

August 2022

Adherence to Best Practices for Stated Preference Valuation within the U.S. Marine Ecosystem Services Literature

Daniel K. Lew

NOAA Fisheries, Alaska Fisheries Science Center

Leif E. Anderson

NOAA Fisheries, Northwest Fisheries Science Center

Douglas W. Lipton

NOAA Fisheries

Tammy B. Murphy

NOAA Fisheries, Northeast Fisheries Science Center

Kristy Wallmo

NOAA Fisheries, Office of Science and Technology

Follow this and additional works at: <https://cbe.miis.edu/joce>



Part of the [Economics Commons](#), and the [Environmental Studies Commons](#)

Recommended Citation

Lew, Daniel K.; Anderson, Leif E.; Lipton, Douglas W.; Murphy, Tammy B.; and Wallmo, Kristy (2022) "Adherence to Best Practices for Stated Preference Valuation within the U.S. Marine Ecosystem Services Literature," *Journal of Ocean and Coastal Economics*: Vol. 9: Iss. 1, Article 3.
DOI: <https://doi.org/10.15351/2373-8456.1159>

This Research Article is brought to you for free and open access by Digital Commons @ Center for the Blue Economy. It has been accepted for inclusion in *Journal of Ocean and Coastal Economics* by an authorized editor of Digital Commons @ Center for the Blue Economy. For more information, please contact ccolgan@miis.edu.

Adherence to Best Practices for Stated Preference Valuation within the U.S. Marine Ecosystem Services Literature

Acknowledgments

The authors wish to thank Robby Fonner and Steve Kasperski for useful comments on an earlier draft of this work. Any remaining errors are the authors' alone. This article and its findings are those of the authors and do not necessarily reflect the views of the National Marine Fisheries Service or the U.S. Department of Commerce.

1. INTRODUCTION

Information about the economic values of ecosystem services (EVES) can be useful in policy and management contexts in which decision-makers are faced with balancing ecological, economic, and socio-cultural priorities. This information provides a means for formal and quantitative trade-off analyses by facilitating comparisons across different types of ecosystem services (ES) and human activities. This is possible since economic values are measured in a common monetary metric. As a result, one can use these values to apply formal policy analytic approaches like benefit-cost analysis (BCA) to evaluate alternative policies or management actions at a variety of spatial scales, potentially involving multiple stakeholder populations and a diversity of ecosystem services. Within the growing body of ecosystem-based management (EBM) approaches like the Millennium Ecosystem Assessment (MA; 2005), IPBES (Diaz et al. 2015), and integrated ecosystem assessments (IEA; Breslow et al. 2016), the ability to conduct trade-off analyses of this sort is a core appeal for decision-makers to adopt such approaches.

TEEB (2010) and Gómez-Baggethun and Barton (2013) highlight several other uses for economic value information of ecosystem services. EVES can help increase public awareness of the importance of an ES to society. EVES can also be used in green accounting efforts to account for natural capital and environmental costs, such as those being conducted as part of the United Nations' System of Environmental Economic Accounting (SEEA) (<https://seea.un.org/>) that attempts to provide a more comprehensive view of the relationship between national-level economies and the natural environment and enables tracking of natural capital value changes over time. EVES may also be used to inform policymakers in their efforts to design management programs involving payments for ecosystem services like user or access fees or determining a project or program budget that does not exceed the value it would have for the public. Lastly, information about EVES is often desired in litigation involving natural resource damages (Kopp and Smith 1989; Barbier 2013).

Economic valuation methods for estimating EVES are well-established in theory and practice (e.g., TEEB 2010; National Research Council 2005). In general, economic values of ecosystem services that are bought and sold in explicit markets (e.g., seafood and minerals) are estimated by analyzing market transactions, while economic values for services not bought or sold in explicit markets (e.g., aesthetic values of nature views, recreation, pollution filtration

services, and nonuse values) must be measured using one of the non-market valuation approaches developed in the environmental economics literature (Freeman et al. 2014). Non-market valuation approaches fall into one of two classes, revealed preference (RP) approaches and stated preference (SP) approaches, which differ in the type of data used to reveal EVES or any other non-market values. RP methods use information on people's behavior to infer EVES, while SP methods use information directly obtained from people through their responses to carefully constructed questions asked in interviews or surveys in which they reveal information about their preferences and values for ecosystem services. Thus, SP methods differentiate themselves from RP methods in their reliance on data about what people say rather than on what they do.

The choice between RP and SP methods for the valuation of non-market ecosystem goods and services depends in large part on the type of ecosystem good or service being valued, the presence of related markets for goods or services that can be directly affected by the ES, and whether or not humans are likely to benefit from the ES independent of any use of it now or in the future (i.e., nonuse value). For ES that are not related to any marketed good or service and are believed to have substantial nonuse value (e.g., threatened or endangered species), stated preference methods are required. In coastal and marine settings, common ES that are valued by SP methods include recreational fishing, coastal and ocean recreation activities, marine protected areas, coastal and marine habitats, and threatened and endangered species and their habitat.

While SP methods have been controversial due to the potential for hypothetical bias (Hausman 2012), the weight of evidence points to them being valid approaches for estimating EVES when properly administered (Arrow et al. 1993; Kling, Herriges, and Zhao 2012; Johnston et al. 2017). To enhance the reliability of SP valuation information, a group of respected SP experts recently developed a set of 23 best practices guidelines (BPGs) for practitioners to follow when conducting a SP study (Johnston et al. 2017). These guidelines are comprehensive in scope, covering aspects from designing and administering a SP survey to analyzing and reporting SP study results. As evidenced by more than 900 citations of the article in Google Scholar (as of June 2022), the guidance has quickly been adopted among economists as a standard reference for development and implementation of new SP studies. These BPGs also provide a means to identify benchmarks against which existing SP research can be evaluated.

In this article, we use the BPGs outlined by Johnston et al. (2017) to assess how well the U.S. coastal and marine ecosystem services valuation literature employing

SP methods have adhered to the guidelines. Assessing the past literature of existing SP-based EVES is important given that these values are often used in other applications for which they were not originally intended, a process called environmental value, or benefits, transfer (Johnston et al. 2015, 2021). De novo (primary) economic valuation studies, particularly SP valuation studies, can be very expensive, time-consuming, and require specialized expertise to properly conduct. As a result, when economic values exist representing the same or a very similar ES, environmental value transfers become a feasible approach for incorporating economic value information. However, successful environmental value transfers depend in part upon the existence of high quality economic value estimates for ES that are sufficiently similar to those to which they would be applied (e.g., Plummer 2009).

Our evaluation of the literature involves the assessment of 82 peer-reviewed journal articles published between 1986 and 2018 that estimate the economic value of one or more coastal or marine ecosystem services using SP methods. Our analysis suggests that adherence to the BPGs is heterogeneous, with no studies adhering to all guidelines and some guidelines being adhered to better than others. We further assess differences in adherence to the guidelines between studies published in different time periods and journal types, studies using different SP methods, and studies valuing different ES. We also evaluate how differences in adherence to best practices may impact the number of citations studies get while controlling for other features of the study.

The remainder of the paper is organized as follows: The next section provides a general overview of SP methodology and then describes the BPGs in Johnston et al. (2017) and the procedure that we developed to measure adherence to these guidelines. The subsequent section enumerates the data we used for the analyses as well as the approaches undertaken to assess adherence to the guidelines and the extent to which adherence affects the uptake of the study within the literature. This is followed by the presentation and discussion of the results. Finally, we discuss the implications of our analysis on the use of the SP-based coastal and marine ES valuation literature for environmental value transfers and point to areas where the literature can improve to better align with the guidelines.

2. MATERIALS AND METHOD

2.1 Measuring Adherence to SP Best Practices

SP data are collected in surveys using carefully-constructed questions (valuation questions) that elicit information about a respondent's underlying preferences and values for the good or service being valued (the valuation good). SP surveys often include many of the following elements: information about the valuation good; questions to help prepare respondents for answering the valuation questions, to evaluate the responses provided, and to check for consistency in the expressed preferences (i.e., auxiliary and supporting questions); and questions to collect information about the individual that can be used in modeling preferences, estimating values, and characterizing and classifying the results across different types of individuals. Due to their complexity, SP surveys are often pretested using focus groups (small group directed discussions), one-on-one interviews, and pilot surveys (small-scale implementations). SP surveys can be administered using a variety of survey modes (mail questionnaires, in-person interviews, web surveys, etc.).

There are several SP methods, with discrete choice experiments (CE) and contingent valuation (CV) being most common in the ecosystem services valuation literature we examine. At a basic level, these methods differ in the format of the valuation questions. CE questions ask an individual to choose between two or more alternatives that are each described in terms of multiple attributes that vary across alternatives. The individual's choice between the alternatives indicates their trade-offs between different attributes presented in the question. In valuation studies, CE questions include a cost attribute, which allows for estimation of economic value information (willingness-to-pay [WTP] or willingness-to-accept [WTA]).¹ Experimental designs of the combinations of attributes and attribute levels seen by respondents in the CE questions are often constructed to optimize the information that can be yielded from choice responses. CE surveys typically include multiple valuation, or choice, questions. The CE data are analyzed using discrete-choice econometric models.

There are a variety of different CV methods that are distinguished by the format of the valuation question. They range from open-ended questions that ask individuals to state the amount they would pay for the valuation good to

¹ Which of WTP and WTA is the appropriate welfare measure depends on property rights (Freeman et al. 2014). The two measures need not correspond, and a number of studies have documented differences between empirical and experimental differences found between them (e.g., Vossler et al. 2020, Kim, Kling, and Zhao 2015).

referendum-style questions that ask whether or not an individual would pay \$X for the valuation good (where X is systematically varied over individuals across a range of feasible values). Open-ended CV responses provide direct measures of WTP (or WTA), while referendum CV responses indicate whether WTP (or WTA) is above or below \$X and, similar to CE data, are typically analyzed using discrete-choice econometric models.

