Water Valuation and the Application of Declared Preferences Method: Contingent Valuation vs a Choice Experiment Approach

Óscar Alfranca
Polytechnic University of Catalonia

In the Declared Preferences methods (DP) questions are developed to different persons in order to establish alternative hypothetical scenarios. The most important methods which are based in a Declared Preferences hypothesis are the Contingent Valuation (CV) method and those which are based in a Choice Experiment (CE) Approach. In DP methods questions are used in order to establish preferences over hypothetical scenarios, which could be different (either goods or services). Each alternative is described through different attributes, and the answers can be used in order to determine the different preferences. Some very relevant methods which are based in the DP hypothesis can be found which are based in CV and CE methods.

Keywords: declared preferences, revealed preferences, experiments of election, choice experiments

INTRODUCTION

If the Declared Preferences (DP) method is valid for the estimation of some environmental values, then these methods could be useful for addressing problems such as, for example, what might be the consequences of holding a referendum on some environmental issue. Of special interest could be, for example, some non-use methods.

Some valuation methods are based on the DP hypothesis. The main methods are based on Contingent Valuation (CV) and Choice Experiments (CE), although another relevant method could be, for example, Contingent Behavior (CB). Questions related to CC focus on potential hypothetical behaviors (e.g., how willing would people be to visit a park if an entrance fee were imposed at a certain price).

Valuation methods based on DP techniques allow to design the approximation of the personal values of goods related to the environment. A principal virtue of these methods lies in their conceptual simplicity. Another important method that is based on DP is the CE. In this case, the CE are described by a finite number of attributes, which can refer to different levels, and thus a good is described by a limited number of characteristics, whose individual attributes can take different levels. A good can be characterized by a limited number of attributes (i.e., the specific alternatives of the different factors)[1]. Water valuation can be carried out by applying any of the DP methods, such as CV and CE, CB, among others. Questions based on the CB method raise questions such as "How many times could a park be visited if a 2 euro entrance fee were imposed?". In contrast to the CV questions, those related to the CE method focus on Hypothetical Behavior.

Among the objectives of some of the studies, conducted using the DP method, are, among others, the collection of data to maximize the validity and reliability of the estimated results. As is commonly reflected

in the literature, the DP-based valuation methods and the sample used should be effectively collected with the desired properties.

In the application of an effective research method based on the DP, (with which to design the instruments that could be used in the environmental valuation), and maintaining the status quo in the sample. In this case we find the selection of a random sample, with a chosen population and the selection of a method with the fundamental environmental conditions.

The conflict between the recommendations of the guidelines, the evidence in contemporary environmental literature and the design of an environmental questionnaire, based on environmental practices and based on the DP method, is of particular interest. The recommendations described here are based on a review of the literature on DP methods, applied in different fields, such as environmental economics, health economics or transport economics.

WATER VALUATION AND DP METHODS

In DP methods, questions are asked to individuals in order to establish their preferences about alternative hypothetical scenarios (either goods or services). Each alternative is described through different attributes, and the answers can be used in order to determine the different preferences. Some very relevant methods which are based in the DP hypothesis can be found which are based in CV and CE methods. Examples based on CV can be, in some cases, controversial[2]. Although the CE method has been frequently applied in the treatment of reclaimed water, this method has not been generally used in water management.

The CE is a quantitative technique based on the DP method (which allows the establishment of individual particularities), which enables to discover how different individuals value various attributes of a program, product or service by means of questions based on the establishment of a choice on hypothetical alternatives. The CE are applied in a variety of methods, which can be related to business planning, resource location or health policy, among others.

The application of the DP technique makes it possible to discover how individuals value the various attributes of a good. The CE is a quantitative technique based on the application of DP. CE is a quantitative method, which can be used in the absence of data on Revealed Preferences (RP). The method involves and develops questions to ask individuals to establish their preferences about alternative hypothetical scenarios (either goods or services). The CE have been applied in diverse methods that include aspects related to health policy, business planning, resource location, or their relationship with the level of income. Although the CE method has been frequently applied in the treatment of reclaimed water, this method has not been generally used in water management.

APPLICATION OF THE DP METHOD

DP-based methods can provide estimates regarding the values associated with changes in welfare. The status quo level makes it possible to propose relative changes, which are usually established to be credible, and thus to answer possible questions. Thus, possible scenarios are presented, based on the DP hypothesis, and in stages that allow the application of a method based on this same hypothesis, and linked to the description of the scenario.

Description of the Scenario
The questionnaires should describe the type and extent of the scenario change, using language that is tailored and understandable to the respondent. The description of this scenario also requires information regarding the overall mechanism (usually a policy, investment or management review), which would generate changes that could be assessed. The mechanism is described at a practical level, (so that it would be consistent and real), which could induce change.

The questionnaires should describe the type and extent of the change, using language that is tailored and understandable to the respondent. The description of the scenario also requires information about the mechanism or set of mechanisms (typically a policy revision, investment or management change) that may

generate the business that should be assessed. In other words, the mechanisms should be described in a practical and consistent way, with a real world, which could be modified[3]. Inaccurate and qualitative terms (such as "high", "medium" and "low") should be avoided until they are clearly defined by those giving the related answers. From the above, intermediate changes (or processes to generate changes), produced by those who give the answers, should also be valued[4].

Baseline conditions and their changes, as well as other scenario assessment components, may include subjective perceptions for the responses[5]. The scenario design should take subjective perceptions into consideration when describing the change to be assessed. This would include the presentation of information in a way that people would understand and accept, and likewise, deliberation about strategies would be used in order to evaluate responses.

Pre-Test Exposition

The Pre-Test method is a comprehensive, specific method based on a guide and a questionnaire[6], based on the DP[7]. Pre-test quality is a necessary component of the design, since it is essential to assess whether the number of interviews conducted is appropriate, and each of them may vary in context. For most applications, a minimum of four to six groups is recommended, and a higher number is suggested for new, unfamiliar, or difficult to quantify assets.

The tests are usually performed using data from pilot studies, which allow the design of DP studies, in which preliminary response rates and behaviors are established, with statistical hypothesis testing.

Any type of trial could be conducted using population-based tests, whenever possible. The design of the study should reach its screening test, including the selection of the chosen groups or the conduct of interviews, which would allow the characteristics of the participants to be obtained, as well as the application of selection methods.

Pre-testing could be documented including types of pre-tests, numbers and characteristics of respondents, and sequences used to manage pre-tests. For quantitative pre-tests, the instruments, for data and surveys, should be retained, and their content should be maintained and documented. Audio and video recordings should be useful when the group's substance or the interview content needs to be reviewed at a later date.

The record should be kept for key decisions, and for each step of the testing process. In addition, the results published and identified, as well as the decisions taken in the design phase, should also be represented.

Approaches Based on the Presence or Absence of Water-Based Attributes

The validity, reliability and applicability of DP studies depend on the explosion of the changes caused in the water. Even if there has been an increase in the use of CE in recent times, it is not clear whether CV or CE, in general, allow a higher valuation to be obtained. Each has advantages and disadvantages[8].

In contrast, CV offers opportunities to estimate values when an element cannot be easily defined in terms of attributes.

The choice and differentiation between CV and CE is complex and should be based on the answers obtained and on the perceptions of the changes in values, the objective decision being considered, and the type of information required. For example, while the valuation application may be conceptualized and communicated as a set of attributes, responses to reports, respondents may or may not think of changes in terms of attributes.

On the other hand, CE may disclose information regarding the value of individual attributes, and their desire to modify some of them. These marginal values of individual attributes, which may remain invisible in a classical CV study.

The two approaches also have different advantages and disadvantages in terms of information provision (at least as commonly designated). Attributes presented using short descriptions, usually in CE, can sometimes simplify important features of the scenario[9].

Given such considerations, the choice of a question format should not preclude the attribution, or non-attribution, of methods based solely on factors such as ease of application or prevalence in the literature.

Some basic considerations are suggested when making decisions to apply the CV or CE format. First of all, the assessment of the specific characteristics of the articles (or the articles as a whole), and what the information needs of the decision-makers are, is modified. In some cases, such as the estimation of damage spills related to oil spills, the question can be linked to the estimation of specific values, and of a set of changes, as a whole.

CV allows for such supports in a decision-making context. In other applications, where a change may affect some attributes and not others, or where a range of different attribute changes need to be evaluated. In these cases, a CE may be more readily applicable because it provides marginal values for individual attributes of interest over a range of possible changes. This can be particularly useful when there is uncertainty about the physical and biological impact. In such cases, the valuation of a range of possible results by different attributes of levels in the CE provides a means of ensuring that the true impact has been assessed, or that values for a range of impacts are available to the decision maker.

Secondly, those who answer, think and value the changes, in terms of individual attributes or as a whole[10]. One consideration is related to the DP structure, although respondents may think about the valuation of change in terms of attributes. The preference structure for these attributes could be more complex than simply linear or additive and separable preferences, which could involve the presentation of a matrix of CE questions and which are typically imposed as the assumptions of an econometric analysis on the response data.

In these cases, considerations would be offered if the CE models were sufficiently subtle to allow estimates of the ways in which different attributes might enter into the utility function of the respondents. In other words, if valid implicit prices could be obtained to represent (the willingness to pay for each attribute), and this attribute could be estimated.

The manner in which the presentation format influences, i.e., the understanding of what is to be valued, should also be considered. If the article is large and complex, it is difficult to describe how to adequately value the changes generated, and a set of attributes is used.

