Department of Economics Working Paper

Number 21-13| October 2021

Estimating the Benefits to Florida Households from Avoiding Another Gulf Oil Spill Using the Contingent Valuation Method: Internal Validity Tests with Probability-based and Opt-in Samples

John C. Whitehead
Appalachian State University

Andrew Ropicki
University of Florida/Florida Sea Grant

John Loomis
Colorado State University

Sherry Larkin
University of Florida

Tim Haab
Ohio State University

Sergio Alvarez
University of Central Florida

Department of Economics
Appalachian State University
Boone, NC 28608
Phone: (828) 262-2148
Fax: (828) 262-6105
www.business.appstate.edu/economics

Estimating the Benefits to Florida Households from Avoiding Another Gulf Oil Spill Using the Contingent Valuation Method: Internal Validity Tests with Probability-based and Opt-in Samples

John C. Whitehead, Appalachian State University

Andrew Ropicki, University of Florida/Florida Sea Grant

John Loomis, Colorado State University

Sherry Larkin, University of Florida

Tim Haab, Ohio State University

Sergio Alvarez, University of Central Florida

October 28, 2021

Corresponding Author: John C. Whitehead, Department of Economics, Appalachian State University, Boone, NC; email: whiteheadjc@appstate.edu; phone: (828)262-6121

This study was funded by the Florida Legislature, Office of Economic and Demographic Research, Project#: 00091735. A previous version of this paper was presented at the 2019 Society for Benefit-Cost Analysis Annual Meeting in Washington, DC.

Estimating the Benefits to Florida Households from Avoiding Another Gulf Oil Spill Using the Contingent Valuation Method: Internal Validity Tests with Probability-based and Opt-in Samples

Abstract

This paper evaluates the importance of contingent valuation method data quality by examining differences in results between probability-based and opt-in internet samples. Our data is from a survey estimating passive use losses associated with the *BP/Deepwater Horizon* oil spill to Florida residents. Several internal tests of validity are conducted. We find that the willingness to pay estimates from the opt-in sample may be biased upwards and only the probability-based sample data pass the scope test. In general, we conclude that the probability-based sample data is of higher quality.

Key words: contingent valuation, scope test, probability-based sample data; opt-in sample data
JEL: Q51

Introduction

The *BP/Deepwater Horizon* (BP/DWH) oil spill began on April 20, 2010. Prior to the capping of the well on September 19, 2010, an estimated 3.19 million barrels of oil were spilled into the Gulf of Mexico. The *BP/Deepwater Horizon* spill was the largest oil spill in U.S. history and approximately 12 times larger than the Exxon Valdez spill in Alaska. The Gulf of Mexico is a complex ecosystem. Near shore estuaries and coastal habitats provide a suite of services that society uses and values directly and indirectly, such as for fisheries, tourism, water management, and amenities to coastal property owners. Part of the losses due to oil spill damage may be the value of offshore natural areas and wildlife. Some of this value lies with the passive-use value that society may hold for knowing these areas and wildlife exist in a healthy state currently and in perpetuity for future generations. As such, all residents can have passive use values for a healthy Gulf marine environment and all the direct and indirect services it can provide to nature and society. This paper arises from a natural resource damage assessment conducted for the State of Florida after the *BP/Deepwater Horizon* oil spill. In Huffaker, Clouser and Larkin et al. (2012) we estimated the passive use value of avoiding damages from another large oil spill in the Gulf of Mexico with the contingent valuation method.

The use of the contingent valuation method (CVM) as a tool for natural resource damage assessment was spurred by two major events. First, the 1989 U.S. Appellate Court opinion, *Ohio v. Department of Interior*, stated that (a) passive use losses were compensable under the Clean Water Act and the Comprehensive, Environmental Response, Compensation and Liability Act (CERCLA), and (b) the Department of Interior's ranking of damage assessment techniques, which had CV at the bottom, was unjustified (Carson, 2000). Second, the passage of the Oil Pollution Act of 1990 led to regulations enacted by the National Oceanic and Atmospheric

Administration (NOAA) which stated: "NOAA believes that the trustee(s) should have the discretion to include passive use values as a component within the natural resource damage assessment determination of compensable values" (Carson et al., 2003).

These events, along with the increased use of stated preference methods to evaluate public policy options, led to the need to standardize and improve the CVM. As a result, NOAA convened a panel of experts - including two Nobel Prize winners in economics - to evaluate the CVM with regards to its ability to accurately value non-market goods (Arrow et al., 1993). The panel concluded that the CVM can produce reliable estimates of passive use values for the purpose of natural resource damage assessments (Carson, 2000). The panel proposed several basic guidelines on the effective use of CVM in valuing non-market goods and passive use values from environmental goods in particular, which have since become common practice. The NOAA Panel "guidelines for value elicitation surveys" included 12 items including that CVM studies should examine WTP responses relative to key determinants (e.g., income, scope). The NOAA Panel recommended that in-person and telephone surveys had advantages over the mail survey mode.[1]

Evaluating the Use of Internet Surveys as a CVM Data Collection Method

Since the NOAA Panel made their recommendation, a number of changes in data collection methods have occurred. Data collection includes mail, in-person, telephone and internet surveys (Champ 2017). There are benefits and costs of each of these survey modes.