Johnston et al. (2017) articulated 23 recommendations for best practices for SP studies that represent a more comprehensive and contemporary guidance than those provided in the NOAA Panel on Contingent Valuation (Arrow et al. 1993), given they draw upon accumulated lessons and experience from the extensive research into SP methods that have been conducted in the intervening quarter century. The BPGs are intended to maximize the validity and reliability of the SP results and include six guidelines related to the survey development and implementation process, seven related to value elicitation, seven about data analysis, two about assessing validity, and one regarding study reporting (Table 1). From these 23 BPGs, we developed 67 evaluation criteria that capture more specific aspects of each guideline and allow us to assess adherence to these 23 BPGs (see Appendix table A-1).

For each evaluation criterion, we measured adherence along two dimensions: (1) the extent to which elements contained in the guideline were acknowledged in the study (Acknowledge score) and (2) the extent to which the elements that were acknowledged in the study were addressed in the study (Address score). Together, these two dimensions capture the inclusion and consideration of the issue(s) contained in a guideline. Both Acknowledge and Address scores were measured on a 5-point scale (see Appendix table A-2).²

The Acknowledge score is based on whether or not an evaluation criterion was discussed or not in the study. This can range from not mentioning or acknowledging the issue at all (score of 1) to accounting for all of the components (if multiple) of the evaluation criterion in the study. Acknowledge scores in between 1 and 5 reflected differing degrees of discussing the multiple elements mentioned in the evaluation criterion.

The Address score measures the extent to which the advice embodied in the guideline was followed. For evaluation criteria that were not mentioned or

² Note that this evaluation approach differs from the yes/no checklists applied to assess economic evaluations in the health economics field, which are focused on evaluating whether or not candidate studies employ “reliable methods” (Drummond et al. 1993; Watts and Li 2019).

acknowledged at all (Acknowledge score of 1), no Address score was possible and a not applicable (N/A) is recorded. For all others, a score between 1 and 5 indicated the extent to which the issue(s) raised in the guideline as reflected in the particular evaluation criterion were addressed in the study. A score of 1 indicated not addressing the issue(s) at all, while a 5 indicated fully addressing the issue(s). Scores in between these indicated different degrees to which the issue(s) were addressed.

Recognizing that the scoring process has components that are inherently subjective, we took several steps to minimize potential biases in the scores and maximize consistency in the application of the scoring rubric. First, all five evaluators were Ph.D. economists with experience conducting SP valuation studies who were well-versed in the SP literature. Second, each study was evaluated by two evaluators independently. Third, a “leveling” meeting between the two evaluators was conducted wherein scores were discussed, any discrepancies were resolved (typically through consulting the scoring rubric and original BPGs, and sometimes by consulting with the full research team), and a single set of final Acknowledge and Address scores for the 67 evaluation criteria were assigned to the study. Lastly, since a number of studies in this literature were authored by one or more coauthors of this article, evaluators were prohibited from evaluating studies they authored.

Table 1. Best Practices Guidelines from Johnston et al. (2017) and number of evaluation criteria per guideline.

Guideline (BPG)	Study stage	Type of guideline	No. evaluation criteria
1		Survey design-related	7
2		Pretesting activities	4
3	Survey development and implementation	Choosing between stated preference approaches	2
4		Experimental design	3
5		Ethics in data collection	4
6		Extent of market, survey mode, sampling, and nonresponse bias	4
7		Willingness to pay versus willingness to accept	2
8	Value elicitation	Valuation question response formats	3
9		No-answer option	2
10		Decision rule	2
11		Payment vehicle	4
12		Auxiliary or supporting questions	3
13		Ex ante procedures to enhance validity	2
14	Data analysis	Choice of econometric estimator	3
15		Modeling preference heterogeneity	2
16		Balancing model parsimony and complexity	2
17		Behavioral response anomalies	1
18		Value estimation	3
19		Using data from auxiliary and supporting questions	2
20		Sample representativeness and value aggregation	6
21	Validity assessment	Conducting and interpreting validity tests	2
22		Weight of evidence in validity testing	2
23	Study reporting	Study reporting	2
Total			67

2.2 U.S. Studies Valuing Marine and Coastal ES

This study focuses on the evaluation of SP studies that value coastal and marine ecosystem services in U.S. waters and coastlines. We limited our analysis to studies involving the following types of ecosystem services: marine recreational fishing, other marine or coastal recreation (e.g., diving, beach recreation, and wildlife viewing), coastal and marine habitats, marine reserves and protected areas (MPAs), protected species (e.g., threatened and endangered species), and other marine cultural services. This set covers the main cultural and supporting/regulating ES that are valued in the SP valuation literature (e.g., Lipton et al. 2014).¹ Studies included in the analysis were limited to peer-reviewed journal articles of U.S. empirical studies that use a SP method to value one or more of the above coastal or marine ecosystem services. This put the focus on primary, or de novo, SP valuation studies and excluded studies that used SP-based economic values from other studies (i.e., those using benefits transfer methods) or did not generate original value estimates using SP methods.

We used a variety of approaches for finding studies meeting these criteria. This included using keyword searches in library databases (e.g., Web of Science), searching tables of contents from economic and marine policy field journals known to publish SP valuation and EVES studies (e.g., *Marine Resource Economics*, *Ecological Economics*, and *Ocean and Coastal Management*), searching through bibliographies of relevant studies we found, and drawing upon our personal libraries of SP studies. The resulting studies were then evaluated to determine if they met all the criteria for inclusion. A total of 82 peer-reviewed journal articles (see Appendix Table A-3) met the criteria and were included in our analysis.

¹ Other types of ES are commonly valued using market-based valuation approaches or RP methods.



Figure 1. Summary of examined studies (N = 82). Distribution of studies by publication date, journal type, study focus, and marine ecosystem services valued.

Figure 1 provides a summary of the 82 studies in terms of the publication year, the study foci, ES valued, and type of journal. The analysis included studies published as early as 1986 and as recently as 2018, with the majority of studies published after 2000.¹ Over 90% of the studies had a primary focus on valuation, and a majority (79%) of the studies had a policy or management focus (studies could have several foci). About one-third of the studies were methodological in nature, defined as testing or proposing a SP methodological innovation as a major aim. Additionally, 17% of the studies utilized more than one valuation method. In terms of the ES valued in these studies, 40% of the studies valued marine recreational fishing, 27% valued protected marine species, 24% valued non-fishing coastal or marine recreation, 15% valued marine or coastal habitat, and 5% valued MPAs. Only one study valued other marine cultural services (Nepal et al. 2018). Note that some studies valued more than one type of ES. In terms of the journal types in which these studies were published, the majority (56%) were published in economics journals, about a quarter were published in policy or management-focused journals (26%), and the remainder were published in other types of journals (e.g., biology or conservation journals, fisheries science journals, etc.). In terms of the SP method used in these studies, about half used the CV method and half used the CE method.

2.3 Assessing Adherence to Best Practices

As discussed in Section 2.1, we evaluated the extent to which the 82 studies adhere to the Johnston et al. (2017) BPGs by analyzing the Acknowledge and Address scores for the 67 evaluation criteria. To this end, we calculated the mean Acknowledge and Address scores for each evaluation criterion over the entire set of studies and examine patterns across evaluation criteria and BPGs. We also calculated these mean scores for subsets of studies to assess the extent to which mean scores differed across different types of studies. In particular, we assessed differences in mean scores temporally (by publication date), by SP method (CV or CE), by ES type, and by journal type.² To determine if the distribution of study-level scores differ along these dimensions, we conducted a series of non-parametric

¹ Although the journal articles we examine include ones published contemporaneously with Johnston et al. (2017) (i.e., seven published in 2017 and one in 2018), none of them reference the Johnston et al. (2017) article.

² We also attempted to analyze differences in scores between studies whose primary focus was on valuation versus those that included valuation as a component of the overall study but was not the focus. However, since only six studies did not have a primary focus on valuation there were insufficient cases to generate reliable mean Address scores to inform the comparison.

Wilcoxon-Mann-Whitney rank tests that assess whether the null hypothesis of equality in the distribution of mean study scores (average of scores across evaluation criteria for each study) was met between paired subsets of studies. These tests were intended to evaluate the extent to which adherence to the 2017 BPGs differs over time as the SP literature has matured, across studies that have valuation as a primary study objective versus others, across the types of journal that may have differing standards and required content, and across SP methodology employed.

2.4 Citation Analysis

We used a count data model to investigate whether a more complete adherence to the best practices guidelines led to increased use of the article, as measured through citation counts. Specifically, we estimated a Poisson model to explore whether the number of citations an article received was an increasing function of the Address and Acknowledge scores, while controlling for other factors that are likely to influence citations. As citation counts are likely to increase the longer the article has been published, we used the number of years published as an exposure variable. An overdispersion test failed to reject the null hypothesis of mean-variance equality, suggesting that the assumptions of the Poisson model specification were not violated.

In addition to the mean Acknowledge and Address scores calculated across the full set of evaluation criteria, the model controlled for a number of other factors that were likely to affect citations. First, we included an indicator variable to control for differences across SP methods. The choice experiment variable measures the effect of using a choice experiment, relative to either a contingent valuation approach or a combination of a choice experiment and contingent valuation. Next, we controlled for the focus of the paper with a set of indicators describing the focus of the article as methodological, primary valuation, or a comparison of multiple valuation methods. These article focus indicator variables were measured relative to a baseline focus of policy and management. Each article was also classified based on the type(s) of ES valued, and these were also included in the Poisson model. The ES types of recreational fishing, endangered species, habitat, and MPAs were measured relative to a baseline that included non-fishing recreation and a single cultural ES type article.

The final model controls were all related to the journal in which the article was published. We controlled for journal type through the inclusion of two indicator variables denoting (1) an economics journal and (2) a policy and management journal. These were measured relative to the baseline of other journal type. The

citation score of the journal was included (as measured through CiteScore³) to estimate the effect of publishing in more highly cited journals. Finally, the number of issues published by the journal each year was included to control for citation rate differences within higher volume journals.