In this type of application, the use of attributes for communicating the changes to be valued can increase the scenario complexity, particularly when the number of attributes is large, or simply the number of attributes is not considered, or simple attribute descriptions are not applied[11]. Complex choices may also trigger responses to cope with simplification, trial-and-error or response strategies inconsistent with decisions that would maximize utility[12]. It is vital to keep in mind that these complexity-related challenges could affect the application of the CV method[13].

Although the respondent can think about how to value changes in terms of attributes, the preference structure of attributes may be more complex than a simple linear, separable, additive preference function, imposed as an econometric analytical assumption in response to the data. In such cases, consideration should be given to whether the CE are sufficiently nuanced to allow estimates of how the different attributes address the response to utility, such that valid implicit prices (i.e., how likely it is that the willingness to pay for each of the attributes can be estimated).

Third, how does the presentation of the information affect the understanding of the good to be valued? If the good is large and complex, it is likely to be difficult to describe how the change is valued adequately, using a parsimonious set of attributes[14]. Complex choices may also trigger trial-and-error responses or response strategies that are not consistent with utility-maximizing decisions, and other methods[15], such as Systematic Conjectures[16], or Selective Inattention[17] and Sequential Effects[18]. The prevalence of anomalies may depend on several factors, including previous responses to knowledge or risks due to complexity[19].

Experimental Design

In this paper we discuss the Experimental Design (ED), in the context of a bidding in a CV exercise, discrete choice and an attribute-level CE assignment. ED in these contexts, and particularly for CE, is complex and elaborate[20].

The ED assumption defines the way in which different treatments are assigned to a question (or sequence of questions within a questionnaire)[21].

The primary objective of an ED, as applied to CV and CE methods, is to develop the means to obtain efficient and unbiased estimators, preference parameters and value estimators. Designs should make use of information obtained from empirical research, and require a pre-testing.

All treatment effects, and interactions with attributes, should be individually or jointly identified.

The ED should generally allow for interactions, (and perhaps other types of non-linear utility attributes), considering both statistical efficiency, cognitive capabilities and their estimates, constraints related to their implausibility, and the level of their attributes, as well as the combination in the design of the use, (which may be consistent in alternative model specifications), and considers how the levels related to the different attributes influence the design of their property.

Among the studies based on the DP method, a priori sources of information used in the formulation of the chosen designs, which can provide evidence to support sources of information, (which can be essential in their a priori formulation), should be distributed and used to design formulations, which can enable the design of evidence, and which provide evidence to enable the chosen design, and inform the different steps in the design of this process.

For a CV-related situation, investigators should choose the cost or quantity bid in consideration for the range and in relation to the space[22]. The Effective Design for these questions ensures that the monetary quantity is credible to the respondents, it trains the estimated welfare in an unbiased way, and minimizes the variance of these estimates[23]. Four components are involved in the basic design of the CE:

(1) Selection of attributes, and the levels for each of them.
(2) Decision on how many alternatives will be needed for each problem, as well as the number of questions for each subject.
(3) Determination on how the levels of each attribute will be combined among the different alternatives.
(4) Determination on how the attributes corresponding to the different questions are combined, can close the step for the presentation of the subsets of the respondents.

Combinations between questions and attributes can be randomly assigned to individual respondents, although this may result in a timely conclusion to increase the efficiency of the estimation.

Estimates of multi-attribute designs statistically enable a welfare that can be consistent with the model specification, and that can ensure the identification of important attribute-related as well as their interaction effects. The effective and statistically efficient multi-attribute design enables the estimated welfare that is consistent with the model specification, as well as ensuring the identification of the main effects and attributes of the interactive effects.

Experimental design decisions are informed by multiple factors, including change to be assessed, prior studies on specific applications, along with intuitions learned through surveys acquired through a design process. Attributes and levels should be selected and based on a combination of values needed to support implementation decision making, response plausibility and statistical efficiency[24]. Levels should be chosen in light of the functional form to be used for the utility. For example, one might expect linear rates of change related to some utility, as well as its dimensions or non-linear changes, among others. The number of levels for each attribute should be sufficient to model these effects.

DP methods are often applied to inform public decisions, allocating knowledge sensitivity on issues related to ethics and research. Familiarity and compliance with recognized standards of ethical research, as well as the protection of human subjects, help to ensure that the stated preference is beyond doubt.

Data Collection and Ethics

DP methods are increasingly being used to inform public decisions, placing knowledge on issues related to ethics and research. The familiarity and conformity with norms or standards related to ethics research and the protection of human subjects can help to ensure that research on stated preferences is beyond doubt.

The DP protocol should be reviewed by universities or other boards to protect individuals, and informed consent should be obtained from them. The rights and welfare of human subjects (i.e., people) should be paramount, and their informed consent should be obtained from them. The study procedures should avoid deception, which could have significant negative consequences for respondents, unintended influence on

the results and validity of the study, or compromise the ability to use the study results to aid in decision making.

Mode Extension: Sampling and Non-Response Bias

DP research relies on multiple applications to collect data and sample data to identify respondents (including internet administration)[25].

Common modes include traditional email and in-person surveys, along with new approaches, such as the use of administration and the internet, mixed surveys with other surveys of the people being considered, mixed surveys and other electronic methods (such as tablets), and workshop evaluation[26]. Each method has advantages and disadvantages, and there is no consensus regarding which method is the best[27]. The use of personal interviews is the method recommended by the NOAA panel, and it allows interviews that can be followed in administrative processes, which could be subject to social desirability, and raise the related effects[28].

The most appropriate data collection is a specific context and the rationale for selecting how it should be documented. For all administration methods there are advantages and disadvantages. However, given the inability to effectively verbalize a complex evaluation of materials (such as a telephone conversation), this survey should be conducted with caution.

Samples should be collected from structures consistent with the populations whose values were estimated, and the respondents should be randomly selected from a structure in which an explicit procedure was used to select the sample. Contemporary approaches should be used in order to identify and mitigate the bias of those sample members who choose not to respond, including survey features and data collection to help identify and characterize the patterns of non-respondents. Wherever possible, the analysts could not rely solely on response rate and demographic data to infer the presence or absence of non-response bias.

Survey design and sampling considerations affect, and are affected by, multiple aspects of survey design, including sample structure and representativeness. A related survey requires a complete list of names/addresses to choose respondents. Those who respond on a website are often recruited via the Internet into a panel (and then a full survey is requested, once they have become panel members). Assuming the initial list, in which the samples are representative of the target population, these two approaches are response bottlenecks. For an e-mail survey, respondents can view the subject line and then decide whether to reply, introducing possibilities that the subject line is related to the same topic. For e-mail survey, the responders can look at the topic and then decide what form of response they will develop, introducing possibilities about the topic and bias.

For the panel survey and recruitment process, it may influence the behavior of the responses and the respondent, who should agree to become a panel that may not reflect the population through observable and unobservable characteristics. Through telephone administration, sample recruitment can be achieved with Random-Digit Dialing (RDD), but nowadays, due to the proliferation of cell phones, the use of dialing numbers has become more complicated, and the response rate has been reduced. Telephone lines established per household can vary based on income and other factors, affecting the likelihood that any household can be reached.

SURVEY MODE AND SAMPLING CONSIDERATIONS

These considerations affect and are affected by multiple aspects of the design, including the form of representativeness. In other words, a mail survey requires a complete list of names and addresses. A web-survey is often contracted to form a panel and then a full panel survey is asked once the members are part of the panel. Assuming the initial list from which the samples are selected and the target population, these two approaches present different impediments to response. For a summary e-mail, respondents can observe, and then decide, whether to replicate or not, introducing possibilities about selection and bias.

For conducting a web-based survey, the panel recruitment process may influence the panel from which those respondents are collected, and respondents who agree to be part of these panels may not reflect the population for all observable and unobservable characteristics. Using telephone administration, sample

recruitment can be achieved by RDD, but the proliferation of cell phones and number blocking has complicated sampling and reduced response rates. Telephone lines per household may vary according to income and other factors, which affect the likelihood that household members will be able to reach it.

The survey method and the way the survey is influenced by multiple design aspects, including representativeness. Recent research suggests that the mode of data collection does not substantially influence DP studies[29].

The market extension is usually defined, in principle, as a group of agents whose welfare would be influenced by the valuation of changes in a DP study. While the market extension could be difficult to identify for non-use values. In addition, differences between jurisdictions and locations of those who are affected may influence the understanding of the permissible market extension. For example, for many national policy decisions, residents of other countries affected by police typically do not hold up legally in a formal Cost-Benefit Analysis[30].

Determining the market extension is important to ensure that the correct market structure has been selected. It is also necessary when expanding (or aggregating), estimating population values (e.g., calculating the CBA). A clearly defined market area also makes it easier to determine whether the sample shape is available and complete (i. e., whether all affected individuals have a known probability of selecting the sample). A selected sample structure is essential for the identification of non-response effects.

Survey Mode and Sampling

Face-to-face interviews have the desirable feature that the completion process can be guided through trained interviews, but this approach is an expensive way to collect data and responses may be affected by unexpected or unanticipated interview-related effects. Telephone surveys are usually relatively inexpensive and are convenient for quick data collection[31].

Many countries provide access to representative sample structures, such as neighborhood address lists for email surveys[32].

