Chapter 4
Contingent Valuation in Practice

Kevin J. Boyle

Abstract Contingent Valuation in Practice takes the reader through each of the basic steps in the design and administration of a contingent-valuation study. The text is written for the novice conducting their first study, decision makers who might want to better understand the method, and experienced researchers. Rich citations to the literature are provided to help a novice reader gain a deeper understanding of the supporting literature and to assist experienced researchers in developing the context for research to enhance contingent valuation in practice. The novice and users of value estimates can rely on the content of this chapter to understand best current practices and researchers can use the content to identify research that addresses uncertainties in the design and conduct of contingent valuation studies. Following the steps outlined in the chapter provides a basis to establish that an empirical study satisfies conditions of content validity.

Keywords Contingent valuation · Stated preference · Nonmarket valuation · Survey research · Hicksian suplus

Contingent valuation is a stated preference method and a survey-based approach to nonmarket valuation. A contingent-valuation question carefully describes a stylized market to elicit information on the maximum a person would pay (or accept) for a good or service when market data are not available. While controversial, as will be discussed below, contingent valuation and choice experiments (Chap. 5)—a close cousin in the stated preference family of valuation methods—are arguably the most commonly used nonmarket valuation methods.

An early contingent-valuation study was conducted by Davis (1963) to estimate the value of big game hunting in Maine. About a decade later, Cicchetti and Smith (1973) and Hammack and Brown (1974) applied contingent valuation to wilderness recreation and waterfowl hunting. Simultaneously, an application to valuing visibility in the Four Corners region of the Southwestern United States represented a turning point after

K.J. Boyle (✉)
Virginia Tech, Blacksburg, VA, USA
e-mail: kjboyle@vt.edu

which contingent valuation gained recognition as a methodology for estimating values for public goods (Randall et al. 1974). The types of environmental applications have continually expanded, and studies have spread to health, cultural, and other types of applications when market data are not available to support decision-making.

Results from early applications of contingent valuation met with skepticism and criticism. One of the more notorious comments was expressed by Scott (1965), who referred to contingent valuation as a "short cut" and concluded, "ask a hypothetical question and you get a hypothetical answer" (p. 37). The contingent-valuation question is hypothetical in the sense that people do not actually make a monetary payment. Some of this criticism was deflected by Bishop and Heberlein's (1979) landmark validity study in which they compared welfare estimates for goose hunting from actual cash transactions, contingent valuation, and travel cost (Chap. 6). This study showed that the contingent-valuation estimate of willingness to pay (WTP) was of similar magnitude to estimates of WTP provided by a travel cost model and the cash transactions. Comparison of the contingent values to the travel cost values is a test of convergent validity, and comparison with cash transactions is a test of criterion validity (Carmines and Zeller 1979).

A workshop sponsored by the U.S. Environmental Protection Agency was the first attempt to synthesize what was known about contingent valuation (Cummings et al. 1986). The notable outcome of this "state-of-the-art assessment" was a set of reference operating conditions for conducting a credible contingent-valuation study. The reference operating conditions motivated much research to evaluate the validity of contingent valuation and to probe the limits of the types of applications where contingent valuation could provide credible welfare estimates.

Perhaps the most substantive contribution to the contingent-valuation literature was made with a book by Mitchell and Carson (1989) that presented detailed discussion on designing a contingent-valuation study. The book provided the broad overview that novices require when conducting their first contingent-valuation study, as well as prescriptive recommendations that set off a whole new wave of validity research. Mitchell and Carson fundamentally shifted the research focus to one that considered the details of study design. Validity, rather than being a global, all-or-nothing criterion, was now viewed as a function of specific aspects of study design; specific design elements were found to create or reduce bias in estimates of central tendency and increase or decrease the dispersion around estimates of central tendency.

