* Improvement *Published:* Yes *Year of Study:* 1998	*Income:* $52,807 *Sample Population*: Residents *Location:* West	$337
×	(Lyke, 1993)	0	Contingent Valuation Method	DC-Single	Usable to Full Use	*Description of Change:* Outputs *Endpoints:* Yes *Disaggregated Endpoints:* No *Narrow Change Description:* No	*Payment Period*: Annual *Change Frame:* Improvement *Published:* No *Year of Study:* 1990	*Income:* $51,270 *Sample Population*: Users *Location*: MidWest	$63

Table 1 continued

In Final Dataset	Study Citation	# of Obs	Valuation Method	Elicitation Method	Extent of Change	Quality Change Factors	Other Study Characteristics	Sample Characteristics	Adj. WTP Values**
✓	(Magat, Huber, Viscusi, & Bell, 2000)	7	Iterative Choices	Sequential Binary Choice Referendum	Usable to Full Use	*Description of Change:* Outputs *Endpoints:* Yes *Disaggregated Endpoints:* Yes *Narrow Change Description:* Yes & No	*Payment Period*: Annual *Change Frame:* Improvement *Published:* Yes *Year of Study:* 1997	*Income:* $54,891 *Sample Population*: Residents & General Public in Area *Location*: National	$124 - $659+
✓	(Moore, Provencher, & Bishop, 2011)	7	Contingent Valuation Method	DC-Single Referendum	Unusable to Usable & Usable to Full Use	*Description of Change:* Outputs *Endpoints:* Yes *Disaggregated Endpoints:* No *Narrow Change Description:* No	*Payment Period*: Annual *Change Frame:* Improvement *Published:* Yes *Year of Study:* 2005	*Income:* $60,826 - $83,147 *Sample Population*: Residents and General Public in Area *Location*: MidWest	$10 - $901+

Table 1 continued

In Final Dataset	Study Citation	# of Obs	Valuation Method	Elicitation Method	Extent of Change	Quality Change Factors	Other Study Characteristics	Sample Characteristics	Adj. WTP Values**
✓	(Randall, DeZoysa, & Yu, 2001)	2	Contingent Valuation Method	DC- Double Bounded Referendum	Unusable to Full Use	*Description of Change:* Outputs *Endpoints:* Yes *Disaggregated Endpoints:* Yes *Narrow Change Description:* Yes	*Payment Period*: Lump Sum *Change Frame:* Prevent Deterioration *Published:* Yes *Year of Study:* 1994	*Income:* $63,316 *Sample Population*: General Public in Area *Location*: MidWest	$115 - $175
✓	(Shresta & Alavalapati, 2004)	2	Choice Experiment	Sequential Binary Choice Referendum	Usable to Full Use	*Description of Change:* Inputs *Endpoints:* No *Disaggregated Endpoints:* No *Narrow Change Description:* Yes	*Payment Period*: Annual *Change Frame:* Improvement *Published:* Yes *Year of Study:* 2002	*Income:* $72,571 *Sample Population*: General Public in Area *Location*: Southeast	$37 - $86

Table 1 continued

In Final Dataset	Study Citation	# of Obs	Valuation Method	Elicitation Method	Extent of Change	Quality Change Factors	Other Study Characteristics	Sample Characteristics	Adj. WTP Values**
✓	(Stumborg, Baerenklau, & Bishop, 2001)	2	Contingent Valuation Method	DC-Multiple Bounded Referendum	Usable to Full Use	*Description of Change:* Outputs *Endpoints:* Yes *Disaggregated Endpoints:* No *Narrow Change Description:* No	*Payment Period*: Annual *Change Frame:* Improvement *Published:* Yes *Year of Study:* 2001	*Income:* $76,847 *Sample Population*: Residents *Location*: MidWest	$70 - $107
✓	(Viscusi, Huber, & Bell, 2008)	2	Conjoint Analysis	Sequential Binary Choice	Unusable to Full Use	*Description of Change:* Outputs *Endpoints:* Yes *Disaggregated Endpoints:* Yes *Narrow Change Description:* Yes	*Payment Period*: Annual *Change Frame:* Improvement *Published:* Yes *Year of Study:* 2004	*Income:* $51,598 *Sample Population*: General Public *Location*: National	$31 - $38

Table 1 continued

In Final Dataset	Study Citation	# of Obs	Valuation Method	Elicitation Method	Extent of Change	Quality Change Factors	Other Study Characteristics	Sample Characteristics	Adj. WTP Values**
✓	(Wey, 1990)	1	Contingent Valuation Method	Payment Card	Unusable to Full Use	*Description of Change:* Outputs *Endpoints:* Yes *Disaggregated Endpoints:* No *Narrow Change Description:* Yes	*Payment Period*: Annual *Change Frame:* Improvement *Published:* No *Year of Study:* 1989	*Income:* $94,361 *Sample Population*: Residents *Location*: Northeast	$67
✓	(Whitehead & Groothuis, 1992)	2	Contingent Valuation Method	Open Ended	Usable to Full Use	*Description of Change:* Outputs *Endpoints:* Yes *Disaggregated Endpoints:* Yes *Narrow Change Description:* Yes	*Payment Period*: Annual *Change Frame:* Improvement *Published:* Yes *Year of Study:* 1991	*Income:* $56,944 *Sample Population*: Users & General Public in Area *Location*: Southeast	$35 - $58

Table 1 continued

In Final Dataset	Study Citation	# of Obs	Valuation Method	Elicitation Method	Extent of Change	Quality Change Factors	Other Study Characteristics	Sample Characteristics	Adj. WTP Values**
✓	(Whitehead, Blomquist, Hoban, & Clifford, 1995)	3	Contingent Valuation Method	DC- Double Bounded Referendum	Usable to Full Use	*Description of Change:* Outputs *Endpoints:* Yes *Disaggregated Endpoints:* No *Narrow Change Description:* No	*Payment Period*: Annual *Change Frame:* Prevent Deterioration *Published:* Yes *Year of Study:* 1990	*Income:* $52,853 - $67,321 *Sample Population*: Users & General Public in Area *Location*: MidAtlantic	$83- $119
×	(Whitehead, 2006)	0	Contingent Valuation Method	DC- Double Bounded	Usable to Full Use & Unusable to Full Use	*Description of Change:* Outputs *Endpoints:* Yes *Disaggregated Endpoints:* No *Narrow Change Description:* No	*Payment Period*: Annual *Change Frame:* Improvement *Published:* Yes *Year of Study:* 1998	*Income:* $95,307 *Sample Population*: Residents *Location*: Southeast	$1- $385

Table 1 continued

In Final Dataset	Study Citation	# of Obs	Valuation Method	Elicitation Method	Extent of Change	Quality Change Factors	Other Study Characteristics	Sample Characteristics	Adj. WTP Values**
✓	(Whittington, Cassidy, Amaral, McClelland, Wang, & Poulos, 1994)	2	Contingent Valuation Method	DC- Double Bounded Referendum	Usable to Full Use	*Description of Change:* Outputs *Endpoints:* Yes *Disaggregated Endpoints:* No *Narrow Change Description:* No	*Payment Period*: Annual *Change Frame:* Improvement *Published:* No *Year of Study:* 1993	*Income:* \$82,680 - \$83,100 *Sample Population*: General Public in Area *Location*: Southeast	\$222 - \$370

* Table for Full Data Set Available in Appendix A.
** Income and WTP values adjusted to 2010 dollars.
\+ Two WTP values identified as outliers and removed from data set.

4.2 Meta-Analysis Model

The stylized meta-regression model for this analysis is

$$WTP_{ij} = ß_0 + ß_1 Quality_{ij} + ß_2 Study_{ij} + ß_3 Sample_{ij} + \mu \qquad \text{(Equation 2)}$$

The subscript *ij* stands for estimate *i* from study *j*. In this model, WTP is mean household willingness to pay, ß is a vector of coefficients on variables in each category, *Quality* represents a vector of quality variables capturing quality change and presentation, *Study* is a vector of study characteristics variables, and *Sample* is a vector of socio-demographic variables representing each study sample.

4.3 Variable Definition and Coding

With the exception of the quality variables, the coding for these variables is straightforward. Binary variables are used in most cases as a simple yes or no for the existence of a given characteristic in the primary study. The definition and means of coding each variable used in the full meta-regression model are detailed below. The variables created for the stylized model and their summary statistics are listed in Table 2.

Table 2: Variable Definitions and Summary Statistics for Reduced Data Set*

Variable	Description	Units of Measurement	Mean	Standard Deviation	Frequency	%
Household WTP**	Annual household willingness to pay for water quality change. WTP for all studies converted into 2010 dollars using University of Oregon conversion table	Dollars (Range: 7 to 581)	145.81	127.83	-	-
Ln of Household WTP	Natural log of annual household willingness to pay for water quality change. WTP for all studies converted into 2010 dollars using University of Oregon conversion table	Natural log of dollars (Range: 1.95 to 6.36)	4.61	0.92	-	-
Extent of Change - Unusable to Usable	Binary variable indicating the extent of quality change in the environmental commodity is from unusable to usable. The reference is a change from usable to higher levels of use. These categories are based on scenario descriptions in the surveys.	Binary (Range: 0 or 1)	-	-	13	19.12
Extent of Change - Unusable to Full Use	Binary variable indicating the extent of quality change in the environmental commodity is from unusable to full use. The reference is a change from usable to higher levels of use. These categories are based on scenario descriptions in the surveys.	Binary (Range: 0 or 1)	-	-	24	35.29
Extent of Change - Usable to Full Use	The reference for the extent of change variables includes usable to medium use and usable to full use.	Binary (Range: 0 or 1)	-	-	31	45.59
Description of Change	Binary variable indicating the change is described in terms of a reduction in inputs (pollutants, nutrients, and sediment). The reference is for change described in terms of outputs or services of the environmental commodity.	Binary (Range: 0 or 1)	-	-	15	22.06
Endpoints	Binary variable indicating the change descriptions identify endpoints or easily understandable outcomes of the change. The reference is for change described generally or in technical language without specific information on environmental commodities that directly affect human use.	Binary (Range: 0 or 1)	-	-	58	85.29

Table 2 continued

Variable	Description	Units of Measurement	Mean	Standard Deviation	Frequency	%
Disaggregated Endpoints	Binary variable indicating the change descriptions identify multiple endpoints and services that will be affected by the change. The reference is for change described using a single endpoint as an indicator of overall change.	Binary (Range: 0 or 1)	-	-	30	44.12
Open vs. Closed Valuation	Binary variable indicating the change valued is described with quantitative measures or linkages to a specific and discrete ecological service. The reference is for change descriptions that allow respondents to value broader ecological services.	Binary (Range: 0 or 1)	-	-	30	44.12
Dichotomous Choice-Multi-Bounded	Binary variable indicating the survey used a double-bounded or multiple-bounded dichotomous choice elicitation question set. The reference is an elicitation method using a single dichotomous choice question set.	Binary (Range: 0 or 1)	-	-	14	20.59
Sequential Binary Choice	Binary variable indicating the survey used a sequential binary choice set with options of price and environmental commodities to elicit stated preference. The reference is an elicitation method using a single dichotomous choice question set.	Binary (Range: 0 or 1)	-	-	25	36.76
SQBC – No Referendum	Binary variable acting as a covariate to control for the sequential binary choice observations not using a referendum decision rule.	Binary (Range: 0 or 1)	-	-	11	16.18
Open Ended	Binary variable indicating the survey used an open-ended question to elicit stated preference. The reference is an elicitation method using a single dichotomous choice question set.	Binary (Range: 0 or 1)	-	-	4	5.88
Payment Card	Binary variable indicating the survey used a payment card method to elicit stated preference. The reference is an elicitation method using a single dichotomous choice question set.	Binary (Range: 0 or 1)	-	-	7	10.29
PC – No Referendum	Binary variable acting as a covariate to control for the payment card observations not using a referendum decision rule.	Binary (Range: 0 or 1)	-	-	3	4.41

Table 2 continued

Variable	Description	Units of Measurement	Mean	Standard Deviation	Frequency	%
Dichotomous Choice-Single	Dichotomous choice – single binary choice is the reference for binary variables on elicitation method for stated preference survey instruments.	Binary (Range: 0 or 1)	-	-	18	26.47
Published	Binary variable indicating the survey results were published in a journal or book, indicating a level of peer review or other outside review. The reference is for grey literature including academic and government reports, conference papers, and dissertations.	Binary (Range: 0 or 1)	-	-	46	67.65
Year Index	Year in which the study was conducted, converted to an index by subtracting 1989.	Year Index (Range: 0 to 16)	9.76	4.58	-	-
Framing	Binary variable indicating the quality change to be valued is an improvement over the current state. The reference is a study that prevents further deterioration.	Binary (Range: 0 or 1)	-	-	57	83.82
Payment Timing	Binary variable indicating the willingness to pay value is a lump sum one-time payment. The reference is for annual payment for a period of 3 to 25 years or a period that is indefinite.	Binary (Range: 0 or 1)	-	-	6	8.82
Income**	The mean or median income of respondents, either as reported in the original study or as imputed from US census medians for the area surveyed.	Dollars (Range: 43,301 to 152,808)	62,316	16,547	-	-
Sample-Users	Binary variable indicating the population surveyed were users of the water body, both local and out of area. The reference is for the general public in the area (county, watershed, or state).	Binary (Range: 0 or 1)	-	-	8	11.76
Sample-Residents	Binary variable indicating the population surveyed were residents or landowners in close proximity or familiar with the water body being valued. The reference is for the general public in the area (county, watershed, or state).	Binary (Range: 0 or 1)	-	-	19	27.94
Sample-General Public in Area	The reference for the binary sample variables. General public in the area is a random sample of households within approximately 200 miles of the water source but with no information on use or interest.	Binary (Range: 0 or 1)	-	-	41	60.29

Table 2 continued

Variable	Description	Units of Measurement	Mean	Standard Deviation	Frequency	%
Location - Northeast or National	Binary variable indicating the location of the water body improvement being valued is in the Northeast. The locations are based on the USDA regional map. The reference is for the Southeast region.	Binary (Range: 0 or 1)	-	-	2	2.94
Location - MidAtlantic	Binary variable indicating the location of the water body improvement being valued is in the Mid- Atlantic region. The locations are based on the USDA regional map. The reference is for the Southeast region.	Binary (Range: 0 or 1)	-	-	17	25
Location - MidWest	Binary variable indicating the location of the water body improvement being valued is in the Mid-West region. The locations are based on the USDA regional map. The reference is for the Southeast region.	Binary (Range: 0 or 1)	-	-	20	29.41
Location - West	Binary variable indicating the location of the water body improvement being valued is in the Mountain Pacific or Western region. The locations are based on the USDA regional map. The reference is for the Southeast region.	Binary (Range: 0 or 1)	-	-	7	10.29
Location - Southeast	The reference for the binary location variables, which is the Southeast region. The locations are based on the USDA regional map.	Binary (Range: 0 or 1)	-	-	22	32.35

* Table of Summary Statistics for Full Data Set available in Appendix A.
** Income and WTP values adjusted to 2010 dollars.

4.3.1 Quality Characteristics

The thirty primary studies analyzed in this paper, while all focusing on the quality of surface water in the U.S., use various methods to capture the aspects of the water quality services and changes. The variables for quality characteristics were systematically and reliably coded into binary variables according to the criteria in Appendix B. The coding decision for the extent of change variable group was more subjective than the other quality description variables because of the wealth of description and the lack of comparable characteristics on which to base a coding decision. The process of coding was similar to quantitative content analysis and the tools commonly used for content analysis were applied to the coding of this variable.[11]

Extent of change - This is the extent of quality change from the degraded state to the improved state presented to survey respondents.[12] The measure of change is highly variable from study to study, if a measure is used at all. Identifying and categorizing the extent of change across studies is critical for two reasons. First, variation of WTP in response to the extent of change is an indicator of validity based on the NOAA scope test, with the expectation that a larger change in quality will result in a larger amount of WTP (Arrow et al., 1993). Second, assuming WTP does vary with the extent of change all effort must be made to control for these effects in order to isolate the quality description effects on WTP raised by Boyd and Krupnick (2009) and Johnston et al. (2012). Obtaining a positive scope test (a positive, statistically significant coefficient on the extent of change variable) is a strong indication that the changes predicted by economic theory have been controlled.

---

[11] "Content analysis is a summarizing, quantitative analysis of messages that relies on the scientific method, including attention to objectivity/intersubjectivity, a priori design, reliability, validity, generalizability, replicability, and hypothesis testing. It is not limited as to the types of messages that may be analyzed, nor as to the types of variables that might be measured" (Neuendorf, 2011, p. 277).

[12] A small number of observations frame the change in terms of preventing degradation, while the majority values an improvement in water quality that has already been degraded. These are not distinguished in the extent of change variable; instead, a separate binary variable is inserted to control for potential differences in WTP based on the framing of the water quality change.

To classify the extent of quality change, this analysis draws on the conclusions of Turner et al. (2003), which hold that the marginal value of a resource improvement is expected to decline as the resource services increase and vice versa over an ecologically determined non-critical range. Farber, Costanza, and Wilson (2002) provide an example of the critical threshold and the critical range relevant to water quality. The gradual degradation in the quality of water (in ecological measures) due to excess nutrients may result in a non-linear change in the economic value when the water body is closed to users – whether recreational users or industrial users. This puts the critical range for economists and respondents at some point between usable (light to moderate pollution) and unusable (moderate to severe pollution) with the critical threshold located at the point at which water quality worsens or improves to prohibit or allow consumptive use. In order to capture the two potential changes over the critical range, the extent of change is divided into improvement from unusable to full use and improvement from unusable to usable. These two variables are binary and the reference for both extent of change variables is the range from usable to full use.

The expected sign on both variables is positive because the size of WTP for improvement from unusable to full use and improvement from unusable to usable is expected to be larger than changes in the non-critical range from usable to full use. This is in line with the results of Johnston et al. (2003), which find that WTP declines as the baseline level of quality increases across studies and affirms the diminishing marginal returns to scale of quality change as the starting point moves to a higher level of quality.