---

[1] Page 30: "The Panel believes it unlikely that reliable estimates of values could be elicited with mail surveys. Face-to-face interviews are usually preferable, although telephone interviews have some advantages in terms of cost and centralized supervision."

Internet surveys provide many advantages over other survey models (Thurston 2006, Lindhjem and Navrud 2011). They are self-paced, allow visual aids and question branching (automatic survey skips based on previous answers) and piping (insertion of previous answers into subsequent questions). Several studies have compared internet survey samples with more traditional survey methods. For example, Berrens et al. (2004) find that the probability of an "advisory vote for ratification" of the Kyoto Protocol is lower for internet samples relative to a telephone sample. Banzhaf et al. (2006) find little difference between mail and internet samples. Boyle et al. (2016) find that internet samples lead to lower willingness to pay values relative to a mail sample. Lindhjem and Navrud (2011) reviewed the stated preference literature and find that internet panel data quality is no lower than more traditional survey modes but internet panel willingness to pay estimates are lower.

Internet surveys can use either probability-based or non-probability (i.e., opt-in, convenience) based samples of respondents. Internet surveys with opt-in panel samples are less expensive than probability-based samples and likely the least expensive of all survey modes. The drawback of opt-in panel data is that it may be of relatively low quality as some lowly compensated opt-in panel respondents pay little attention to the details of the valuation questions (e.g., Giguere, Moore and Whitehead 2020). Johnston et al. (2017) assert that high quality samples use probability-based sampling and the Dillman method, with repeated contacts, for internet surveys. Probability-based internet panels are more expensive but respondents may pay more attention to the surveys and may generate higher quality data. Recently, Sandstrom et al. (2021) compare two opt-in panels, MTurk and Qualtrics, with a mixed mode mail/internet sample. They find that each sample produces valid results but there are differences in the survey responses to the program cost and scope variables across samples.

Since the CVM is based on responses to hypothetical valuation questions, there have been concerns about the accuracy of value estimates (Bishop and Boyle 2019). Accuracy of a measure of a theoretical construct (e.g., willingness to pay) is comprised of validity and reliability. Validity is the extent to which a valuation method generates a measure that is unbiased, that is, provides an estimate centered around the true value, if it were known. There are several types of validity that have been considered in the CVM literature. Theoretical validity, also known as internal validity because the tests conducted are internal to the data, begins with economic theory. Comparative static results are derived from the willingness to pay function as derived from the assumed underlying preference structure. For example, willingness to pay can be shown to increase with income (for normal goods) and the scope of the policy. Standard rational choice theory suggests that the probability of a referendum vote in favor of a policy should be sensitive to the cost of the policy.

In this paper we evaluate the importance of data quality by examining differences in internal validity tests between probability-based and opt-in panel data in an internet survey. We use data collected for a natural resource damage assessment conducted for the State of Florida after the *BP/Deepwater Horizon* oil spill. We consider three internal validity tests. The first is to determine the sensitivity of hypothetical referendum votes to the cost of the policy. Theory predicts an inverse relationship between the dollar amount households are asked to pay and their likelihood of voting for the program. We test this by comparing the probability of referendum votes in favor of the policy, the coefficients in simple logistic regression models and the willingness to pay estimates and their standard errors. The second internal validity test is for sensitivity to the scope of the policy, specifically that respondents' likelihood of voting for the program increases with the proposed level of protection. To test for scope effects, we compare scope elasticities across samples. The third internal validity test is a comparison of income elasticities of willingness to pay across samples.

We find that both samples (probability-based and opt-in) pass the cost test but the willingness to pay estimates are higher in the opt-in sample. We find that only the probability-based sample passes the scope test. We find that the income elasticity is larger for the probability-based sample but the differences are not statistically significant. In general, we conclude that the probability-based sample data exhibits greater internal validity.

Questionnaire and Data

The development of the survey took place in a number of stages. The first stage involved gathering information on the *BP/Deepwater Horizon* oil spill and its effects on the Gulf of Mexico. The second stage involved holding an initial round of focus groups around the state to determine Floridians awareness of, and sentiment towards, the oil spill and its effects on the Gulf of Mexico. The focus groups were also used to identify the most effective type of hypothetical scenario and the scope of the program to adopt for the valuation question. In addition, we also conducted a mall intercept survey to focus on details of the hypothetical program and bid vehicle. A second round of focus groups were then conducted in population centers away from areas most heavily impacted by the spill in order to get feedback from more dispassionate residents. At the conclusion of the focus groups we decided that an Internet survey was the only viable mode of implementation given the sequential and graphical nature of the questions and supporting materials. Once a questionnaire designed for the Internet was developed, the last stage involved formal pre-testing with the completion of several cognitive interviews to test for any logical issues with the instrument and then a formal pilot survey with 543 opt-in sample respondents to finalize the bid values.