Through the first 25 or more years of the use of contingent valuation, critiques of this methodology seemed to ebb and flow without a specific focal point of attack. This all changed when contingent-valuation estimates began to be used in legal cases as the basis of damage payments by parties responsible for large-scale pollution under the Comprehensive Environmental Response, Compensation and Liability Act (CERCLA) of 1980 (Ward and Duffield 1992). The controversy became particularly heated after the settlement of the Natural Resources Damage claim for the Exxon Valdez oil spill. Exxon supported the publication of a book that critiqued the fundamental premises of contingent valuation (Hausman 1993), and the National Oceanic and Atmospheric Administration (NOAA) responded with a blue ribbon panel to evaluate the credibility of using contingent valuation to

estimate passive-use values (Arrow et al. 1993). Passive-use values are those values held by individuals for the item being valued itself and not for any personal use of the item (see Chap. 2, Eq. (2.26)) and are also commonly referred to as nonuse values.[1]

The NOAA panel provided specific recommendations on how a contingent-valuation study should be designed and conducted to develop "reliable" estimates of passive-use values. The panel's recommendations set off another wave of research designed to investigate the credibility of contingent valuation, particularly in the context of applying the NOAA Panel's recommendations to the estimation use, passive-use values, and total values.[2]

From 1993 when the NOAA Panel report was published through 2003 when the first edition of this chapter was published, the contingent-valuation literature was populated with methodological studies designed to better understand the relative strengths and weaknesses of this valuation method in the context of the Panel's recommendations. Moving forward from 2003 to the present, a review of the contingent-valuation literature still reveals methodological studies, but there are many more applications to a variety of issues around the world. Johnston et al. (2017) present contemporary guidance for stated-preference studies, including contingent valuation, that addresses all common applications, not just the estimation of passive use values for litigation.

Recently, the Journal of Economic Perspectives published a special section on contingent valuation as an update on what has been learned in the past 20 years. Kling et al. (2012) concluded that a "carefully constructed number(s) based on stated preference analysis is now likely to be more useful than no number," and "the remarkable progress that stated preference researchers have made … serves as a model for the evaluation of other co-critical techniques" (p. 23). In contrast, Hausman (2012) continued to focus on the issues of hypothetical bias, disparity between willingness to pay and willingness to accept, and embedding/scope problems.[3] An element that is not clearly addressed in either of these papers is the standard that should be used to judge the credibility of contingent-valuation estimates. Kling et al. laid out four criteria where the strongest—criterion validity—asks if "the estimate generated by stated preference methods [is] the same as a willingness-to-pay value that would be generated if real payment was made" (p. 14). The question that is not well addressed is whether real payments are the

---

[1]The term "item" is used generically to refer to the object being valued. Due to the diversity of potential contingent-valuation applications, I did not want to implicitly restrict discussion to imply that studies are restricted to goods or bads, or commodities or resources, or quality or quantity, etc.

[2]The use of contingent valuation to estimate use values, which was the focus of the Bishop and Heberlein (1979) study, was not called into question. Some people have subsequently maintained that use-value studies must also meet the conditions set out by the panel, and some researchers have investigated the panel's recommendations in the context of estimating use values.

[3]Hypothetical bias and scope are discussed in Sect. 4.2.2 in the discussion on validity. The discussion in this chapter focuses on estimating willingness to pay. Readers should see Chap. 2, Sect. 2.1.6 for a discussion of willing to pay and willingness to accept.

conceptually correct counterfactual to judge the credibility of contingent-valuation estimates and whether market contexts can affect real payments as survey design features might affect contingent-valuation estimates. Chapter 4 does not delve into these issues of validity, but some of these issues are covered in Chap. 12.

This chapter provides an overview of the steps in designing and implementing a contingent-valuation study. This chapter does cover the major considerations in designing and implementing a contingent-valuation study. Careful attention to the design features discussed in this chapter can enhance the overall credibility of any contingent-valuation study.

Valuation of groundwater quality is used as an example to help explain the steps involved in designing and conducting a contingent-valuation study. As contingent valuation is applied to a wide array of issues, the design features presented in this chapter are broadly appropriate for all applications. Methodological citations will be drawn from all types of applications published in the peer-reviewed literature, not just groundwater valuation applications. Design features unique to certain applications are not discussed here.

4.1 Steps in Conducting a Contingent-Valuation Study

Most of the action in designing a contingent-valuation study occurs in the development of the survey instrument in terms of scenario wording, choice of a valuation question, and bid selection where appropriate (Table 4.1). The valuation scenario in the survey is the foundation for value estimation, and the auxiliary questions in the survey provide respondent-specific data for statistical analyses. As choices in the survey design and data analysis stages of a study can affect welfare estimates, careful design of a contingent-valuation survey and careful analysis of the resultant data are crucial to the estimation of credible welfare estimates.