Response Rates and Non-Response Biases

Most high quality surveys continue to be administered in an established manner as a method of increasing response rates[33]. Many studies have found a weak relationship between response rates and non-response bias[34] and the effort to increase the response rate[35].

STIMULATION OF VALUE

Methodological choices related to the stimulation of value can be classified into five categories:
1) Attribute and non-attribute based methods are applied.
2) The type of welfare measured to stimulate value.
3) The answer frame chosen to select the format of the chosen question.
4) The choice of payment method.
5) The use of ancillary questions and other design elements to support and assess validity.

The first of these is classified under the discussion of a survey, which has followed the proposed design. Categories 2-5 are listed here. Subsequently properties and incentives are discussed, along with other elements required to ensure reliable welfare estimation. The choice of approach depends on the contextual issue to be assessed, including the type of good, the nature of the decision being evaluated and the associated institutions (such as, for example, the existence of property rights or a tax system).

The literature shows that, in dealing with Public Goods, the ideal design includes choices, where the valued question is clearly understood, the payoff is complex if the proposed change is implemented, the respondent perceives his/her answers as influencing the provision of a thing that is being valued, and other aspects of the format's stimulus (such as, for example, the number of questions and alternatives per question), incentivize true preference revelation.

Willingness to Pay vs. Willingness to Accept

There are multiple reasons to explain the possible differences between Willingness to Pay (WTP) and Willingness to Accept to Pay (WTA), along with the relationships between these empirical estimators and Hicksian welfare measures[36].

From conceptual and theoretical perspectives of change under consideration[37] at the same time, established empirical difficulties associated with WTA estimation lead most studies to estimate WTP. Some recent research attempts to address these empirical challenges, for example, via the use of stimulant mechanisms[38].

The decision about a general context determines whether WTP or WTA is the most appropriate measure for welfare, from a conceptual perspective, but the choice of welfare would be motivated by a combination of theoretical and empirical considerations, and the motivation for this choice should be explained. Since WTA estimation is often exposed to empirical challenges, such as difficulty in shaping compatible questions and increasing rejection rates of the WTP estimation scenario will lead to estimates with superior empirical properties, it may over-estimate or under-estimate true welfare measures if WTA is conceptually appropriate. However, the WTP estimate would not always be considered a default and the WTA estimate would be applied when appropriate and reliable.

In the work proposed by the NOAA panel[39], WTP estimation is recommended. However, with improvements since then, both in the understanding of why WTP and WTA should be differentiated and in the design of SP-based methods, this recommendation should not necessarily be satisfied[40].

In cases where payment reductions are institutionally feasible, questions of incentive compatibility may be designated. Examples include tax reductions associated with environmental changes[41].

The choice of WTP vs. WTA can be used and understood from the literature on this subject, supplemented by information from pre-testing[42]. At the same time, established empirical difficulties associated with estimating WTA lead most studies to estimate WTP. WTA estimation can also be practical and appropriate.

The generalized decision context determines which of the two measures (either WTP or WTA) is the more suitable welfare measure from a conceptual perspective. The final welfare measure chosen, however, would be motivated by a combination of theory and empirical combinations, and the motivation for such a choice could be explained. Given the fact that estimation of WTA is often challenging, (such as difficulty in constituting compatible questions, and an increasing rate in a rejection scenario), estimation of WTP will often lead to estimates with superior empirical properties, but may overestimate or underpredict the true measure if WTA is conceptually appropriate. However, WTP estimation may not always be considered a default, and WTA estimation should be performed when appropriate and feasible.

The NOAA panel recommended the WTP estimate[43]. In some situations where payment decreases are institutionally feasible, situations that represent incentives or payment decreases could and can be generated[44].

Response Question Evaluation Forms

There are multiple response formats for CV and CE methods[45]. Questions in the form of responses and which configure contingent valuations include, but are not limited to, iterative bidding, open-ended extraction, payoff letters, and binary or dichotomous choice.
These questions include several types of binary choice and multinomial format, along with several types of ranking, and which can refer to the best and worst type of good[46], although there are inconsistencies in the literature, as far as formats are concerned, under the estimation of the DP technique[47].

DATA ANALYSIS

Improvement in econometric data analysis is one of the main topics in the literature concerning DP methods. Improvements in econometric analysis are a prominent theme in the literature. There are many parametric, semi-parametric and non-parametric methods modeling alternatives. In particular, for a DP

function, the most appropriate estimators would be selected according to the data and the questions to be answered. Best practices for estimation and interpretation of models with similar data structure.

The recommendations focus on specific questions for DP data analysis or because it requires specific considerations. Advances in econometric analysis are a prominent theme in the literature. There are many parametric, semi-parametric and non-parametric alternatives related to modeling[48]. Multiple classes of DP data are possible, and each poses a challenge that requires specific econometric methods. While econometric methods for analyzing some classes of DP data are virtually untested (e.g., inconclusive data), others have been subject to rapid development (e.g., discrete choice data). Current modeling often employs specific approximations to achieve a response pattern[49].

For example, the estimation often allows preference (and/or scale), heterogeneity and correlated responses with various types of conditional, mixed or generalized logit model[50]. It should be noted that there is no particular model, or set of models, that can be recommended for all DP modeling. The selection of econometric estimators should reflect the specific aspects of the data to be analyzed, the hypotheses to be investigated and how the estimation will result in decision making. The exchange considered in the selection of the estimator should be explicitly documented. Modeling would be a selector by utility structure, by the theoretical structure assumed to motivate behavior in a DP context. Theory-utility, behavioral, statistical and other selection and specification models should be made explicit.

PRINCIPAL DIFFERENCES BETWEEN CV AND CE

Although the section emphasizes the differences between CV and CE, it is important to recognize that these approaches are close relatives in the Semi-Parametric valuation genre and could in principle be used to estimate equivalent values[51]. These studies provide mixed results, but the large differences between the structure of the questions related to CV and the CE-type assessment are not always clear[52].

The difficulty of assimilating the effects of individual CV versus CE design elements, when comparing these approaches, speaks to the importance of collection formats as a whole. While the use of focus groups, cognitive interviews or other forms of pre-testing may help in the choice of an attribute, or the design of non-attributes[53].

ESTIMATION OF VALIDITY

Most of the literature published regarding methods for DP research focuses on the validity and reliability of the response value of the data. Although most research has shown that DP data are reliable[54]. The main contention question has been validated, and the DP methods have been subjected to an extensive deployment of validity tests[55].

The prevalence of DP studies suggests a consensus that such analyses are important. As a precursor to our recommendations regarding this issue, this topic provides a brief introduction to the validity of the calculation on stated preference studies. Some individuals note a hypothetical bias about the validity of DP methods; of great interest have been DP studies, the value of which in some cases exceeds the value of the studies themselves[56]. Most of the work has been conducted in the context of CV, but similar results have been found in CE work[57]. The conclusion of a hypothetical bias is largely based on cash transactions in which an actual transaction (usually for a private good in an experimental setting), provides a benchmark against which DP estimators are compared. There is disagreement about the implications of these findings. Therefore, some authors emphasize that few of these investigations satisfy the compatibility requirements on the true revelation of preferences, and that it is not clear how experimental payoffs compare to a true payoff (value or criterion)[58]. Many valuation studies use an unimportant experimental treatment to obtain a behavior from which to compare cash transactions. There is an overwhelming evidence that this framework does not characterize response incentives for many responses, and validity tests lead in more than one representative field, toward biased experiments[59].
In addition, research that is developed using voting-related behavior could generate a hypothetical bias, which is not universal[60].

STUDY REPORTS

Study reports are a critical element of the scientific method, enhancing the understanding of a paper and providing the basis for its replication. An effective communication is the cornerstone for future research in order to move forward with an empirical method and for the use and reuse of results from a given outcome, e.g., related to benefit transfer.

All studies, whether applied (to assist in decision making) or methodological (to assist in the evolution of research), should fully design the document, study design, implementation, analysis and results. Such transparency is essential for the scientific credibility of the studies and for their proper interpretation and utilization of the results. Documentation can also help to use the results for future decision making, even if it was not foreseen in the initial part of the study.

CONCLUSIONS

This paper describes the main steps for the application of a valuation process using the DP method. DP models can be combined with RP models, although this is only possible when they come from the same data-generating function. An area that has proven its conceptual importance in this literature is related to the examination of convergence or divergence in parameter estimation for discrete models.

This method is particularly useful in measuring the natural capital of ecosystem services. These services contribute to the measurement of human welfare, either through Direct or Indirect Methods[61].