The questionnaire contained five sections. The first described the *BP/Deepwater Horizon* oil spill. The second section of the questionnaire provided a detailed description of the hypothetical program that was developed to elicit respondents' stated preferences for a restored Gulf environment. The two main goals of the program were to quickly stop future leaks and continuous monitoring for surface and subsurface oil. Then the means to achieve this goal were described: (a) procurement of five ships to be operated by the U.S. Coast Guard and designed to quickly stop and clean up spills based on information learned from the "2010 Gulf oil spill" (a picture of a similar ship was shown) and (b) oil monitoring and detection, (pictures of two types of equipment were shown). The amount of detail was considered necessary given the comments received from pre-survey focus groups. Respondents were told that the U.S. Coast Guard would be designated the lead agency in addressing oil spill monitoring and clean up and that all personnel would be trained to fulfill these new responsibilities.

In order to evaluate whether WTP increased with the scale or scope of the stated outcomes of the program respondents were shown estimates of how effective scientists estimated the program to be. Effectiveness was defined as the percentage reduction in environmental impacts, relative to the 2010 oil spill, and this percentage was randomly selected to be either 20%, 45%, 70% or 90%. This information was summarized in a box with the number of dead animals (birds, sea turtles and marine mammals) and miles of oiled coastline (presented earlier), one of the randomly selected effectiveness levels (i.e., percentage reduction of impacts), and the corresponding number of coastline miles that would not have been oiled and number of animals that would not have died if the program had already been established. In order to put the benefits of the program into context and provide neutrality in the description, respondents were reminded that "the number of most animals it would protect is small compared to their total numbers in the

Gulf of Mexico." To test the sensitivity to the different levels of cost of the program respondents were told that the federal government is considering a range of programs that differ based on how effective they are expected to be and how much they would cost. Respondents were then shown a bar graph with the four percentages, including the effectiveness level they were asked to evaluate, and zero percent effective for no program.

Lastly, the discussion moved toward program funding and the payment vehicle. The description of what is and is not required of oil companies under the Oil Pollution Act (OPA) was summarized to convince respondents that the program is feasible. In particular, we stated that it is not legally possible to make oil companies pay for the upfront costs to establish the program (i.e., purchase the U.S. Coast Guard equipment) but oil leasing fees could be quickly increased and maintained in the long run to cover the ongoing maintenance costs. Thus, if a vote of the general public approved of the establishment of this program in November 2012, all federal income tax filers would be assessed a one-time fee (which would be identical for everyone) payable directly to the U.S. Coast Guard. These funds would cover the cost of the new equipment and training only; the ongoing maintenance costs would be paid with the higher oil lease fees. Payment of ongoing maintenance costs through higher oil lease fees was included in the program to decrease protests from respondents; pre-survey focus groups indicated that some respondents believed oil companies should share responsibility for financing the program.

In order to stress that there were no "right or wrong" answers to the referendum vote question, the questionnaire provided respondents with a table of valid reasons for deciding whether to vote in favor or against the program. Specifically, the table listed seven reasons for and seven reasons against. Respondents were then asked if any of the reasons included how they felt about the proposed program. Before asking how they would vote, respondents were provided with "cheap talk," which is a script designed to encourage them to give an unbiased response. Respondents were then provided an opportunity to review information presented earlier via pop-up boxes, and they were reminded that there is "no right or wrong answer". Respondents were also reminded that when deciding how to vote on the program at the cost they were asked to pay to consider: (a) their income and budget; and (b) all the other environmental causes that they currently or plan to support. The contingent valuation question was worded as follows:

> *If an election were held today, would you vote for, or would you vote against the funding of a U.S. Coast Guard program to reduce environmental impacts of another large Gulf oil spill by X% if a <u>one-time</u> payment of $A would be added to your household's federal income tax?*

The effectiveness level, X%, was randomly selected from one of the four levels defined earlier (i.e., 20%, 45%, 70% or 90%) and the bid level, $A, was randomly selected among eight levels ranging from $10 to $385 that were chosen based on results of a pre-test.

This section of the questionnaire ended with a sequence of questions (seven in total) designed to determine what they thought about future oil spills and various aspects of the proposed program. These questions also address the perceived consequences of the program and objectiveness of the survey and include (1) their best estimate of the likelihood of another large spill happening, (2) their opinion of the environmental impacts of another large spill without the program, (3) their belief of how effective the program would be compared to the scientists' estimate that they were asked to consider, (4) how often they thought they would have to make the payment, (5) whether they thought the survey pushed them to vote one way, (6) whether and, if so, how strongly they believed the survey results would affect oil spill monitoring or clean up decisions by the U.S. Coast Guard in the Gulf of Mexico, and (7) how much confidence they have in the Federal government's ability to reduce the impacts of oil spills.