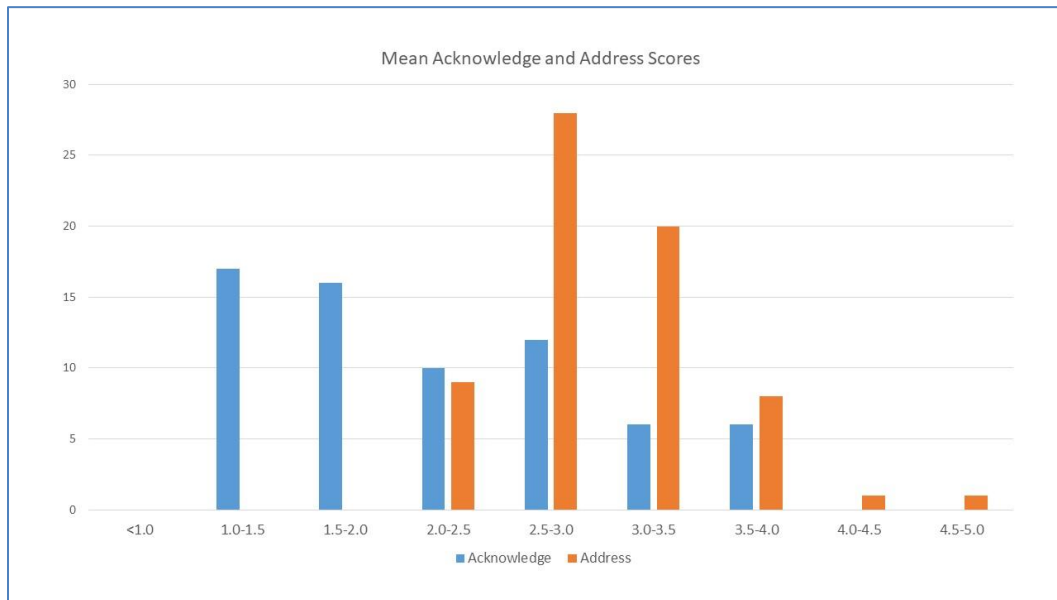


Figure 2. Distribution of mean address and acknowledge scores (across all studies) for the 67 evaluation criteria.

3. RESULTS

3.1 Overall Assessment of Best Practices Adherence

The distribution of the 67 mean Acknowledge and mean Address scores -- calculated over the 82 studies -- are presented in Figure 2. As the figure shows, the mean Acknowledge scores had a lower distribution than the mean Address scores, suggesting that across the 82 studies the evaluation criteria Acknowledge scores tended to be lower than the Address scores. In fact, the mean (median) of the mean Acknowledge scores—over the 67 evaluation criteria--was 2.20 (2.08) and the

³ For details, see <https://www.elsevier.com/connect/editors-update/citescore-a-new-metric-to-help-you-choose-the-right-journal>.

mean (median) of the mean Address scores was 3.03 (2.98).⁴ Moreover, the mean Acknowledge scores across BPGs ranged from 1.06 to 3.81, while the Address scores ranged from 2.53 to 4.34. These indicate that studies in this literature did not always acknowledge all the elements of the best practices guidelines, but elements that were acknowledged were at least somewhat addressed in the study.

The study-level mean Acknowledge and Address scores (calculated over the 67 evaluation criteria for a single study) provide a similar signal about how well individual studies generally adhere to the BPGs. Study-level mean scores ranged from 1.45 to 3.06 for Acknowledge and 1.93 to 4.08 for Address, with corresponding mean (median) scores of 2.19 (2.19) and 3.07 (3.08), respectively. These study-level metrics provide additional evidence that studies in the literature had lower Acknowledge scores than Address scores.

Figure 3 displays the mean Acknowledge and Address scores for the 23 BPGs, where the mean score for a specific guideline is the average of the mean scores for all evaluation criteria associated with that guideline. The figure shows that for each guideline, the mean Address score was higher than the mean Acknowledge score.

⁴ The standard errors of the mean Acknowledge and Address scores are 0.12 and 0.24, respectively. A simple t-test for the difference in means supports the mean Address score being statistically greater than the mean Acknowledge score.



Figure 3. Mean address and acknowledge scores by best practices guideline across all studies. Mean address scores are calculated over all 82 studies. Mean address scores are calculated over studies for which at least some of the issues related to the best practice guideline are acknowledged (excludes “N/A”).

Nine BPGs had mean Acknowledge scores of less than 2.0, which corresponds to less than “very few elements are acknowledged.” Of these nine, the lowest score was for BPG5 (ethics in data collection), which relates to whether or not the study design and protocols were reviewed by a university or review board charged with ensuring adherence to ethical standards of human subjects, study participants provided informed consent before participating, and the study design avoided deception. This BPG’s Acknowledge score was so low due to very few studies in this literature mentioning human subjects review, informed consent, or efforts to avoid deception in study design (less than 4% of the studies). Similarly, a low number of studies mentioned why WTP or WTA was the appropriate measure of economic value to measure in the study (BPG7) or why a no-answer option was used (BPG9) (less than 20% of studies in each case), leading to low Acknowledge scores for these guidelines. Other low mean Acknowledge scores corresponded to guidelines related to the type of valuation question response format to use (BPG8); steps to take to enhance the validity of the results through use of incentive compatible and consequential SP scenarios (BPG13); the potential for and investigation of anomalous behavior (BPG17); the use of supplemental information in modeling and analysis (BPG19); conducting validity tests (BPG21); and applying a weight-of-the-evidence approach to validity testing (BPG22).

Three notably high mean Acknowledge scores (those above 3.0, which indicate that studies on average at least acknowledged “some, but not all, elements” of the BPG), were all associated with guidelines relating to the data analysis stage. They include the guidelines related to the choice of econometric estimator (BPG14), modeling preference heterogeneity (BPG15), and welfare estimation (BPG18).

The Address scores indicated the extent to which the studies addressed elements of the 23 BPGs. The mean Address scores were only calculated for the studies that acknowledged those elements. Thus, for guidelines like BPG5, BPG7, and BPG9 that have few studies that acknowledge elements of those guidelines, their mean Address scores were averaged over those few studies. The six BPGs that had the lowest mean Address scores were all below 2.8, below the “somewhat addressed” level (3.0), and were related to the choice of valuation question response format (BPG8); justifying the use of a no-answer option (BPG9); applying ex ante procedures to enhance validity (BPG13); using data from auxiliary and supporting questions in the model and validity testing (BPG19); addressing challenges of ensuring sample representativeness and appropriate value aggregation (BPG20);

and the documentation of study design, implementation, analyses, and results to ensure transparency and replicability (BPG23).

The highest mean Address scores (above 3.5) were associated with guidelines that relate to data analysis, specifically the choice of econometric estimator (BPG14) and welfare estimation (BPG18). Note that these two guidelines also received high Acknowledge scores, indicating that these are particular priorities to researchers in this literature.

Table 2. Wilcoxon-Mann-Whitney Rank Tests. Test statistics for non-parametric evaluation of the differences in distributions of mean study scores across journal type, publication date, SP method, and type of ecosystem service valued.

Test of equality of mean study score distributions	Acknowledge	Address
Economics vs non-economics journal	0.9437	0.2616
Policy/management vs other journal	0.1009	0.2390
Published pre-2000 vs 2000 or after	1.5348	1.3443
Published in 2010s vs before 2010	3.0704***	1.7301*
Contingent valuation vs non-contingent valuation	2.3656**	1.8646*
Threatened and endangered species studies vs all other studies	1.4652	0.8268
Recreation studies vs all other studies	2.5867***	2.6441***

Note: *, **, *** denote rejection of null hypothesis of equal mean study score distributions at the 10%, 5%, and 1% levels, respectively.

3.2 The Role of Study Characteristics on Adherence Scores

Acknowledge and Address scores were also used to assess differences in adherence to the BPGs between subsets of the 82 studies in terms of journal type, publication date, SP method utilized in the study, and ecosystem service type valued. Studies were first categorized by the type of journal in which they were published. Three categories were used: economic journals (46 studies; e.g., *Marine Resource Economics*, *Ecological Economics*, and *Land Economics*), policy/management

journals (21 studies; e.g., *Marine Policy*, *Ocean and Coastal Management*, and *Journal of Environmental Management*), and other journals (15 studies; includes all other disciplinary journals, inclusive of general sciences). Mean scores for each BPG by journal type are presented in Figure 4. Mean Acknowledge scores were fairly similar across journal types, but mean Address scores depart somewhat in terms of SPG5 for studies in the other journal group, though this is largely driven by the fact that studies in this group were generally the only ones to explicitly address aspects of the ethics of data collection. Studies in economics and policy/management journals had similar mean Address scores, with ones in economics journals having slightly higher Address scores for the BPGs related to value elicitation and data analysis. However, a Wilcoxon-Mann-Whitney rank test indicated that there were no statistically significant differences between overall mean study-level scores between studies published in economics and all other studies or between studies published in policy/management journals and all other studies (Table 2).

To evaluate whether older studies differed from newer ones, we divided studies into three time periods: pre-2000 (18 studies), 2000-2009 (32 studies), and 2010-2018 (32 studies) (Figure 4). Mean Acknowledge scores related to survey development and implementation (BPG1-BPG4) were higher for more recent studies, as well as some aspects of the data analysis stage (BPG14, BPG15, and BPG18 in particular). This makes sense given that the guidelines include current best practices that sometimes involve techniques or data and survey design approaches that had not yet been developed or vetted in earlier studies. As a result, later studies are more likely to acknowledge those things while early studies would not have. The mean Address scores by time period varied across the BPGs without a discernible pattern. However, the Wilcoxon-Mann-Whitney rank tests comparing the pre-2010 studies to 2010-2018 studies suggested the later period studies have statistically different study-level mean Acknowledge and Address scores at the 1% and 10% levels, respectively (Table 2).