4.1.1 Identifying the Change in Quantity or Quality to Be Valued

The first step in conducting a contingent-valuation study involves developing a theoretical model of the value(s) to be estimated, which is based on the difference between the baseline utility with the current environmental condition and the utility with the new environmental condition (Step 1 in Table 4.1).[4] Following the notation protocol established in Chap. 2 (see Eq. 4.2), the value for a program to protect groundwater from contamination can be defined as

---

[4]"Change" is used as a generic reference to an increase or decrease in quality or quantity valued.

Table 4.1 Steps in conducting a contingent-valuation study

Step 1	Identify the change(s) in quantity or quality to be valued.
Step 2	Identify whose values are to be estimated
Step 3	Select a data collection mode
Step 4	Choose a sample size
Step 5	Design the information component of the survey instrument
Step 5.1	Describe the item to be valued
Step 5.2	Select a provision mechanism
Step 5.3	Select a payment vehicle
Step 5.4	Select a decision rule
Step 5.5	Select a time frame of payment
Step 5.6	Include substitutes and budget constraint reminders
Step 6	Design the contingent-valuation question
Step 6.1	Select a response format
Step 6.2	Allow for values of $0
Step 6.3	Address protest and other types of misleading responses
Step 7	Develop auxiliary questions for statistical analyses of valuation responses
Step 8	Pretest and implement the survey
Step 9	Analyze data
Step 10	Report study

$$v\left(P^0, Q^1, y\right) = v\left(P^0, Q^0, y - \text{WTP}\right), \qquad (4.1)$$

where $v(\cdot)$ is indirect utility function, P is the price of drinking water obtained from groundwater, y is income, WTP is a Hicksian measure of value, and Q is groundwater quality (Q^0 denotes current water quality and Q^1 denotes degraded quality if the protection program is not implemented).[5,6,7] The value of interest is that of protecting groundwater from contamination ($Q^0 > Q^1$). Note that P is assumed to remain constant. Assuming the change does not affect drinking water (e.g., water utility purification protects drinking water supplies), the value that would be estimated here is a passive-use value associated with a change in groundwater quality (ΔQ). The purpose here is not to delve into details of theoretical definitions of changes in welfare, which is the domain of Chap. 2, but to

---

[5]In Chap. 2, Flores uses C to designate WTP, and here WTP is directly used as the symbol.

[6]All other arguments in the indirect utility function are assumed to be constant between the status quo (Q^0) and the condition with diminished groundwater quality (Q^1), orthogonal to p and Q, and are suppressed for notational convenience. It is assumed here that some activity is undertaken to remediate the contamination such that the price of providing potable water increases ($p^0 < p^1$), but the activity is not fully effective so there is a residual value to be estimated.

[7]"Willingness to pay" is used in the presentation of this chapter while fully recognizing that many decisions actually require estimates of willingness to accept (Chap. 2, Eq. (2.3)).

discuss aspects of the value definition that must be considered in the design of any contingent-valuation study.

The theoretical definition of value is fundamental to three key components of any contingent-valuation study. First, the definition of value plays a central role in developing the description of the conditions with and without the item to be valued (Step 5.1 in Table 4.1).[8] Second, the definition frames the statistical analysis of the contingent-valuation responses (Step 9). Third, having a theoretical definition allows for clear interpretations of value estimates in specific decision-making contexts (Step 10). Decision-making is used in a very general sense to refer to any condition that can change utility and for which welfare estimates are required. This broad definition includes any public or private decision-making process where nonmarket values would be used.

The difficult part of Step 1 is identifying the physical change and how it affects utility. Economists depend on physical and biological information to establish status quo conditions (Q^1) and predictions of future conditions to be avoided (Q^0) or in other cases attained. This information may not be known with certainty, and applied valuation further requires information about this uncertainty. For example, a definition of value for a policy that reduces the probability that groundwater will become contaminated is as follows:

$$\pi_1 v\left(P^0, Q^1, y - op\right) + (1 - \pi_1) v\left(P^0, Q^0, y - op\right)$$
$$= \pi_0 v\left(P^0, Q^1, y\right) + (1 - \pi_0) v\left(P^0, Q^0, y\right), \tag{4.2}$$

where π is the probability of contamination, $\pi_0 > \pi_1$, and op is willingness to pay (option price as defined in Eq. 2.32, Chap. 2) to reduce the probability of contamination (Bishop 1982). Here, π_0 is the probability that the current water quality, Q^1, will not be degraded in the absence of a policy, and π_1 is the enhanced probability that Q^1 will not be degraded if the policy is enacted. Thus, op is maximum WTP to increase the probability that the current water quality will be maintained. Applying contingent valuation to conditions that change the probability of an outcome is a common practice.