The identification of the three nominal categories – usable to full use, unusable to usable, and unusable to full use – was based on multiple reviews of the before and after change data available in each primary study. For this analysis, the messages on the extent of change are gathered from the background section and valuation question in the primary survey, when possible. In cases where the survey was not available, details on how the change was presented to respondents is gathered from the journal article or report on the study. The text describing the change scenario was presented in a spreadsheet for review, identification of nominal categories, and coding. Due to the subjective nature of this coding process, a panel of two independent coders was selected. In keeping with the recommendations of Neuendorf (2011), the two coders did not possess expert knowledge on water quality measures nor were they informed of a true or "right" coding decision. Instead, the purpose of the coding was explained as grouping the studies according to the characteristics common to each nominal category provided on a separate sheet. Immediately after the first round of coding, all observations on which the coders did not agree were discussed and consensus reached on the preferred coding.[13]

The intercoder reliability, based on the initial codes returned, was found using Cohen's kappa. Cohen's kappa is a widely used indicator of intercoder agreement for nominal data that controls for chance agreement. Values for this indicator above 0.60 are considered acceptable levels for intercoder agreement (Neuendorf, 2011; Lombard, Snyder-Duch, & Bracken, 2002). The Cohen's kappa value for intercoder reliability of this variable is 0.692, above the threshold for acceptability.[14]

Description of Quality Change – This variable captures how the water quality change is described. Van Houtven et al. (2007) identify two paradigms for water quality change: 1) estimating WTP for reductions in one or more pollutants and 2) estimating WTP for enhancements to ecosystem services. Surveys describing the change in terms of reducing pollutants are categorized as input changes and surveys describing the change in terms of ecosystem services are categorized as output changes. The description of the change is of interest because the input description puts the focus on the cause of the degradation rather than the effect of its removal, potentially allowing respondents to independently determine the commodity to be valued. This commodity may be services highly valued by the respondent or other services with value and effects unknown to the respondent. This precisely mirrors the

---

[13] The coding criteria for the extent of change variables are laid out in Table B1 in Appendix B.

[14] An online intercoder reliability calculator operated by Deen Freelon from American University was used. The calculator is available at freelon.org. Full results for the intercoder reliability scores are available in Table B2 in Appendix B.

problematic utility function presented by Johnston et al. (2012) in which respondents are asked to value the input X rather than the environmental services E of which they are familiar. When available, the specific valuation question is the source of information for this variable. For the studies in which the valuation question is not available, evidence from the description of the study and the change scenario is adequate to determine the focus of the study.[15]

This is a binary variable and the reference is studies described in terms of outputs. The sign on this variable is not known because of the competing directions: higher WTP due to the requirement that respondents make their own assumptions and bring expansive priors to the relationship between E and X, a reduction in WTP due to the absence of services or even resources to be valued, or an insignificant variable due to variation in both directions across the studies.

Endpoints – Based on the Boyd and Krupnick (2009) report, studies are separated according to those that use endpoints to describe the proposed program or policy change and those that do not. The definition of ecological endpoints is "meaningful biophysical outputs that do not require expert knowledge of biophysical production functions in order to determine their economic value." Boyd and Krupnick (2009) provide examples of endpoints including fish and bird abundance, water clarity, and odor. This is in contrast to technical measures, requiring expert knowledge on how dissolved oxygen affects water clarity or fish and wildlife abundance. This is a binary variable, in which the reference is for studies that do not use endpoints. The expected sign for this variable is positive because endpoints have the effect of translating change into understandable services for which a person can express a preference and sufficiently determine the effect on their utility (Boyd & Krupnick, 2009).

There is a potential for correlation between *endpoints* and *description of quality change* and the correlation table does indicate a high degree of correlation between the two variables (Appendix C). Both variables are kept because they are measuring two

---

[15] Two examples of each type of valuation question are available in Table B3 in Appendix B.

distinct aspects of the quality characteristics. This is because not all studies valuing an output change describe the output in terms of endpoints and some studies valuing an input change also include related services with specific endpoint descriptions. By including both of these variables, *endpoints* and *description of quality change,* the goal is to break apart specific quality information effects that may well push WTP in different directions. The two variables are not used in the same model.

Disaggregated Endpoints – In the critiques of both Boyd and Krupnick (2009) and Johnston et al. (2012), the remedy to invalid valuation results is the unbundling and clear listing of environmental characteristics. Studies describing and valuing attributes whether using CV, CE, or CJ are much closer to this recommendation than those valuing a general description of an environmental commodity. The distinction, however, is not as simple as the use or not of attributes. Indeed, the attributes used in CE and CJ may be as aggregated as descriptions used in CV.

To avoid a complicated process of sorting through survey text on endpoints, this analysis identifies disaggregated endpoints when a study presents water quality in regard to two or more ecosystem service endpoints. The purpose of requiring two detailed endpoints is to avoid those studies that use one endpoint, such as water clarity or fish abundance, as the only indicator of water quality improvement. By definition, this is an aggregated endpoint.[16]

Studies are divided according to those using two or more endpoints and those that do not. This is a binary variable and the reference is studies not using disaggregated endpoints. There is not a clear expectation on the sign of this variable. Boyd and Krupnick (2009) did not identify an optimal level of disaggregation, noting instead that unbundled endpoints open the way for expansive priors while endpoints that are too disaggregated or specific may decrease interest or attention. Separating the disaggregated endpoints according to interest is outside the scope of this analysis;

[16] The criteria for distinguishing between bundled and disaggregated endpoints are listed in Table B4 in Appendix B.

therefore, this analysis only seeks to determine whether or not disaggregated endpoints have a systematic effect on WTP.

Open versus Closed Valuation – An alternative to the focus on disaggregated endpoints is a deeper investigation into the valuation scenario's linkage to a specific quality attribute versus a scenario with an open description that, in effect, is valuing a general water quality process improvement. This variable identifies studies that present the environmental change in an open or vague manner and those that present the change in a narrow or precise manner. The inclusion of this variable is based on the findings of Johnston et al. (2012) and Czajkowski, Buszko-Briggs, and Hanley (2009) in which the value of broad natural processes is larger than the value for species-specific changes. The expectation is that this variable will have a negative sign, indicating that studies with closed valuation scenarios have lower WTP values than studies with open valuation scenarios.

The information used for determining this variable is from the background and valuation scenario in the survey or, if unavailable, from the primary study's description of the environmental change.[17] The reference for this binary variable is studies with an open valuation process.

This variable differs from both the *description of quality change* and *disaggregated endpoints* in a few important ways. The *description of quality change* and *disaggregated endpoints* variables do not distinguish between studies that value specific quality changes and those that value an overall environmental condition. For example, *description of quality change* distinguishes only between the initial framing of the change – input versus output – whereas *open versus closed valuation* examines the overall scenario description and the linkage between the description and the valuation question. The *disaggregated endpoints* variable is far closer to the *open versus closed valuation* variable and the correlation table (Appendix C) shows a high enough level of correlation to preclude using the variables in the same regression

[17] The criteria used to distinguish between open versus closed studies are available in Table B5 in Appendix B.

model. However, the variables are not measuring the same thing. The *open versus closed valuation* variable excludes many of the observations included in the *disaggregated endpoints* variable because disaggregated endpoints are not always specific, lacking any quantification and often including many potential changes.

4.3.2 Study Characteristics

The study characteristics considered in this meta-analysis are the elicitation method, referendum-decision rule, valuation framing, timing of WTP payment, and two study quality indicators (published and year index).

Elicitation Method – The effect elicitation methods have on the WTP values has been heavily researched and, for the most commonly used types of elicitation, a clear set of expectations developed. The elicitation methods employed by studies in this meta-analysis fall into five categories: dichotomous choice (DC), multiple bounded DC, sequential binary choice, open-ended, and payment card.[18] DC or single binary choice questions are the most commonly used elicitation method due to its incentive compatibility, which was further validated by the NOAA panel recommendations (Carson & Groves, 2007). DC is expected to yield higher values than payment card and open-ended elicitation questions, due to yea-saying and anchoring (Boyle, 2003; Carson & Groves, 2007; Champ & Bishop, 2006).[19] Higher

---

[18] For the purposes of this meta-analysis, the elicitation methods are defined by the following primary characteristics. Open-Ended: The question asks respondents how much they are willing to pay and leaves a blank for the respondent to fill in a number. Payment Card: Lists prices (often 5 to 10) and asks respondent to circle the price closest to their willingness to pay. Dichotomous Choice: Single binary choice approach, asks if respondent is willing to pay more for the good than the single amount presented. This may be presented as a referendum (would the respondent vote for a policy if the cost was X). Multiple bounded binary choice: This includes multiple bounded discrete choice, in which single binary choice questions are repeatedly asked on a list of prices (laid out as in a payment card system) and a yes/no response is required on each amount listed. A related approach is the double bounded binary choice, in which two binary choice questions are asked with the cost raised or lowered once based on the previous response. A variation on this method is an iterative binary choice that updates a series of questions with different prices to identify upper and lower bounds of willingness to pay. Sequential Binary Choice: Lists multiple choice sets with variations in attribute levels and price (Boyle, 2003; Brown, 2003).

[19] Though discrete choice elicitation formats consistently estimate higher WTP values than continuous format (open-ended and payment card), the continuous format is never expected to elicit true WTP.

values compared to payment card and open-ended elicitation questions are expected to carry through in all discrete choice elicitation methods, including multiple bounded DC (Welsh & Poe, 1998).

The relationship between discrete choice elicitations – single DC, double or multiple bounded DC, and sequential binary choice – has not been widely established (Bateman & Jones, 2003; Brown, 2003; Van Houtven et al., 2007). Though preliminary, the empirical results do show significant and substantial directional differences between the discrete choice results and in order to prevent omitted variable bias in the meta-regression the three discrete choice methods will be separated. McNair et al. (2011) found repeated binary choice formats (those used in CJ and CE) resulted in smaller WTP estimates. This result is similar to Racevskis and Lupi's (2008) findings on multiple bounded CV estimates and Carson and Groves's (2007) review, which reports the responses from a single binary choice question are higher than those from a double bounded question. The reasons for such findings are numerous, ranging from strategic misrepresentation (rejecting a higher price if a low price is offered initially) and respondent revision of value based on a weighted average of values presented (McNair et al., 2011). An additional reason to separate the discrete choice elicitation methods is to identify all CE and CJ studies based on the primary distinguishing factor from CV, the sequential binary choice elicitation method.

The elicitation method is divided into four binary variables: DC- multiple bounded, sequential binary choice, open-ended, and payment card. The reference for these variables is single DC.

Decision Rule - Referendum – The decision rule or the point at which the respondent's commitment will result in the provision of the hypothetical service is considered a critical factor in the incentive compatibility of the elicitation mechanism and responsible for respondents' accurate representation of their WTP (Boyle, 2003). The most common decision rule in stated preference studies is a referendum format,

---

Instead, the mechanism is incentive incompatible because it incentivizes the respondent to respond strategically rather than truthfully (Carson & Groves, 2007).

which couches the elicitation question in a vote on which a respondent may only answer yes or no to both the referendum and the price. The 1993 NOAA Panel recommended the use of single discrete binary choice (dichotomous choice) elicitation questions with a referendum setting, aiming to create an incentive compatible scenario for the respondent (Carson & Groves, 2007).

Though the importance of referenda decision rule in promoting incentive compatibility is widely accepted, the literature does not provide much information on the direction of willingness to pay estimates when a referendum is used. This is in no small part due to the importance the context, payment vehicle, and elicitation method have in facilitating the incentive compatibility of the referenda decision rule (Carson & Groves, 2007). As a result, the expected sign on the referendum variable is not known. Rather, the reason to include this variable is to control for the effect a referendum decision rule will have on variations in WTP.[20]

The studies used in this meta-analysis exhibit different levels of research, review, and overall quality. To capture heterogeneity caused by quality differences in the studies, this meta-analysis uses two binary variables, *published* and *year index*, to capture potential and otherwise unexplained variation in WTP estimates.

Published – This is a binary variable identifying published studies that have undergone a peer-review process. The reference variable is for gray literature in this case dissertations, academic research or conference papers, and government reports. There is not an expectation on the sign of this variable; indeed, the purpose for including it is to control for potential heterogeneity.

Year Index – This is a continuous variable from 0 to 15 indicating the number of years from the reference year in which the study was conducted. The reference year is 1989, which is the earliest date a study in this meta-analysis was conducted. A number of meta-analyses in environmental economics use a similar variable with the

[20] The studies within the dataset for this meta-analysis complicate the process of controlling for referendum use. A full discussion of the steps taken to control for referendum use is found in the preliminary analysis section; however, a separate binary variable for referendum use was not used.

expectation that later studies will be of a higher quality (Nelson & Kennedy, 2009). This variable is included to control for potential quality changes over the past two decades but the sign of the variable is uncertain.

Framing – Framing effects in stated preference surveys have been well documented. A basic effect found across stated preference methods is that respondents value losses more highly than a commensurate gain when compared to the status quo (Knetsch, 2010; Kragt & Bennett, 2012). This variable distinguishes between studies that present respondents with the choice of paying for an improvement in water quality and those studies that frame the choice as a payment to prevent degradation in water quality. The reference for this binary variable is loss framing or studies valuing the prevention of degraded water quality. The expectation is that the sign on this variable will be negative, indicating respondents value the degradation in water quality more highly than the improvement in water quality.

Payment Timing – The timing of payment refers to the length of time for which a respondent is told the annual WTP amount will be paid. Three categories are used in stated preference surveys: 1) indefinite, 2) short term from 3 to 25 years, and 3) one-time lump sum payment. Other meta-analyses with a variety of payment times in their collected studies use one or two means for controlling for differences. The first is the creation of a binary variable identifying lump sum versus indefinite annual payments; the second is the conversion of annual short-term payments (typically 3 to 5 years) into a single lump sum figure. This meta-analysis uses the first approach rather than the second. The WTP values from studies limiting the time frame for annual WTP payments are not turned into lump sum payments. This is for two reasons. First, Stevens et al. (1997) in studying the difference between periodic and lump sum payments found significantly different discount rates between payments elicited as lump sums and those elicited as periodic payments. Turning a periodic payment (over 3, 5, or 25 years) into a lump sum payment simply by adding the values is not accurate and leads to a questionable comparison with other lump sum amounts and periodic amounts. Second, this analysis expects that all periodic payments whether for a limited or undefined period will be comparable because the undefined period is likely to vary among respondents and studies.

A single binary variable distinguishing between lump sum and periodic payments is used in this meta-analysis. The reference for the variable is studies using periodic payments.

4.3.3 Sample Characteristics

Socio-economic characteristics of each study sample (or subsample) are typically reported in the primary study and are taken at face value when entered into the meta-analysis dataset.

Income – In a theoretical analysis of income and quality change in CV studies, Whitehead (1995) presents comparative statics to show, assuming water quality is a normal good, the marginal cost of utility will be greater with a degraded quality level and the income effect on WTP will be positive. In a meta-analysis on the income effect in stated preference studies, Schlapfer (2006) found minimal income effect and any effect diminished when studies with dichotomous choice (DC) elicitation were used. Most techniques using attribute-based methods (CE and CJ) do not include income in the function used to estimate mean or median WTP (Boyle, 2003). This combined with the findings of Schlapfer (2006) means the expected effect of income in this meta-analysis is uncertain. Income is retained as a variable due to the theoretical expectation of a positive effect on WTP.

Income, either the mean or median of the sample, is reported in 20 of the 30 studies. When income is not reported, the median household income from the closest year and closest geographical location is obtained from census or county/state level sources. The mean or median household income of the sample or of the closest geographical region is converted to 2010 dollars and inserted as a continuous variable in the dataset.

Users, Residents, and General Public – Users and those living in close proximity to the resource are expected to be willing to pay more for improvement than nonusers (Van Houtven et al., 2007). This is complicated, however, by users from outside the area with available substitutes. Another complicating factor is that individuals familiar with the resource may bring their own perceptions about quality and not factor in the information presented in the survey instrument (Magat et al., 2000; Viscusi et al., 2008; Whitehead, 2006).

This variable separates respondents into three categories: users (both local and out-of-area), residents in close proximity to the resource, and the general public typically within 200 miles of the resource. The reference for the binary variable is the general public in the area. The reason to separate residents and users is due to the possibility that residents, though in close proximity to the resource, do not use the resource and their WTP value may be lower than the WTP of users. This is due to a lower or potentially nonexistent use value for the resource on the part of residents.

The signs on the resident and user variables are expected to be positive, but the complications may reduce the significance of the effect depending on variation within and perceptions of the study population.

4.4 Model Specification

4.4.1 Variable Specification

As shown in the stylized regression model (equation 2), the variables used in the meta-analysis fit into three categories: quality, study, and sample characteristics. All study and sample variables (Table 2) are used in the full model. The six quality variables described above are *Extent of Change – Unusable to Usable* (EC-UU), *Extent of Change – Unusable to Full Use* (EC-UNFU), *Description of Change* (D_CHNGE), *Open versus Closed Valuation* (OP_CL), *Endpoints* (Endpts), and *Disaggregated Endpoints* (DG_Endpts). The three quality variable combinations are below.

The quality variables accounting for extent of change are included in all models with the aim of controlling for the change in WTP values predicted by economic theory, also called the scope test. The alternate pairings of the remaining quality variables are shown below.

Model 1: *$Quality_{ij}$ = Extent of Change Variables, Description of Change & Open versus Closed Valuation*

Model 2: *$Quality_{ij}$ = Extent of Change Variables, Endpoints & Open versus Closed Valuation*

Model 3: *$Quality_{ij}$ = Extent of Change Variables, Endpoints & Disaggregated Endpoints*

The variable pairings are chosen according to explanatory value and to avoid placing two variables with a high correlation value in the same model. The three models are run throughout the preliminary and final analyses. Akaike's Information Criterion is used in the final data analysis to compare the three models' relative ability to describe the quality characteristics of importance with a minimum of information loss.