Another dimension in which adherence to the BPGs may possibly differ is across the type of SP method used in each study. Mean Acknowledge scores for studies that employed CE methods (42 studies) were notably higher than for studies that employed CV methods (40 studies) for most survey development and implementation guidelines (BPG1-BPG4), for two value elicitation guidelines (BPG10 and BPG11), and several data analysis guidelines (BPG14-BPG16) (Figure 5). Differences between mean Address scores suggest CE studies generally scored higher for data analysis, but lower than the CV studies in terms of most value

elicitation, validity assessment, and reporting guidelines. The Wilcoxon-Mann-Whitney rank tests for equality of the distributions of the mean study-level scores suggests there were statistically significant differences between the study-level mean Acknowledge (at the 5% level) and Address (at the 10% level) scores.

The final dimension we examined for differences in adherence to the BPGs was type of ES valued. For this, we divided studies into those valuing recreation ES (51 studies), threatened or endangered species ES (22 studies), and all other ES (15 studies) (Figure 5).¹ Mean Acknowledge scores for recreation ES studies were lower than those valuing other ES for a handful of BPGs related to survey development and testing (BPG1-BPG3), value elicitation (BPG8 and BPG9), and data analysis (BPG19 and BPG20). Mean Address scores for recreation ES studies were also lower than for studies valuing other ES for the same set of BPGs, as well as for additional value elicitation (BPG11-BPG13) and data analysis BPGs (BPG17). Non-parametric Wilcoxon-Mann-Whitney rank tests rejected (at the 1% level) the null hypotheses that the distributions of study-level mean Acknowledge and Address scores for the recreation studies were the same as for other studies (Table 2), while failing to reject the null hypothesis that threatened and endangered species study-level mean scores were the same as those from all other studies.

¹ Six studies valued multiple types of ES and were included in all applicable categories.

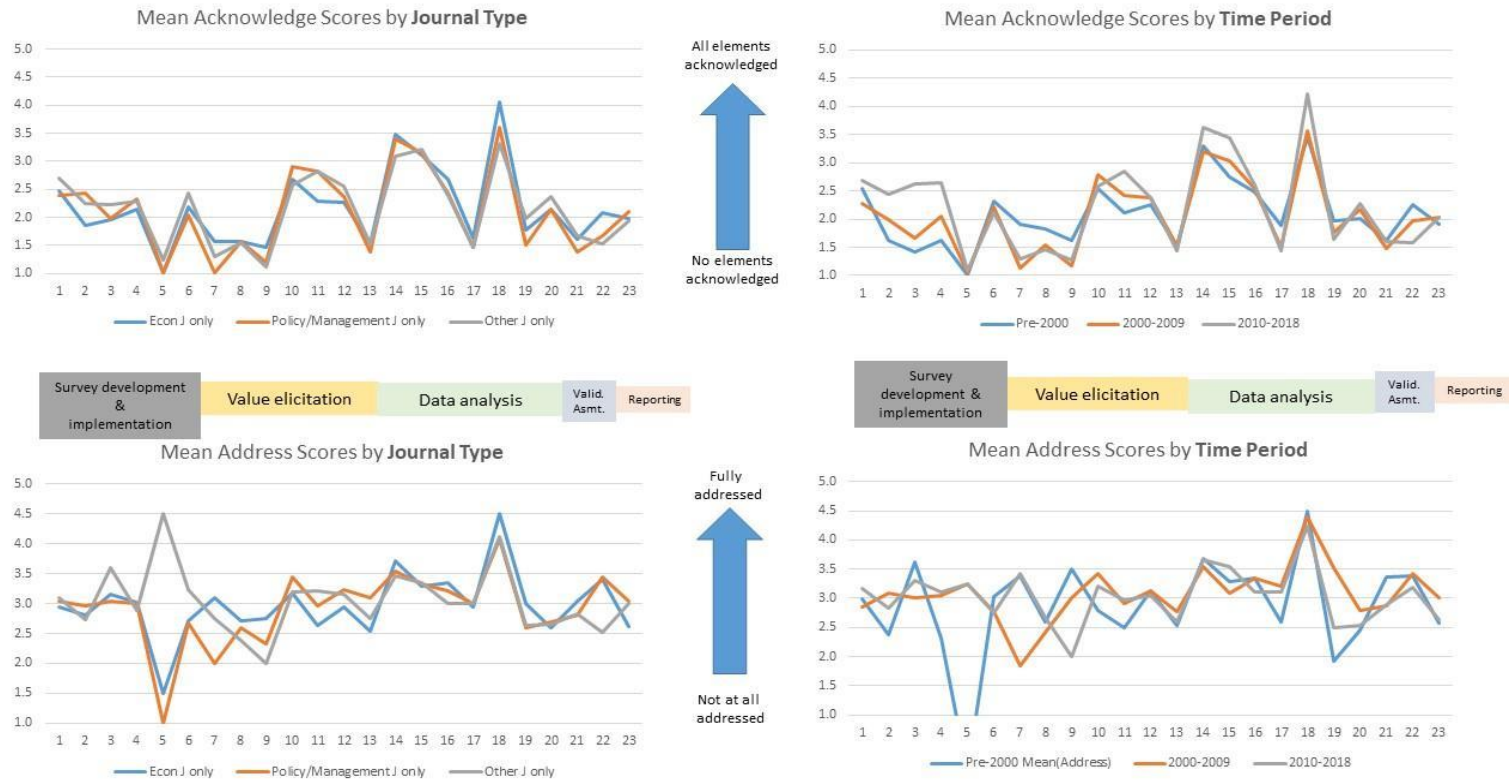


Figure 4. Mean address and acknowledge scores by best practices guideline across applicable studies (by journal type and time period). Mean address scores are calculated over studies for which at least some of the issues related to the best practice guideline are acknowledged (excludes “N/A”).

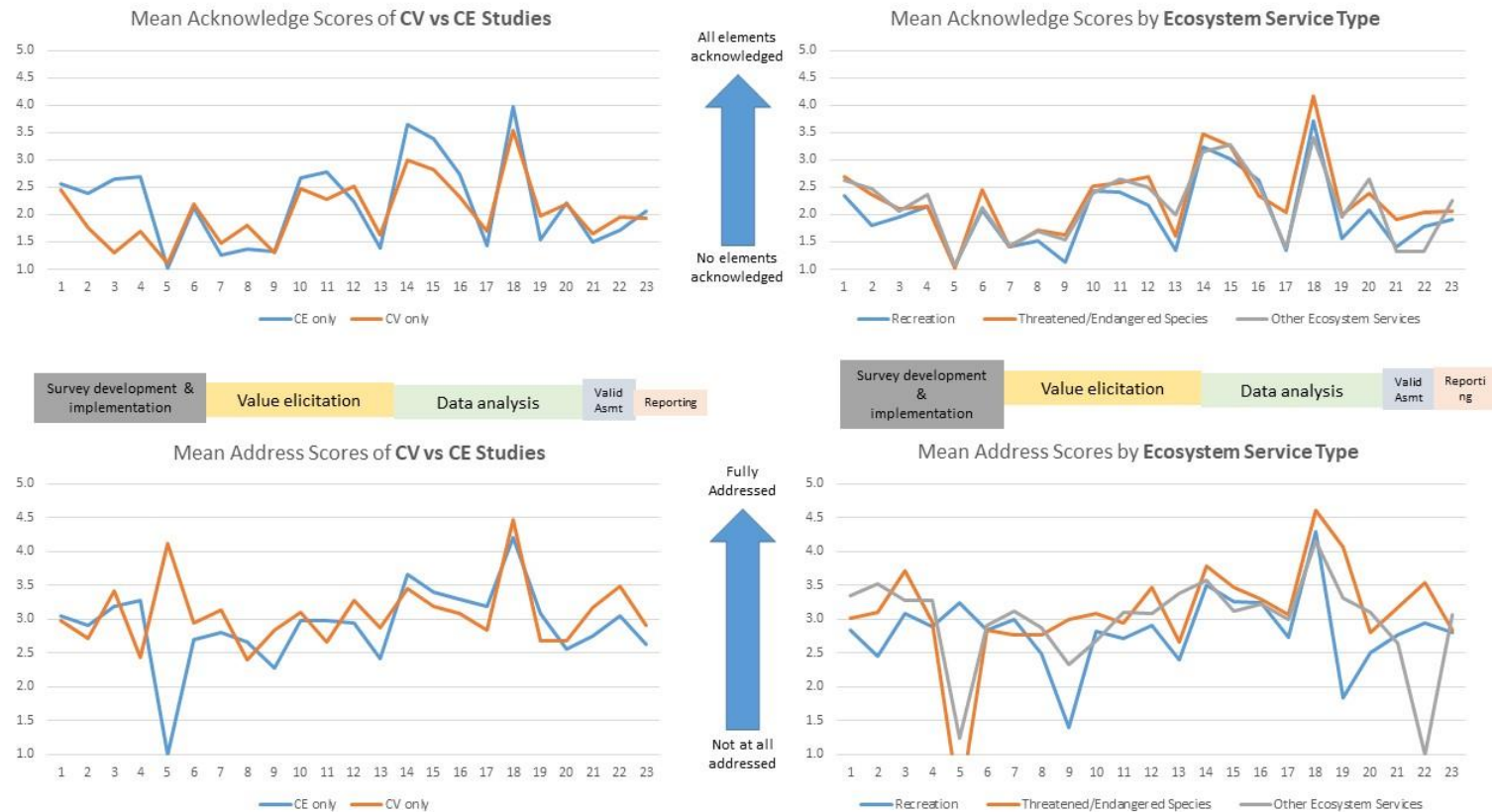


Figure 5. Mean address and acknowledge scores by best practices guideline across applicable studies (by SP methodology and by ecosystem service type). SP methodology distinguishes between contingent valuation (CV) and choice experiment (CE) studies. ES types considered are recreation, threatened/endangered species, and all other ecosystem services. Mean address scores are calculated over studies for which at least some of the issues related to the best practice guideline are acknowledged (excludes “N/A”).