The fundamental information transmitted in the survey is the description of the change to be valued. For example, if a policy will protect groundwater from contamination, the survey will explain current conditions in the aquifer (Q^1) and what would happen to quality if the protection policy were not implemented (Q^0).

Describing changes in quality in contingent-valuation surveys is an area where there has been considerable progress in recent years in two dimensions. First, there have been advances in the information on environmental quality available to study designers. These advances have evolved from interdisciplinary research projects where economists are interacting with physical and biological scientists at very early stages in the research process (Johnston et al. 2012), the expanding availability of GIS data (geographical information systems; Bateman et al. 2002), and the use of simulated images to portray new conditions (Madureira et al. 2011). The

---

[8]"Item" is used as a generic term to reference a quantity or quality change valued.

second evolution is the use of Internet surveys where the presentation of information can be carefully presented to respondents, such as allowing respondents to reference information as they complete the survey and limiting respondent's ability to read ahead or change answers as they complete the survey (Lindhjem and Navrud 2011). These advances have improved the realism of the information in contingent-valuation scenarios to respondents and helps to more carefully link value estimates to decision-making.

Sometimes, in the absence of clear information on the change to be valued, survey designers will focus on the action that affects the change in services people enjoy. With only a description of the action and not a description of the change, survey respondents are left to make two assumptions: what will be the change to value and how this change will affect the services they receive. This issue has not been directly and extensively addressed in the contingent-valuation literature and deserves more consideration. The key design consideration is to avoid descriptions that are vague or confusing. In the absence of clearly understood descriptions of the change to be valued, survey respondents are left to their own assumptions regarding what the change will accomplish. This is a problem because different respondents could make different assumptions about what to value, and those assumptions might not be consistent with the change the study is designed to value.

It could be the case that a contingent-valuation study is being conducted in advance of an anticipated decision, and the details of the action and the consequent effects are not known. Contingent-valuation studies are often designed to estimate an array of values that represent plausible conditions that might occur when the physical information becomes known or the actual decision is finalized. Thus, the design of a contingent-valuation study proceeds using a "best knowledge" of the actual change that will be implemented and of the ultimate effects with and without the action. Here is where a careful theoretical definition of value estimated is crucial; the definition allows an ex post interpretation of whether the value estimate matches the realized change it is intended to value.

In addressing this step, it is important to recognize that there are two types of contingent-valuation studies. Many studies in the peer-reviewed literature focus on methodological contributions and are not designed to address a specific decision-making issue (e.g., Murphy et al. 2010), while other studies have been designed to address a specific decision-making issue (Whittington and Pagiola 2012). This difference is crucial because methodological studies can use whatever definition of value is convenient. In a specific decision-making context, however, the value definition must be linked to the specific change in utility that will occur if the proposed change is implemented. Methodological studies still require theoretical definitions of estimated values to guide data analyses. Theoretical definitions of value can also allow estimates from methodological studies to be useful for benefit transfers (see Chap. 11).

The discussion in this section has focused on Hicksian measures of value, which is the theoretical concept typically measured in contingent-valuations studies. Step 1 involves an explanation of the theoretical definition of value presented within the survey instrument and the change that respondents are asked to value.

4.1.2 Identify Whose Values Are to Be Estimated

Once the change to be valued has been specified, the affected population can be identified (Step 2). This step involves identifying the affected population and selecting a sampling procedure (see Chap. 3, Fig. 3.1). This information is important in selecting a sample frame so that the contingent-valuation survey will be administered to a representative sample of affected individuals. It is also important to determine if the contingent-valuation study will estimate values on a per-capita or per-household basis. It is necessary to know the size of the affected population to expand the per-unit values to a population value. The affected population and unit of measurement could indicate a desirable mode of data collection or modes of data collection that are acceptable for the particular application.

In the example where groundwater quality is threatened, the affected population might constitutes those individuals who use the aquifer as a source of drinking water and those who hold passive-use values for groundwater in the aquifer. If everyone in a community obtains water for household usage that is drawn from the aquifer, it is fairly clear that the affected population includes everyone in the community. However, the affected population might also include people who live outside of the city and hold passive-use values. In this case it may be more difficult to identify the population holding passive-use values, and a uniform list might not be available to draw a sample from.