4.4.2 Functional Form

Meta-analysis literature offers mixed practice on functional form. Many meta-analyses employ various functional forms, relying on both theoretical justification and the statistical performance of the different models. Meta-analyses using log-level and log-log functional form are frequently employed in models with critical variables in continuous form, such as water quality change, and is justified by noting a key benefit to the form: WTP approaches zero as the change in water quality approaches zero (Bergstrom & Taylor, 2006; Johnston et al., 2005; Smith & Osborne, 1996; Van Houtven et al., 2007). Meta-analyses using linear functional form are also justified based on statistical performance and typically note the overwhelming number of qualitative dummy variables in the model (Bateman & Jones, 2003; Poe, Boyle, & Bergstrom, 2001; Rosenberger & Loomis, 2000b).

Level-level and log-level functional forms are used throughout the preliminary data analysis to capture variations in the empirical model from full to reduced form.

The decision on the prefered form is made based on statistical performance, using Ramsey's regression specification error test (RESET), in the preliminary analysis.

4.5 Preliminary Analysis – Phase 1 - Data Review

Before proceeding to a full analysis, a two-phase preliminary analysis is used to address variable correction, outliers, heteroskedasticity, and correlation between observations from the same study. The first stage of the preliminary analysis focuses on dataset revisions needed to correct a key variable and address outliers. The resulting revisions reduce the dataset by seven observations from 75 to 68 total observations.

4.5.1 Referendum Decision Rule – Data Set Review

As mentioned earlier, the referendum variable attempts to identify studies using a referendum decision rule in the elicitation question. Though recommended for dichotomous choice questions, a review of the studies in this meta-analysis found a number of studies using referenda formats in single dichotomous choice, double and multi-bounded dichotomous choice, payment card, and sequential binary choice. A separate binary variable marking studies that use a referendum was unsuccessful because of the small number of single dichotomous choice observations not using the referendum decision rule. The referendum variable was highly significant in the initial regression but severely impacted the elicitation method variables because the reference group was virtually eliminated by having a separate referendum variable assumed to be zero. Attempts to insert separate covariates to control for referendum use in the single DC and multiple bounded DC failed (the F statistic was not reported for the model), because of the small number of observations (four for multiple DC and one for single DC) against which the covariates were compared.

In order to retain a control on referenda use, the five observations from the two dichotomous choice (both single and multiple bounded) studies that did not use referenda were dropped from the analysis. For studies using payment card and sequential binary choice elicitation methods, nineteen observations used referenda and fourteen did not. To control for the different referenda usage in these two elicitation formats, covariates for payment card without referenda and sequential binary choice without referenda were included in the Models 1-3c.

4.5.2 Outlier Analysis

Potential outliers in this dataset were identified using a series of graphical and diagnostic tests. The studies and context from which the potential outliers originated were examined to determine whether potential outliers should be removed from the dataset (Bollen & Jackman, 1985). The first graphical tool employed is a simple plot of the distribution of household WTP, used to identify entry errors or observations with exceptionally high or low WTP values. A more refined graph of standardized residuals offers a second graphical look at the WTP values and the ability of the full linear model to predict fitted values. Both graphs identify two observations as potential outliers, one from the Magat et al. (2000) study and one from the Moore et al. (2011) study.

The second set of diagnostic tools is designed to measure both deviant residuals and extreme leverage points, using DFITs[21] and Cook's D_i[22] (Bollen & Jackman, 1985). This goes beyond the standard measures that identify either extreme residuals or outsized leverage. Both measures were applied to the full model and four potential influential outliers identifed based on their values above the conservative cut off for DFITs and Cook's D_i. values. After reviewing the context for the four values, two observations, also identified in the graphical analysis, are removed as outliers for

---

[21] DFITs, an indicator of leverage and high studentized residuals, measures how much an observation influences the regression model as a whole or how much the predicted value changes as a result of including and excluding a particular observation. The conservative value at which an observation should be flagged is $2\sqrt{(k/n)}$, where k is the number of explanatory variables including the intercept and n is the sample size. The less conservative cut off identifying only the most extreme observations is $\sqrt{(k)}$ (Belsley, Kuh, & Welsch, 1980; Bollen & Jackman, 1985).

[22] Cook's D_i is a popular alternative to DFITs. Both DFITs and Cook's D_i typically rank observations in similar order. A difference noted in the literature is that DFITs gives a greater weight to outliers than does Cook's D_i. The conservative cutoff for Cook's D_i is 4/n and the high cut off is 1 (Belsley et al., 1980; Bollen & Jackman, 1985).

reasons below. The listing of the DFITs and Cook's D_i values as well as the graphical representations are presented in Appendix D.

The first deleted outlier is one of eight observations from the Magat et al. (2000) study. The value of this observation is remarkably larger than the other observations from the same study on the order of between $535 and $202. The outlier observation differs from the other observations in the study because it investigates the sensitivity of WTP to the type of pollution, in this case industrial toxic wastes versus agricultural pollutants. This type of characterization is not controlled for in the meta-analysis because the expectation is that the quality change is either caused by specific pollutants or, if not, only poses a threat to convenience and use rather than an immediate threat to human life. The study remarks on the high value assigned to the removal of this pollutant but fails to note that the category of industrial toxic waste, left undefined, holds a rather ominous range of pollutants with the implication of severe harm to human health. The other studies in the meta-analysis address a mix of pollutants both agricultural and industrial, but none of the other studies cite industrial toxic waste and those dealing with industrial waste, such as PCBs, do so in areas familiar with the risks. The quality variables are not structured to capture the percieved change from catastrophic pollution to pristine. Instead, the variables are formulated to capture a far more narrow range of quality change. For this reason, the observation is removed from the dataset.

The second outlier observation is drawn from Moore et al. (2011) and is also one of eight observations. The observation value is also far above the other observations from the study in the range of $320 to $891 higher than the other values. The study employs a new technique, geospatial referencing (GSR), to estimate WTP based on the quality of water closest to the respondent. As with the above outlier, this extreme difference cannot be accounted for in the meta-analysis model because it is not simply water quality change driving the difference. Instead, this value from the GSR model appears to be affected by the perception of high value to what is the same quality change across the water body (four feet of water clarity). This does not affect the other observations from the same model so the other observations are retained. However, this observation is consistently identified as a potential outlier and a closer look at the study confirms a variation that cannot be controlled for in this meta-analysis and so is removed from the dataset.

4.6 Preliminary Analysis – Phase 2 – Simple OLS Regression & Diagnostics

The second phase of the preliminary analysis is used to identify the appropriate regression method and model for the data set. The three core models, each run with an alternative functional form on the reduced data set, are used in the preliminary analysis for a total of six sets of results.[23] OLS regression results showing several statistically significant and expected patterns affecting WTP for water quality are presented in Table 3.[24]

---

[23] The performance of the reduced dataset is strong and the remaining data analysis proceeds with the reduced data set; results of the comparison are available in Appendix D.

[24] Heteroskedasticity is a broadly recognized concern in the data used for meta-analyses (Nelson & Kennedy, 2009). The Breusch-Pagan test, which has a null hypothesis assuming homoskedasticity, is applied to the regression results of the six models in Table 3. In order to reject the null hypothesis and find evidence of heteroskedasticity, a p-value < 0.01 must be returned by the test. In all six models the Breusch-Pagan test returns a p-value > 0.466, showing evidence that the data used for this meta-analysis is homoskedastic, having a constant error variance across the different explanatory variables.

Table 3: Preliminary Analysis OLS Regression Results for Reduced Data Set

Regression Method	OLS	OLS	OLS	OLS	OLS	OLS
Variables	Model 1a	Model 1a_ln	Model 2a	Model 2a_ln	Model 3a	Model 3a_ln
Dependent Variable:	HH WTP	Ln HH WTP	HH WTP	Ln HH WTP	HH WTP	Ln HH WTP
Extent of Change – Unusable to Usable	-97.78* (56.277)	-0.90* (0.436)	-122.44** (57.904)	-1.04** (0.438)	-111.29* (58.880)	-0.96** (0.446)
Extent of Change – Unusable to Full Use	-2.98 (55.029)	0.16 (0.427)	-7.43 (57.675)	0.14 (0.436)	-6.65 (59.156)	0.16 (0.448)
Description of Quality Change	-113.90** (43.261)	-0.59** (0.335)	-	-	-	-
Endpoints	-	-	76.04 (52.233)	0.36 (0.395)	89.41 (56.733)	0.52 (0.430)
Disaggregated Endpoints	-	-	-	-	-14.38 (46.263)	-0.24 (0.350)
Open vs. Closed Valuation	-69.31* (38.242)	-0.54* (0.296)	-58.73 (40.238)	-0.49 (0.304)	-	-
Multiple Bounded - DC	-133.97** (59.539)	-0.75 (0.461)	-123.64* (63.925)	-0.68 (0.484)	-150** (62.686)	-0.90* (0.475)
Sequential Binary Choice	-46.48 (56.148)	0.13 (0.435)	-84.46 (59.004)	-0.07 (0.446)	-110.66* (58.103)	-0.25 (0.440)
SQBC- No Referendum	-2.61 (80.936)	-0.49 (0.627)	19.76 (84.442)	-0.37 (0.639)	7.87 (86.405)	-0.44 (0.654)

Table 3 continued

Regression Method	OLS	OLS	OLS	OLS	OLS	OLS
Variables	Model 1a	Model 1a_ln	Model 2a	Model 2a_ln	Model 3a	Model 3a_ln
Dependent Variable:	HH WTP	Ln HH WTP	HH WTP	Ln HH WTP	HH WTP	Ln HH WTP
Open Ended	-223.49** (103.395)	-1.34 (0.801)	-245.13** (110.916)	-1.43* (0.839)	-319.63*** (100.880)	-1.99** (0.764)
Payment Card	-128.57 (104.576)	-0.96 (0.811)	-152.12 (109.086)	-1.08 (0.825)	-165.37 (111.996)	-1.15 (0.848)
PC- No Referendum	-136.92 (114.061)	-1.32 (0.884)	-127.58 (121.599)	-1.28 (0.920)	-138.52 (124.025)	-1.36 (0.939)
Published	-33.73 (39.151)	-0.36 (0.303)	-22.93 (41.708)	-0.31 (0.316)	-21.93 (42.841)	-0.29 (0.324)
Year Index	-11.80** (4.919)	-0.08** (0.038)	-10.22* (5.105)	-0.07* (0.039)	-11.98** (5.164)	-0.09** (0.039)
Framing	38.46 (56.001)	0.12 (0.434)	63.05 (57.494)	0.25 (0.435)	82.20 (57.446)	0.39 (0.435)
Payment Timing	-13.63 (55.267)	0.01 (0.428)	-15.14 (58.386)	.0003 (0.442)	-4.72 (59.721)	0.10 (0.452)
Income	.003** (.001)	.00002 (.00001)	.003 * (.002)	.00002 (.00001)	.003** (.002)	.00002 (.00001)
Sample-Users	63.62 (48.082)	0.43 (0.373)	67.24 (50.406)	0.45 (0.381)	74.02 (51.281)	0.51 (0.388)
Sample-Residents	145.96** (41.761)	0.64* (0.324)	123.88*** (43.412)	0.52 (0.328)	107.96** (44.716)	0.43 (0.339)
Location - NE or National	-273.72* (151.991)	-0.85 (1.178)	-219.70 (161.322)	-0.53 (1.221)	-276.54* (163.880)	-1.08 (1.241)
Location-Mid-Atlantic	-68.85 (59.315)	-0.12 (0.460)	-53.23 (61.759)	-0.04 (0.467)	-38.43 (68.905)	0.004 (0.522)
Location- Mid West	0.28 (44.103)	0.18 (0.342)	-17.28 (48.040)	0.09 (0.363)	-14.46 (49.806)	0.09 (0.377)

Table 3 continued

Regression Method	OLS	OLS	OLS	OLS	OLS	OLS
Variables	Model 1a	Model 1a_ln	Model 2a	Model 2a_ln	Model 3a	Model 3a_ln
Dependent Variable:	HH WTP	Ln HH WTP	HH WTP	Ln HH WTP	HH WTP	Ln HH WTP
Location-West	-1.54 (79.178)	0.30 (0.614)	5.43 (85.340)	0.35 (0.646)	-40.18 (82.823)	0.04 (0.627)
Constant	191.05* (105.247)	5.11*** (0.816)	86.66 (122.537)	4.61*** (0.927)	46.27 (128.216)	4.39*** (0.971)
Statistical Performance	# Obs. = 68 P > F = 0.002 R^2 = 0.563 Adj R^2 = 0.363	# Obs. = 68 P > F = 0.018 R^2 = 0.49 Adj R^2 = 0.26	# Obs. = 68 P > F = 0.008 R^2 = 0.519 Adj R^2 = 0.299	# Obs. = 68 P > F = 0.034 R^2 = 0.465 Adj R^2 = 0.22	# Obs. = 68 P > F = 0.014 R^2 = 0.50 Adj R^2 = 0.27	# Obs. = 68 P > F = 0.061 R^2 = 0.44 Adj R^2 = 0.19
F Test (Location)	p-value = 0.323	p-value = 0.856	p-value = 0.608	p-value = 0.948	p-value = 0.485	p-value = 0.87
B-P Test	p-value = 0.466	p-value = 0.75	p-value = 0.527	p-value = 0.787	p-value = 0.50	p-value = 0.85

* Significant at 10% level
** Significant at 5% level
***Significant at 1% level

Though OLS is used in the preliminary analysis in which diagnostic tests and sensitivity analyses are run, this study follows Nelson and Kennedy's (2009) recommendation to place a high priority on correcting for potential correlation among observations drawn from the same study. In reviewing regression methods used to address this correlation - multilevel, panel, or cluster modeling, Nelson and Kennedy (2009) avoid identifying a preferred method; instead, the review notes the importance of a method that produces cluster-robust standard errors. By using this form of robust standard errors for inference, the form of correlation found in meta-analyses is addressed (Nelson & Kennedy, 2009; Cameron & Trivedi, 2005).

To account for this likely correlation, two separate methods are explored: cluster command in OLS regression and panel regression. The regression using OLS clustering shows the expected R^2 value and strong F-test for the model. To test for panel effects in the data, the Breusch-Pagan Lagrange multiplier statistic test is used. Rosenberger and Loomis (2000a) recommend over specifying the model used for the random effects panel regression when applying the Lagrange multiplier statistic test, therefore, the full model including location variables is used for the random effects panel regression.[25] The null hypothesis for this test is that cross-sectional correlation and heteroskedasticity among the panels are not present, and OLS is an appropriate regression method. In order to reject the null hypothesis and find panel effects, the Lagrange multiplier statistic test must yield a p-value < 0.05. The data yields a p-value of 0.12, meaning the null hypothesis for equal effects is not rejected, so OLS is the preferred regression method. OLS clustering by primary study is used for the remaining data analysis and offers an econometric method capable of correcting for potential correlation among study observations.[26]

---

[25] The aim is to increase the power of the Breusch-Pagan Lagrange multiplier statistic to detect latent panel effects (Rosenberger and Loomis, 2000a).

[26] Two forms of weighting, often used in meta-analyses, are not employed in this study. One form of weighting seeks to correct for the variations in efficiency (and in so doing heteroskedasticity) by weighting each WTP estimate with the inverse of its variance (Lipsey & Wilson, 2001; Nelson & Kennedy, 2009). Since this information is not available in all studies, an alternative is to weight each WTP observation by the number in the study sample that was used to derive the WTP estimate. A

Diagnostics and comparisons of model results are used to identify the appropriate specification of the final core models. The three core models with the full list of explanatory variables, excluding location variables, are regressed using OLS clustering. The results for the three models in both level-level and log-level functional forms are presented in Table 4.

The results across the functional forms are quite similar with the prime difference evidenced in the level of significance (1%, 5%, or 10%) for certain explanatory variables. In order to identify a preferred functional form, Ramsey's regression specification error test (RESET) is used. Though not effective at definitively finding models with omitted variables, the test is useful for identifying functional form misspecification. As suggested by Wooldridge (2013), the RESET statistic is reported for each model's results and serves as a comparison between the level-level and log-level functional forms for Models 1, 2, and 3. The null hypothesis for the RESET test is that the model is correctly specified. As seen in Table 4, the three models in log-level functional forms have a RESET p-value greater than 50% and the level-level functional forms have a p-value of 1% or lower. This shows strong evidence that the level-level functional form is misspecified while the log-level functional form is not. As a result, the log-level functional form is used for the final data analysis.

Variance Inflation Factor (VIF) is used to measure correlation between multiple explanatory variables. The test is run after a regression and provides the factor by which the coefficient on each explanatory variable is inflated. The rule of thumb is that a VIF value above 4 requires investigation and a VIF value above 10 is the sign of serious multicollinearity, requiring correction. VIF mean values are

---

second form of weighting multiplies the effect size estimate (WTP) by the inverse of the number of observations drawn from a single study (Nelson & Kennedy, 2009). Doubts about using weights in meta-analysis abound with the most basic criticism summed up as follows: the process values precision over other factors such as unbiasedness (Bergstrom & Taylor, 2006). Rather than using weights, this analysis takes the modeling approach advocated by Bateman and Jones (2003), which is one that explicitly deals with the lack of independence between observations from the same study and with heteroskedasticity. Bateman and Jones (2003) use multilevel modelling and this analysis uses clustering.

reported for each regression model (in Tables 4, 5, & 6) and consistently show a mean VIF of 2.24 and lower and individual VIF values of 3.55 and lower. As a result, no action is taken to correct for multicollinearity in any of the reported models.