Table 3. Poisson model estimates of article citation counts

Variable	Coeff.	Std. Err.	p-value
Constant	-0.4856***	0.1179	0.000
Contingent Valuation	0.2239***	0.0285	0.000
CiteScore	0.1526***	0.0076	0.000
Focus: Methodological	0.8760***	0.0292	0.000
Focus: Primary Valuation	0.6670***	0.0462	0.000
Focus: Multivaluation	0.2147***	0.0353	0.000
Journal Type: Economics	-0.0650	0.0430	0.130
Journal Type: Policy & Management	-0.1256**	0.0498	0.012
Journal Issues Per Year	-0.0115***	0.0015	0.000
ES Type: Recreational Fishing	0.2645***	0.0351	0.000
ES Type: Endangered Species	0.5322***	0.0385	0.000
ES Type: Habitat	0.1593***	0.0385	0.000
ES Type: MPAs	-0.2288***	0.0822	0.005
Acknowledge Score	-0.5795***	0.0561	0.000
Address Score	0.4723***	0.0493	0.000

Exposure Variable: Years Published

Sample Size: 80

LL at Zero: -5040.80

LL at Convergence: -1936.07

Note: *, **, *** denote significance at the 10%, 5%, and 1% levels, respectively.

3.3 Citation Analysis Results

Poisson regression estimates of citation counts indicated that both measures of adherence to BPGs (Acknowledge score and Address score) as well as many of the included study- and journal-related controls had statistically significant impacts on the number of citations received (Table 3).

First, an examination of the study-related controls showed that the use of the CV method was positively related to the number of citations, relative to the use of a CE approach or a combination of methods. Among the different article foci, articles comparing multiple valuation methods and those with either a primary valuation or methodological focus were shown to have increased citations, relative to a baseline of policy and management. For ES type, we found significant positive effects for endangered species, recreational fishing, and habitat, and a negative effect for the ES type of MPAs.

At the journal level, we found that articles published in journals that received more citations (as measured through CiteScore) were cited more often. We also identified differences by journal type, with articles published within a policy and management journal receiving a lower number of citations, relative to the baseline journal type of other. There was a negative effect of publishing within a journal with a larger number of issues per year.

After controlling for all of these article and journal characteristics, our count model found that the two measures of adherence to best practices had significant effects on citations, although they were in opposite directions. The estimated effect of Acknowledge score was negative; articles that described a greater number of the BPGs – without also addressing these criteria – produced less cited articles. Conversely, the effect of Address score was positive, implying that studies that more fully addressed the BPGs received a higher number of citations.

4. DISCUSSION

Our analysis showed that the existing U.S.-focused SP marine and coastal ecosystem services valuation literature is fairly heterogeneous in its adherence to the BPGs outlined in Johnston et al. (2017). We measured adherence along two

dimensions to capture both the extent to which studies mentioned or acknowledged elements of the BPGs and the extent to which they addressed them. Studies in general received higher Address scores than Acknowledge scores, suggesting that studies in the literature more often than not made an effort to address the elements from the BPGs identified within the study.

Unsurprisingly, no study we examined perfectly adhered to the BPGs. In part, this can be explained by the fact that the studies we examined were all published prior to, or, in the case of a handful of studies, around the same time as Johnston et al. (2017). None of the most recent studies (published in 2017 or 2018) referenced the Johnston et al. study or the BPGs. Thus, our assessment cannot speak to the impact the BPGs have had on adherence to best practices in the literature since their publication. Rather, it is primarily an assessment of how well the past literature adhered to current best practices. However, since the BPGs are largely a documentation of the best practices used in the literature (based on an accumulation of knowledge and experience), our evaluation is still important for assessing the literature's adoption of best practices up to that point. Since time, money, expertise, and other resource constraints often preclude the estimation of EVES information via original SP valuation work, researchers and analysts are often in a position where they must turn to benefit transfer methods, which draw upon the existing literature. For this reason, our assessment is also important for helping inform the viability of applying the benefits transfer approach in coastal and marine policy contexts, given the reliance of the approach on the existence of high-quality EVES information in the literature. We leave assessing the impact of Johnston et al. (2017) on studies that were performed after the BPGs were published to future work.

By examining scores for different types of studies, we gained several important insights. First, more recent studies tended to adhere to best practices more than older studies did. This should not come as too much of a surprise given research aimed at improving the validity and reliability of SP-based values has been a very active research area since the 1990s (e.g., Smith 2000; Kube et al. 2018). Thus, it stands to reason that as empirical and conceptual lessons have been learned and methodologies have improved, SP researchers would adopt them over practices and techniques that were found to be deficient or at least suboptimal. It is also the case that some methods and issues that are mentioned in the BPGs were not on the radar for SP researchers in earlier studies because they had yet to be developed or even identified. In part, this reflects technological advances, such as improved computing power that over time has enabled the development and use of more

sophisticated econometric techniques. It also reflects the emergence of web-based data collection that has opened the door to new ways of collecting, archiving, and sharing data and results, while also introducing a host of additional challenges related to survey-related biases and their repercussions on the study results and its generalizability (Menegaki, Olson, and Tsagarakis 2016). Furthermore, some of the BPGs (e.g., BPG17) reflect the growing integration of concepts and methods from behavioral and experimental economics that has brought more realism to the modeling of individual decision-making behavior in environmental and resource economics (Shogren and Taylor 2008).

Second, we found statistically significant differences between CE and CV studies' adherence scores, with evidence pointing to CE studies generally having higher scores. This can partly be explained by the fact that the CE method is a newer SP valuation method that has been increasingly used in the past two decades but was absent from the literature we studied prior to 2001. In contrast, half of the CV studies in the examined literature were published before 2001 and only six (of 40) were published in the 2010s. Given the discussion above, this temporal distribution of CV and CE studies likely influenced the imbalance in adherence we observed. It may also suggest a movement away from CV in favor of CE in the future for valuing coastal and marine ES using SP methods.

And finally, studies valuing recreation-related ES tended to adhere less to the BPGs than studies valuing other ES. The reasons for this are unclear, but it should be noted that recreation is one of the few coastal and marine ES that can be valued using both RP and SP approaches since the economic values are largely, if not exclusively, composed of use values (Johnston et al. 2006). As a result, it is possible that SP studies valuing coastal or marine recreation activities pay less attention to BPGs intended to mitigate the role of hypothetical bias (Loomis 2014). Moreover, recent research has highlighted how combining RP and SP data to value recreation helps minimize pitfalls of using RP or SP data alone (e.g., Whitehead and Lew 2020), which could potentially shift research on recreation-based EVES to employ more RP-SP modeling approaches, further raising the bar of BPGs for researchers conducting valuation studies.

The Acknowledge and Address scores for individual BPGs reflect the relative priorities placed on different aspects of the SP research process by authors, editors, and reviewers in the peer-reviewed literature, which themselves are influenced by trends in the literature and norms in the scientific community. For instance, the high Address scores for numerous data analysis and validity assessment BPGs may reflect an emphasis on empirical methods and applied econometrics in the broader

economics literature (Hamermesh 2013; Kube et al. 2018). Moreover, there was almost-universally low Acknowledge scores for BPG5 (ethics in data collection). This could suggest that almost no attention has been paid in the literature to ethics in data collection, such as getting informed consent from study participants and review by human subjects committees or review boards. However, the low scores may instead obscure the fact that studies may have been conducted following these ethics in human research principles, but these details were not reported in the published paper. Such non-reporting could be a function of the constraints placed on the article by the journal (e.g., manuscript length restrictions), editorial standards that for many journals typically value innovation and contributions over providing every detail of the research process, and idiosyncratic demands of reviewers in the peer-review process. All of these factors may disincentivize researchers from describing some details that were undertaken or considered, but are not viewed as critical to the communication of the research contribution, design, analysis, and results. Our analysis of citations provides some evidence that authors are often rewarded for not trying to “check all the boxes”, as the analysis suggests studies with lower Acknowledge scores received more citations, all else equal. Additionally, articles that thoroughly addressed issues raised in the study (i.e., having high Address scores) tended to receive more citations than ones that were less rigorous in addressing identified issues.

In recent years, there has been a call for bringing more transparency in economic research (e.g., Ferraro and Shukla 2018), which in part motivates the study reporting best practice guideline (BPG23) in Johnston et al. (2017). While we found that this best practice was only moderately followed by the past literature (on average), it seems likely that as the BPGs become more established and accepted by researchers, and expected to be followed by editors and reviewers, more complete study reporting and adherence to the guidelines will occur.

A few final caveats about our results are important. First, our analysis did not involve any source tracing. That is, we did not look for past peer-reviewed articles, technical reports, book chapters, working papers, or other materials that may have held details that were withheld from the studies we examined. We examined each peer-reviewed article in isolation, but acknowledge that, at least in some cases, additional information that could have resulted in higher Acknowledge or Address scores may have been contained in other documents. Second, our assessment of the SP literature valuing coastal and marine ES was limited to U.S.-focused peer-reviewed journal articles. This precluded consideration of adherence to SP best practices of studies reported in government reports, dissertations, working papers,

book chapters, or other “grey literature”, not to mention studies valuing ES outside of the United States. As such, the lessons about the literature gleaned from our analysis should be viewed within this somewhat narrow lens. Assessing the broader SP literature and the extent to which adherence in other countries and in the wider literature is left for future work. Third, we do not attempt to set a scoring threshold that indicates a “good” or “high-quality” study. Obviously, higher scores are better, but our analysis focused on relative scores when comparing across BPGs or across types of studies rather than absolute ones.^{9,10}

So what does this assessment suggest about the ability of the SP-based EVES literature to be a source of high-quality estimates for benefits transfers involving U.S. coastal and marine ES? First, our examination suggests that most of the attention in the literature has been on a small subset of ecosystem service values, particularly coastal and marine recreation and threatened and endangered species. This speaks to the availability of certain types of EVES information, but also suggests gaps in the literature for many other types of EVES and thus opportunities for SP researchers. For benefits transfers involving some of the lesser studied EVES, the paucity of available studies may pose problems for finding appropriate values to use that comply with best practices associated with the benefits transfer method (Johnston et al. 2021). Second, our assessment points to better adherence by more recent studies and to a likely upward trend in adherence for future studies. This suggests that researchers wishing to conduct benefits transfers should rely, if at all possible, on using EVES information from the more recent literature. Using recent estimates from the literature is also supported by research on how preferences and values may change over time, with recent studies of the temporal stability of SP values suggesting SP values may only be stable (unchanged) for a period of five years or less (Skourtos et al. 2010; Lew and Wallmo 2017). While our results suggest SP-focused adherence to best practices would be enhanced by using EVES from the more recent literature, this is but one criterion to consider when conducting benefits transfer. The recent best practices guidance for benefits transfer (Johnston et al. 2021) points to a variety of other considerations that are important to consider as well.