Sample Design

All households in the State of Florida constitute the study population. At the time of the survey, Knowledge Networks (KN) maintained a panel of respondents that had been scientifically recruited to represent the Florida population.[2] For purposes of this survey, KN invited 2,088 panelists of which 1,280 (61.3%) clicked the link to begin the questionnaire. KN obtained an additional 767 responses from an "opt-in" sample for a total of 2,047 respondents. KN did not disclose the source of the opt-in panel respondents to the researchers. Knowledge Networks gave no indication as to the content of the survey at the time that the probability-based and opt-in samples were recruited. Respondents were not told who funded the survey, but were directed to contact Knowledge Networks at a toll-free number that was provided if they wanted the contact information for the investigators; however, no calls were received. All respondents completed the survey with a median time of 25 minutes. All surveys were completed between September 20 and September 29, 2011, and all respondents received a cash-equivalent $5 incentive due to the relatively long duration of the survey.[3]

In order to detect any complete rejection of the premise of the program, two survey questions were asked to obtain information on the perceived probability of another large oil spill

---

[2] Knowledge Networks became known as GfK Knowledge Panel. GfK's Knowledge Panel was acquired by Ipsos in 2018. For the remainder of the paper we will refer to Knowledge Networks.

[3] Knowledge Networks provided weights for aggregating to the population. Since this analysis is a comparison between the opt-in and probability-based samples, and Knowledge Networks stated that their weights were not valid for analysis of the separate samples, we forgo their use here.

in the Gulf of Mexico. Respondents who voted against the government program were asked: "Why did you decide to vote against the program?" Those who answered "I don't believe another large spill will happen because companies will voluntarily improve" were removed from the analysis below. Respondents were also asked: "With oil drilling resumed in the Gulf and continuing to move into deeper waters, what is your best guess of the chances of another large oil spill in the Gulf of Mexico in the next 10 years?" Those who answered "0%; I don't think there is any chance of another large spill" were also removed from the analysis. In total, 130 observations (6.4% of the full sample) were deleted for indicating that they believed there was no possibility of a future spill, as these were considered protest no votes or a rejection of the scenario. With other observations dropped for incompleteness of the respondent's survey, the sample size used for analysis is 1840 with 37% percent of those in the opt-in sample.

Results

Table 1 describes the variables and Table 2 summarizes the data used in this analysis. The number who voted in favor of the program is 58% for the opt-in sample and 51% for the probability-based sample. The proportions of votes are statistically different ($p<0.01$). The average bid amount is $173 and $166 for the opt-in and probability-based samples. These variables will be discussed further in Table 3.

The average scope level is 57% and 56% for the opt-in and probability-based samples. To account for how effective the respondents believed the program would be, we asked: "Scientists estimate that the program you evaluated would reduce the environmental impacts by [X%] from another large oil spill in the Gulf of Mexico." Respondents who believed that the proposed program would be "A lot more effective than stated" or "Somewhat more effective than stated" are coded as 1 for the "more scope" variable and 0 otherwise. Respondents who answered

"Somewhat less effective than stated" or "A lot less effective than stated" are coded as 1 for the "less scope" variable and zero otherwise. Thirty-two percent of opt-in respondents felt that the program would be more effective than scientists estimate in reducing a future spill (more scope) while only 18% of the probability-based sample felt that the proposed program would be more effective ($p<0.01$). Twenty-seven percent of the probability-based sample thought that the program would be less effective while 22% of the opt-in sample felt this way ($p<0.01$).

The opt-in respondents' best estimate of the likelihood of another large spill happening is 54%. The probability-based respondents best estimate of the likelihood of another large spill happening is similar, 50%, but the difference is statistically significant using a χ^2 test for differences in frequencies ($p<0.01$) and also a t-test when treating the levels of probability in the survey question as continuous ($p<0.01$). The proportion of respondents who visit the Gulf of Mexico for the purpose of saltwater-based recreation is about two-thirds and this is not different across samples.

The income variable measured the respondents' annual household income. This information was collected by Knowledge Networks and was provided in a closed-ended response format that included 19 ranges from "$0 to $4,999" through "more than $175,000." For the analysis, respondent income was assumed to be at the midpoint of their income range (e.g., $2,500 for range 1). For the highest category (more than $175,000) income was assumed to be $175,000. The average income is $48 thousand for the opt-in sample and $59 thousand for the probability-based sample. This difference is statistically significant at the $p<0.01$ level.

To address consequentiality we investigated whether respondents believed their responses to the survey could impact policy decisions, which the CVM literature has shown to be important for valid responses (e.g., Carson and Groves, 2007, Herriges et al., 2010). To measure consequentiality we asked respondents to indicate their disagreement or agreement with the following statement: "I believe the results of this survey will affect decisions about oil spill monitoring and cleanup by the U.S. Coast Guard in the Gulf of Mexico." The results show that a majority of respondents believed the survey results would affect decisions. Although 35% of respondents were unsure of the impact the survey would have, only 10% felt the survey results would have little or no impact. We find that 66% of the opt-in respondents perceive the survey to be consequential while only 53% of the probability-based respondents find the survey to be consequential. The frequencies are statistically different at the $p<0.01$ level.