ENDNOTES

1. A recent example from among the CE can be found, for example, in Sarkar et al., 2019.
2. Some important examples of its use are those of the Exxon Valdez oil tanker, applied to assess the damage generated (Carson et al., 2003), as well as the activities related to the Kakadu National Park in Australia (Bennett, 1996). CV could provide, in some cases, a controversial view of environmental problems. This vision would be extended to various authors such as Hanemann, 1994. Bateman et al., 1997a; 1997b. Bateman, Munro & Poe, 2008. Carson, 2012.
3. From the above, the baseline and changes could be presented, in an adjusted, measurable and interpretable way, (Sarkar et al., 2012), and thus reflect results that would allow the assessment of utility and consequences that could be identified with those making the responses. Boyd and Krupnick, 2013. Johnston et al., 2016.
4. That is, these responses should be able to identify the links between intermediate changes and the final changes directly responsible for well-being, as proposed by authors such as Sarkar et al., 2012. Boyd y Krupnick, 2009; 2013. Boyd et al., 2016. Johnston et al., 2016.
5. For example, those that encourage a reaction from the adversary to avoid the rejection of a given scenario. Adamowicz et al., 1997. 2014. Cameron, DeShazo & Johnson, 2011.
6. A summary of the literature can be found in Presser et al., 2004.
7. The lack of guidance, and the quality of the test instruments, are a central component in the content validity of the test. Smith, 2006. Carson, 2012.
8. For example, an attractive feature of CE is the ability to estimate marginal values, based on the attributes and the study design. However, this feature usually involves a disadvantageous trade-off, such as high complexity and potential loss with incentive compatibility, depending on the structure of the CE. Vossler, Doyon & Rondeau, 2012.
9. Conversely, the response may have difficulties identifying and distinguishing effects explained using narrative texts common in the CV method (Hoehn, Lupi & Kaplowitz, 2010). Both approaches, depending on the design, can present complex information.
10. For example, if the respondent thinks about a landscape or an ecosystem as a whole, then attributing this form could be inconsistent. In cases like this, the syntactic analysis of the totality as a whole could incorrectly characterize the way in which the components of the attributes or characteristics are understood or enhanced (Madureira, Nunes & Santos, 2005). On the other hand, individuals might think about changes, (e.g., in recreational locations), in terms of consistent traits, with a tabular, or attribute-based presentation (Hoehn, Lupi, & Kaplowitz, 2010).

11. As, for example, proposed by authors such as: Arentze et al., 2003. Hensher, 2006b. Islam, Louviere & Burke, 2007. Balcombe & Fraser, 2011. Burton & Rigby, 2012. Dellaert, Donkers & Van Soest, 2012. Alemu et al., 2013.

12. This type of explanation is suggested by authors such as: Kahneman, Slovic & Tversky, 1982. Mazzotta y Opaluch 1995. Gigerenzer & Todd 1999. Swait & Adamowicz 2001a; 2001b. Hensher 2006a; 2006b. Boxall, Adamowicz y Moon, 2009. McFadden, 2014. Meyerhoff, Oehlmann & Weller, 2015. Olsen & Meyerhoff, 2016.

13. This could lead to behavior based on simplification by discovery. Hoehn, Lupi & Kaplowitz, 2010.

14. In such applications, the use of attributes is usually large or simple descriptors, as proposed by authors such as:_Arentze et al., 2003. Hensher, 2006b. Islam, Louviere & Burke, 2007. Balcombe y Fraser, 2011. Burton y Rigby, 2012. Dellaert, Donkers & Van Soest, 2012. Alemu et al., 2013.

15. As proposed by authors such as: Kahneman, Slovic & Tversky, 1982. Mazzotta & Opaluch 1995. Gigerenzer & Todd 1999. Swait & Adamowicz 2001a; 2001b. Hensher 2006a; 2006b. Boxall, Adamowicz & Moon, 2009. McFadden, 2014. Meyerhoff, Oehlmann & Weller, 2015. Olsen & Meyerhoff, 2016.

16. Proposed by authors such as: Hensher & Rose 2009. Hensher & Layton 2010.

17. Some examples of this method can be found in: Scarpa et al., 2009. Balcombe, Burton & Rigby, 2011. Balcombe, Fraser & McSorley, 2015.

18. This method is considered to address potential anomalies that may arise in managing the response to simplicity levels. Day et al., 2012.

19. Some literature examples about the importance of complexity in the CV method can be found in authors such as: Sandorf, Campbell & Hanley, 2017. Hoehn, Lupi & Kaplowitz, 2010.

20. As stated by authors such as: Sandor & Wedel, 2002. Johnson et al., 2006; 2013. Ferrini & Scarpa, 2007. Rose et al., 2008. Scarpa & Rose, 2008. Rose & Bliemer, 2009; .2014. Bliemer & Rose, 2010. Vermeulen et al., 2011.

21. The ED could represent, for example, how some blocks of questions are assigned to different versions of a survey (as could be, for example, bidding different levels of attributes). ED could be as simple as developing the supply structure for a simple, binary choice question, or as complex as a choice matter with multiple attributes, multiple alternatives within a question of choice, with multiple attributes, multiple alternatives within a question, multiple questions within a questionnaire, and multiple versions of survey instruments. Caussade et al., 2005.

22. The ED assumption defines the way in which different treatments are assigned to a question (or sequence of questions within a questionnaire). The ED could represent, for example, how some blocks of questions are assigned to different versions of a survey (as could be, for example, bidding different levels of attributes). ED could be as simple as developing the supply structure for a simple, binary choice question, or as complex as a choice matter with multiple attributes, multiple alternatives within a question of choice, with multiple attributes, multiple alternatives within a question, multiple questions within a questionnaire, and multiple versions of survey instruments. Caussade et al., 2005.

23. Some examples can be found in: Cooper & Loomis, 1992. Cooper, 1993. Kanninen, 1993a; 1993b; 1995. Alberini, 1995. Haab & McConnell, 2002.

24. As proposed by, for example: Johnson et al., 2013.

25. In this case, the compilation of methods that were not known at the time of the NOAA panel should be noted. Arrow et al., 1993.

26. As suggested by Macmillan et al., 2002.

27. For a general discussion, one could follow, for example: Dillman et al., 2014.

28. Some examples of this type of work are: Boyle & Bishop, 1988. Andreoni, 1989. Blamey et al., 1999. Leggett et al., 2003.

29. Despite the fact that the results usually work as a mix, in a specific context. Bell, Huber & Viscusi, 2011. Lindhjem & Navrud, 2011a; 2011b. Windle & Rolfe, 2011. Šĉasný & Alberini, 2012. Boyle et al., 2016. Sandorf, Aanesen & Navrud, 2016.

30. These individuals are therefore excluded from both the market and the sample structure. Loomis, 1996.

31. Despite these characteristics, though they have experienced declining response rates (Brick & Williams, 2013) and have been forced for counting an increasing proportion of the population that could be reached via cell phones (e.g., Blumberg et al., 2012). In addition, cell phone numbers may not be immediately assigned to specific geographic areas, which would make matching responses to neighbor characteristics more complex.

32. However, the problem of random sampling of individuals at random among neighbors remains a challenge. Link et al., 2008.
    The Internet adoption has been one of the major recent developments for the gathering of data collections. Tourangeau, Conrad & Cooper, 2013.
33. Although in some cases, this indicator may be considered a poor indicator of non-response bias. Groves, 2006. Tourangeau & Plewes, 2013.
34. Some examples can be found in: Curtin, Presser & Singer, 2000.
35. This impulse causes in some cases the worsening of the non-response bias, and at the same time the effort to improve the response rate, which in some cases worsens the non-response bias. Peytchev, 2009. Lundquist & Särndal, 2013.
36. Younjun Kim, Catherine L. Kling, and Jinhua Zhao Kim, review the literature concerning the differences between estimators explaining WTP and WTA, considering theoretical and empirical explanations. Kim, Kling & Zhao, 2015.
37. The logical reference condition of the response, as indicated by: Carson, Flores, & Meade, 2001. Kim, Kling & Zhao, 2015.
    This choice of welfare measures does not necessarily refer to legal subsidies, as indicated by Knetsch, 2007.
38. This is the case of Bush et al., 2013.
39. The NOAA panel recommended the estimation of WTP (Arrow et al., 1993). However, with the advances that have developed since then, both in terms of the possible divergence of WTP and WTA (Brown & Gregory, 1999; Horowitz, McConnell & Murphy, 2013; Ericson & Fuster, 2014; Kim, Kling & Zhao, 2015), and in the design of the DP methods themselves, this recommendation would not necessarily hold in all cases. In some situations, for which the decrease in payments is institutionally feasible, situations that constitute incentives or decreases in payments may be generated.
40. Some authors who support this principle are: Brown & Gregory, 1999. Horowitz, McConnell & Murphy, 2013. Ericson & Fuster, 2014. Kim, Kling & Zhao, 2015.
41. As well as the use of financial incentives, to modify procedures related to behavior and health, Promberger, Dolan & Marteau, 2012.
    A proposal in this sense can be found in Johnston, Swallow & Weaver, 1999.
42. This paper essentially discusses incentive behavior, and can be found in the considerations summarized by Kim, Kling & Zhao, 2015.
43. However, with the advances that have developed since then, both in the possible divergence of WTP and WTA (Brown & Gregory, 1999. Horowitz, McConnell & Murphy, 2013. Ericson & Fuster, 2014. Kim, Kling & Zhao, 2015), and in the design of DP methods, this recommendation would not necessarily hold for all subjects.
44. Notable examples include tax payments with environmental management (Johnston, Swallow & Weaver, 1999.), or the relationship with health (such as Promberger, Dolan, & Marteau, 2012.).
45. For example, some relevant cases are: Mitchell & Carson, 1989. Adamowicz et al., 1998. Carson et al., 2001. Hanley, Mourato & Wright, 2001. Bateman et al., 2002. Carson & Louviere, 2011.
46. As proposed by: Haab & McConnell, 2002. Scarpa & Alberini, 2005. Train, 2009. Hess & Daly, 2014.
47. As suggested by Carson & Louviere, 2011.
48. Some examples can be found in: Haab & McConnell, 2002. Scarpa & Alberini, 2005. Train, 2009. Hess & Daly, 2014.
49. As proposed by: Lewbel, 2000. Swait & Adamowicz, 2001a; 2001b. Scarpa & Alberini, 2005. Scarpa, Thiene, & Marangon, 2008. Train, 2009. Watanabe, 2010. Lewbel, McFadden, & Linton 2011. Boeri, Scarpa, & Chorus 2014. Hess & Daly, 2013; 2014.
50. Some examples in the literature are: Revelt & Train, 1998. Scarpa & Alberini, 2005. Train, 2009. Fiebig et al., 2009.
51. Multiple studies have investigated the convergence validity of the estimates from these two formats, such as: Adamowicz et al., 1998. Hanley et al., 1998. Cameron et al., 2002. Foster & Mourato, 2003. Ryan, 2004. Jin, Wang & Ran, 2006. Mogas, Riera & Bennett, 2006. Goldberg & Rosen, 2007.
52. For example, the use of text against tabular displays could make it difficult to make clean and controlled comparisons, as suggested by Tourangeau, Couper & Conrad, 2004.
53. See, as an example, the works of: Coast, 1999; Bennett & Blamey, 2001. Coast & Horrocks, 2007. Coast et al., 2012. Riera et al., 2012.
54. Some reference examples are the works of: Loomis, 1989. Teisl et al., 1995. Bliem, Getzner & Rodiga-Laßnig, 2012. Mørkbak & Olsen, 2015.