Table 4: Preliminary Analysis OLS – Cluster Regression Results – Reduced Dataset

Regression Method	OLS - Cluster	OLS - Cluster	OLS - Cluster	OLS - Cluster	OLS - Cluster	OLS - Cluster
Variables	Model 1a	Model 1b	Model 2a	Model 2b	Model 3a	Model 3b
Dependent Variable:	HH WTP	Ln HH WTP	HH WTP	Ln HH WTP	HH WTP	Ln HH WTP
Extent of Change – Unusable to Usable	-118.81** (54.306)	-0.86* (0.450)	-134.38** (55.209)	-0.96** (0.448)	-146.17*** (52.368)	-1.04** (0.439)
Extent of Change – Unusable to Full Use	-21.87 (55.845)	0.14 (0.372)	-22.78 (55.653)	0.12 (0.373)	-40.61 (57.720)	0.03 (0.362)
Description of Quality Change	-76.42** (36.090)	-0.49* (0.244)	-	-	-	-
Endpoints	-	-	47.54 (44.254)	0.36 (0.310)	47.33 (51.293)	0.42 (0.388)
Disaggregated Endpoints	-	-	-	-	-5.98 (34.684)	-0.24 (0.282)
Open vs. Closed Valuation	-68.48** (32.991)	-0.54** (0.226)	-55.30* (31.690)	-0.45** (0.218)	-	-
Multiple Bounded - DC	-105.24* (52.158)	-0.71* (0.354)	-102.63* (52.312)	-0.70* (0.354)	-100.36* (56.246)	-0.73* (0.388)
Sequential Binary Choice	-45.67 (49.475)	0.02 (0.321)	-75.68 (53.344)	-0.17 (0.328)	-84.62 (55.279)	-0.21 (0.363)
SQBC- No Referendum	-43.55 (78.583)	-0.62 (0.493)	-13.91 (83.921)	-0.43 (0.526)	-9.96 (81.063)	-0.45 (0.516)

Table 4 continued

Regression Method	OLS - Cluster	OLS - Cluster	OLS - Cluster	OLS - Cluster	OLS - Cluster	OLS - Cluster
Variables	Model 1a	Model 1b	Model 2a	Model 2b	Model 3a	Model 3b
Dependent Variable:	HH WTP	Ln HH WTP	HH WTP	Ln HH WTP	HH WTP	Ln HH WTP
Open Ended	-226.73***	-1.46***	-246.26***	-1.59***	-292.82***	-1.93***
	(68.757)	(0.420)	(74.362)	(0.426)	(86.724)	(0.442)
Payment Card	-199.62***	-1.27***	-200.52***	-1.26***	-176.55***	-1.18***
	(47.163)	(0.326)	(49.329)	(0.330)	(59.498)	(0.378)
PC- No Referendum	-96.92	-1.17***	-115.63**	-1.26***	-137.25*	-1.36***
	(61.595)	(0.270)	(53.986)	(0.232)	(76.680)	(0.353)
Published	-46.31	-0.45	-38.56	-0.39	-37.10	-0.38
	(57.002)	(0.399)	(62.234)	(0.425)	(63.210)	(0.432)
Year Index	-6.48*	-0.06***	-6.62*	-0.06***	-6.87*	-0.07***
	(3.339)	(0.021)	(3.356)	(0.021)	(3.602)	(0.022)
Framing	56.47	0.20	72.93	0.29	83.73*	0.38
	(50.269)	(0.322)	(50.487)	(0.304)	(48.595)	(0.299)
Payment Timing	-10.41	0.03	-15.62	.005	-7.23	0.10
	(63.054)	(0.440)	(63.903)	(0.448)	(58.875)	(0.442)
Income	.002	.00001*	.002*	.00001*	.002	.00001*
	(.001)	(6.65e-06)	(.001)	(6.88e-06)	(.001)	(7.40e-06)
Sample-Users	62.57	0.44	68.94	0.48	72.06	0.51
	(49.223)	(0.365)	(52.711)	(0.372)	(56.671)	(0.373)
Sample-Residents	112.10*	0.52	97.92*	0.42	91.94	0.38
	(51.043)	(0.304)	(53.083)	(0.324)	(55.035)	(0.345)
Constant	208.89*	5.29**	137.27	4.79***	108.75	4.75**
	(100.197)	(0.685)	(125.827)	(0.858)	(137.733)	(0.930)

Table 4 continued

Regression Method	OLS - Cluster	OLS - Cluster	OLS - Cluster	OLS - Cluster	OLS - Cluster	OLS - Cluster
Variables	Model 1a	Model 1b	Model 2a	Model 2b	Model 3a	Model 3b
Dependent Variable:	HH WTP	Ln HH WTP	HH WTP	Ln HH WTP	HH WTP	Ln HH WTP
Statistical Performance	# Obs. = 68 P > F = 0.000 R^2 = 0.52	# Obs. = 68 P > F = 0.000 R^2 = 0.58	# Obs. = 68 P > F = 0.000 R^2 = 0.49	# Obs. = 68 P > F = 0.000 R^2 = 0.46	# Obs. = 68 P > F = 0.000 R^2 = 0.46	# Obs. = 68 P > F = 0.000 R^2 = 0.43
RESET	p-value = 0.002	p-value = 0.68	p-value = 0.004	p-value = 0.96	p-value = 0.0003	p-value = 0.51
F Test (Users, Residents)	p-value = 0.094	p-value = 0.22	p-value = 0.14	p-value = 0.30	p-value = 0.17	p-value = 0.32
Mean VIF+	2.24	2.24	2.20	2.20	2.24	2.24

+ No evidence of multicollinearity. All individual variable VIF values < 3.55.
* Significant at 10% level
** Significant at 5% level
*** Significant at 1% level

For the final two sets of regressions, two iterations of reduced models are used. In both reduced models, variables not significant at the 10% level or less are excluded. The variables dropped from the restricted model are *published, framing, payment timing*, and the two sample variables – *users and residents*.[27] All quality, quality description, and elicitation variables are kept for conceptual reasons even if the individual variables are not significant at the 10% level or less. The results for the first reduced models are presented in Table 5.

The model results shown in Table 5 use the two covariates for referendum control, payment card without referendum and sequential binary choice without referendum. The number of observations for each covariate is small (three for payment card without referendum and 11 for sequential binary choice without referendum). The small number of observations indicates a limited explanatory value. This combined with the high correlation values between payment card without referendum covariate and the binary variable for payment card with referendum (0.6348) and sequential binary choice without referendum and the binary variable for sequential binary choice (0.5617), resulted in the decision to find an alternative model. The final reduced models (1-3d), which do not control for referenda, are shown below and the results are available in Table 6.[28] Models 1-3d are used for the final data analysis and conclusions.

---

[27] *Users* and *residents* are jointly insignificant in the three models using the log-level functional form. The results are presented in Table 4.

[28] In order to determine the stability of Models 1-3d's results, two additional models with alternative variables used to control for elicitation method were run. The reason behind these alternatives is to test the sensitivity of Models 1-3d's results to elicitation method, which is the most important determinant of WTP as shown in Section 4.3.2. The results for these alternative models are presented in Appendix E in Tables E1 and E2 and show stability in the results for the quality characteristics, meaning Models 1-3d have, to the extent possible, successfully controlled for different elicitation methods and the decision rule. Table E1 shows Models 1-3e without referenda covariates and the open-ended elicitation method. The open-ended elicitation variable is eliminated by removing the four open-ended observations. Models 1-3f categorize all elicitation methods regardless of decision rule as a single binary variable (single DC with referenda as the base against all other elicitation methods with and without referenda). The results for the f model are in Table E2. Both e and f model sets show nearly identical results to model set d presented in Table 6. This stability across the three model sets shows the explanatory value of removing the referenda covariates with small numbers of observations.

Table 5: Reduced Models with Referendum Controls

Regression Method	OLS - Cluster	OLS - Cluster	OLS - Cluster
Variables	Model 1c	Model 2c	Model 3c
Dependent Variable:	Ln HH WTP	Ln HH WTP	Ln HH WTP
Extent of Change – Unusable to Usable	-0.45 (0.277)	-0.64** (0.253)	-0.73*** (0.253)
Extent of Change – Unusable to Full Use	0.46* (0.227)	0.32 (0.215)	0.23 (0.209)
Description of Quality Change	-0.50** (0.205)	-	-
Endpoints	-	0.62** (0.229)	0.67** (0.300)
Disaggregated Endpoints	-	-	-0.24 (0.299)
Open vs. Closed Valuation	-0.48* (0.277)	-0.42* (0.247)	-
Multiple Bounded - DC	-0.83** (0.315)	-0.82** (0.302)	-0.83** (0.347)
Sequential Binary Choice	0.19 (0.211)	-0.01 (0.171)	-0.04 (0.229)
SQBC- No Referendum	-0.81 (0.536)	-0.57 (0.511)	-0.61 (0.503)
Open Ended	-1.27*** (0.400)	-1.37*** (0.381)	-1.62*** (0.381)
Payment Card	-1.12*** (0.241)	-1.06*** (0.231)	-0.98*** (0.269)
PC- No Referendum	-0.95*** (0.264)	-0.88*** (0.280)	-0.94** (0.376)
Year Index	-0.05** (0.023)	-0.05** (0.022)	-0.05** (0.024)
Income	.000015* (8.17e-06)	.000015* (7.61e-06)	.0000133 9.03e-06
Constant	4.90*** (0.611)	4.35*** (0.634)	4.42*** (0.774)
Statistical Performance	# Obs. = 68 P > F = 0.000 R^2 = 0.37	# Obs. = 68 P > F = 0.000 R^2 = 0.38	# Obs. = 68 P > F = 0.000 R^2 = 0.35
Mean VIF[+]	1.95	1.92	2
AIC	175.0416	174.1974	177.071

+ No evidence of multicollinearity. All individual variable VIFs <2.57.
* Significant at 10% level
** Significant at 5% level
*** Significant at 1% level

Model 1d: $Ln(WTP) = \beta_0 + \beta_1 EC\_UU + \beta_2 EC\text{-}UNFUS + \beta_3 D\_CHNGE + \beta_4 OP\_CL + \beta_5 Elic\_DCMB + \beta_6\ Elic\_SQBC + \beta_7\ Elic\_OP + \beta_8\ Elic\_PC + \beta_9 YearIndex + \beta_{10} Income + \mu$

Model 2d: $Ln(WTP) = \beta_0 + \beta_1 EC\_UU + \beta_2 EC\text{-}UNFUS + \beta_3 Endpts + \beta_4 OP\_CL + \beta_5 Elic\_DCMB + \beta_6\ Elic\_SQBC + \beta_7\ Elic\_OP + \beta_8\ Elic\_PC + \beta_9 YearIndex + \beta_{10} Income + \mu$

Model 3d: $Ln(WTP) = \beta_0 + \beta_1 EC\_UU + \beta_2 EC\text{-}UNFUS + \beta_3 Endpts + \beta_4 DG\_Endpts + \beta_5 Elic\_DCMB + \beta_6\ Elic\_SQBC + \beta_7\ Elic\_OP + \beta_8\ Elic\_PC + \beta_9 YearIndex + \beta_{10} Income + \mu$

5 Empirical Results

5.1 Model Results

The OLS clustering regression results for Model 1d, Model 2d, and Model 3d show several statistically significant determinants of WTP in Table 6. The meta-analysis results show WTP does systematically vary according to the key quality description variables identified in the Boyd and Krupnick (2009) and Johnston et al. (2012) reviews. Each category of variables included in the three models, quality characteristics and study characteristics, are discussed below along with a brief look at the sample variable included in the reduced model.

Table 6: Fully Reduced Models – Final Analysis

Regression Method	OLS - Cluster	OLS - Cluster	OLS - Cluster
Variables	Model 1d	Model 2d	Model 3d
Dependent Variable:	Ln HH WTP	Ln HH WTP	Ln HH WTP
Extent of Change – Unusable to Usable	-0.52 (0.328)	-0.70** (0.284)	-0.80*** (0.286)
Extent of Change – Unusable to Full Use	0.08 (0.326)	0.01 (0.324)	-0.11 (0.322)
Description of Quality Change	-0.41 (0.291)	-	-
Endpoints	-	0.71*** (0.233)	0.76** (0.295)
Disaggregated Endpoints	-	-	-0.21 (0.361)
Open vs. Closed Valuation	-0.48 (0.329)	-0.44 (0.286)	-
Multiple Bounded - DC	-0.69** (0.319)	-0.75** (0.302)	-0.74** (0.357)
Sequential Binary Choice	-0.04 (0.277)	-0.17 (0.217)	-0.21 (0.240)
Open Ended	-1.10*** (0.355)	-1.26*** (0.349)	-1.53*** (0.361)
Payment Card	-1.24*** (0.403)	-1.17*** (0.371)	-1.08** (0.409)
Year Index	-0.04 (0.028)	-0.05* (0.025)	-0.05 (0.030)
Income	.00001* (6.96e-06)	.00001** (6.56e-06)	.00001 (8.36e-06)
Constant	4.93*** (0.651)	4.34*** (0.643)	4.38*** (0.810)
Statistical Performance	# Obs. = 68 P > F = 0.0007 R^2 = 0.303	# Obs. = 68 P > F = 0.0001 R^2 = 0.335	# Obs. = 68 P > F = 0.0001 R^2 = 0.303
Mean VIF+	1.62	1.66	1.66
AIC	177.7691	174.497	177.7504

+ No evidence of multicollinearity. All individual variable VIFs <2.02.
* Significant at 10% level
** Significant at 5% level
***Significant at 1% level

5.1.1 Quality Characteristics

The variables covering the quality characteristics in these three models are best separated into three groups for review: extent of change, ambiguity of valuation, and translation of services. Each group captures the major quality change and description effects described in the critiques by Boyd and Krupnick (2009) and Johnston et al. (2012). An interpretation of the models is undertaken in the following section.

Extent of Change: Both the *extent of change – unusable to usable* and *extent of change – unusable to full use* variables were expected to be statistically significant and positive. In the final model results, the only significant variable is *extent of change – unusable to usable*, which is negative and statistically significant at the 1% level. *Extent of change – unusable to full use* is statistically insignificant in all three models. A few preliminary conclusions may be drawn from this result. First, the significance of *extent of change – unusable to usable* in two of the models indicates that potential scope effects are being controlled but also shows the complicated relationship of scope to these quality description variables. The four quality description variables - *description of quality change, endpoints, disaggregated endpoints,* and *open vs. closed valuation*, affect the significance of the different extent of change variables, highlighting the fact that scope is not an absolute or isolated characteristic. Instead, it appears to be highly sensitive to the scenario in which the quality change is set. Second, the most striking conclusion to take from these results is that the characterization of quality change according to a modified critical ecological range is not of critical significance to respondents. The results are not in keeping with those of Turner et al. (2003) or Johnston et al. (2003), which found WTP dropped as the baseline level of quality increases. Instead, these results indicate that higher value is placed on the range of uses that the water quality allows rather than the absolute level of change in water quality. All the studies in this meta-analysis are valuing water with consumptive uses often for recreation or agriculture. The negative results for *extent of change – unusable to usable* indicate water quality that is merely usable is worth less than water quality that has a higher level of use, regardless of the starting point for the water quality level. The same holds true for the statistical insignificance of *extent of change – unusable to full use*. For both variables the comparison is with *extent of change – usable to full use,* so one potential explanation for these results is that the critical range is not over unusable to usable. Instead, the critical range is over the full use range, which would explain the statistical insignificance of *extent of change – unusable to full use* and the negative sign on *extent of change – unusable to usable*.

Ambiguity of Valuation: The two variables, *disaggregated endpoints* and *open vs. closed valuation,* measure the open-ended nature or vagueness of the water quality change. As explained earlier, this ambiguity is expected to allow respondents to make assumptions about what is being valued and bring in expansive priors. Neither variable is statistically significant at the 10% level in any of the models. This suggests two possible conclusions for these variables. The first is simply that ambiguity in the valuation scenario does not have a systematic effect on WTP estimates. However, this conclusion is not in keeping with earlier research on valuation differences between species specific changes and overall or system wide environmental processes (Czajkowski et al., 2009; Johnston et al., 2012). Though not focused on ambiguity, the research does show valuation differences between different types and sizes of environmental commodities. The potential that the two ambiguity of valuation variables are relevant but not capturing the expected effects points to the second possibility, which is additional research is needed to identify the precise mechanism of ambiguity and the effect on WTP.

Specifically, these results highlight the possibility that the solution to invalid stated preference results proposed by both Boyd and Krupnick (2009) and Johnston et al. (2012) is far more complicated in application than would first appear. The *disaggregated endpoints* variable is modeled directly on the suggestions of both critiques. The variable was included to determine whether attempts to unbundle and disaggregate environmental commodities are successful at narrowing respondents' focus and, therefore, make results easier to interpret. The meta-analysis results suggest not. In fact, the statistical insignificance of the variable provides a certain amount of empirical backing to observations made in the coding process. That is, a listing of a number of endpoints rather than the purposeful selection anticipated in the two critiques is often undertaken in a casual, list-all-possible approach in the studies. In order to truly embark on a precise commodity definition, additional guidelines must be proposed for how to design and present these disaggregated endpoints.

Another means of capturing ambiguity was attempted using the *open vs. closed valuation* variable, which applies two observations made by Johnston et al. (2012) – quantify the change being made and make a distinction, even if implicit, between the change being valued and a larger environmental process. For example, in one of the studies boaters are asked to value the water quality improvement for one tier of a five-tiered water quality scale. The scale has tiers from poor to excellent with each level defined in terms of boating activity (Lipton, 2004). Though this variable was also statistically insignificant, there is a distinct possibility that further efforts to break apart and identify the key factors in ambiguity, specifically cues enabling or disabling the use of expansive priors, would be the next step in determining ambiguity's effect on WTP.

Translation of Services: The two variables, *description of quality change* and *endpoints*, measure the familiarity and immediacy of the change in the environmental resource and related services. The sign on the *description of quality change* was not predicted due to uncertainty over the way it would be interpreted. The results in Table 6 show *description of water quality* is not statistically significant at the 10% level in Model 1. The *endpoints* variable is positive and statistically significant at the 5% level and lower in Models 2 and 3. The sign and statistical significance of *endpoints* matches initial expectations and may be interpreted as showing that providing information on resources and services to make them tangible and understandable systematically raises WTP. These results support Boyd and Krupnick's (2009) contention that including endpoints for the water quality change will facilitate the process of relating the change to a respondent's perceived loss or gain in utility. This also fits with earlier findings on the information effect and may indicate the higher value is closer to "true WTP" because respondents will have a far better understanding of the services impacted by a program and how water quality change affects them (Milon & Scrogin, 2006).