Statistical Tests

Table 3 shows the number and proportion of respondents that voted in favor of the policy at each bid amount (Vote in favor = 1). For both samples the proportion of votes in favor decline with the bid amount ($A) ($p<0.01$) but the relationship is stronger for the probability-based sample ($\chi^2 = 43.72[7\ df]$) than the opt-in sample ($\chi^2 = 23.12[7\ df]$). A negative relationship between the cost and support for the policy is a simple validity test that is easily passed when considering the entire range of costs presented to respondents. Two deeper issues are non-monotonicity (Haab and McConnell, 2002) and fat tails (Parsons and Myers, 2016). Non-monotonicity exists when the support for the policy rises once or more as the cost increases in pairwise comparisons. The opt-in data exhibits two instances of non-monotonicity (at A = $235, $385) while the probability-based data exhibits one instance (at A = $285). Fat tails exists when support for the policy at the upper end of the cost distribution is not responsive to the level of costs. It appears that the fat-tails problem exists in both datasets but it is slightly worse in the opt-in data with a flat portion of the bid curve over the $185 to $385 bid range. The probability-based data flattens over the $235 to $385 bid range.

To further test for the sensitivity of the referendum votes to the bid amounts, we estimate simple logistic regression linear and log-linear models (Table 4). In each of these models the bid amount is negative and statistically different from zero. The log-linear model provides better model fit with the opt-in data with higher model χ^2 statistic and higher Pseudo R^2. The model fit diagnostics are similar with the probability-based data across functional form. A likelihood ratio test indicates that the coefficients are jointly different across the opt-in and probability-based samples for the linear ($\chi^2 = 9.77[2\ df]$) and log-linear ($\chi^2 = 10.46[2\ df]$) models.

The willingness to pay (WTP) estimates are presented in Table 5. We report the mean and truncated mean WTP estimates from the linear model and the median WTP from the log-linear model (Hanemann 1984). The standard errors are developed with the Delta Method (Cameron 1991). The mean WTP and truncated WTP estimates from the opt-in data are 66% and 34% higher than those from the probability-based data, respectively. The median WTP estimate from the opt-in data is 149% higher than that from the probability-based data. The difference in the mean WTP estimates that allows for zero WTP is statistically different from zero at the $p<0.05$ level ($t=2.60$). The difference in the truncated mean WTP estimates that do not allow for zero WTP is not statistically different from zero ($t=1.29$). The difference in the median WTP estimate from the log-linear model is statistically different from zero at the $p<0.10$ level ($t=1.90$).

One reason that the large difference in the truncated mean WTP estimates is not statistically different from zero is the non-monotonicities and fat tails exhibited in both data sets.

Both problems decrease the precision surrounding the point estimate of willingness to pay.[4] As observed above, these two problems are somewhat greater with the opt-in data. This leads to standard errors of WTP that are disproportionately greater for the opt-in data. The standard errors from the opt-in WTP estimates are 113%, 109% and 327% greater than those from the probability-based WTP estimates from the mean, truncated mean and median estimates, respectively.

To test for sensitivity of responses to several factors that are *a priori* expected to have an impact on the responses, multivariate logistic regression analysis was next performed. The logit model estimated in this analysis included seven independent variables: two variables to capture bid and scope sensitivity and five others designed to capture other attitudinal, behavioral, and socio-economic variables expected to influence a respondents' vote (Table 5). We used the log-linear functional form for two reasons. First, the log-linear model is statistically preferred with the opt-in data. Second, it simplifies the calculation of scope and income elasticities.

The results show that five of the seven independent variables in the opt-in data model were statistically significant at the $p<0.05$ level and the effects on referendum votes have expected signs. Six of the seven independent variables are statistically significant at the $p<0.05$ level or lower in the probability-based model with expected signs. The remaining coefficient on visits is statistically significant at the $p<0.10$ level in the probability-based model. The probability-based model exhibits a greater model χ^2 statistic but this is likely due to the greater

[4] Note that the smaller sample size is another reason for the larger standard errors with the opt-in data. A full investigation of the relative contributions of sample size and data quality to the standard error is beyond the scope of this paper.

sample size. In both models the coefficient on the log of the bid amount is negative and statistically different from zero.

Sensitivity to the scope of policy exists if willingness to pay is nondecreasing in quality or quantity. In Table 6, the coefficient on the scope variable is statistically different from zero at the $p<0.01$ level (one-tailed test) and positive in the probability-based model but not statistically different from zero in the opt-in model.[5] This indicates that opt-in survey respondents do not exhibit sensitivity to the effectiveness of the proposed government program. In each model the coefficients on the perceived effectiveness of the program are statistically different from zero at the $p<0.01$ level. Using the coefficients on scope in Table 6 on “More scope” and “Less Scope”, respondents who believe that the program would be more effective than asserted in the questionnaire are 3.91 and 2.77 times more likely to vote in favor of the program in the opt-in and probability-based models, respectively. Respondents who believe that the program would be less effective than asserted in the questionnaire are 2.23 and 3.40 times less likely to vote in favor of the program in the opt-in and probability-based models.