55. See, for example, the work of: Smith, 2006. Kling, Phaneuf & Zhao, 2012. Carson, Groves & List, 2014.
56. As proposed by: List & Gallet, 2001. Little & Berrens, 2004. Murphy et al., 2005.
57. For example: Lusk & Schroeder, 2004. Moser, Raffaelli, & Notaro, 2013.
58. A discussion regarding this topic can be found in Kling, Phaneuf & Zhao, 2012, and Carson, Groves & List, 2014.
59. As indicated by authors such as: Carson, Groves & List, 2014.
60. Some examples are: Vossler & Kerkvliet, 2003. Johnston, 2006. Vossler & Watson, 2013.
61. This is the case of authors such as Costanza, D'Arge, Groot, Farber, among others, who estimate the Economic Value of 17 ecosystem services for 16 biomasses, based on published studies and a few original calculations. In order to calculate the full value of the biosphere, (most of which is obtained with results gathered outside the market), the estimated value is calculated at around \$16-54 trillion per year, with an average of \$33 trillion per year. Costanza et al., 1997.

REFERENCES

Adamowicz, W., Boxall, P., Williams, M., & Louviere, J. (1998). Stated preference approaches for measuring passive use values: Choice Experiments and Contingent Valuation. *American Journal of Agricultural Economics*, *80*(1), 64–75. https://doi.org/10.2307/3180269

Adamowicz, W., Dickie, M., Gerking, S., Veronesi, M., & Zinner. D. (2014). Household decision making and valuation of environmental health risks to parents and their children. *Journal of the Association of Environmental and Resource Economists*, *1*, 481–519. http://dx.doi.org/10.1086/679255

Adamowicz, W., Swait, J., Boxall, P., Louviere, J., & Williams, M. (1997). Perceptions versus Objective Measures of Environmental Quality in Combined Revealed and Stated Preference Models of Environmental Valuation. *Journal of Environmental Economics and Management*, *32*(1), 65–84. https://doi.org/10.1006/jeem.1996.0957

Alberini, A. (1995). Optimal designs for discrete choice contingent valuation surveys: Single-bound, double-bound, and bivariate models. *Journal of Environmental Economics and Management*, *28*(3), 287–306. https://doi.org/10.1006/jeem.1995.1019

Alemu, M.H., Mørkbak, M.R., Søren, B.O., & Carsten L.J. (2013). Attending to the reasons for attribute non-attendance in choice experiments. *Environmental and Resource Economics*, *54*, 333–359. https://doi.org/10.1007/s10640-012-9597-8

Andreoni, J. (1989). Giving with impure altruism: Applications to charity and Ricardian equivalence. *Journal of Political Economy*, *97*(6), 1447–1458. https://doi.org/10.1086/261662

Arentze, T., Borgers, A., Timmermans, H., & DelMistro, R. (2003). Transport stated choice responses: Effects of task complexity, presentation format and literacy. *Transportation Research Part E: Logistics and Transportation Review*, *39*, 229–244. https://doi.org/10.1016/S1366-5545(02)00047-9

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*, 4601–14.

Balcombe, K., & Fraser, I. (2011). A general treatment of "don't know" responses from choice experiments. *European Review of Agricultural Economics*, *38*, 171–191. https://doi.org/10.1093/erae/jbr010

Balcombe, K., Burton, M., & Rigby, D. (2011). Skew and attribute non-attendance within Bayesian mixed logit model. *Journal of Environmental Economics and Management*, *62*(3), 446–461. https://doi.org/10.1016/j.jeem.2011.04.004

Balcombe, K., Fraser, I., & McSorley, E. (2015). Visual Attention and Attribute Attendance in Multi-Attribute Choice Experiments. *Journal of Applied Econometrics*, *30*(3), 447–467. https://doi.org/10.1002/jae.2383

Bateman, I.J., Carson, R.T., Day, B.H., Hanemann, W.M., Hanley, N., Hett, T., . . . Pearce, D.W. (2002). *Economic valuation with stated preference techniques: A manual*. Cheltenham (United Kingdom), Edward Elgar Publishing. https://doi.org/10.4337/9781781009727

Bateman, I.J., Munro, A., & Poe, G. (2008). Decoy effects in choice experiments contingent valuation: Asymmetric dominance. *Land Economics*, *84*(1), 115–127. https://doi.org/10.3368/le.84.1.115

Bateman, I.J., Munro, A., Rhodes, B., & Starmer, C., & Sudgen, R. (1997b). A test of the theory of reference-dependent preferences. *Quarterly Journal of Economics*, *112*, 479–505. https://doi.org/10.1162/003355397555262

Bateman, I.J., Munro, A., Rhodes, B., Starmer, C., & Sugden, R. (1997a). Does partwhole bias exist? An experimental investigation. *Economic Journal*, *107*(441), 322–332. https://doi.org/10.1111/j.0013-0133.1997.160.x

Bell, J., Huber, J., & Viscusi, W.K. (2011). Survey mode effects on valuation of environmental godos. *International Journal of Environmental Research and Public Health*, *8*, 1222–1243. https://doi.org/10.3390/ijerph8041222

Bennett, J. (1996). The contingent valuation: A post Kakadu assessment. *Agenda, A Journal of Policy Analysis and Reform*, *3*, 185–194. https://doi.org/10.22459/AG.03.02.1996.06

Bennett, J., & Blamey, R. (Eds.) (2001). *The choice modelling approach to environmental valuation*. Cheltenham (United Kingdom), Edward Elgar Publishing.

Blamey, R.K., Bennett, J.W., & Morrison, M.D. (1999). Yea-saying in contingent valuation surveys. *Land Economics*, *75*(1), 126–141. https://doi.org/10.2307/3146997

Bliem, M., Getzner, M., & Rodiga-Laßnig, P. (2012). Temporal stability of individual preferences for river restoration in Austria using a choice experiment. *Journal of Environmental Management*, *103*, 65–73. https://doi.org/10.1016/j.jenvman.2012.02.029

Bliemer, M.C.J., & Rose, J.M. (2010). Construction of experimental designs for mixed logit models allowing for correlation across choice observations. *Transportation Research Part B: Methodological*, *44*(6), 720–734. https://doi.org/10.1016/j.trb.2009.12.004

Blumberg, S.J., Luke, J.V., Ganesh, N., Davern, M.E., & Boudreaux, M.H. (2012). *Wireless substitution: State-level estimates from the National Health Interview Survey, 2010–2011*. National Health Statistics Reports, 61. Hyattsville (MD, United States), Centers for Disease Control and Prevention.

Boeri, M., Scarpa, R., & Chorus, C. (2014). Stated choices and benefit estimates in the context of traffic calming schemes: Utility maximization, regret minimization, or both? *Transportation Research Part A: Policy and Practice*, *61*, 121–135. https://doi.org/10.1016/j.tra.2014.01.003

Boxall, P.C., Adamowicz, W.L., & Moon, A. (2009). Complexity in choice experiments: Choice of the status quo alternative and implications for welfare measurement. *Australian Journal of Agricultural and Resource Economics*, *53*(4), 503–519. https://doi.org/10.1111/j.1467-8489.2009.00469.x

Boyd, J., & Krupnick, A. (2009). *The definition and choice of environmental commodities for nonmarket valuation*. RFF Discussion Paper 09–35, Washington, DC (USA), Resources for the Future. https://doi.org/10.2139/ssrn.1479820

Boyd, J., & Krupnick, A. (2013). Using ecological production theory to define and select environmental commodities for nonmarket valuation. *Agricultural and Resource Economics Review*, *42*(1), 1–32. https://doi.org/10.1017/S1068280500007590

Boyd, J., Ringold, P., Krupnick, A., Johnston, R.J., Weber, M.A., & Hall, K. (2016). Ecosystem services indicators: Improving the linkage between biophysical and economic analyses. *International Review of Environmental and Resource Economics*, *8*(3–4), 359–443. https://doi.org/10.1561/101.00000073

Boyle, K., Paterson, R., Carson, R., Leggett, C., Kanninen, B., Molenar, J., & Neumann, J. (2016). Valuing shifts in the distribution of visibility in national parks and wilderness areas in the United States. *Journal of Environmental Management*, *173*, 10–22. https://doi.org/10.1016/j.jenvman.2016.01.042