The results for *description of quality change* suggest two possibilities. The first, as with the ambiguity of valuation variables, is that an input focus in the valuation scenario does not systematically affect WTP. However, this would seem to be a rash conclusion particularly given the consistent statistical significance of the *endpoints* variable. As implied by the results for *endpoints*, the translation or identification of the specific services does systematically affect WTP. The reason to consider further research on the *description of quality change* variable is that this variable captures the effects of both translation of services and ambiguity of valuation. Though the logic laid out by Johnston et al. (2012) for the effect an input focus may have on utility measurement is strong and compelling, the ability of this variable to capture effects with competing directions on WTP may well result in the lack of statistical significance. Further research identifying a means of separating out these effects may be the next step for investigating this phenomenon.

5.1.2 Study Characteristics

Elicitation effects – *multiple bounded DC, sequential binary choice, open ended,* and *payment car*d – exhibit the expected signs and, with the exception of *sequential binary choice,* are statistically significant at the 5% level or lower in all three models. The results are evidence that the meta-analysis model is sufficient and the results, robust. As noted earlier, the relationship between the continuous methods, payment card and open-ended, and single DC are well established. The results reflect these empirically based expectations that open-ended and payment card elicitation methods have consistently smaller WTP values than single DC. The results for *multiple bounded DC* are also supported by a more limited number of studies, finding that multiple or double bounded DC reports a lower WTP estimate than single DC. The *sequential binary choice* variable is not statistically significant in any of the three models. This result, though not predicted, may well be due to other elements found in sequential binary choice studies not controlled for in this meta-analysis. This includes varying levels of choice complexity and attribute explanations as well as different sizes in the scope of the projects being valued in this group of studies. Indeed, the number of sequential binary choice studies is large and the similarity between these studies is limited to the theoretical structure of the elicitation questions. The application of this elicitation structure varies from study to study, and this variation may well be the cause of the statistical insignificance for the *sequential binary choice* variable.

The *yearindex* variable included in the reduced model is negative and statistically significant at the 10% level in Model 2 only. The results match the expectation noted by Nelson and Kennedy (2009). The *yearindex* coefficient is negative, meaning more recent studies have lower WTP than do older studies. The lack of a consistent finding across models does indicate this result is highly sensitive to the quality characteristics. This is likely a result of a heavier concentration of certain quality characteristics in more recent studies.

5.1.3 Sample Characteristics

Income is the only sample characteristic variable used in the reduced model. The variable is positive and statistically significant at the 10% level in Models 1 and 2. The variable's impact on WTP is very small, which is in keeping with the expectation noted earlier.

5.2 Model Selection

In addition to identifying quality related WTP variation across stated preference studies, this meta-analysis seeks to identify the set of quality description factors most able to account for variation across studies. Each of the three models uses a different pairing of the quality description variables to determine WTP. The three pairings outlined earlier are as follows:

Model 1d: *Description of Change* and *Open vs. Closed Valuation*

This pairing attempts to capture the different directional effects the description factors have on WTP by using variables that measure the effects of an input focus in the form described by Johnston et al. (2012) and of a narrow change scenario explicitly linked to valuation.

Model 2d: *Endpoints* and *Open vs. Closed Valuation*

This pairing tries to explain the quality description differences as a function of complete information on the service change as recommended by Boyd and Krupnick (2009) and of a narrow change scenario explicitly linked to valuation.

Model 3d: *Endpoints* and *Disaggregated Endpoints*

This pairing covers the quality description differences using two related factors with different effects on the WTP value. The quality description provides complete information on the service change and an unbundled set of services both recommended by Boyd and Krupnick (2009) and Johnston et al. (2012).

The diagnostic used to determine the models' relative ability to describe the data is Akaike's Information Criterion (AIC) (reported in Tables 5 and 6). The criterion is a relative measure of a model's information loss, so no information is offered on the absolute ability of the model to describe data. The criterion is useful for comparing models with the same dependent variable and same number of parameters. The smaller the AIC value the better the model is able to describe the data. The results for the three models show Model 2 is far better at describing the study data than Models 1 or 3.

This result provides preliminary guidance for researchers and those designing stated preference surveys. The variable results of the meta-analysis confirm a quality description factor, *endpoints*, has a systematic effect on WTP across stated preference methods. However, the model results also provide tentative conclusions on the type of quality factors that, when combined, have a stronger impact on survey results. Specifically, the model results show the focus may need to start on the valuation scenario both in terms of describing the outputs affected by a program change and quantitatively linking this change to the valuation question.

6 Conclusion

This meta-analysis investigates the concerns raised by Boyd and Krupnick (2009) and Johnston et al. (2012) on commodity definition in stated preference valuation studies. The results show that the way a commodity is described has a statistically significant effect on WTP estimates in a direction consistent with most of the claims of both critiques. The results on the scope of quality change, as shown in the *extent of change* variables, also show a systematic effect on WTP but in a direction counter to earlier findings. Beyond a simple validation of the criticisms and reviews of the two research groups, this meta-analysis points to the need for additional research into descriptive mechanisms and the effect the interaction of quality descriptions have on WTP and the perceived scope of quality change.

For this analysis, the focus was primarily on the effect individual information variables, all of which related to quality change, have on WTP. The results point to a critical need for further research into the joint effects of these information variables on WTP estimates. An alternative approach would be to focus on a few of the phenomena identified in the critiques, rather than simply trying to identify specific textual constructs in the survey. For example, the concept of expansive priors is one that crosses through all the quality characteristics identified in this meta-analysis – *extent of change, description of quality change, endpoints, disaggregated endpoints,* and *open vs. closed valuation*. Framing the analysis as the study and identification of the variables likely to enable the use of expansive priors, such as *extent of change* variables in the full use range and *endpoints*, would be valuable for survey design. Identifying variables that represent textual constructs that restrict the use of expansive priors would be a valuable companion to such an analysis. This approach would take the results of this meta-analysis beyond merely a confirmation that quality characteristics systematically affect WTP and move the research into an area that identifies the mechanisms for this effect and, perhaps more importantly, provides clear guidance to researchers on survey design.

These results also highlight the need for additional research into scope and the perceived value of resources with consumptive uses. The initial take away from this analysis is that the scope effect is not isolated from the resource or the services attached to that resource. Indeed, the results indicate the same phenomenon affecting the processing of survey information, expansive priors, is also potentially at work in how respondents understand the scope of the quality change. For example, the scope range that was not expected to be critical in this analysis, *extent of change* from usable to full use, may well be the area of highest value to respondents. The reason for this is potentially tied to the type of resource, water with consumptive uses, and the accompanying expansive priors respondents bring when considering the range of recreational and industrial uses high quality water enables. This result points to an interesting avenue of research that would evaluate these quality scope ranges alongside specific quality characteristics that are tied to the resource being valued.

Literature Cited

Abdul-Mohsen, A. (2005). *Economic efficiency and income distribution evaluation of toxics and dam removal using contingent valuation.* Columbus: Ohio State University.

Adamowicz, W., Boxall, P., Williams, M., & Louviere, J. (1998 February). Stated preference approaches for measuring passive use values: Choice experiments and contingent valuation. *American Journal of Agricultural Economics , 80*, 64-75.

Ajzen, I., Brown, T., & Rosenthal, L. (1996). Information bias in contingent valuation: Effects of personal relevance, quality of information, and motivational orientation. *Journal of Environmental Economics and Management, 30*, 43-57.

Arrow, K., Solow, R., Portney, P., Leamer, E., Radner, R., & Schuman, H. (1993). *Report of the NOAA Panel on Contingent Valuation.* Washington DC: Federal Register.

Azevedo, C., Herriges, J., & Kling, C. (2001). *Valuing preservation and improvements of water quality in Clear Lake.* Center for Agricultural and Rural Development. Iowa State University.

Barkmann, J., Glenn, K., Keil, A., Leemhuis, C., Dietrich, N., Gerold, G., et al. (2008). Confronting unfamiliarity with ecosystem functions: The case for an ecosystem service approach to environmental valuation with stated preference methods. *Ecological Economics, 65*, 48-62.

Bateman, I., & Jones, A. (2003 May). Contrasting conventional with multilevel modeling approaches to meta-analyses: Expectation consistency in UK woodland recreation values. *Land Economics, 79*(2), 235-258.

Belsley, D., Kuh, E., & Welsch, R. (1980). *Regression diagnostics: Identifying influential data and sources of collinearity*. New York: John Wiley.

Bergstrom, J., & Taylor, L. (2006). Using meta-analysis for benefits transfer: Theory and practice. *Ecological Economics, 60*, 351-360.

Bergstrom, J., Stoll, J., & Randall, A. (1990). The impact of information on environmental commodity valuation decisions. *American Journal of Agricultural Economics, 72*, 614–621.

Berrens, R., Bohara, A., Jenkins-Smith, H., Silva, C., & Weimer, D. (2003). The advent of internet surveys for political research: A comparison of telephone and internet samples. *Political Analysis, 11*(1), 1-23.

Bishop, R. C., Breffle, W. S., Lazo, J. K., Rowe, R. D., & Wytinck, S. M. (2000). *Restoration scaling based on total value equivalency: Green Bay natural resource damage assessment.* Stratus Consulting. Boulder: Prepared for US Fish and Wildlife Service; US Department of Interior; US Department of Justice.

Blomquist, G., & Whitehead, J. (1998). Resource quality information and validity of willingness to pay in contingent valuation. *Resource and Energy Economics, 20*, 179–196.

Bollen, K., & Jackman, R. (1985 May). Regression diagnostics: An expository treatment of outliers and influential cases. *Sociological Methods Research, 13*(4), 510-542.

Boxall, P., Adamowicz, W., Swait, J., Williams, M., & Louviere, J. (1996). A comparison of stated preference methods for environmental valuation. *Ecological Economics, 18*, 243-253.

Boyd, J., & Krupnick, A. (2009). *The definition and choice of environmental commodities for nonmarket valuation.* RFF DP 09-35, Resources for the Future.

Boyle, K. (1989). Commodity specification and the framing of contingent-valuation questions. *Land Economics, 65*(1), 57-63.

Boyle, K. (2003). Contingent valuation in practice. In P. Champ, K. Boyle, & T. Brown (Eds.), *A primer on nonmarket valuation* (pp. 111-169). Drodrecht: Kluwer Academic Publishers.

Brander, L., Florax, R., & Vermaat, J. (2006). The empirics of wetland valuation: A comprehensive summary and a meta-analysis of the literature. *Environmental and Resource Economics, 33*, 223-250.

Brouwer, R., Dekker, T., Rolfe, J., & Windle, J. (2010). Choice certainty and consistency in repeated choice experiements. *Environmental and Resource Economics, 46*, 93-109.

Brown, T. (2003). Introduction to stated preference methods. In P. Champ, K. Boyle, & T. Brown, *A primer on nonmarket valuation* (pp. 99-110). Dordrecht: Kluwer Academic Publishers.

Cameron, A., & Trivedi, P. (2005). *Microeconometrics: methods and applications.* Cambridge University Press.

Carson, R., & Groves, T. (2007). Incentive and informational properties of preference questions. *Environmental and Resource Economics, 37*, 181–210.

Carson, R., & Louviere, J. (2011). A common nomenclature for stated preference elicitation approaches. *Environmental and Resource Economics, 49*(4), 539-559.

Carson, R., Flores, N., & Mitchell, R. (1999). The theory and measurement of passive-use value. In I. Bateman, & K. Willis, *Valuing environmental preferences* (pp. 97-130). New York: Oxford University Press.

Carson, T. (1998). Valuation of tropical rainforests: Philosophical and practical issues in the use of contingent valuation. *Ecological Economics, 24*, 15-29.

Champ, P. (2003). Collecting survey data for nonmarket valuation. In P. Champ, K. Boyle, & T. Brown, *A primer on nonmarket valuation* (pp. 59-98). Dordrecht, Netherlands: Kluwer Academic Publishers.

Champ, P., & Bishop, R. (2006 May). Is willingness to pay for a public good sensitive to the elicitation format? *Land Economics, 82*(2), 162-173.

Champ, P., & Welsh, M. (2007). Survey methodologies for stated choice studies. In B. Kanninen, *Valuing environmental amenities using stated choice studies: A common sense approach to theory and practice* (pp. 21-42). Dordrecht: Springer.

Champ, P., Flores, N., Brown, T., & Chivers, J. (2002 November). Contingent valuation and incentives. *Land Economics, 78*(4), 591-604.

Christie, M., Hanley, N., Warren, J., Murphy, K., Wright, R., & Hyde, T. (2006). Valuing the diversity of biodiversity. *Ecological Economics, 58*, 304-317.

Collins, A., Benson, M., Borisova, T., & D'Souza, G. (2006). Differences among watershed sub-populations in willingness to pay for water quality improvements: The impact of TMDL development. *No. 21195.* American Agricultural Economics Association.

Collins, A., Rosenberger, R., & Fletcher, J. (2005). The economic value of stream restoration. *Water Resources Research, 41*, 1-9.

Czajkowski, M., Buszko-Briggs, M., & Hanley, N. (2009 October). Valuing changes in forest biodiversity. *Ecological Economics, 68*(12), 2910-2917.

de Groot, R., Wilson, M., & Boumans, R. (2002). A typology for the classification, description and valuation of ecosystem functions, goods and services. *Ecological Economics, 41*, 393–408.

Eisen-Hecht, J., & Kramer, R. (2002). A cost-benefit analysis of water quality protection in the Catawba Basin. *Journal of the American Resources Association, 38*(2), 453-465.

Farber, S., & Griner, B. (2000). Valuing watershed quality improvements using conjoint analysis. *Ecological Economics, 34*, 63-76.

Farber, S., Costanza, R., & Wilson, M. (2002). Economic and ecological concepts for valuing ecosystem services. *Ecological Economics*, 375-392.

Fischer, A., & Hanley, N. (2007). Analysing decision behaviour in stated preference surveys: A consumer psychological approach. *Ecological Economics, 61*, 303-314.

Flores, N. (2003). Conceptual framework for nonmarket valuation. In P. Champ, K. Boyle, & T. Brown, *A primer on nonmarket valuation* (pp. 27-58). Dordrecht, Netherlands: Kluwer Academic Publishers.

Flores, N., & Strong, A. (2008). Estimating the economic benefits of acidic rock drainage clean up using cost shares. *Ecological Economics*, 348-355.

Freeman, A. M. (2003). Economic valuation: What and why. In P. Champ, K. Boyle, & T. Brown, *A primer on nonmarket valuation* (pp. 1-26). Dordrecht: Kluwer Academic Publishers.

Gregory, R., Lichtenstein, S., & Slovic, P. (2006). Valuing environmental resources: A constructive approach. In S. Lichtenstein, & P. Slovic, *The construction of preference* (pp. 609-628). New York: Cambridge University Press.

Griffiths, C., Klemick, H., Massey, M., Moore, C., Newbold, S., Simpson, D., Walsh, P., Wheeler, W. (2012). U.S. Environmental Protection Agency valuation of surface water quality improvements. *Review of Environmental Economics and Policy, 6*(1), 130–146.

Hannemann, W. (1994). Valuing the environment through contingent valuation. *Journal of Economic Perspectives*, 19-43.

Heberling, M. (2000). *Valuing public goods using the stated choice method.* University Park: Pennsylvania State University.

Hite, D. (2002). *Willingness to pay for water quality improvements: The case of precision application technology.* Department of Agricultural Economics and Rural Sociology. Auburn: Auburn University.

Holmes, T., Bergstrom, J., Huszar, E., Kask, S., & Orr, F. (2004). Contingent valuation, net marginal benefits, and the scale of riparian ecosystem restoration. *Ecological Economics, 49*, 19 – 30.

Hoyos, D. (2010). The state of the art of environmental valuation with discrete choice experiments . *Ecological Economics, 69*, 1595-1603.

Huang, J., Haab, T., & Whitehead, J. (1997). Willingness to pay for quality improvements: Should revealed and stated preference data be combined? *Journal of Environmental Economics and Management*, 240-255.

Hushak, L., & Bielen, M. (1999). *Valuing the Ottawa River: The economic values and impacts of recreational boating.* Ohio Sea Grant College Program Technical Bulletin Series Publication.

Johnston, R., Besedin, E., & Wardwell, R. (2003). Modeling relationships between use and nonuse values for surface water quality: A meta-analysis. *Water Resources Research, 39*(12), 2-1- 2-9.

Johnston, R., Besedin, E., Iovanna, R., Miller, C., Wardwell, R., & Ranson, M. (2005). Systematic variation in willingness to pay for aquatic resource improvements and implications for benefit transfer: A meta-analysis. *Canadian Journal of Agricultural Economics, 53*, 221-248.

Johnston, R., Schultz, E., Segerson, K., Besedin, E., & Ramachandran, M. (2012 February). Enhancing the content validity of stated preference valuation: The structure and function of ecological indicators. *Land Economics, 88*(1), 102-120.

Kaoru, Y. (1993). Differentiating use and nonuse values for coastal pond water quality improvements. *Environmental and Resource Economics*, 487-494.

Knetsch, J. (2010). Values of gains and losses: Reference states and choice of measure. *Environmental and Resource Economics, 46*, 179-188.

Kragt, M., & Bennett, J. (2012). Attribute framing in choice experiments: How do attribute level descriptions affect value estimates? *Environmental and Resource Economics, 51*, 43-59.

Lichtkoppler, F., & Blaine, T. (1999). Environmental awareness and attitudes of Ashtabula County voters concerning the Ashtabula River area of concern: 1996-1997. *Journal of Great Lakes Resources*, 500-514.

Lindsey, G. (1994). Market models, protest bids, and outliers in contingent valuation. *Journal of Water Resources Planning and Management*, 121-129.

Lipsey, M., & Wilson, D. (2001). *Practical meta-analysis.* Thousand Oaks: Sage Publications.

Lipton, D. (2004). *Value of improved water quality to Chesapeake Bay boaters.* Department of Agricultural and Resource Economics. College Park: University of Maryland.

Lombard, M., Snyder-Duch, J., & Bracken, C. C. (2002). Content analysis in mass communication: Assessment and reporting of intercoder reliability. *Human Communication Research, 28*, 587–604.

Loomis, J. K., Strange, L., Fausch, K., & Covich, P. (2000). Measuring the total economic value of restoring ecosystem services in an impaired river basin: Results from a contingent valuation survey. *Ecological Economics*, 103-117.