Since willingness to pay is shorthand for willingness and ability to pay, it could be expected that the probability of a vote in favor of the government program will be positive if the government program is a normal good (McConnell 1990). The coefficients on the log of household income are both positive and statistically different from zero. A vote in favor of the program should increase with the perceived probability that another spill will occur. The

[5] In probability-based data models with randomly chosen sample sizes of n=680 (the sample size of the opt-in sample) the mean coefficient on ln(scope) is statistically different from zero at the p=0.0503 level in a one-tailed test (n=40 random draws).

probability of a vote in favor of the government program is increasing with the perceived probability of a future oil spill in both models. If the WTP to avoid another spill includes both use and passive use values, then resource users may be willing to pay more than non-users. Respondents who visit the Gulf of Mexico for the purpose of saltwater-based recreation are 1.26 times more likely to vote in favor of the program in the probability-based data model. There are no apparent use values in the opt-in data model.

Scope and income elasticities are presented in Table 7. Whitehead (2016) proposes a measure of scope elasticity to assess the plausibility of scope estimates: $e_S = \frac{\%\Delta WTP}{\%\Delta Scope}$. In the logistic regression model, the ratio of the log of scope coefficient to the log of bid coefficient can be interpreted as the scope elasticity. The scope elasticity is 0.77 and is statistically different than zero in the probability-based data model and is 0.27, but not statistically different from zero, in the opt-in data model. The probability-based data scope elasticity estimate suggests that for each 10% increase in scope, WTP increases by 7.7%.

Hanemann (1991) develops a theoretical model of the income elasticity of willingness to pay (defined as a virtual price). Flores and Carson (1997) find that this income elasticity is similar to the income elasticity of demand but are unable to develop bounds similar to the income elasticity of demand (e.g., income elasticity of demand greater than one is considered a luxury good). Kristrom and Riera (1996) find that the income elasticity of WTP is less than one for a number of contingent valuation data sets. The income elasticities, measured as the ratio of the income and bid coefficients, are 0.63 and 1.08 in opt-in and probability-based data models. The opt-in income elasticity is consistent with the findings from Kristrom and Riera. But, neither elasticity estimate is statistically different from one.

Conclusions

The goal of this paper is to provide a test of the internal validity of two different but commoly used sources of data for a CVM survey: probability-based samples and opt-in samples. The data was originally developed to estimate the passive use losses suffered by Floridians due to the 2010 *BP/Deepwater Horizon* oil spill. A questionnaire was developed and implemented that adhered to several standard principles resulting from both the NOAA Blue Ribbon panel, state-of-the-art recommendations from Carson, et al. (2004), and the pre-2011 peer-reviewed literature on the CVM. Data collection was performed using Intenet surveys conducted by what was then known as Knowledge Networks, Inc. The data includes Knowledge Networks probability sample as well as an opt-in sample. These two types of Internet survey data allows for testing the performance of these two types on several key elements of internal validity of CVM responses.

Our results indicate a sensitivity of the referendum vote responses to the dollar bid amounts respondents were asked to pay in both types of Internet surveys, However, the percentage of responses in favor of the government program is greater in the opt-in sample. This leads to larger willingness to pay estimates relative to the probability-based data. While both samples suffer from non-monotonicities and fat tails, the problems are slightly worse with the opt-in sample data. These problems increase the standard errors associated with the willingness to pay estimates.

The probability-based logistic regression model exhibits greater internal validity relative to the opt-in model. In particular the probability-based model exhibits sensitivity to scope while the opt-in panel data does not. Somewhat surprisingly, we find that the opt-in sample respondents find the survey to have greater consequentiality than the probability-based sample respondents. Therefore, the differences described above are in spite of greater consequentiality which is typically considered to enhance the validity of stated preference studies.

One limitation of our study is that the comparison is based on two data sets that have different sample sizes. This was due to the original data collection constraints. The lower sample size for the opt-in data leads to larger standard errors and complicates comparisons for the internal validity tests. Future research that compares probability-based and opt-in samples should address this shortcoming by considering sample size as part of the study design. Nonetheless our analysis and findings are of growing relevance as opt-in samples have been increasingly used due to their low cost (Johnston, 2021).

Further studies are needed to determine if our findings are prevalent in other CVM studies. Given their low cost and accessibility to researchers without deep pockets, it is likely that opt-in data will continue to be used. These opt-in panel data are useful to explore methodological issues in stated preference studies as well as to demonstrate new empirical estimation methods. Several innovations in stated preference research since the data used in this paper were collected can be used to increase the validity of models estimated with opt-in data. For example, Giguere, Moore and Whitehead (2020) find that repeated questions and attribute non-attendance methods uncover sensitivity to scope with opt-in panel data. In addition, respondent attention check and pre-screening questions should be used to eliminate “speeders” and other respondents who do not take the survey task seriously. More attention to these innovations is warranted.