Boyle, K.J., & Bishop, R.C. (1988). Welfare measurements using contingent valuation: A comparison of techniques. *American Journal of Agricultural Economics*, *70*, 20–28. https://doi.org/10.2307/1241972

Brick, J.M., & Williams, D. (2013). Explaining rising nonresponse rates in cross-sectional surveys. *Annals of the American Academy of Political and Social Science*, *645*(1), 36–59. https://doi.org/10.1177/0002716212456834

Brown, T.C., & Gregory, R. (1999). Why the WTA-WTP disparity matters. *Ecological Economics*, *28*(3), 323–335. https://doi.org/10.1016/S0921-8009(98)00050-0

Burton, M., & Rigby, D. (2012). The self selection of complexity in choice experiments. *American Journal of Agricultural Economics*, *94*(3), 786–800. https://doi.org/10.1093/ajae/aas015

Bush, G., Hanley, N., Moro, M., & Rondeau, D. (2013). Measuring the local costs of conservation: A provision point mechanism for eliciting willingness to accept compensation. *Land Economics*, *89*(3), 490–513. https://doi.org/10.3368/le.89.3.490

Cameron, T.A., DeShazo, J.R., & Johnson, E.H. (2011). Scenario adjustment in stated preference research. *Journal of Choice Modelling*, *4*(1), 9–43. https://doi.org/10.1016/S1755-5345(13)70017-4

Cameron, T.A., Poe, G.L., Ethier, R.G., & Schulze, W.D. (2002). Alternative nonmarket value-elicitation methods: Are the underlying preferences the same? *Journal of Environmental Economics and Management*, *44*(3), 391–425. https://doi.org/10.1006/jeem.2001.1210

Carson, D., Gilmore, A., Perry, C., & Grønhaug, K. (2001). *Qualitative Market Research*. London (United Kingdom), Sage Publications Ltd. https://doi.org/10.4135/9781849209625

Carson, R. (2012). Contingent valuation: A practical alternative when prices aren't available. *Journal of Economic Perspectives*, *26*, 27–42. https://doi.org/10.1257/jep.26.4.27

Carson, R.T., & Louviere, J.L. (2011). A common nomenclature for stated preference elicitation approaches. *Environmental and Resource Economics*, *49*, 539–559. https://doi.org/10.1007/s10640-010-9450-x

Carson, R.T., Flores, N.F., & Meade, N.F. (2001). Contingent valuation: Controversies and evidence. *Environmental and Resource Economics*, *19*, 173–210. https://doi.org/10.1023/A:1011128332243

Carson, R.T., Groves, T., & List, J.A. (2014). Consequentiality: A theoretical and experimental exploration of a single binary choice. *Journal of the Association of Environmental and Resource Economists*, *1*, 171–207. https://doi.org/10.1086/676450

Carson, R.T., Mitchell, R.C., Hanemann, M., Kopp, R.J., Presser, S., & Ruud, P.A. (2003). Contingent valuation and lost passive use: Damages from the Exxon Valdez oil spill. *Environmental and Resource Economics*, *25*, 257–289. https://doi.org/10.1023/A:1024486702104

Caussade, S., Ortúzara, J., Rizzia, L.I., & Hensherb, D.A. (2005). Assessing the influence of design dimensions on stated choice experiment estimates. *Transportation Research Part B: Methodological*, *39*(7), 621–640. https://doi.org/10.1016/j.trb.2004.07.006

Coast, J. (1999). The appropriate uses of qualitative methods in health economics. *Health Economics*, *8*(4), 345–353. https://doi.org/10.1002/(SICI)1099-1050(199906)8:4<345::AID-HEC432>3.0.CO;2-Q

Coast, J., & Horrocks, S.A. (2007). Developing attributes and levels for discrete choice experiments using qualitative methods. *Journal of Health Services Research and Policy*, *12*(1), 25–30. https://doi.org/10.1258/135581907779497602

Coast, J., Hareth A-J., Sutton, E.J., Horrocks, S.A., Vosper, A.J., Swancutt, D.R., & Flynn, T. (2012). Using qualitative methods for attribute development for discrete choice experiments: Issues and recommendations. *Health Economics*, *21*, 730–741. https://doi.org/10.1002/hec.1739

Cooper, J.C. (1993). Optimal bid selection for dichotomous choice contingent valuation surveys. *Journal of Environmental Economics and Management*, *24*(1), 25–40. https://doi.org/10.1006/jeem.1993.1002

Cooper, J.C., & Loomis, J. (1992). Sensitivity of willingness-to-pay estimates to bid design in dichotomous choice contingent valuation models. *Land Economics*, *68*(2), 211–224. https://doi.org/10.2307/3146775

Costanza, R., d'Arge, R., de Groot, R., Farber, S., Grasso, M., Hannon, B., . . . van den Belt, M. (1997). The value of the world's ecosystem services and natural capital. *Nature*, *387*, 253–260. https://doi.org/10.1038/387253a0

Curtin, R., Presser, S., & Singer, E. (2000). The effects of response rate changes on the Index of Consumer Sentiment. *Public Opinion Quarterly*, *64*(4), 413–428. https://doi.org/10.1086/318638

Day, B., Bateman, I.J., Carson, R.T., Dupont, D., Louviere, J.J., Morimoto, S., . . . Wang, P. (2012). Ordering effects and choice set awareness in repeat-response stated preference studies. *Journal of Environmental Economics and Management*, *63*, 73–91. https://doi.org/10.1016/j.jeem.2011.09.001

De Bekker-Grob, E., Ryan, M., & Gerard, K. (2012). Discrete choice experiments in health economics: A review of the literatura. *Health Economics*, *i*, 145–172. https://doi.org/10.1002/hec.1697

Dellaert, B.G.C., Donkers, B., & Van Soest, A. (2012). Complexity effects in choice experiment based models. *Journal of Marketing Research*, *49*(3), 424–434. https://doi.org/10.1509/jmr.09.0315

Dillman, D.A., Smyth, J.A., & Christian, L.M. (2014). *Internet, phone, mail and mixed-mode surveys: The tailored design method* (4th edition). Hoboken (NJ, United States), John Wiley & Sons.

Ericson, K., & Fuster, A. (2014). The endowment effect. *Annual Review of Economics*, *6*, 555–579. https://doi.org/10.1146/annurev-economics-080213-041320

Ferrini, S., & Scarpa, R. (2007). Designs with a priori information for nonmarket valuation with choice experiments: A Monte Carlo study. *Journal of Environmental Economics and Management*, *53*(3), 342–363. https://doi.org/10.1016/j.jeem.2006.10.007

Fiebig, D., Keane, M.P., Louviere, J., & Wasi, N. (2009). The Generalized Multinomial Logit Model: Accounting for Scale and Coefficient Heterogeneity. *Marketing Science*, *29*(3), 393–421. https://doi.org/10.1287/mksc.1090.0508

Foster, V., & Mourato, S. (2003). Elicitation format and sensitivity to scope. *Environmental and Resource Economics*, *24*, 141–160. https://doi.org/10.1023/A:1022856329552

Gigerenzer, G., Todd, P.M., & the ABC Research Group. (1999). *Simple heuristics that make us Smart*. New York (United States), Oxford University Press.

Goldberg, I., & Roosen, J. (2007). Scope insensitivity in health risk reduction studies: A comparison of choice experiments and the contingent valuation method for valuing safer food. *Journal of Risk and Uncertainty*, *34*(2), 123–144. https://doi.org/10.1007/s11166-007-9006-9

Groves, R.M. (2006). Nonresponse rates and nonresponse bias in household surveys. *Public Opinion Quarterly*, *70*(5), 646–675. https://doi.org/10.1093/poq/nfl033

Groves, R.M., & Peytcheva, E. (2008). The impact of nonresponse rates on nonresponse bias: A meta-analysis. *Public Opinion Quarterly*, *72*(2), 167–189. https://doi.org/10.1093/poq/nfn011

Haab, T.C., & McConnell, K. (2002). *Valuing environmental and natural resources. The econometrics of non-market valuation*. Cheltenham (United Kingdom), Edward Elgar Publishing. https://doi.org/10.4337/9781843765431

Hanemann, W.M. (1994). Valuing the environment through contingent valuation. *Journal of Economic Perspectives*, *8*(4), 19–43. https://doi.org/10.1257/jep.8.4.19

Hanley, N., Mourato, S., & Wright, R.E. (2001). Choice modelling approaches: A superior alternative for environmental valuation? *Journal of Economic Surveys*, *15*, 435–462. https://doi.org/10.1111/1467-6419.00145

Hensher, D.A. (2006a). How do respondents process stated choice experiments? Attribute consideration under varying information load. *Journal of Applied Econometrics*, *21*(6), 861–878. https://doi.org/10.1002/jae.877

Hensher, D.A. (2006b). Revealing differences in willingness to pay due to the dimensionality of stated choice designs: An initial assessment. *Environmental and Resource Economics*, *34*(1), 7–44. https://doi.org/10.1007/s10640-005-3782-y

Hensher, D.A., & Layton. D. (2010). Parameter transfer of common-metric attributes in choice analysis: Implications for willingness to pay. *Transportation*, *37*(3), 473–490. https://doi.org/10.1007/s11116-010-9260-6

Hensher, D.A., & Rose, J.M. (2009). Simplifying choice through attribute preservation or nonattendance: Implications for willingness to pay. *Transportation Research Part E: Logistics and Transportation Review*, *45*, 583–590. https://doi.org/10.1016/j.tre.2008.12.001

Hess, S., & Daly, A. (Eds.) (2013). *Choice modelling: The state of the art and the state of practice*. Cheltenham (United Kingdom), Edward Elgar Publishing.