⁹ This is partly to acknowledge that despite efforts to avoid subjectivity and evaluator bias, the scoring system is somewhat subjective and therefore there may be some idiosyncratic errors in the scores.

¹⁰ Nevertheless, we illustrate the implication of setting a threshold score. One might try to set a study-level score threshold like 3.0 or higher. However, that would mean only 2.4% of studies meet this standard for the Acknowledge score, while the majority (55%) of studies would qualify for the Address score.

5. CONCLUSION

This article provides a first assessment of the extent to which the existing U.S. coastal and marine ecosystem service valuation literature adheres to a recently established set of best practice guidelines for stated preference research. Aggregating over the multiple criteria that underlie the best practices to provide a comparison at the study-level, we find significant variation in the extent to which different studies have adhered to the guidelines. Through this lens, our assessment can be used as a yardstick against which individual studies can be measured; better adherence to established best practices should yield more robust estimates of economic values, which are better suited for informing the management of U.S. coastal and marine ecosystems. Comparing adherence across the individual best practices criteria, we also find significant variation, suggesting that the literature has prioritized a more explicit treatment of some guidelines over others and highlighting distinct criteria that may warrant additional consideration in future studies. Overall, this assessment provides a framework to understand ways in which future studies could better incorporate best practices while also providing context to assess the potential reliability of the values found in the existing literature, as a means to inform applications of benefits transfer.

REFERENCES

- Arrow, K., Solow, R., Portney, P. R., Leamer, E. E., Radner, R., & Schuman, H., 1993. Report of the NOAA panel on contingent valuation. *Federal Register* 58(10): 4601-4614.
- Barbier, E. B., 2013. “Valuing ecosystem services for coastal wetland protection and restoration: Progress and challenges.” *Resources* 2(3): 213-230.
- Breslow, S. J., Sojka, B., Barnea, R., Basurto, X., Carothers, C., Charnley, S., Coulthard, S., Dolsak, N., Donatuto, J., Garcia-Quijano, C., Hicks, C., Levine, A., Mascia, M., Norman, K., Poe, M., Satterfield, T., St. Martin, K., & Levin, P. S., 2016. “Conceptualizing and operationalizing human wellbeing for ecosystem assessment and management.” *Environmental Science & Policy* 66: 250-259.
- Díaz, S., Demissew, S., Carabias, J., Joly, C., Lonsdale, M., Ash, N., Larigauderie, A., Adhikari, J., Arico, S., Baldi, A., Bartuska, A., Baste, I., Bilgin, A., Brondizio, E., Chan, K., Figueroa, V., Duraiappah, A., Fischer, M., &

- Zlatanova, D., 2015. “The IPBES Conceptual Framework—connecting nature and people.” *Current Opinion in Environmental Sustainability* 14: 1-16.
- Drummond, M., Brandt, A., Luce, B., & Rovira, J., 1993. “Standardizing methodologies for economic evaluation in health care: practice, problems, and potential.” *International Journal of Technology Assessment in Health Care* 9(1): 26-36.
- Ferraro, P. J., & Shukla, P. 2020. “Is a Replicability Crisis on the Horizon for Environmental and Resource Economics?” *Review of Environmental Economics and Policy* 14(2): 339-351.
- Freeman III, A. M., Herriges, J. A., & Kling, C. L., 2014. *The Measurement of Environmental and Resource Values: Theory and Methods*. Routledge.
- Gómez-Baggethun, E., & Barton, D. N., 2013. “Classifying and valuing ecosystem services for urban planning.” *Ecological Economics* 86: 235-245.
- Hamermesh, D. S., 2013. “Six decades of top economics publishing: Who and how?” *Journal of Economic Literature* 51(1): 162-72.
- Hausman, J., 2012. “Contingent valuation: from dubious to hopeless.” *Journal of Economic Perspectives* 26(4): 43-56.
- Johnston, R. J., Boyle, K. J., Adamowicz, W., Bennett, J., Brouwer, R., Cameron, T. A., Hanemann, W.M., Hanley, N., Ryan, M., Scarpa, R., Tourangeau, R., & Vossler, C. A., 2017. “Contemporary guidance for stated preference studies.” *Journal of the Association of Environmental and Resource Economists* 4(2): 319-405.
- Johnston, R. J., Boyle, K. J., Loureiro, M. L., Navrud, S., & Rolf J., 2021. “Guidance to Enhance the Validity and Credibility of Environmental Benefit Transfers.” *Environmental and Resource Economics* 79(3): 575-624
- Johnston, R. J., Ranson, M. H., Besedin, E. Y., & Helm, E. C., 2006. “What determines willingness to pay per fish? A meta-analysis of recreational fishing values.” *Marine Resource Economics* 21(1): 1-32.
- Kim, Y., Kling, C. L., & Zhao, J., 2015. “Understanding behavioral explanations of the WTP-WTA divergence through a neoclassical lens: Implications for environmental policy.” *Annual Review of Resource Economics* 7(1): 169-187.

- Kling, C. L., Phaneuf, D. J., & Zhao, J., 2012. "From Exxon to BP: Has some number become better than no number?" *Journal of Economic Perspectives* 26(4): 3-26.
- Kopp, R. J., & Smith, V. K., 1989. "Benefit estimation goes to court: The case of natural resource damage assessments." *Journal of Policy Analysis and Management* 8(4): 593-612.
- Kube, R., Löschel, A., Mertens, H., & Requate, T., 2018. "Research trends in environmental and resource economics: Insights from four decades of JEEM." *Journal of Environmental Economics and Management* 92: 433-464.
- Lew, D. K., & Wallmo, K., 2017. "Temporal stability of stated preferences for endangered species protection from choice experiments." *Ecological Economics* 131: 87-97.
- Lipton, D., Lew, D. K., Wallmo, K., Wiley, P., & Dvarskas, A., 2014. "The evolution of non-market valuation of US coastal and marine resources." *Journal of Ocean and Coastal Economics* 2014(1): 6.
- Loomis, J. B., 2014. "2013 WAEA keynote address: Strategies for overcoming hypothetical bias in stated preference surveys." *Journal of Agricultural and Resource Economics* 39(1): 34-46.
- MA, 2005. *Millennium Ecosystem Assessment*. Washington, DC: New Island, 13, 520.
- Menegaki, A. N., Olsen, S. B., & Tsagarakis, K. P., 2016. "Towards a common standard – A reporting checklist for web-based stated preference valuation surveys and a critique for mode surveys." *Journal of Choice Modelling* 18: 18-50.
- National Research Council, 2005. *Valuing Ecosystem Services: Toward Better Environmental Decision-Making*. National Academies Press.
- Nepal, N., Steltzer, E., Bohara, A. K., & Cullen, K., 2018. "Public values on offshore wind farm." *Environmental Economics and Policy Studies* 20(1): 225-240.
- Plummer, M. L., 2009. "Assessing benefit transfer for the valuation of ecosystem services." *Frontiers in Ecology and the Environment* 7(1): 38-45.
- Shogren, J. F., & Taylor, L. O., 2008. "On behavioral-environmental economics." *Review of Environmental Economics and Policy* 2(1): 26-44.

- Skourtos, M., Kontogianni, A., & Harrison, P. A., 2010. "Reviewing the dynamics of economic values and preferences for ecosystem goods and services." *Biodiversity and Conservation* 19(10): 2855-2872.
- Smith, V. K., 2000. "JEEM and non-market valuation: 1974–1998." *Journal of Environmental Economics and Management* 39(3): 351-374.
- TEEB, 2010. *The Economics of Ecosystems and Biodiversity Ecological and Economic Foundations*. Edited by Pushpam Kumar. Earthscan: London and Washington.
- Vossler, C.A., Bergeron, S., Doyon, M. & Rondeau, D., 2020. "Revisiting the gap between the willingness-to-pay and willingness-to-accept for public goods." CIRANO Working Papers 2020s-48, CIRANO.
- Watts, R. D., & Li, I. W., 2019. "Use of checklists in reviews of health economic evaluations, 2010 to 2018." *Value in Health* 22(3): 377-382.
- Whitehead, J. C., & Lew, D. K., 2020. "Estimating recreation benefits through joint estimation of revealed and stated preference discrete choice data." *Empirical Economics* 58(4): 2009-2029.

APPENDIX

Table A-1. Evaluation Criteria.