Most studies use geopolitical boundaries such as a city, county, or state in the United States to define the relevant study population. In the above example, the study population might be extended from the specific city to the county where the city is located if the aquifer of interest extends beyond the city boundary. The literature provides little guidance for selecting study populations (those who are affected by the change), but two points of consideration are useful. First, geopolitical boundaries are useful for identifying locations affected by the change being valued and those who will pay for the change to be implemented. Second, a spatially referenced sample will allow for an investigation of how value estimates change with distance from the affected area. If values decrease or increase with distance, this can affect the magnitude of aggregate welfare calculations (Bateman et al. 2006; Hanley et al. 2003; Mansfield et al. 2012).

Chapter 3 distinguishes between the affected population and the sample frame (those who are eligible for selection in the sample). Sample frame is important because this determines whether the selected sample is representative of the affected population.

Another issue is the unit of measurement for values. Some contingent-valuation surveys elicit values for individuals (e.g., Bateman et al. 1995), and others elicit household values (e.g., Poe and Bishop 2001). It is important that the framing of contingent-valuation questions makes clear whether a household valuation or an individual valuation is being requested. Mitchell and Carson (1989) asserted that "payments for most pure public goods are made at the household level. … When this is the case, the appropriate sampling procedure is to allow any adult who claims

to be a household head to be a spokesperson for the household—the current practice of the U.S. Census Bureau" (pp. 265-266). (Mitchell and Carson cited Becker's (1981) *Treatise on the Family* in support of this assertion.)

However, the issue is more complicated. Quiggin (1998) argued that only when intra-household altruism does not exist or is paternalistic will the sum of individual values equal household values, yielding the same aggregate measure of welfare change. In the absence of these conditions, he argued that sampling and aggregating over individuals will yield larger measures of aggregate values than sampling and aggregating over households. In contrast, Munro (2005) argued that individual and household values will be the same only if household members pool their incomes. Supporting these potential limitations, Bateman and Munro (2009) and Lindhjem and Navrud (2012) found significant differences between individual and household values. These results would appear to violate the unitary model where households act like a person and there is only one utility function. However, Chiappori and Ekeland's (2009) collective model assumed income sharing among household members where it is possible to isolate individual welfare measures based on an allocation of resources to household members. In this theoretical framework, the potential contradictions mentioned above may not, in fact, be contradictions.

The lesson here is that restrictions on household member utility functions and interactions are necessary to secure equality between the sum of individual household member values and household values, and these assumptions might not always be testable. Further, it is not appropriate to assume that household values are a conservative approach. Thus, when choosing individual or household sampling units, it is important to ask several questions:

- Do households make individual or group decisions?
- If they make group decisions, do households pool their income?
- If they make group decisions, do decision-makers adequately account for the preferences of all household members?
- Do some households make group decisions and others make individual decisions?

Answers to these questions have important implications for individual and aggregate welfare estimation. A decision on individual versus household units of measurement should be based on the specific decision-making context being valued, and this selection influences what payment vehicle is selected (see Sect. 4.1.5.3), pretesting the survey instrument, the availability of a sampling frame, and other dimensions of the study design. A recent study suggests that people in a household act similarly, so this might not be an issue if the finding can be generalized to other applications (Adamowicz et al. 2014).

The bottom line for both sample frame and sampling units is that the sample frame should be known, and each unit in the sample should have a known

probability of selection if value estimates are to be used to support decision-making.[9] Sample frame and unit of observation decisions should be clearly documented and any limitations or uncertainties noted in reporting study results.

4.1.3 Select a Data Collection Mode

A contingent-valuation study requires the collection of primary data (Step 3). The various modes of data collection are discussed in detail in Chap. 3 of this volume, so this section focuses on insights for contingent-valuation applications.

The most common way to implement contingent-valuation surveys is by mail (Schneemann 1997). Recently, Internet survey implementation has increased. Each method has its relative strengths and weaknesses (Dillman 2007). For a discussion of the alternative survey modes, see Chap. 3.

The primary reason that mail surveys are commonly used is that they have been the least expensive way to collect data. A recent call to a premier survey research firm suggests that the cost of a completed survey ranges from less than $25 for Internet implementation to $50–75 for a mail survey, $75–150 for a telephone survey, and $1,000–2,000 for personal interviews. These numbers are for national probability samples where an existing sample frame is assumed for the Internet implementation, and a national area probability sample is used for the personal interviews. Thus, with a limited budget, it is easy to see why Internet surveys are gaining popularity.