Hess, S., & Daly, A. (Eds.). (2014). *Handbook of choice modelling*. Cheltenham (United Kingdom), Edward Elgar Publishing. https://doi.org/10.4337/9781781003152

Hoehn, J.P., Lupi, F., & Kaplowitz, M.-D. (2010). Stated choice experiments with complex ecosystem changes: The effect of information formats on estimated variances and choice parameters. *Journal of Agricultural and Resource Economics*, *35*(3), 568–590. Retrieved from https://www.jstor.org/stable/23243072

Horowitz, J., McConnell, K., & Murphy, J. (2013). Behavioral foundations of environmental economics and valuation. In J.A. List, & M.K. Price (Eds.), *Handbook on experimental economics and the environment* (pp. 115–156). Cheltenham (United Kingdom), Edward Elgar Publishing. https://doi.org/10.4337/9781781009079

Islam, T., Louviere J., & Burke, P.F. (2007). Modeling the effects of including/excluding attributes in choice experiments on systematic and random components. *International Journal of Research in Marketing*, *24*(4), 289–300. https://doi.org/10.1016/j.ijresmar.2007.04.002

Jin, J., Wang, Z., & Ran, S. (2006). Comparison of contingent valuation and choice experiment in solid waste management programs in Macao. *Ecological Economics*, *57*(3), 430–441. https://doi.org/10.1016/j.ecolecon.2005.04.020

Johnson, F.R., Kanninen, B., Bingham, M., & Özdemir, S. (2006). Experimental design for stated choice studies. In B.J. Kanninen (Eds.), *Valuing environmental amenities using stated choice studies* (pp. 159–202). New York (United States), Springer. https://doi.org/10.1007/1-4020-5313-4_7

Johnson, F.R., Lancsar, E., Marshall, D., Kilambi, V., Mühlbacher, A., Regier, D.A., . . . Bridges, J.F.P. (2013). Constructing experimental designs for discrete choice experiments: Report of the ISPOR conjoint analysis experimental design good research practices task forcé. *Value in Health*, *16*(1), 3–13. https://doi.org/10.1016/j.jval.2012.08.2223

Johnston, R. (2006). Is hypothetical bias universal? Validating contingent valuation responses using a binding referendum. *Journal of Environmental Economics and Management*, *52*, 469–481. https://doi.org/10.1016/j.jeem.2005.12.003

Johnston, R., Schultz, E.T., Segerson, K., Besedin, E.Y., & Ramachandran M. (2016). Biophysical causality and environmental preference elicitation: Evaluating the validity of welfare analysis over intermediate outcomes. *American Journal of Agricultural Economics*, *99*(1), 163–185. https://doi.org/10.1093/ajae/aaw073

Johnston, R., Swallow, S., & Weaver, T. (1999). Estimating Willingness to Pay and Resource Tradeoffs with Different Payment Mechanisms: An Evaluation of a Funding Guarantee for Watershed Management. *Journal of Environmental Economics and Management*, *38*(1), 97–120. https://doi.org/10.1006/jeem.1999.1077

Kahneman, D., Slovic, P., & Tversky, A. (Eds.) (1982). *Judgement under uncertainty: Heuristics and biases*. Cambridge (United Kingdom), Cambridge University Press. https://doi.org/10.1017/CBO9780511809477

Kanninen, B. (1993a). Design of sequential experiments for contingent valuation studies. *Journal of Environmental Economics and Management*, *25*(1), S1–S11. https://doi.org/10.1006/jeem.1993.1029

Kanninen, B. (1993b). Optimal experimental design for double-bounded dichotomous choice contingent valuation. *Land Economics*, *69*(2), 138–146. Retrieved from https://www.jstor.org/stable/3146514

Kanninen, B. (1995). Bias in discrete response contingent valuation. *Journal of Environmental Economics and Management*, *28*, 114–125. https://doi.org/10.1006/jeem.1995.1008

Keeter, S., Kennedy, C., Dimock, M., Best, J., & Craighill, P. (2006). Gauging the impact of growing nonresponse on estimates from a national RDD telephone survey. *Public Opinion Quarterly*, *70*, 759–779. https://doi.org/10.1093/poq/nfl035

Keeter, S., Miller, C., Kohut A., Groves, R.M., & Presser, S. (2000). Consequences of reducing nonresponse in a national telephone survey. *Public Opinion Quarterly*, *64*(2), 125–148. https://doi.org/10.1086/317759

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. http://dx.doi.org/10.1146/annurev-resource-100913-012501

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. https://doi.org/10.1257/jep.26.4.3

Knetsch, J.L. (2007). Biased valuations, damage assessments, and policy choices: The choice of measure matters. *Ecological Economics*, *63*(4), 684–689. https://doi.org/10.1016/j.ecolecon.2007.02.012

Leggett, C.G., Kleckner, N.S., Boyle, K.J., Dufield, J.W., & Mitchell, R.C. (2003). Social Desirability Bias in Contingent Valuation Surveys Administered Through In-Person Interviews. *Land Economics*, *79*(4), 561–575. https://doi.org/10.2307/3147300

Lewbel, A. (2000). Semiparametric qualitative response model estimation with unknown heteroscedasticity or instrumental variables. *Journal of Econometrics*, *97*(1), 145–177. https://doi.org/10.1016/S0304-4076(00)00015-4

Lewbel, A., McFadden, D., & Linton, O. (2011). Estimating features of a distribution from binomial data. *Journal of Econometrics*, *162*(2), 170–188. https://doi.org/10.1016/j.jeconom.2010.11.006

Lindhjem, H., & Navrud, S. (2011a). Are internet surveys an alternative to face-to-face interviews in contingent valuation? *Ecological Economics*, *70*(9), 1628–1637. https://doi.org/10.1016/j.ecolecon.2011.04.002

Lindhjem, H., & Navrud, S. (2011b). Using internet in stated preference surveys: A review and comparison of survey modes. *International Review of Environmental and Resource Economics*, *5*, 309–351. https://doi.org/10.1561/101.00000045

Link, M.W., Battaglia, M.P., Frankel, M.R., Osborn, L., & Mokdad, A.H. (2008). A comparison of address-based sampling (ABS) versus random digit-dialing (RDD) for general population surveys. *Public Opinion Quarterly*, *72*(1), 6–27. https://doi.org/10.1093/poq/nfn003

List, J.A., & Gallet, C.A. (2001). What experimental protocol influence disparities between actual and hypothetical stated values? Evidence from a meta-analysis. *Environmental and Resource Economics*, *20*, 241–254. https://doi.org/10.1023/A:1012791822804

Little, J., & Berrens, R. (2004). Explaining disparities between actual and hypothetical stated values: Further investigation using meta-analysis. *Economics Bulletin*, *3*(6), 1–13.

Loomis, J. (1989). Test–retest reliability of the contingent valuation method: A comparison of general population and visitor responses. *American Journal of Agricultural Economics*, *71*(1), 76–84. https://doi.org/10.2307/1241776

Loomis, J. (1996). How large is the extent of the market for public goods: Evidence from a nationwide contingent valuation survey. *Applied Economics*, *28*, 779–782. https://doi.org/10.1080/000368496328209

Lundquist, P., & Särndal, C.-E. (2013). Aspects of responsive design with applications to the Swedish living conditions survey. *Journal of Official Statistics*, *29*, 557–582. https://doi.org/10.2478/jos-2013-0040

Lusk, J.L., & Schroeder, T.C. (2004). Are choice experiments incentive compatible? A test with quality differentiated beef steaks. *American Journal of Agricultural Economics*, *86*(2), 467–482. https://doi.org/10.1111/j.0092-5853.2004.00592.x

Macmillan, D.C., Philip, L., Hanley, N., & Alvarez-Farizo, B. (2002). Valuing the nonmarket benefits of wild goose conservation: A comparison of interview and group-based approaches. *Ecological Economics*, *43*, 49–59. https://doi.org/10.1016/S0921-8009(02)00182-9

Madureira, L., Nunes, L.C., & Santos, J.M.L. (2005). *Valuing multi-attribute environmental changes: Contingent valuation and choice experiments*. Paper presented at 14th annual conference of the European Association of Environmental and Resource Economists, Bremen, June 23–26.

Mazzotta, M.J., & Opaluch, J.J. (1995). Decision making when choices are complex: A test of Heiner's hypothesis. *Land Economics*, *71*(4), 500–515. https://doi.org/10.2307/3146714

McFadden, D. (2014). The new science of pleasure: Consumer behavior and the measurement of well-being. In S. Hess, & A. Daly (Eds.), *Handbook of choice modelling* (pp. 7–48). Cheltenham (United Kingdom), Edward Elgar Publishing. https://doi.org/10.4337/9781781003152.00007

Merkle, D., & Edelman, M. (2002). Nonresponse in exit polls: A comprehensive analysis. In R.M. Groves, D.A. Dillman, J.L. Eltinge, & J.A. Roderick (Eds.), *Survey nonresponse* (pp. 243–258). New York (United States), Wiley.