Table 1. Variables	
Label	Description
Vote	Equal to 1 if respondent answered "For" to the referendum vote question, 0 otherwise
Bid	Randomly selected one-time tax payment added to the household's federal income tax; $10, $45, $85, $135, $185, $235, $285, or $385
Scope	Estimated impacts of another similar size Gulf oil spill with the proposed program would be X% lower, X% equals 20%, 45%, 70% or 90%
More scope	Equal to 1 if the respondent believed the program would be a lot or somewhat more effective than stated, 0 otherwise
Less scope	Equal to 1 if the respondent believed the program would be a lot or somewhat less effective than stated, 0 otherwisee
Future spill	Response to the question "… what is your best guess of the chances of another large oil spill in the Gulf of Mexico in the next 10 years?" The possible responses used in this study were 25%, 50%, 75% and 100%
Visit	Equal to 1 if the respondent spent one or more days at coastal areas on the Gulf of Mexico for saltwater-based recreation, 0 otherwise
Income	Household income ($1000)

Table 2. Data Summary				
	Opt-in		Probability-based	
	Mean	Std dev	Mean	Std dev
Vote (1=in favor)	0.58	0.49	0.51	0.50
Bid ($)	172.71	121.13	165.75	119.23
Scope (%)	57.09	26.09	56.34	26.04
More scope (0,1)	0.32	0.47	0.18	0.39
Less scope (0,1)	0.22	0.42	0.27	0.45
Future spill (%)	53.53	23.94	49.85	23.44
Visit (0,1)	0.67	0.47	0.64	0.48
Income ($1000)	48.00	34.94	58.68	39.67
Sample Size	680		1160	

Table 3. Distribution of responses for the vote in favor variable (For) by bid value ($)

	Opt-in			Probability-based		
Bid	For	Total	%For	For	Total	%For
10	67	89	75.28	100	151	66.23
45	59	87	67.82	98	160	61.25
85	44	73	60.27	80	144	55.56
135	54	91	59.34	81	148	54.73
185	37	74	50.00	70	149	46.98
235	43	83	51.81	53	134	39.55
285	46	95	48.42	54	132	40.91
385	45	88	51.14	56	142	39.44
Total	395	680	58.09	592	1,160	51.03

Table 4. Logistic Regression of Determinants of Referendum Votes: Dependent Variable = For						
	Opt-in					
	Coef.	Std. Err.	t-stat	Coef.	Std. Err.	t-stat
Constant	0.800	0.141	5.69	1.936	0.366	5.28
Bid	-0.0027	0.00065	-4.13			
Ln(Bid)				-0.338	0.074	-4.55
$\chi^2(1)$	17.37			22.31		
Pseudo R^2	0019			0.024		
Sample size	680			680		
	Probability-based					
	Coef.	Std. Err.	z	Coef.	Std. Err.	z
Constant	0.561	0.103	5.42	1.609	0.266	6.04
Bid	-0.0031	0.00051	-6.15			
Ln(Bid)				-0.334	0.055	-6.07
$\chi^2(1)$	39.14			38.83		
Pseudo R^2	0.024			0.024		
Sample size	1160			1160		

Table 5. Willingness to Pay Estimates						
	Opt-in			Probability-based		
	Coef.	Std. Err.	z	Coef.	Std. Err.	z
Mean	296.63	40.95	7.24	178.84	19.19	9.32
Truncated Mean	434.27	78.11	5.56	322.88	37.44	8.62
Median	307.74	94.62	3.25	123.46	22.15	5.57

Table 6. Logistic Regression of Determinants of Referendum Votes: Dependent Variable = For						
	Opt-in			Probability-based		
	Coef.	Std. Err.	z	Coef.	Std. Err.	z
Constant	0.138	0.826	0.17	-1.637	0.638	-2.57
Ln(Bid)	-0.383	0.080	-4.81	-0.379	0.059	-6.39
Ln(Scope)	0.104	0.152	0.68	0.291	0.116	2.51
More scope	1.365	0.213	6.41	1.019	0.183	5.57
Less scope	-0.803	0.215	-3.74	-1.226	0.158	-7.77
Future spill	0.0108	0.0036	3.00	0.0153	0.0028	5.42
Visit	-0.019	0.182	-0.11	0.233	0.135	1.72
Ln(Income)	0.240	0.097	2.47	0.408	0.083	4.92
$\chi^2(7)$	165.56			214.51		
Pseudo R^2	0.179			0.133		
Sample size	680			1160		

Table 7. Elasticity Estimates						
	Opt-in			Probability-based		
	Coef.	Std. Err.	z	Coef.	Std. Err.	z
Scope	0.272	0.402	0.68	0.768	0.325	2.36
Income	0.626	0.282	2.22	1.076	0.266	4.04

References

Arrow, K., R. Solow, P.R. Portney, E.E. Leamer, R. Radner, and H. Schuman. 1993. *Report of the NOAA Panel on Contingent Valuation*. Washington, D.C.

Banzhaf, H. Spencer, Dallas Burtraw, David Evans, and Alan Krupnick. "Valuation of natural resource improvements in the Adirondacks." Land Economics 82, no. 3 (2006): 445-464.

Berrens, Robert P., Alok K. Bohara, Hank Jenkins-Smith, Carol Silva, and David L. Weimer. "The advent of Internet surveys for political research: A comparison of telephone and Internet samples." Political Analysis 11, no. 1 (2003): 1-22.

Bishop, Richard C., and Kevin J. Boyle. "Reliability and validity in nonmarket valuation." Environmental and Resource Economics 72, no. 2 (2019): 559-582.