Another factor that has traditionally affected the choice of a survey mode is the expected survey response rate and the potential for item nonresponse to key questions in the survey (Messonnier et al. 2000). Even with careful design, Schneemann (1997) showed that response rates to contingent-valuation surveys are affected by factors that are not under the control of the study design. For example, decisions that involve specific user groups (e.g., anglers or hunters) are likely to result in higher response rates than general population surveys. Schneemann also found that survey response rates are related to the study application, the affected population, and the components of the contingent-valuation question itself, but this is not unique to contingent-valuation surveys. Rather, it is an issue to consider in the design of any survey (Groves and Peytcheva 2008). Thus, even with a good design process, it is important to recognize that features of the study application, as well as the design of the contingent-valuation component of the survey, affect survey response rates across all survey modes.

Response rates have been an issue of concern for Internet surveys. A credible Internet survey requires recruitment of individuals to represent the demographics of a known population, (e.g., the population of the country where the sample is drawn

---

[9]If the only purpose of a study is experimentation, the only important criteria may be random stratification, and sample representativeness generally is not a concern.

[see Chap. 3]). After this selection, the survey response rate can be less than 10% (Boyle et al. 2016; Manfreda et al. 2008). However, focusing on survey response rates as an indicator of survey quality and nonresponse bias has been shown to be misleading (Groves and Peytcheva 2008; Keeter et al. 2000). As a consequence, other measures of survey representativeness have been proposed (Groves et al. 2008; Schouten et al. 2009). Thus, while the expected survey response rate should be a design consideration, it is not necessarily a good metric to use to select the appropriate survey mode for implementing a contingent-valuation survey.

Conveying information on the item being valued is the fundamental component of a contingent-valuation survey. Personal interviews may be the most effective survey mode for conveying information because visual information can be provided and an interviewer is present to explain the information and answer questions.[10] An Internet survey allows for control of the information respondents read and can provide live links if a respondent wants to review information already presented, but an interviewer is not available to assist, answer questions, and monitor administration of the survey. A mail survey is more limited because no interviewer is present to explain the visual information and no control is provided on how respondents proceed through the survey instrument. The ability to provide information in a telephone survey is much more limited because no visual information is available. A mixed-mode survey where a telephone interview is conducted after respondents have received written and visual information in the mail, similar to the content of a mail survey, is one way to overcome the informational deficiencies of telephone interviews (Hanemann et al. 1991).

Other methods of implementing contingent-valuation surveys include on-site intercepts (Boyle et al. 1994) and convenience samples of students or other groups who are brought to a central location to participate in the study (Cummings and Taylor 1998). Davis and Holt (1993) noted that "behavior of (real) decision makers has typically not differed from that exhibited by more standard … student subject pools"[11] (p. 17; see also Smith et al. 1988). Maguire et al. (2003) found that college students behave similarly to a sample of all adults. These alternative survey modes with convenience samples are best used for implementing methodological experiments and not for studies designed to estimate values generalizable to a specific population to support decision-making. When experiments are conducted, it is still important that study participants are exogenously selected, and people are not allowed to voluntarily opt into the sample.

It is important to consider the advantages of each mode when implementing a contingent-valuation survey. A mail survey does not always dominate just because of cost advantages, and personal interviews do not always dominate just because of

---

[10]Personal interviews can also result in interviewer bias in responses to the contingent-valuation question. While Boyle and Bishop (1988) found interviewer effects with graduate students doing the interviewing, these effects might not persist if professional interviewers are used.

[11]Davis and Holt (1993) were referring to induced-value studies, and questions remain regarding whether contingent-valuation estimates derived from samples of convenience can be used to develop population inferences of gains and losses in economic welfare.

informational advantages. In reviewing studies that compared contingent-valuation studies implemented via the Internet versus mail, phone, and in-person implementation, Boyle et al. (2016) found that the ratio of willingness-to-pay estimates based on Internet surveys to estimates based on data from other modes averaged 0.92.[12]

4.1.4   Choose a Sample Size

Concerns about sample sizes (Step 4) are important for two reasons. First, the precision of estimated values affects their usefulness in decision-making. An estimate with large error bounds can leave questions about whether benefits really exceed costs or what a specific damage payment should be if a study were used to assess damages. Second, statistical precision affects the ability to detect differences among value estimates in methodological studies designed to investigate the validity and reliability of contingent-valuation estimates.