Meyerhoff, J., Oehlmann, M., & Weller, P. (2015). The influence of design dimensions on stated choices in an environmental context. *Environmental and Resource Economics*, *61*, 385–407. https://doi.org/10.1007/s10640-014-9797-5

Mitchell, R.C., & Carson, R.T. (1989). *Using surveys to value public goods: The contingent valuation method*. Washington, DC (United States), Resources for the Future.

Mogas, J., Riera, P., & Bennett, J. (2006). A comparison of contingent valuation and choice modelling with second-order interactions. *Journal of Forest Economics*, *12*(1), 5–30. https://doi.org/10.1016/j.jfe.2005.11.001

Mørkbak, M.R., & Olsen, S. (2015). A within-sample investigation of test- retest reliability in choice experiment surveys with real economic incentives. *Australian Journal of Agricultural and Resource Economics*, *59*, 375–392. https://doi.org/10.1111/1467-8489.12067

Moser, R., Raffaelli, R., & Notaro, S. (2013). Testing hypothetical bias with a real choice experiment using respondents' own money. *European Review of Agricultural Economics*, *41*(1), 25–46. https://doi.org/10.1093/erae/jbt016

Murphy, J.J., Allen, P.G., Stevens, T.H., & Weatherhead, D. (2005). A Meta-analysis of Hypothetical Bias in Stated Preference Valuation. *Environmental and Resource Economics*, *30*, 313–325. https://doi.org/10.1007/s10640-004-3332-z

Olsen, S.B., & Meyerhoff, J. (2016). Will the alphabet soup of design criteria affect discrete choice experiment results? *European Review of Agricultural Economics*, *44*(2), 309–336. https://doi.org/10.1093/erae/jbw014

Peytchev, A. (2009). Survey breakoff. *Public Opinion Quarterly*, *73*(1), 74–97. https://doi.org/10.1093/poq/nfp014

Presser, S., Couper, M.P., Lessler, J.T., Martin, E., Martin, J., Rothgeb, J.M., & Singer, E. (2004). Methods for Testing and Evaluating Survey Questions. *Public Opinion Quarterly*, *68*(1), 109–130. https://doi.org/10.1093/poq/nfh008

Promberger, M., Dolan, P., & Marteau. (2012). Pay them if it works: Discrete choice experiments on the acceptability of financial incentives to change health related behaviour. *Social Science and Medicine*, *75*(12), 2509–2514. https://doi.org/10.1016/j.socscimed.2012.09.033

Revelt, D., & Train, K. (1998). Mixed Logit with Repeated Choices: Households' Choices of Appliance Efficiency Level. *The Review of Economics and Statistics*, *80*(4), 647–657. https://doi.org/10.1162/003465398557735

Riera, P., Signorello, G., Thiene, M., Mahieu, P.-H., Navrud, S., Kaval, P., . . . Dragoi, S. (2012). Non-market valuation of forest goods and services: Good practice guidelines. *Journal of Forest Economics*, *18*(4), 259–270. https://doi.org/10.1016/j.jfe.2012.07.001

Rose, J.M., & Bliemer, M.C.J. (2009). Constructing efficient stated choice experimental designs. *Transport Reviews*, *29*(5), 587–617. https://doi.org/10.1080/01441640902827623

Rose, J.M., Bliemer, M.C.J., Hensher, D.A., & Collins, A.T. (2008). Designing efficient stated choice experiments in the presence of reference alternatives. *Transportation Research Part B: Methodological*, *42*(4), 395–406. https://doi.org/10.1016/j.trb.2007.09.002

Ryan, M. (2004). A comparison of stated preference methods for estimating monetary values. *Health Economics*, *13*(3), 291–296. https://doi.org/10.1002/hec.818

Sandor, Z., & Wedel, M. (2002). Profile construction in experimental choice designs for mixed logit models. *Marketing Science*, *21*, 455–475. https://doi.org/10.1287/mksc.21.4.455.131

Sandorf, E.D., Aanesen, M., & Navrud, S. (2016). Valuing unfamiliar and complex environmental goods: A comparison of valuation workshops and internet panel surveys with videos. *Ecological Economics*, *129*, 50–61. https://doi.org/10.1016/j.ecolecon.2016.06.008

Sandorf, E.D., Campbell, D., & Hanley, N. (2017). Disentangling the influence of knowledge on attribute non-attendance. *Journal of Choice Modelling*, *24*, 36–50. https://doi.org/10.1016/j.jocm.2016.09.003

Sarkar, A., Aronson, K., Patil, S., Hugar, L., & Vanloon, G. (2012). Emerging health risks associated with modern agriculture practices: A comprehensive study in India. *Environmental Research*, *115*, 37–50. https://doi.org/10.1016/j.envres.2012.03.005

Sarkar, A., Sensarma, S.R., & van Loon, G.W. (Eds) (2019). *Sustainable Solutions for Food Security*. Dordrecht (Holland), Springer. https://doi.org/10.1007/978-3-319-77878-5

Scarpa, R., & Alberini, A. (Eds.). (2005). *Applications of simultaneous methods in environmental and resource economics*. The Economics of Non-Market Goods and Resources. Dordrecht (Holland), Springer. https://doi.org/10.1007/1-4020-3684-1

Scarpa, R., & Rose, J.M. (2008). Design efficiency for non-market valuation with choice modelling: How to measure it, what to report and why. *Australian Journal of Agricultural and Resource Economics*, *52*, 253–282. https://doi.org/10.1111/j.1467-8489.2007.00436.x

Scarpa, R., Gilbride, T.J., Campbell, D., & Hensher, D.A. (2009). Modelling attribute non-attendance in choice experiments for rural landscape valuation. *European Review of Agricultural Economics*, *36*(2), 151–174. https://doi.org/10.1093/erae/jbp012

Scarpa, R., Thiene, M., & Marangon, F. (2008). Using Flexible Taste Distributions to Value Collective Reputation for Environmentally Friendly Production Methods. *Canadian Journal of Agricultural Economics/Revue Canadienne d'agroeconomie*, *56*(2), 145–162. https://doi.org/10.1111/j.1744-7976.2008.00122.x

Šĉasný, M., & Alberini, A. (2012). Valuation of mortality risk attributable to climate change: Investigating the effect of survey administration modes on a VSL. *International Journal Environmental Research and Public Health*, *9*, 4760–4781. https://doi.org/10.3390/ijerph9124760

Smith, V.K. (2006). Judging quality. In B.J. Kanninen (Ed.), *Valuing environmental amenities using stated choice studies: A common sense approach to theory and practice* (pp. 297–333). Dordrecht (Holland), Springer. https://doi.org/10.1007/1-4020-5313-4

Swait, J., & Adamowicz, W. (2001a). Choice environment, market complexity, and consumer behavior: A theoretical and empirical approach for incorporating decision complexity into models of consumer choice. *Organizational Behavior and Human Decision Processes*, *86*(2), 141–167. https://doi.org/10.1006/obhd.2000.2941

Swait, J., & Adamowicz, W. (2001b). The influence of task complexity on consumer choice: A latent class model of decision strategy switching. *Journal of Consumer Research*, *28*(1), 135–148. https://doi.org/10.1086/321952

Teisl, M.F., Boyle, K.J., McCollum, D.W., & Reiling, S.D. (1995). Test-retest reliability of contingent valuation with independent sample pretest and posttest control groups. *American Journal of Agricultural Economics*, *77*(3), 613–619. https://doi.org/10.2307/1243229

Tourangeau, R., & Plewes, T.J. (Eds.). (2013). *Nonresponse in social science survey*. Washington, DC (United States), The National Academies Press. Retrieved from https://researchsociety.com.au/documents/item/1496

Tourangeau, R., Couper, M.P., & Conrad, F. (2004). Spacing, position, and order: Interpretive heuristics for visual features of survey questions. *Public Opinion Quarterly*, *68*, 368–393. https://doi.org/10.1093/poq/nfh035

Train, K.E. (2009). *Discrete choice methods with simulation*. New York (United States), Cambridge University Press. https://doi.org/10.1017/CBO9780511805271

Vermeulen, B., Goos, P., Scarpa, P., & Vandebroek, M. (2011). Bayesian conjoint choice designs for measuring willingness to pay. *Environmental and Resource Economics*, *48*(1), 129–149. https://doi.org/10.1007/s10640-010-9401-6

Vossler, C.A., & Kerkvliet, J. (2003). A criterion validity test of the contingent valuation method: Comparing hypothetical and actual voting behavior for a public referendum. *Journal of Environmental Economics and Management*, *45*, 631–649. https://doi.org/10.1016/S0095-0696(02)00017-7

Vossler, C.A., & Watson, S.B. (2013). Understanding the consequences of consequentiality: Testing the validity of stated preferences in the field. *Journal of Economic Behavior and Organization*, *86*, 137–147. https://doi.org/10.1016/j.jebo.2012.12.007

Vossler, C.A., Doyon, M., & Rondeau, D. (2012). Truth in consequentiality: Theory and field evidence on discrete choice experiments. *American Economic Journal Microeconomics*, *4*(4), 145–171. https://doi.org/10.1257/mic.4.4.145

Watanabe, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research*, *11*, 3571–3594.

Windle, J., & Rolfe, J. (2011). Comparing responses from internet and paper-based collection methods in more complex stated preference environmental valuation surveys. *Economic Analysis and Policy*, *41*(1), 83–97. https://doi.org/10.1016/S0313-5926(11)50006-2