Boyle, Kevin J., Mark Morrison, Darla Hatton MacDonald, Roderick Duncan, and John Rose. "Investigating Internet and mail implementation of stated-preference surveys while controlling for differences in sample frames." Environmental and Resource Economics 64, no. 3 (2016): 401-419.

Cameron, T.A. 1991. "Interval Estimates of Non-market Resource Values from Referendum Contingent Valuation Surveys." *Land Economics* 67: 413-421.

Carson, R.T. 2000. "Contingent Valuation: A User's Guide." *Environmental Science & Technology* 34(8): 1413-1418.

Carson, R.T., M.B. Conaway, W.M. Hanemann, J.A. Krosnick, R.C. Mitchell, and S. Presser. 2004. *Valuing Oil Spill Prevention: A Case Study of California's Central Coast.* Boston: Kluwer Academic Press.

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

Carson, R.T., R.C. Mitchell, M. Hanemann, R.J. Kopp, S. Presser, and P.A. Ruud. 2003. "Contingent Valuation and Lost Passive Use: Damages from the Exxon Valdez." *Environmental and Resource Economics* 25: 257-286.

Champ, Patricia A. "Collecting survey data for nonmarket valuation." Chapter 3 in Champ, Patricia, Kevin Boyle, and Thomas Brown (Eds.) A Primer on Nonmarket Valuation, 2nd Edition, Springer, Netherlands (2017), pp. 59-98, 2017.

Cummings, R., and L. Taylor. 1999. "Unbiased Value Estimates for Environmental Goods: A Cheap Talk Design for the Contingent Valuation Method." *American Economic Review* 89(3): 649-65.

Deepwater Horizon Natural Resource Damage Assessment Trustees. 2016. *Deepwater Horizon oil spill: Final Programmatic Damage Assessment and Restoration Plan and Final Programmatic Environmental Impact Statement*. Retrieved from http://www.gulfspillrestoration.noaa.gov/restoration-planning/gulf-plan

Flores, Nicholas E., and Richard T. Carson. "The relationship between the income elasticities of demand and willingness to pay." Journal of Environmental Economics and Management 33, no. 3 (1997): 287-295.

Giguere, Christopher, Chris Moore, and John C. Whitehead. "Valuing hemlock woolly adelgid control in public forests: scope effects with attribute nonattendance." Land Economics 96, no. 1 (2020): 25-42.

Haab, T.C., and K.E. McConnell. 2002. *Valuing Environmental and Natural Resources: The Econometrics of Non-market Valuation*. Northampton, MA: Edward Elgar.

Hanemann, W. Michael. "Welfare evaluations in contingent valuation experiments with discrete responses." American journal of agricultural economics 66, no. 3 (1984): 332-341.

Hanemann, W. Michael. "Willingness to pay and willingness to accept: how much can they differ?." The American Economic Review 81, no. 3 (1991): 635-647.

Herriges, J.A., C.L. Kling, C.-C. Liu, and J. Tobias. 2010. "What Are the Consequences of Consequentiality?" *Journal of Environmental Economics and Management* 59(1): 67-81.

Huffaker, R.G., R.L. Clouser, and S.L. Larkin, "Contract for Analytical Services Related to the Deepwater Horizon Disaster: Estimation of lost indirect and passive use economic values to Floridians," Food and Resource Economics Department, University of Florida, Final Report.

Johnston, Robert, "Do You Know Who's Answering Your Survey? Expanding Threats to the Integrity of Online Panel Data in Environmental and Resource Economics", paper presented at the 2021 AERE Summer Conference.

Johnston, R.J., Boyle, K.J., Adamowicz, W., Bennett, J., Brouwer, R., Cameron, T.A., Hanemann, W.M., Hanley, N., Ryan, M., Scarpa, R. and Tourangeau, R "Contemporary guidance for stated preference studies." Journal of the Association of Environmental and Resource Economists 4, no. 2 (2017): 319-405.

Kristrom, Bengt, and Pere Riera. "Is the income elasticity of environmental improvements less than one?." Environmental and resource Economics 7, no. 1 (1996): 45-55.

Lindhjem, H., and S. Navrud. "Using internet in stated preference surveys: a review and comparison of survey modes." International Review of Environmental and Resource Economics 5 (2011): 309-351.

McConnell, Kenneth E. "Models for referendum data: the structure of discrete choice models for contingent valuation." Journal of environmental economics and management 18, no. 1 (1990): 19-34.

Parsons, George R., and Kelley Myers. "Fat tails and truncated bids in contingent valuation: An application to an endangered shorebird species." Ecological Economics 129 (2016): 210-219.

Sandstrom, Kaitlynn, Frank Lupi, Hyunjung Kim and Joseph A. Herriges, “Comparing Water Quality Valuation Across Probability and Non-Probability Samples,” Selected Paper prepared for presentation at the 2021 Agricultural and Applied Economics Association Annual Meeting, Austin, TX, August 1-3 (2021).

Thurston, Hale W. "Non-market valuation on the internet," Chapter 12 in Alberini, Anna, and James R. Kahn (eds). Handbook on Contingent Valuation. Edward Elgar Publishing, UK (2006).