Evaluation criterion	eval_id
Does the survey instrument clearly explain the baseline (or status quo) conditions?	1.1
Is the mechanism of change clearly conveyed in the survey instrument in a way that is consistent with a plausible real-world action?	1.2
Is the change(s) to be valued clearly presented/described in an accurate, measurable, and interpretable way that avoids imprecise or qualitative terms (unless those terms are clearly defined)?	1.3
Are there survey questions that elicit evidence that information on the baseline, mechanism for change, and change(s) to be valued are understood, accepted, and viewed as credible by respondents?	1.4
Does the scenario design (and survey generally) take subjective perceptions into consideration when describing the actual change to be valued and incorporate means to minimize the role of subjective perceptions in interpreting survey/scenario information? This means considering the impacts of (a) provided information on subjects' responses to valuation questions; (b) framing of valuation questions; (c) sequencing in sequential valuation question formats; (d) respondents' prior experience and knowledge.	1.5
Is the cost (bid) amount clearly described (voluntary/mandatory, frequency of payment, duration of payment, and method of payment)? Is the payment vehicle and cost amounts plausible and salient to respondents? Is it consistent with the mechanism described that would bring about the change(s) valued?	1.6
When risk or uncertainty is an important aspect of the baseline or change being valued, do scenarios communicate this information in terms readily understood by respondents?	1.7
Were qualitative pretesting activities (focus groups, one-on-one interviews, verbal protocols) undertaken to evaluate respondent understanding, perceptions, and other design components of the survey? Were at least 4-6 focus groups or interview sessions (one-on-one, cognitive or verbal protocol interviews) conducted (more is better)?	2.1
Were peer reviews by other scientists/economists done?	2.2

If this was a large or high-stakes survey where aggregate values may be large and value estimates controversial, was quantitative pretesting with a smaller sample from the target population conducted (pilot survey)?	2.3
Were pretesting activities documented, including types of pretesting, number and characteristics of respondents, and scripts used to conduct the pretests (i.e., were pretest materials and data maintained and documented for quantitative pretesting; recordings of interviews and focus groups kept, and records of key decisions made at each step in the pretesting process)?	2.4
Was the choice between stated preference method (CE vs CV) based on respondent perceptions of the change being valued (attributes or as a whole), the decision objective being considered, and the type of information required (e.g., complexity)?	3.1
Was the decision/process for determining whether or not an attribute-based method was appropriate or not clearly documented?	3.2
Were attributes and attribute levels selected based on a combination of the values needed to support decision making, feasibility of implementation, plausibility to respondents, and statistical efficiency?	4.1
Does the experimental design allow for interactions (and perhaps other types of nonlinear-in-attributes utility functions), consider both statistical efficiency and respondents' cognitive abilities and attention budgets, employ constraints on implausible attribute levels and combinations, use designs that are robust to alternative model specifications, and consider how the levels chosen for each attribute influence design properties?	4.2
Was the experimental design evaluated using qualitative pretesting or pilot (pretest) studies?	4.3
Was the study design/SP protocols reviewed by a university or review board tasked with protection of human subjects?	5.1
Was informed consent obtained from subjects (or provided in a cover letter or on the survey itself regarding the voluntary nature of the study, for example)?	5.2
Did survey procedures avoid deception that could have significant negative consequences for respondents, unintended influences on study outcomes or validity, or compromise the ability to use study results to support decision making?	5.3
Were steps taken to protect confidentiality and sensitive data, including use of standard practices for data storage and study reporting?	5.4
Was the rationale for the mode of data collection documented?	6.1
Was the sample(s) drawn from known frames that are consistent with the population from which values are to be estimated (incl. extent of market considerations such as clearly defined market area)?	6.2

Were respondents randomly selected from the sampling frame using an explicit sampling procedure?	6.3
Were efforts made to identify/assess and then mitigate (if present) nonresponse bias, such as survey design features and data collections (nonresponse surveys) to aid in identifying and characterizing nonresponse patterns?	6.4
Is the justification for choice between WTP and WTA welfare measures provided?	7.1
Is the choice between WTP and WTA motivated by theoretical and empirical considerations?	7.2
Is a single binary choice question format used with one of the choice alternatives being the status quo or opt-out alternative?	8.1
If a single binary choice format is not used, are the reasons for this choice explained and the potential implications for welfare estimates discussed?	8.2
If more than one choice question is asked, did the study discuss or evaluate the trade-offs in bias, efficiency, and evolution of choice heuristics? Was the question order randomized across respondents? Was pretesting done to evaluate complexity, difficulty of choices, and the potential for presence of choice heuristics?	8.3
Does the response format include a no-answer option distinct from the status quo alternative?	9.1
If a no-answer option is used, is this justified or explored during pretesting activities?	9.2
Is the decision rule realistic and binding on respondents?	10.1
If collective decision rules do not apply (e.g., choosing between trips or purchase a private good), is the decision-making frame plausible and its use justified?	10.2
Is the payment vehicle realistic, credible, familiar, and binding for all respondents to the maximum extent possible?	11.1
Is the payment vehicle on that is viewed as fixed and unmalleable?	11.2
Was the payment vehicle evaluated/tested in qualitative pretesting?	11.3
Is the payment vehicle adequately described and the potential for it to under or over-estimate values discussed?	11.4

Are auxiliary or supplemental questions asked to enhance the validity of the SP study, including for the purpose of (a) breaking up flow of long sections of text; (b) helping engage respondents in processing information; (c) evaluating understanding and/or acceptance of information; (d) identifying protest responses or other motivations for value elicitation responses; (e) providing information to evaluate validity; (f) evaluating respondent's perceptions of survey instrument/experience (e.g., difficulty, neutrality); (g) understanding respondent's attitudes, opinions, behaviors, knowledge, and experiences; and (h) identifying household and individual characteristics, including demographics?	12.1
Did questions have a specific purpose that was established ex ante?	12.2
Were questions pretested to ensure they served the intended purposes?	12.3
Were valuation scenarios and valuation questions designed to enhance incentive compatibility and encourage truthful responses?	13.1
Were approaches that enhance the (policy and payment) consequentiality of valuation scenarios used?	13.2
Does the selection of econometric estimator reflect the unique aspects of the data to be analyzed, the hypotheses to be investigated, and how the estimation results will be used to support decision making?	14.1
Does the study document the justification for the econometric models used and the trade-offs involved in their use?	14.2
Is the econometric model guided by utility theory and considerations of behavioral and statistical properties, and other assumptions?	14.3
Does the study consider whether and how heterogeneity may be relevant to consistent estimates of preference parameters, interpretation of estimation results, and the use of point estimates to compute aggregate welfare measures under historical or new conditions?	15.1
Are the researcher choices made about model selection and estimation justified and documented?	15.2
Does the study include the simplest, most parsimonious specifications with maintained hypotheses consistent with the basic axioms of choice and properties of the data (e.g., conditional logit models)?	16.1
Does the study include more complex models that impose additional investigator assumptions on the structure of the responses?	16.2
If indicated by prior research (e.g., similar studies) or pretesting, were anomalous behaviors (strategic response, protesting, use of heuristics, outliers, anchoring, scenario rejection, etc.) investigated to an appropriate degree (both in design and estimation)?	17.1
Are the methods used to calculate welfare transparent and are the welfare estimates consistent with theory and statistically well-defined?	18.1

Are both measures of central tendency and dispersion of welfare estimates reported?	18.2
If multiple models are estimated, is the preferred model for computing welfare estimates identified and reasons for its selection documented?	18.3
Were the supporting or debriefing questions used in analysis justified by theoretical, survey design, or empirical arguments that explain their use?	19.1
When used in model estimation, is consideration of potential endogeneity (e.g., variables related to valuation responses/questions) and related concerns such as measurement error considered appropriately?	19.2
Is the generalizability of value estimates documented?	20.1
For analyses seeking to produce decision or policy-relevant estimates, are there assessments to support the generalizability of value estimates to the sampled population?	20.2
For analyses not seeking aggregate values, is there sufficient information provided that would allow someone to generalize to the population (e.g., during a benefits transfer) or alternatively a caution about an inability to link to general population preferences?	20.3
Are respondent characteristics documented in terms of standard socioeconomic characteristics and key study-specific characteristics (often used to evaluate sample representativeness)?	20.4
If aggregate values are computed, do they recognize potential effects of sample selection, preference heterogeneity, and the extent of the market?	20.5
Are modifications in value estimates to address issues related to generalizing from the sample documented?	20.6
Does the study include formal tests of construct validity and evaluations of content validity? (For example, scope/embedding tests, analyses of consequentiality, binding payments, and plausible scenarios; attribute non-attendance, protest responses, serial nonparticipation.)	21.1
Does the study consider potential confounding influences, effects of study design, and role of investigator-imposed assumptions in assessment of the validity?	21.2
Does the study include multiple validity tests targeted to specific issues or concerns that might arise within the application?	22.1
If the study assesses the validity of the results within the study, does it rely on the weight of evidence from multiple study-specific tests and the body of evidence in the literature?	22.2
Are key aspects of the study design, implementation, analyses, and results clearly documented to ensure sufficient information is provided to fully understand the study and results?	23.1

Does the study provide detailed archival reporting and documentation necessary for replication? (May be important at the study level versus journal article level)	23.2
--	------

Table A-2. Acknowledge and Address Score Levels.

Score	Acknowledge Score <i>“Does the study acknowledge/mention the issue(s) underlying criterion...”</i>	Address Score <i>“For those acknowledged/mentioned, how completely are they addressed in the study?”</i>
1	No elements of the issue(s) are acknowledged	Not at all addressed
2	Very few elements are acknowledged	Only a little addressed
3	Some, but not all, elements are acknowledged	Somewhat addressed
4	Most elements are acknowledged	Mostly addressed
5	All elements are acknowledged	Fully addressed

Table A-3. Full list of studies evaluated (N=82).