Lyke, A. (1993). *Discrete choice models to value changes in environmental quality: A Great Lakes case study.* Madison: University of Wisconsin.

Magat, W., Huber, A., Viscusi, K., & Bell, J. (2000). An iterative choice approach to valuing clean lakes, rivers, and streams. *Journal of Risk and Uncertainty*, 7-43.

McConnell, K. (1990). Models for referendum data: The structure of discrete choice models for contingent valuation. *Journal of Environmental Economics and Management, 18*, 19-34.

McNair, B., Bennett, J., & Hensher, D. (2011). A comparison of responses to single and repeated discrete choice questions. *Resource and Energy Economics, 33*, 554-571.

Milon, J., & Scrogin, D. (2006). Latent preferences and valuation of wetland ecosystem restoration. *Ecological Economics, 56*, 162– 175.

Moore, R., Provencher, B., & Bishop, R. (2011). Valuing a spatially variable environmental resource: Reducing non-point-source pollution in Green Bay, Wisconsin. *Land Economics, 87*(1), 45–59.

Nelson, J., & Kennedy, P. (2009). The use (and abuse) of meta-analysis in environmental and natural resource economics: An assessment. *Environmental Resource Economics, 42*, 345–377.

Neuendorf, K. (2011). Content analysis—a methodological primer for gender research. *Sex Roles*, 276–289.

Nunes, P., & van den Bergh, J. (2001). Economic valuation of biodiversity: Sense or nonsense? *Ecologocal Economics, 39*, 203-222.

Poe, G., Boyle, K., & Bergstrom, J. (2001). A preliminary meta analysis of contingent values for ground water quality revisited. In J. Bergstrom, K. Boyle, & G. Poe, *The economic value of water quality*. Northampton, MA: Edward Elgar Publishers.

Powe, N., Garrod, G., & McMahon, P. (2005). Mixing methods within stated preference environmental valuation: Choice experiments and post-questionnaire qualitative analysis. *Ecological Economics, 52*, 513-526.

Racevskis, L., & Lupi, F. (2008). Incentive compatibility in an attribute-based referendum model. *Paper Presented at American Agricultural Economics Association Annual Meeting*, *27-29 July*. Orlando.

Randall, A., DeZoysa, D., & Yu, S. (2001). Ground water, surface water, and wetlands valuation in Ohio. In J. Bergstrom, K. Boyle, & G. Poe, *The Economic value of water quality* (pp. 83-99). Northampton, MA: Edward Elgar.

Rosenberger, R., & Loomis, J. (2000a). Panel stratification in meta-analysis of economic studies: An investigation of its effects in the recreation valuation literature. *Journal of Agricultural and Applied Economics, 32*(3), 459-470.

Rosenberger, R., & Loomis, J. (2000b). Using meta-analysis for benefit transfer: In-sample convergent validity tests of an outdoor recreation database. *Water Resources Research, 36*(4), 1097-1107.

Samples, K., Dixon, J., & Gowen, M. (1986). Information disclosure and endangered species valuation. *Land Economics, 62*, 306–312.

Schlapfer, F. (2006). Survey protocal and income effects in contingent valuation of public goods: A meta-analysis. *Ecological Economics, 57*, 415-429.

Schwarz, N. (1997). Cognition, Communication, and Survey Measurements: Some Implications for Contingent Valuation Surveys. In R. Kopp, W. Pommerehne, & N. Schwarz, *Determining the Value of Non-Marketed Goods* (pp. 165-188). Boston: Kluwer Academic Publishers.

Shresta, R. K., & Alavalapati, J. (2004). Valuing environmental benefits of silvopasture practice: A case study of the Lake Okeechobee watershed in Florida. *Ecological Economics, 49*, 349-359.

Smith, V., & Osborne, L. (1996 November). Do contingent valuation estimates pass a "scope" test? A meta-analysis. *Journal of Environmental Economics and Management, 31*(3), 287-301.

Smith, V., & Pattanayak, S. (2002). Is meta-analysis a Noah's Ark for non-market valuation? *Environmental and Resource Economics, 22*, 271-296.

Stevens, T., DeCoteau, N., & Willis, C. (1997 Feb). Sensitivity of contingent valuation to alternative payment schedules. *Land Economics, 73*(1), 140-148.

Stumborg, B., Baerenklau, K., & Bishop, R. (2001). Nonpoint source pollution and present values: A contingent valuation of Lake Mendota. *Review of Agricultural Economics, 23*(1), 120-132.

Turner, R., Paavola, J., Cooper, P., Farber, S., Jessamy, V., & Georgiou, S. (2003). Valuing nature: Lessons learned and future research directions. *Ecological Economics, 43*, 493-510.

Van Houtven, G., Powers, J., & Pattanayak, S. (2007). Valuing water quality improvements in the united states using meta-analysis: Is the glass half-full or half-empty for national policy analysis? *Resource and Energy Economics, 29*, 206-228.

Viscusi, W., Huber, J., & Bell, J. (2008). The Economic value of water quality. *Environmental and Resource Economics, 41*, 169–187.

Welsh, M., & Poe, G. (1998). Elicitation effects in contingent valuation: Comparisons to a multiple bounded discrete choice approach. *Journal of Environmental Economics and Management, 36*, 170-185.

Wey, K. (1990). *Social welfare analysis of congestion and water quality of Great Salt Pond, Block Island, Rhode Island.* Providence, RI: University of Rhode Island.

Whitehead, J. (1995). Willingness to Pay for quality improvements: Comparative statics and interpretation of contingent valuation results. *Land Economics, 71*(2), 207–215.

Whitehead, J. (2006). Improving willingness to pay estimates for quality improvements through joint estimation with quality perceptions. *Southern Economic Journal, 73*(1), 100-111.

Whitehead, J., & Groothuis, P. (1992). Economic benefits of improved water quality: A case study of North Carolina's Tar Pamlico River. *Rivers, 3*, 170-178.

Whitehead, J., Blomquist, G., Hoban, T., & Clifford, W. (1995). Assessing the validity and reliability of contingent values: A comparison of on-site users, off-site users, and non-users. *Journal of Environmental Economics and Management*, 238-51.

Whittington, D., Cassidy, G., Amaral, D., McClelland, E., Wang, H., & Poulos, C. (1994). *The economic value of improving the environmental quality of Galveston Bay.* The Galveston Bay National Estuary Program. Chapel Hill: University of North Carolina.

Wooldridge, J. (2013). *Introductory econometrics: A modern approach* (Fifth Edition ed.). Mason, OH: South-Western, Cengage Learning.

Appendices

Appendix A: Studies and Summary Statistics from Full Data Set

Table A1: Studies Analyzed in Meta-Analysis for Full Data Set

In Final Dataset	Study Citation	# of Obs	Valuation Method	Elicitation Method	Extent of Change	Quality Change Factors	Other Study Characteristics	Sample Characteristics	Adj. WTP Value*
✓	(Abdul-Mohsen, 2005)	1	Contingent Valuation	DC-Single Referendum	Unusable to Full Use	*Description of Change:* Outputs *Endpoints:* Yes *Disaggregated Endpoints:* Yes *Narrow Change Description:* No	*Payment Period*: Lump Sum *Change Frame:* Improvement *Published:* No *Year of Study:* 2003	*Income*:* $47, 538 *Sample Population:* General Public *Location:* MidWest	$119

Table A1 continued

In Final Dataset	Study Citation	# of Obs	Valuation Method	Elicitation Method	Extent of Change	Quality Factors	Other Study Characteristics	Sample Characteristics	Adj. WTP Value *
✓	(Azevedo, Herriges, & Kling, 2001)	5	Contingent Valuation	DC-Single Referendum	Unusable to Usable & Unusable to Full Use	*Description Of Change:* Outputs *Endpoints:* Yes *Disaggregated Endpoints:* Yes *Narrow Change Description:* Yes	*Payment Period*: Annual *Change Frame:* Improvement/Prevent Deterioration *Published:* No *Year of Study:* 2000	*Income:* $50, 633 - $56, 962 *Sample Population:* General Public & Users *Location:* Mountain Pacific	$22 - $144
✓	(Bishop, Breffle, Lazo, Rowe, & Wytinck, 2000)	5	Choice Experiment	Sequential Binary Choice Referendum	Unusable to Usable & Unusable to Full Use	*Description Of Change:* Inputs *Endpoints:* Yes *Disaggregated Endpoints:* Both *Narrow Change Description:* Both	*Payment Period*: Annual *Change Frame:* Improvement *Published:* No *Year of Study:* 1999	*Income:* $56, 283 *Sample Population:* Residents *Location:* MidWest	$31 - $262

Table A1 continued

In Final Dataset	Study Citation	# of Obs	Valuation Method	Elicitation Method	Extent of Change	Quality Factors	Other Study Characteristics	Sample Characteristics	Adj. WTP Value*
✓	(Collins, Benson, Borisova, & D'Souza, 2006)	3	Contingent Valuation	Payment Card Referendum	Unusable to Full Use	*Description Of Change:* Inputs *Endpoints:* Yes *Disaggregated Endpoints:* No *Narrow Change Description:* No	*Payment Period*: Annual *Change Frame:* Improvement *Published:* No *Year of Study:* 2005	*Income:* $52, 228 - $61, 795 *Sample Population:* Residents & General Public *Location:* MidAtlantic	$50 - $89
✓	(Collins, Rosenberger, & Fletcher, 2005)	3	Choice Experiment	Sequential Binary Choice	Unusable to Full Use	*Description Of Change:* Outputs *Endpoints:* Yes *Disaggregated Endpoints:* Yes *Narrow Change Description:* Yes	*Payment Period*: Annual *Change Frame:* Improvement *Published:* Yes *Year of Study:* 2003	*Income:* $50,948 *Sample Population:* Residents & General Public *Location:* MidAtlantic	$51 - $72

Table A1 continued

In Final Dataset	Study Citation	# of Obs	Valuation Method	Elicitation Method	Extent of Change	Quality Factors	Other Study Characteristics	Sample Characteristics	Adj. WTP Value*
✓	(Eisen-Hecht & Kramer, 2002)	2	Contingent Valuation Method	DC-Single Referendum	Usable to Full Use	*Description Of Change:* Inputs *Endpoints:* No *Disaggregated Endpoints:* No *Narrow Change Description:* No	*Payment Period*: Annual *Change Frame:* Prevent Deterioration *Published:* Yes *Year of Study:* 1998	*Income:* $74,172 *Sample Population:* Residents *Location:* Southeast	$186 - $259
✓	(Farber & Griner, 2000)	3	Conjoint Analysis	Sequential Binary Choice	Unusable to Full Use & Unusable to Usable	*Description Of Change:* Outputs *Endpoints:* Yes *Disaggregated Endpoints:* No *Narrow Change Description:* No	*Payment Period*: Annual *Change Frame:* Improvement *Published:* Yes *Year of Study:* 1996	*Income:* $52,083 *Sample Population:* Residents *Location:* MidAtlantic	$37- $105

Table A1 continued

In Final Dataset	Study Citation	# of Obs	Valuation Method	Elicitation Method	Extent of Change	Quality Factors	Other Study Characteristics	Sample Characteristics	Adj. WTP Value*
✓	(Flores & Strong, 2008)	1	Contingent Valuation Method	DC-Multiple Bounded Referendum	Unusable to Full Use	*Description Of Change:* Outputs *Endpoints:* Yes *Disaggregated Endpoints:* No *Narrow Change Description:* No	*Payment Period*: Lump Sum *Change Frame:* Improvement *Published:* Yes *Year of Study:* 2004	*Income:* $80,831 *Sample Population:* Residents *Location:* West	$64
✓	(Heberling, 2000)	3	Choice Experiment	Sequential Binary Choice	Unusable to Full Use	*Description Of Change:* Outputs *Endpoints:* Yes *Disaggregated Endpoints:* Yes *Narrow Change Description:* Yes	*Payment Period*: Annual *Change Frame:* Improvement *Published:* No *Year of Study:* 2000	*Income:* $44,301 – $44,667 *Sample Population:* Residents *Location:* MidAtlantic	$316 - $363

Table A1 continued

In Final Dataset	Study Citation	# of Obs	Valuation Method	Elicitation Method	Extent of Change	Quality Factors	Other Study Characteristics	Sample Characteristics	Adj. WTP Value*
✓	(Hite, 2002)	2	Contingent Valuation Method	DC-Single Referendum	Usable to Full Use	*Description Of Change:* Inputs *Endpoints:* No *Disaggregated Endpoints:* No *Narrow Change Description:* No	*Payment Period*: Lump Sum *Change Frame:* Prevent Deterioration *Published:* Yes *Year of Study:* 1999	*Income:* $56,805 *Sample Population*: General Public *Location:* Southeast	$61 - $65
✓	(Holmes, Bergstrom, Huszar, Kask, & Orr, 2004)	2	Contingent Valuation Method	DC- Double Bounded Referendum	Unusable to Full Use & Unusable to Usable	*Description Of Change:* Outputs *Endpoints:* Yes *Disaggregated Endpoints:* Yes *Narrow Change Description:* Yes	*Payment Period*: Annual *Change Frame:* Improvement *Published:* Yes *Year of Study:* 2000	*Income:* $56,962 *Sample Population*: Residents *Location:* Southeast	$7 - $68

Table A1 continued

In Final Dataset	Study Citation	# of Obs	Valuation Method	Elicitation Method	Extent of Change	Quality Factors	Other Study Characteristics	Sample Characteristics	Adj. WTP Value*
✓	(Huang, Haab, & Whitehead, 1997)	1	Contingent Valuation Method	DC- Double Bounded Referendum	Unusable to Full Use	*Description Of Change:* Outputs *Endpoints:* Yes *Disaggregated Endpoints:* No *Narrow Change Description:* No	*Payment Period*: Annual *Change Frame:* Improvement *Published:* Yes *Year of Study:*1995	*Income:* $45,445 *Sample Population*: Residents *Location:* Southeast	$269
✓	(Hushak & Bielen, 1999)	2	Contingent Valuation Method	Payment Card	Unusable to Full Use	*Description Of Change:* Inputs *Endpoints:* No *Disaggregated Endpoints:* No *Narrow Change Description:* No	*Payment Period*: Annual *Change Frame:* Improvement *Published:* No *Year of Study:*1998	*Income:* $88,715 *Sample Population*: Users *Location:* MidWest	$29 - $51
✓	(Kaoru, 1993)	1	Contingent Valuation Method	Open Ended	Unusable to Full Use	*Description of Change:* Outputs *Endpoints:* Yes *Disaggregated Endpoints:* No *Narrow Change Description:* Yes	*Payment Period*: Annual *Change Frame:* Improvement *Published:* Yes *Year of Study:*1989	*Income:* $152,808 *Sample Population*: Residents *Location:* Northeast	$230

Table A1 continued

In Final Dataset	Study Citation	# of Obs	Valuation Method	Elicitation Method	Extent of Change	Quality Factors	Other Study Characteristics	Sample Characteristics	Adj. WTP Value*
✓	(Lichtkoppler & Blaine, 1999)	1	Contingent Valuation Method	DC- Double Bounded Referendum	Unusable to Usable	*Description Of Change:* Inputs *Endpoints:* No *Disaggregated Endpoints:* No *Narrow Change Description:* No	*Payment Period*: Annual *Change Frame:* Improvement *Published:* Yes *Year of Study:* 1997	*Income:* $43,910 *Sample Population*: General Public *Location:* MidWest	$44
✓	(Lindsey, 1994)	1	Contingent Valuation Method	Payment Card Referendum	Usable to Full Use	*Description Of Change:* Inputs *Endpoints:* No *Disaggregated Endpoints:* No *Narrow Change Description:* No	*Payment Period*: Annual *Change Frame:* Improvement *Published:* Yes *Year of Study:* 1989	*Income:* $64,183 *Sample Population*: Residents *Location:* MidAtlantic	$74

Table A1 continued

In Final Dataset	Study Citation	# of Obs	Valuation Method	Elicitation Method	Extent of Change	Quality Factors	Other Study Characteristics	Sample Characteristics	Adj. WTP Value*
✓	(Lipton, 2004)	1	Contingent Valuation Method	Open Ended	Usable to Full Use	*Description Of Change:* Outputs *Endpoints:* Yes *Disaggregated Endpoints:* No *Narrow Change Description:* Yes	*Payment Period*: Annual *Change Frame:* Improvement *Published:* Yes *Year of Study:* 2001	*Income:* $70,936 *Sample Population*: Users *Location:* MidAtlantic	$67
✓	(Loomis, Strange, Fausch, & Covich, 2000)	1	Contingent Valuation Method	DC-Single Referendum	Usable to Full Use	*Description Of Change:* Outputs *Endpoints:* Yes *Disaggregated Endpoints:* Yes *Narrow Change Description:* Yes	*Payment Period*: Annual *Change Frame:* Improvement *Published:* Yes *Year of Study:* 1998	*Income:* $52,807 *Sample Population*: Residents *Location:* West	$337

Table A1 continued

In Final Dataset	Study Citation	# of Obs	Valuation Method	Elicitation Method	Extent of Change	Quality Factors	Other Study Characteristics	Sample Characteristics	Adj. WTP Value*
✓	(Lyke, 1993)	1	Contingent Valuation Method	DC-Single	Usable to Full Use	*Description Of Change:* Outputs *Endpoints:* Yes *Disaggregated Endpoints:* No *Narrow Change Description:* No	*Payment Period*: Annual *Change Frame:* Improvement *Published:* No *Year of Study:* 1990	*Income:* $51,270 *Sample Population*: Users *Location*: MidWest	$63
✓	(Magat, Huber, Viscusi, & Bell, 2000)	8	Iterative Choices	Sequential Binary Choice Referendum	Usable to Full Use	*Description Of Change:* Outputs *Endpoints:* Yes *Disaggregated Endpoints:* Yes *Narrow Change Description:* Yes & No	*Payment Period*: Annual *Change Frame:* Improvement *Published:* Yes *Year of Study:* 1997	*Income:* $54,891 *Sample Population*: Residents & General Public *Location*: National	$124 - $659

Table A1 continued

In Final Dataset	Study Citation	# of Obs	Valuation Method	Elicitation Method	Extent of Change	Quality Factors	Other Study Characteristics	Sample Characteristics	Adj. WTP Value*
✓	(Moore, Provencher, & Bishop, 2011)	8	Contingent Valuation Method	DC-Single Referendum	Unusable to Usable & Usable to Full Use	*Description Of Change:* Outputs *Endpoints:* Yes *Disaggregated Endpoints:* No *Narrow Change Description:* No	*Payment Period*: Annual *Change Frame:* Improvement *Published:* Yes *Year of Study:* 2005	*Income:* $60,826 - $83,147 *Sample Population*: Residents and General *Location*: MidWest	$10 - $901
✓	(Randall, DeZoysa, & Yu, 2001)	2	Contingent Valuation Method	DC- Double Bounded Referendum	Unusable to Full Use	*Description Of Change:* Outputs *Endpoints:* Yes *Disaggregated Endpoints:* Yes *Narrow Change Description:* Yes	*Payment Period*: Lump Sum *Change Frame:* Prevent Deterioration *Published:* Yes *Year of Study:* 1994	*Income:* $63,316 *Sample Population*: General Public *Location*: MidWest	$115 - $175

Table A1 continued

In Final Dataset	Study Citation	# of Obs	Valuation Method	Elicitation Method	Extent of Change	Quality Factors	Other Study Characteristics	Sample Characteristics	Adj. WTP Value*
✓	(Shresta & Alavalapati, 2004)	2	Choice Experiment	Sequential Binary Choice Referendum	Usable to Full Use	*Description Of Change:* Inputs *Endpoints:* No *Disaggregated Endpoints:* No *Narrow Change Description:* Yes	*Payment Period*: Annual *Change Frame:* Improvement *Published:* Yes *Year of Study:* 2002	*Income:* $72,571 *Sample Population*: General Public *Location*: Southeast	$37 - $86
✓	(Stumborg, Baerenklau, & Bishop, 2001)	2	Contingent Valuation Method	DC-Multiple Bounded Referendum	Usable to Full Use	*Description Of Change:* Outputs *Endpoints:* Yes *Disaggregated Endpoints:* No *Narrow Change Description:* No	*Payment Period*: Annual *Change Frame:* Improvement *Published:* Yes *Year of Study:* 2001	*Income:* $76,847 *Sample Population*: Residents *Location*: MidWest	$70 - $107

Table A1 continued

In Final Dataset	Study Citation	# of Obs	Valuation Method	Elicitation Method	Extent of Change	Quality Factors	Other Study Characteristics	Sample Characteristics	Adj. WTP Value*
✓	(Viscusi, Huber, & Bell, 2008)	2	Conjoint Analysis	Sequential Binary Choice	Unusable to Full Use	*Description Of Change:* Outputs *Endpoints:* Yes *Disaggregated Endpoints:* Yes *Narrow Change Description:* Yes	*Payment Period*: Annual *Change Frame:* Improvement *Published:* Yes *Year of Study:* 2004	*Income:* $51,598 *Sample Population*: General Public *Location*: National	$31 - $38
✓	(Wey, 1990)	1	Contingent Valuation Method	Payment Card	Unusable to Full Use	*Description Of Change:* Outputs *Endpoints:* Yes *Disaggregated Endpoints:* No *Narrow Change Description:* Yes	*Payment Period*: Annual *Change Frame:* Improvement *Published:* No *Year of Study:* 1989	*Income:* $94,361 *Sample Population*: Residents *Location*: Northeast	$67

Table A1 continued

In Final Dataset	Study Citation	# of Obs	Valuation Method	Elicitation Method	Extent of Change	Quality Factors	Other Study Characteristics	Sample Characteristics	Adj. WTP Value*
✓	(Whitehead & Groothuis, 1992)	2	Contingent Valuation Method	Open Ended	Usable to Full Use	*Description Of Change:* Outputs *Endpoints:* Yes *Disaggregated Endpoints:* Yes *Narrow Change Description:* Yes	*Payment Period*: Annual *Change Frame:* Improvement *Published:* Yes *Year of Study:* 1991	*Income:* $56,944 *Sample Population*: Users & General Public *Location*: Southeast	$35 - $58
✓	(Whitehead J. , Blomquist, Hoban, & Clifford, 1995)	3	Contingent Valuation Method	DC- Double Bounded Referendum	Usable to Full Use	*Description Of Change:* Outputs *Endpoints:* Yes *Disaggregated Endpoints:* No *Narrow Change Description:* No	*Payment Period*: Annual *Change Frame:* Prevent Deterioration *Published:* Yes *Year of Study:* 1990	*Income:* $52,853 - $67,321 *Sample Population*: Users & General Public *Location*: MidAtlantic	$83- $119

Table A1 continued

In Final Dataset	Study Citation	# of Obs	Valuation Method	Elicitation Method	Extent of Change	Quality Factors	Other Study Characteristics	Sample Characteristics	Adj. WTP Value*
✓	(Whitehead J. , 2006)	4	Contingent Valuation Method	DC- Double Bounded	Usable to Full Use & Unusable to Full Use	*Description Of Change:* Outputs *Endpoints:* Yes *Disaggregated Endpoints:* No *Narrow Change Description:* No	*Payment Period*: Annual *Change Frame:* Improvement *Published:* Yes *Year of Study:* 1998	*Income:* $95,307 *Sample Population*: Residents *Location*: Southeast	$1- $385
✓	(Whittington , Cassidy, Amaral, McClelland, Wang, & Poulos, 1994)	2	Contingent Valuation Method	DC- Double Bounded Referendum	Usable to Full Use	*Description Of Change:* Outputs *Endpoints:* Yes *Disaggregated Endpoints:* No *Narrow Change Description:* No	*Payment Period*: Annual *Change Frame:* Improvement *Published:* No *Year of Study:* 1993	*Income:* $82,680 - $83,100 *Sample Population*: General Public in Area *Location*: Southeast	$222 - $370

* Income and WTP values adjusted to 2010 dollars.

Table A2: Variable Definitions and Summary Statistics for Full Data Set

Variable	Description	Units of Measurement	Mean	Standard Deviation	Frequency	%
Household WTP*	Annual household willingness to pay for water quality change. WTP for all studies converted into 2010 dollars using University of Oregon conversion table	Dollars (Range: 1 to 901)	161.73	164.77	-	-
Ln of Household WTP	Natural log of annual household willingness to pay for water quality change. WTP for all studies converted into 2010 dollars using University of Oregon conversion table	Natural log of dollars (Range: -.5978 to 6.5778)	4.60	1.10	-	-
Extent of Change - Unusable to usable	Binary variable indicating the extent of quality change in the environmental commodity is from unusable to usable. The reference is a change from usable to higher levels of use. These categories are based on scenario descriptions in the surveys.	Binary (Range: 0 or 1)	-	-	14	18.67
Extent of Change - Unusable to Full Use	Binary variable indicating the extent of quality change in the environmental commodity is from unusable to full use. The reference is a change from usable to higher levels of use. These categories are based on scenario descriptions in the surveys.	Binary (Range: 0 or 1)	-	-	25	33.33
Extent of Change - Usable to Full Use	The reference for the extent of change variables includes usable to medium use and usable to full use.	Binary (Range: 0 or 1)	-	-	36	48
Description of Change	Binary variable indicating the change is described in terms of a reduction in inputs (pollutants, nutrients, and sediment). The reference is for change described in terms of outputs or services of the environmental commodity.	Binary (Range: 0 or 1)	-	-	16	21.33

Table A2 continued

Variable	Description	Units of Measurement	Mean	Standard Deviation	Frequency	%
Endpoints	Binary variable indicating the change descriptions identify endpoints or easily understandable outcomes of the change. The reference is for change described generally or in technical language without specific information on environmental commodities that directly affect human use.	Binary (Range: 0 or 1)	-	-	65	86.67
Disaggregated Endpoints	Binary variable indicating the change descriptions identify multiple endpoints and services that will be affected by the change. The reference is for change described using a single endpoint as an indicator of overall change.	Binary (Range: 0 or 1)	-	-	31	41.33
Open vs. Closed Valuation	Binary variable indicating the change valued is described with quantitative measures or linkages to a specific and discrete ecological service. The reference is for change descriptions that allow respondents to value broader ecological services.	Binary (Range: 0 or 1)	-	-	30	40
Dichotomous Choice-Multi-Bounded	Binary variable indicating the survey used a double-bounded or multiple-bounded dichotomous choice elicitation question set. The reference is an elicitation method using a single dichotomous choice question set.	Binary (Range: 0 or 1)	-	-	18	24
Sequential Binary Choice	Binary variable indicating the survey used a sequential binary choice set with options of price and environmental commodities to elicit stated preference. The reference is an elicitation method using a single dichotomous choice question set.	Binary (Range: 0 or 1)	-	-	26	34.67
SQBC – No Referendum	Binary variable acting as a covariate to control for the sequential binary choice observations not using a referendum decision rule.	Binary (Range: 0 or 1)	-	-	11	15.71
Open Ended	Binary variable indicating the survey used an open-ended question to elicit stated preference. The reference is an elicitation method using a single dichotomous choice question set.	Binary (Range: 0 or 1)	-	-	4	5.33

Table A2 continued

Variable	Description	Units of Measurement	Mean	Standard Deviation	Frequency	%
Payment Card	Binary variable indicating the survey used a payment card method to elicit stated preference. The reference is an elicitation method using a single dichotomous choice question set.	Binary (Range: 0 or 1)	-	-	7	9.33
PC – No Referendum	Binary variable acting as a covariate to control for the payment card observations not using a referendum decision rule.	Binary (Range: 0 or 1)	-	-	3	4.29
Dichotomous Choice-Single	Dichotomous choice – single binary choice is the reference for binary variables on elicitation method for stated preference survey instruments.	Binary (Range: 0 or 1)	-	-	20	26.67
Published	Binary variable indicating the survey results were published in a journal or book, indicating a level of peer review or other outside review. The reference is for grey literature including academic and government reports, conference papers, and dissertations.	Binary (Range: 0 or 1)	-	-	52	69.33
Year Index	Year in which the study was conducted, converted to an index by subtracting 1989.	Year Index (Range: 0 to 16)	9.67	4.55	-	-
Framing	Binary variable indicating the quality change to be valued is an improvement over the current state. The reference is for an improvement over the current state.	Binary (Range: 0 or 1)	-	-	64	85.33
Payment Timing	Binary variable indicating the willingness to pay value is a lump sum one-time payment. The reference is for annual payment for a period of 3 to 25 years or a period that is indefinite.	Binary (Range: 0 or 1)	-	-	6	8
Income*	The mean or median income of respondents, either as reported in the original study or as imputed from US census medians for the area surveyed.	Dollars (Range: 43,301 to 152,808)	64,107	17,656	-	-
Sample-Users	Binary variable indicating the population surveyed were users of the water body, both local and out of area. The reference is for the general public in the area (county, watershed, or state).	Binary (Range: 0 or 1)	-	-	8	10.67

Table A2 continued

Variable	Description	Units of Measurement	Mean	Standard Deviation	Frequency	%
Sample-Residents	Binary variable indicating the population surveyed were residents or landowners in close proximity or familiar with the water body being valued. The reference is for the general public in the area (county, watershed, or state).	Binary (Range: 0 or 1)	-	-	24	32
Sample-General Public in Area	The reference for the binary sample variables. General public in the area is a random sample of households within approximately 200 miles of the water source but with no information on use or interest.	Binary (Range: 0 or 1)	-	-	43	57.3
Location - Northeast or National	Binary variable indicating the location of the water body improvement being valued is in the Northeast. The locations are based on the USDA regional map. The reference is for the Southeast region.	Binary (Range: 0 or 1)	-	-	2	2.67
Location - MidAtlantic	Binary variable indicating the location of the water body improvement being valued is in the Mid- Atlantic region. The locations are based on the USDA regional map. The reference is for the Southeast region.	Binary (Range: 0 or 1)	-	-	17	22.67
Location - MidWest	Binary variable indicating the location of the water body improvement being valued is in the Mid-West region. The locations are based on the USDA regional map. The reference is for the Southeast region.	Binary (Range: 0 or 1)	-	-	22	29.33
Location - West	Binary variable indicating the location of the water body improvement being valued is in the Mountain Pacific or Western region. The locations are based on the USDA regional map. The reference is for the southeast region.	Binary (Range: 0 or 1)	-	-	7	9.33
Location - Southeast	The reference for the binary location variables, which is the Southeast region. The locations are based on the USDA regional map.	Binary (Range: 0 or 1)	-	-	27	36
* Income and WTP values adjusted to 2010 dollars.						

Appendix B: Quality Characteristics - Variable Coding Criteria

Table B1: Extent of Change Variable – Coding Criteria

Nominal Category	Value	Attributes/Criteria for Coding	
		Description of Water Quality Before Change	Description of Water Quality After Change
Unusable/ Unsupported to Usable/ Supported	1	*Possible attributes of unusable/unsupported water body** - High levels of toxic pollutants or runoff that have fundamentally altered the ecosystem of the water body. - High levels of turbidity or limited (3 feet or less of water clarity). - Swimming or wading is frequently prohibited or cautioned against. - Catching/eating of fish is frequently prohibited – fish catch advisories in place. - Limited number of aquatic species - flora or fauna – able to survive.	*Usable/Supported Water will have a change to 1 or more of the criteria below*** - Reduction in pollutants or runoff in the water body to allow recovery of some ecosystem functions. - Improvement in water clarity, permitting recreation or ecosystem function. - Swimming or wading permitted - Fish catch advisories removed - Diverse aquatic species – flora or fauna – able to survive. - Studies measuring change in miles or acres will be in the usable category if the number of miles/acres is less than 50%.
Unusable/ Unsupported to Fully Usable/ Supported	2	*Possible attributes of unusable/unsupported water body** - High levels of toxic pollutants or runoff that have fundamentally altered the ecosystem of the water body. - High levels of turbidity or limited (3 feet or less of water clarity). - Swimming or wading is frequently prohibited or cautioned against. - Catching/eating of fish is frequently prohibited – fish catch advisories in place. - Limited number of aquatic species - flora or fauna – able to survive.	*Fully Usable/Supported Water will have a change to all of the criteria below*** - Significant reduction in pollutant or runoff amounts to bringing ecosystem function to previously held levels. - Significant improvement in water clarity, permitting recreation and normal ecosystem function. - Swimming or wading permitted - Fish catch advisories removed - Diverse aquatic species flora or fauna – able to survive. - Studies measuring change in miles or acres will be in the usable category if the number of miles/acres is less than 50%.

Table B1 continued

<table>
<tr><th rowspan="2">Nominal Category</th><th rowspan="2">Value</th><th colspan="2">Attributes/Criteria for Coding</th></tr>
<tr><th>Description of Water Quality Before Change</th><th>Description of Water Quality After Change</th></tr>
<tr><td>Usable/Supported to Fully Usable/Supported</td><td>3</td><td>Possible attributes of fully usable/supported water body*
- Moderate levels of runoff hinder water clarity at peak times.
- Swimming or wading permitted.
- Occasional fish catch advisories in place.
- Limited numbers of aquatic species flora or fauna – able to survive.</td><td>Fully Usable/Supported Water will have a change in 1 or more of the criteria below**
- Significant improvement in water clarity, permitting recreation and normal ecosystem function.
- Swimming or wading permitted
- Fish catch advisories removed
- Diverse aquatic species flora or fauna – able to survive.</td></tr>
<tr><td colspan="4">* Typical changes occur over a period of 5 to 15 years. If the change described takes longer, then lower the rating from fully supported/ usable to usable/supported. The logic for this is that if the results of the policy or program take 20 to 40 years to materialize, the respondent's ability or willingness to link the projected changes to the specific policy or program will be less than were the changes to occur within 15 years.
** The number of criteria that must change ultimately depends on the scenario. If the scenario only considers water clarity, then a substantial change in water clarity will be considered as a change from unusable to fully usable.</td></tr>
</table>

Table B2: Inter Coder Reliability Results

	Percent Agreement	Scott's Pi	Cohen's Kappa	Krippendorff's Alpha (nominal)	N Agreements	N Disagreements	N Cases	N Decisions
Extent of Change Variable	81.40%	0.692	0.692	0.694	48	11	59	118

Table B3: Description of Quality Change – Coding Criteria

Nominal Category	Value	Example
Input	1	If a local referendum were held to vote on an annual tax for 30 years to finance the dredging and disposal of contaminated sediments from the Ashtabula River, where firm estimates of the total cost of dredging and disposal are not available, how much would you be willing to pay annually to complete this project for the next 30 years.[29]
Output	2	To fund these actions a South Platte river restoration fund has been proposed. All citizens along the front range from Denver to Fort Collins would be asked to pay an increased water bill (or rent if water is included in your rent) to: (1) purchase water from farmers to increase water for fish and wildlife from 17% shown in the top pie chart to 42% as shown on the lower pie chart (point to); (2) to manage the South Platte river as shown in the increased ecosystem services … along the 45 miles of the South Platte river shown on the map … The funds collected can only be used to restore natural vegetation along 45 miles of the South Platte river and purchase water from willing farmers to increase instream flow to improve habitat for six native fish so they are not in danger of extinction. If the South Platte river restoration fund was on the ballot in the next election and it cost your household $– each month in a higher water bill would you vote in favor or against?[30]

[29] Lichtkoppler & Blaine, 1999, p. 514.
[30] Loomis et al., 2000, p. 111.

Table B4: Disaggregated Endpoints – Coding Criteria

Nominal Category	Value	Attributes/Criteria for Coding
Bundled Endpoint	1	• Endpoint is single proxy or indicator for overall condition of water quality
Disaggregated Endpoint	2	• Two or more endpoints are used to describe overall condition of water quality. • Two or more attributes or characteristics are valued separately but all are presented to respondent as the available endpoints or ecosystem services from the environmental change.