Year	Author(s)	Journal	DOI
1986	Samples, K.C., Dixon, J.A. and M.M. Gowen	Land Economics	10.2307/3146394
1989	D. Huppert	Marine Resource Economics	10.1086/mre.6.2.42629010
1989	Trudy Ann Cameron and Daniel D. Huppert	Journal of Environmental Economics and Management	10.1016/0095-0696(89)90018-1
1989	Neal S. Johnson and Richard M. Adams	Marine Resource Economics	10.1086/mre.6.1.42629002
1990	John C. Bergstrom, John R. Stoll, John P. Titre, Vernon L. Wright	Ecological Economics	10.1016/0921-8009(90)90004-e
1990	Richard Carson, Michael Hanemann, and Dan Steinberg	Journal of Behavioral Economics	10.1016/0090-5720(90)90017-2
1991	Thomas H. Stevens, Jaime Echeverria, Ronald J. Glass, Tim Hager and Thomas A. More	Land Economics	10.2307/3146546
1992	Lindsay, Bruce E., John M. Halstead, Helen C. Tupper, and Jerry J. Vaske	Coastal Management	10.1080/08920759209362179
1992	John C. Whitehead	Applied Economics	10.1080/00036849200000075
1992	Silberman, J., D.A. Gerlowski, and N.A. Williams	Land Economics	10.2307/3146776
1992	Trudy Ann Cameron	Land Economics	10.2307/3146378
1993	John C. Whitehead	Marine Resource Economics	10.1086/mre.8.2.42629055
1994	John B. Loomis and Douglas M. Larson	Marine Resource Economics	10.1086/mre.9.3.42629085
1996	Brian Roe, Kevin Boyle, Mario Tiesl	Journal of Environmental Economics and Management	10.1006/jeeem.1996.0037
1996	John Loomis	Water Resources Research	10.1029/95WR03243
1998	Tkac, J.	American Journal of Agricultural Economics	10.2307/1244227
1998	Jeffrey D. Kline and Stephen K. Swallow	Coastal Management	10.1080/08920759809362351
1999	K.L. Giraud, J.B. Loomis, and R.L. Johnson	Journal of Environmental Management	10.1006/jema.1999.02777

2000	R.P. Berrens, A.K. Bohara, C.L. Silva, D. Brookshire, and M.McKee	Journal of Environmental Management	10.1006/jema.1999.0308
2000	Kotchen, Matthew J. and Stephen D. Reiling	Ecological Economics	10.1016/s0921-8009(99)00069-5
2001	Robert J. Johnston, James J. Opaluch, Thomas A. Grigalunas, and Marisa J. Mazzotta	Growth and Change	10.1111/0017-4815.00161
2001	David F Layton	American Journal of Agricultural Economics	10.1111/0002-9092.00284
2001	John C. Whitehead, William B. Clifford, and Thomas J. Hoban	Marine Resource Economics	10.1086/mre.16.3.42629318
2002	T. Park, J.M. Bowker, and V.R. Leeworthy	Journal of Environmental Management	10.1006/jema.2002.0552
2002	Hamel, C., Herrmann, M., Lee, S.T., Criddle, K., and Geier, H.T.	Annals of Regional Science	10.1007/s001680200085
2002	Kelly Giraud, Branka Turcin, John Loomis, and Joseph Cooper	Marine Policy	10.1016/S0308-597X(02)00025-8
2003	D. Gillig, R. Woodward, T. Ozuna Jr. and W. Griffin	Agricultural and Resource Economics Review	10.1017/s1068280500005980
2003	Bell, Kathleen P., Daniel Huppert, and Rebecca L. Johnson	Marine Resource Economics	10.1086/mre.18.1.42629381
2003	Criddle, K., Herrmann, M., Lee, S.T. and Hamel, C.	Marine Resource Economics	10.1086/mre.18.4.42629404
2003	Landry, Craig E., Andrew G. Keeler, Warren Kriesel	Marine Resource Economics	10.1086/mre.18.2.42629388
2004	Douglas M. Larson, Sabina L. Shaikh, and David F. Layton	American Journal of Agricultural Economics	10.1111/j.0092-5853.2004.00580.x
2004	Solomon, Barry D., Cristi M. Corey-Luse, and Kathleen E. Halvorsen	Ecological Economics	10.1016/j.ecolecon.2004.03.025
2004	Robert N. Cantrell, Marissa Garcia, PingSun Leung, David Ziemann	Fisheries Research	10.1016/j.fishres.2004.01.003
2004	Lipton, Douglas	Marine Resource Economics	10.1086/mre.19.2.42629432
2005	Chi-Ok Oh, Robert B. Ditton, Brad Gentner, and Robin Reichers	Human Dimensions of Wildlife	10.1080/10871200591003427
2006	John R. Stoll and Robert B. Ditton	Human Dimensions of Wildlife	10.1080/10871200500470969
2006	J. Walter Milon and David Scrogin	Ecological Economics	10.1016/j.ecolecon.2005.01.009

2006	Oh, C. and R. B. Ditton	Leisure Sciences	10.1080/01490400600745886
2006	Whitehead, John C.	Applied Economics	10.1080/00036840500427130
2006	David F. Layton and S. Todd Lee	Environmental and Resource Economics	10.1007/s10640-005-3784-9
2006	Edward R. Morey and William S. Breffle	American Journal of Agricultural Economics	10.1111/j.1467-8276.2006.00844.x
2006	Massey, D.M., Newbold, S.C. and B. Gentner	Journal of Environmental Economics and Management	10.1016/j.jeem.2006.02.001
2007	Gwendolyn A. Aldrich, Kristine M. Grimsrud, Jennifer A. Thacher, Matthew J. Kotchen	Environmental and Resource Economics	10.1007/s10640-006-9054-7
2007	Nancy A. Connelly, Tommy L. Brown, and Jonathan W. Brown	Journal of the American Water Resources Association	10.1111/j.1752-1688.2007.00083.x
2007	Huang, J., Poor, J. and Zhao, M.	Marine Resource Economics	10.1086/mre.22.3.42629556
2007	Oh, Chi-Ok, Robert B. Ditton, and Robin Riechers	Environmental Management	10.1007/s00267-006-0010-7
2008	Wallmo, K. and S. Edwards	Marine Resource Economics	10.1086/mre.23.3.42629620
2008	Chi-Ok Oh, Anthony W Dixon, James W Mjelde, and Jason Draper	Ocean and Coastal Management	10.1016/j.ocecoaman.2008.09.003
2009	John C. Whitehead, Peter A. Groothuis, Rob Southwick, Pat Foster-Turley	Journal of Great Lakes Research	10.1016/j.jglr.2009.03.005
2009	Oh, C., Draper, J., and Dixon, A.W.	Coastal Management	10.1080/08920750802701128
2010	Lew, Daniel K., David F. Layton, and Robert D. Rowe	Marine Resource Economics	10.5950/0738-1360-25.2.133
2010	Giraud, Kelly, and Branka Valcic	Journal of International Wildlife Law and Policy	10.1080/13880290490480167
2010	Oh, C., Draper, J. and Dixon, A.W.	Ocean and Coastal Management	10.1016/j.ocecoaman.2010.04.007
2011	Kristy Wallmo and Daniel K. Lew	Journal of Environmental Management	10.1016/j.jenvman.2011.02.012
2011	Daniel K. Lew and Kristy Wallmo	Environmental and Resource Economics	10.1007/s10640-010-9394-1
2012	Anthony W. Dixon, Chi-Ok Oh, and Jason Draper	Journal of Travel Research	10.1177/0047287512451136
2012	John Duffield, Chris Neher, Stewart Allen, David Patterson, Brad Gentner	Marine Resource Economics	10.5950/0738-1360-27.4.343

2012	Wallmo, K. and D.K. Lew	Conservation Biology	10.1111/j.1523-1739.2012.01899.x
2012	Daniel K. Lew and Douglas M. Larson	North American Journal of Fisheries Management	10.1080/02755947.2012.681012
2012	Carter, David W. and Christopher Liese	North American Journal of Fisheries Management	10.1080/02755947.2012.675943
2013	John Loomis and Luis Santiago	Coastal Management	10.1080/08920753.2012.749754
2013	Oh, Chi-Ok	Ocean and Coastal Management	10.1016/j.ocecoaman.2012.10.004
2013	Leif E. Anderson, S. Todd Lee, and Phillip S. Levin	Land Economics	10.3368/le.89.2.371
2013	Leif E. Anderson and S. Todd Lee	Marine Resource Economics	10.5950/0738-1360-28.2.175
2014	Jerrold Penn, Wuyang Hu, Linda Cox, and Lara Kozloff	Ocean and Coastal Management	10.1016/j.ocecoaman.2014.06.002
2014	Daniel R. Petrolia, Matthew G. Interis and Joonhyun Hwang	Marine Resource Economics	10.1086/676289
2014	Lew, Daniel K., and D. M. Larson	Fisheries Research	10.1016/j.fishres.2014.04.005
2015	Geoffrey S. Shideler, David W. Carter, Christopher Liese, Joseph E. Serafy	Fisheries Research	10.1016/j.fishres.2014.07.009
2015	Kristy Wallmo, Daniel K. Lew	Frontiers in Marine Science	doi.org/10.3389/fmars.2015.00055
2015	Lew, D.K. and Larson, D.M.	Marine Policy	10.1016/j.marpol.2015.07.007
2016	Shanna Grafeld, Kirsten Oleson, Michele Barnes, Marcus Peng, Catherine Chan, Mariska Weijerman	Ecological Economics	10.1016/j.ecolecon.2016.05.005
2016	Interis, M.G. and D.R. Petrolia	Land Economics	10.3368/le.92.2.292
2016	Wallmo, K and D.K. Lew	Journal of Environmental Management	10.1016/j.jenvman.2016.04.053
2016	Jerrold Penn, Wuyang Hu, Linda Cox, and Lara Kozloff	Marine Resource Economics	10.1086/683795
2017	Kristy Wallmo and Rosemary Kosaka	Marine Policy	10.1016/j.marpol.2017.05.034
2017	Liese, C. and D. Carter	Marine Fisheries Review	10.7755/MFR.79.3-4.1
2017	Marcus Peng and Kirsten L.L. Oleson	Ecological Economics	10.1016/j.ecoecon.2017.02.0030921-8009
2017	Lew, D. and K. Wallmo	Ecological Economics	10.1016/j.ecolecon.2016.08.009
2017	Lew, D.K., and Larson, D.M.	Marine Fisheries Review	10.7755/MFR.79.3-4.2

2017	Min-Yang Lee, Scott Steinback, and Kristy Wallmo	Marine Resource Economics	10.1086/690676
2017	Jorge Holzer and Kenneth McConnell	Journal of the Association of Environmental and Resource Economists	10.1086/691976
2018	Naresh Nepal, Eric Stelzer, Alok K. Bohara and Kelly Cullen	Environmental Economics and Policy Studies	10.1007/s10018-017-0191-8