Table B5: Open vs. Closed Valuation – Coding Criteria

Nominal Category	Value	Attributes/Criteria for Coding
Open Valuation	1	• Improvement described in general terms – no linkage to specific activity or quantification of change
Closed Valuation	2	• Quantitative metric or scale used to describe change • Multiple attributes tied to resource are valued • Change directly linked to discrete activity

Appendix C: Correlation Table

Table C1: Variable Correlations

	hhwtp	lnwtp	EC-Unusable to Usable	EC- Unusable to Full Use	Description of Change	Endpts	Disaggregated Endpts	Open vs. Closed Valuation
hhwtp	1							
lnwtp	0.8036	1						
EC-Unusable to Usable	-0.0463	-0.14	1					
EC- Unusable to Full Use	-0.0901	0.0090	-0.3388	1				
Description of Change	-0.0801	-0.0462	0.1682	-0.1611	1			
Endpts	0.1738	0.1261	0.0872	0.1109	-0.7532	1		
Disag. Endpts	-0.0903	-0.0299	0.0978	0.1732	-0.2923	0.3203	1	
Open vs. Closed Valuation	-0.2188	-0.1578	0.0978	0.2309	-0.1594	0.1601	0.7222	1
DC- Multiple Bounded	-0.0979	-0.1367	-0.0288	-0.0662	-0.2164	0.1286	-0.2039	-0.2039
Seq. Binary Choice	0.0423	0.1021	0.0105	0.1387	0.1678	0.1209	0.3774	0.2631
Sequential Binary Choice -NoRef	-0.0693	-0.0577	-0.1019	0.5064	-0.2159	0.1626	0.2770	0.2770
Open-Ended	-0.0932	-0.0627	-0.1137	-0.0420	-0.1236	0.0931	0.0485	0.2907
Payment Card	-0.1947	-0.1539	-0.1537	0.3565	0.1686	-0.2787	-0.2620	-0.1684
Payment Card_ No Ref	-0.1406	-0.1435	-0.0978	0.2887	0.2259	-0.3203	-0.1667	-0.0278
Published	0.0828	-0.0045	-0.2751	-0.2658	-0.1478	-0.0907	-0.1062	-0.1062
YearIndex	0.1233	-0.0060	0.1642	0.0898	-0.0481	0.0579	0.0301	-0.0181
Framing	0.1043	-0.0100	0.0052	0.1333	-0.1521	0.2809	0.0308	0.0308
PaymentTiming	-0.1115	-0.0220	-0.0151	0.1043	0.0864	-0.1735	0.0602	-0.0401
Income	0.1531	0.0277	-0.1550	-0.0395	0.0062	-0.1156	-0.4829	-0.1785
Users	-0.0652	-0.0185	0.0562	-0.1527	0.0309	-0.1186	0.0705	0.0705
Residents	0.2449	0.0652	0.1115	0.2425	-0.0084	0.2691	0.0233	0.1400
Loc_NE	-0.0134	0.0329	-0.0793	0.2341	-0.0862	0.0649	-0.1351	0.2027
Loc_MA	-0.1340	-0.0403	-0.1776	0.3603	-0.2042	0.1187	-0.0520	0.0130
Loc_MW	0.1383	0.0968	0.2175	-0.0828	0.2364	-0.0057	-0.2272	-0.2869
Loc_West	-0.0818	-0.0550	0.4345	-0.1296	-0.1671	0.1258	0.2994	0.2994

Table C1 continued

	DC-Multiple Bounded	Sequential Binary Choice	Sequential Binary Choice - NoRef	Open-Ended	Payment Card	Payment Card_ No Ref	Published	YearIndex
DC- Multiple Bounded	1							
Sequential Binary Choice	-0.4093	1						
Sequential Binary Choice - NoRef	-0.2330	0.5691	1					
Open-Ended	-0.1334	-0.1729	-0.0984	1				
Payment Card	-0.1803	-0.2337	-0.1330	-0.0762	1			
Payment Card_ No Ref	-0.1147	-0.1487	-0.0846	-0.0485	0.6362	1		
Published	0.2383	-0.0016	0.0305	0.1579	-0.3830	-0.3069	1	
YearIndex	-0.2904	0.0910	0.1642	-0.2979	-0.0169	-0.1658	-0.0299	1
Framing	-0.2083	0.3020	0.1719	0.0984	0.1330	0.0846	-0.1122	0.2783
PaymentTiming	0.1795	-0.2148	-0.1223	-0.0700	-0.0946	-0.0602	0.0895	0.0109
Income	0.2362	-0.4098	-0.3452	0.2747	0.1651	0.3088	0.1258	-0.1397
Users	-0.0930	-0.1609	-0.1433	0.3024	0.0376	0.1499	-0.0512	-0.1276
Residents	0.0161	0.1610	0.2004	-0.0356	-0.0236	0.0058	-0.1636	0.1963
Loc_NE	-0.0930	-0.1206	-0.0686	0.3290	0.2314	0.3885	-0.0694	-0.3544
Loc_MA	-0.0805	0.2079	0.5857	0.0132	0.2642	-0.1105	-0.0543	-0.0376
Loc_MW	-0.0192	-0.1616	-0.2671	-0.1529	-0.0054	0.1674	-0.1431	0.2616
Loc_West	-0.0730	-0.2337	-0.1330	-0.0762	-0.1029	-0.0655	-0.2836	0.1151

Table C1 continued

	Framing	Payment Timing	Income	Users	Residents	Loc_NE	Loc_MA	Loc_MW
Framing	1							
PaymentTiming	-0.4334	1						
Income	0.0727	-0.0449	1					
Users	-0.2230	0.0573	-0.0694	1				
Residents	0.2844	-0.2023	0.2684	-0.2370	1			
Loc_NE	0.0686	-0.0488	0.5613	-0.0572	0.2413	1		
Loc_MA	-0.0456	-0.1596	-0.2948	0.0193	0.1065	-0.0896	1	
Loc_MW	0.1015	0.1339	0.0794	-0.1278	0.1230	-0.1066	-0.3488	1
Loc_West	-0.1261	0.0743	-0.1123	0.1861	-0.2201	-0.0531	-0.1737	-0.2067

Appendix D: Outlier Analysis

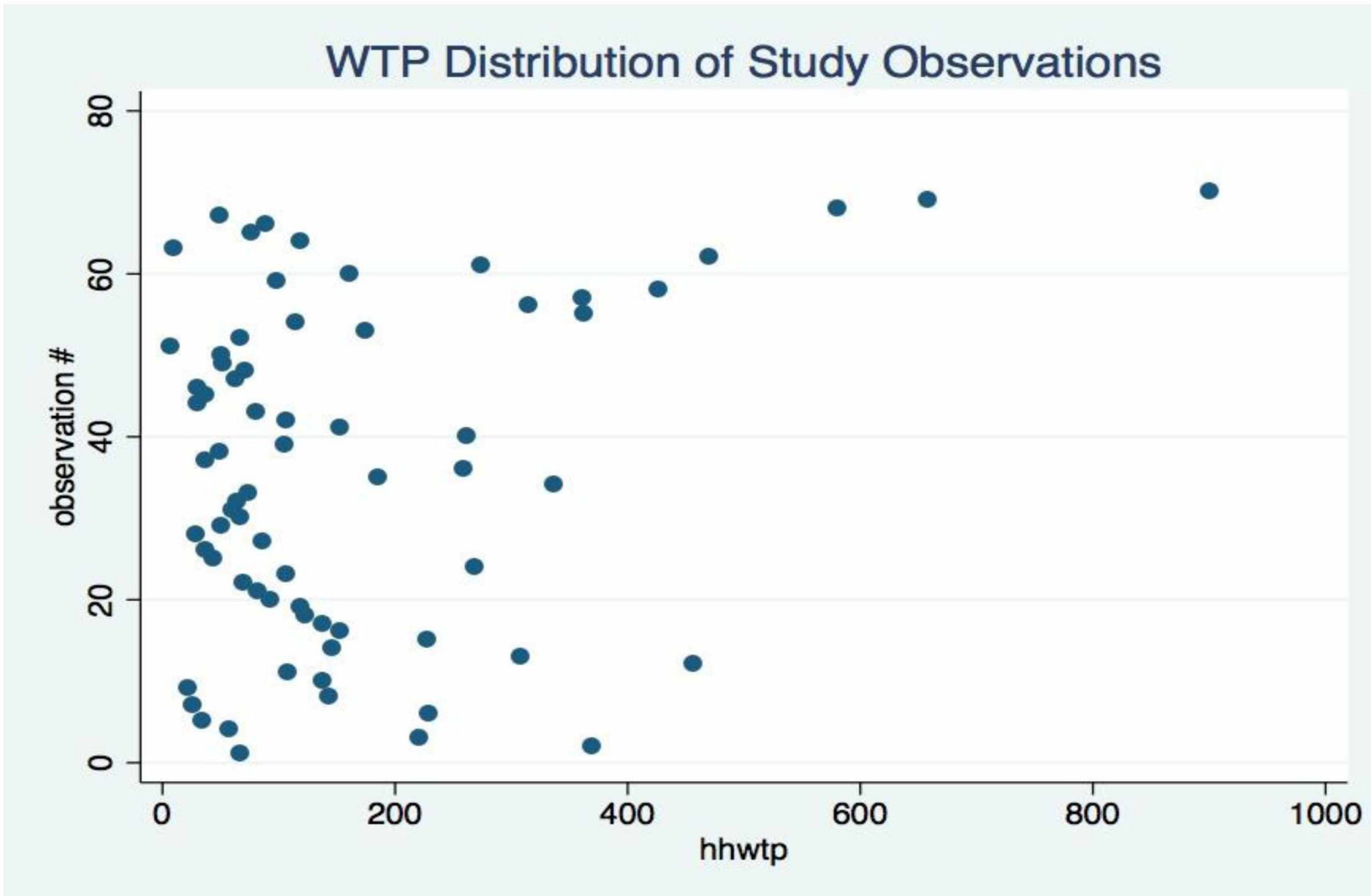

Figure D1: Willingness to Pay Distribution by Study

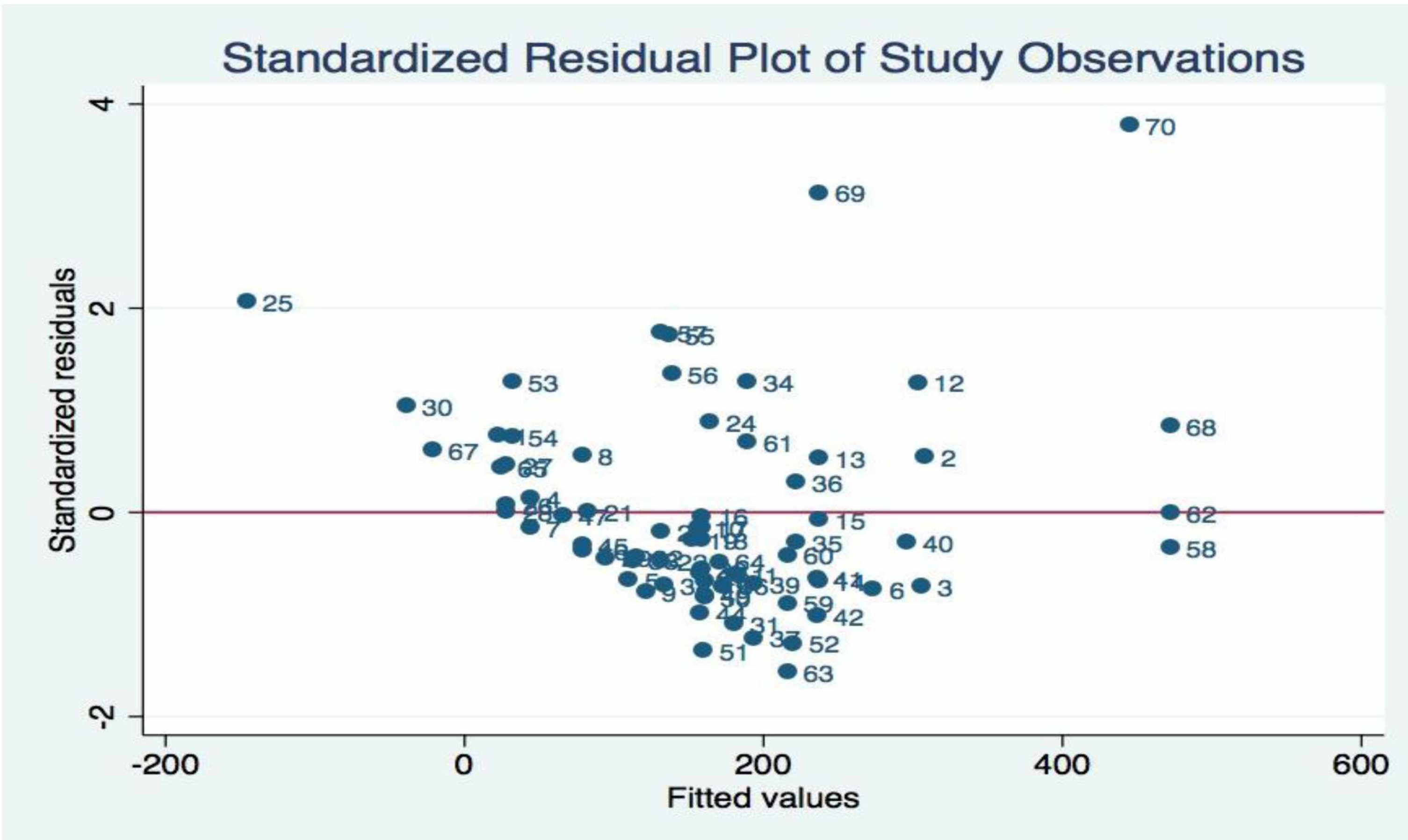

Figure D2: Standardized Residual Plot of Willingness to Pay Values by Study

Table D1: Observations Exceeding Conservative Outlier Diagnostic Value

Observation #	DFITs	Cut off > (2√(k/n))	Cook's D_i	Cut off > (4/n)
1	1.684155		0.1362787	
6			0.1362787	
25	2.574333	1.121	0.29418	0.057
69	1.279457		0.0637003	
70	2.902198		0.2896725	

Table D2: Review of Potential Outliers in Data Set

Obs #	Study	HH WTP	Decision	Justification
1	Wey (1990)	$67	Keep	Critical element of quality controlled for in study and heterogeneity needed.
6	Kaoru (1993)	$230	Keep	Critical element of quality controlled for in study and heterogeneity needed.
25	Lichtkoppler & Blaine (1999)	$44	Keep	Critical element of quality controlled for in study and heterogeneity needed.
69	Magat et al. (2000)	$659	Remove	Unable to account for variation in WTP with available variables.
70	Moore et al. (2011)	$901	Remove	Unable to account for variation in WTP with available variables.

Table D3: Statistical Performance of Full Model with and without Outliers

	Full Model -All Observations	Full Model - 2 Outliers Removed
Number of obs	70	68
F	2.05	2.36
Prob > F	0.0202	0.0076
R-squared	0.4730	0.5187
Adj R-squared	0.2424	0.2990

Appendix E: Alternative Elicitation Method Regressions

Table E1: Fully Reduced Model – Open-Ended Observations Removed

Regression Method	OLS - Cluster	OLS - Cluster	OLS - Cluster
Variables	Model 1e	Model 2e	Model 3e
Dependent Variable:	Ln HH WTP	Ln HH WTP	Ln HH WTP
Extent of Change – Unusable to Usable	-0.51 (0.422)	-0.70* (0.355)	-0.82** (0.344
Extent of Change – Unusable to Full Use	0.10 (0.425)	0.01 (0.409)	-0.14 (0.380)
Description of Quality Change	-0.42 (0.312)	-	-
Endpoints	-	0.71*** (0.237)	0.75** (0.296)
Disaggregated Endpoints	-	-	-0.19 (0.363)
Open vs. Closed Valuation	-0.49 (0.348)	-0.44 (0.297)	-
Multiple Bounded - DC	-0.73* (0.377)	-0.77** (0.347)	-0.74* (0.399)
Sequential Binary Choice	-0.06 (0.282)	-0.19 (0.217)	-0.23 (0.239)
Payment Card	-1.26** (0.533)	-1.18** (0.488)	-1.05** (0.494)
Year Index	-0.05 (0.034)	-0.05 (0.031)	-0.05 (0.033)
Income	.00001 (0.00002)	.00001 (0.00001)	.00001 (0.00002)
Constant	5.02*** (0.980)	4.46*** (0.982)	4.52*** (1.120)
Statistical Performance	# Obs. = 64 P > F = 0.0033 R-S = 0.28	# Obs. = 64 P > F = 0.0002 R-S = 0.316	# Obs. = 64 P > F = 0.0006 R-S = 0.280

*Significant at 10% level
** Significant at 5% level
***Significant at 1% level

Table E2: Fully Reduced Model – Single Dichotomous Choice versus All Other Elicitation Methods

Regression Method	OLS - Cluster	OLS - Cluster	OLS - Cluster
Variables	Model 1f	Model 2f	Model 3f
Dependent Variable:	Ln HH WTP	Ln HH WTP	Ln HH WTP
Extent of Change – Unusable to Usable	-0.58 (0.351)	-0.72** (0.335)	-0.79** (0.349)
Extent of Change – Unusable to Full Use	-0.10 (0.314)	-0.13 (0.296)	-0.21 (0.313)
Description of Quality Change	-0.31 (0.336)	-	-
Endpoints	-	0.75*** (0.248)	0.71*** (0.248)
Disaggregated Endpoints	-	-	-0.09 (0.329)
Open vs. Closed Valuation	-0.33 (0.230)	-0.35 (0.235)	-
All Elicitation (except Single DC)	-0.36 (0.261)	-0.45* (0.227)	-0.47* (0.251)
Year Index	-0.02 (0.023)	-0.02 (0.023)	-0.02 (0.024)
Income	3.09e-06 (8.65e-06)	4.36e-06 (7.65e-06)	3.96e-06 (8.13e-06)
Constant	5.25*** (0.719)	4.61*** (0.679)	4.58*** (0.744)
Statistical Performance	# Obs. = 68 P > F = 0.206 R-S = 0.167	# Obs. = 68 P > F = 0.0012 R-S = 0.225	# Obs. = 68 P > F = 0.002 R-S = 0.194

*Significant at 10% level
** Significant at 5% level
***Significant at 1% level