As noted in Chap. 3, selection of a sample size (Step 4 in Table 4.1) is a matter of choosing an acceptable level of precision within a given budget. That is, the standard error of mean WTP is

$$\mathrm{se}_{\overline{\mathrm{WTP}}} = \frac{\sigma}{\sqrt{n}}, \tag{4.3}$$

where σ is the standard deviation and n is the number of completed surveys used in the analysis. Thus, for a given variance, the absolute value of the standard error can be reduced by increasing the sample size. The larger is σ, the larger the required sample size to attain the desired level of precision. Additionally, for a fixed σ, a sample size will represent a larger percentage of the population for a small population than it will for a large population (Salant and Dillman 1994).

For applications where there have been a lot of contingent-valuation studies conducted (e.g., recreational fishing), there may be sufficient information to develop a reasonable estimate of σ for a new application. Where existing studies are not available, an estimate of σ can be obtained through a field test of the survey instrument. In practice, most studies choose the largest sample size possible given the available budget.

Selecting a sample size also involves consideration of the expected response rate and the expected item nonresponse to the contingent-valuation question and other

---

[12]Caution is warranted when interpreting these ratios as the comparisons are confounded by differences in sample frames and survey modes for all but one study. In the sole study where this confound was not present, the ratio was 0.86 (Linhjem and Navrud 2011); for the Boyle et al. (2016) study, which also controlled for sample frame, the ratio was 0.78.

variables that will be used to analyze contingent-valuation responses.[13] Other considerations include whether the sample will be stratified into subsamples for analysis and data reporting.

4.1.5 Design the Information Component of the Survey Instrument

Step 5 in Table 4.1 focuses on the information provided to respondents in the survey instrument. This includes telling respondents what it is they are being asked to value, how it will be provided, and how they will pay for it. Economics does not provide guidance for selecting these design features, and careful consideration must be given to these design features in the pretesting of the survey instrument. The design of the valuation scenario should be developed so that respondents believe the change to be valued can be feasibly accomplished and that the features selected do not unduly influence welfare estimates.

There are no hard and fast rules in the elements of this portion of the design phase. Thus, the use of cognitive interviews, focus groups, and pilot surveys is crucially important to understanding respondents' comprehension of the information presented and how they react to the information.

4.1.5.1 Describe the Item to Be Valued

The description of the item to be valued (Step 5.1 in Table 4.1) tells respondents what is being valued and presents the change in the quantity, quality, or probability to be valued (Step 1). The information set should include a description of the item being valued, baseline conditions, and the new conditions. This does not require an extensive, detailed valuation scenario, but a scenario that is clear so that respondents understand the change they are being asked to value.

The information is presented in written or verbal form and is accompanied by graphs, pictures, and other visual stimuli to facilitate respondent understanding. The information scenario is not a marketing or sales pitch but a neutral and fair description of the change to be valued, and it can include statements about why some people would desire the change and why others would not (Alberini et al. 2005). This information should also include an explanation of why the change is being valued.

---

[13]Consider a case where the desired sample size for statistical analyses is 500. Further consider that a mail survey will be used with a sample list that has 10% bad addresses, an expected response rate of 60%, and 10% item nonresponse to a valuation question. The initial sample size would need to be at least 1,029 [$500 \times (1/0.9) \times (1/0.6) \times (1/0.9)$].

While the description of the item to be valued is the fundamental component in the design of any contingent-valuation study, it seems that the information is rarely complete in terms of the baseline condition(s) and the new condition(s) that will result as a consequence of the change (e.g., the levels of Q defined in Eqs. (4.1) and (4.2)). This problem arises most frequently in the estimation of values for marginal changes and to a lesser extent when total values are estimated. If respondents must infer the change being valued, it is likely that different respondents will use different subjective perceptions.

The appropriate quantity and quality of information required for respondents to provide valid value responses is a matter of much interest. While some have concluded that a credible contingent-valuation study requires that survey respondents be provided extensive, detailed information, the literature does not support such a conclusion. The literature does indicate that specific types of information should be provided to respondents. While only a small number of studies have investigated the effects of information on contingent-valuation estimates, these studies collectively tell an important story.

Samples et al. (1986) confirmed the obvious: You must tell people what it is they are being asked to value. Boyle (1989), in a study of trout fishing, found that providing respondents with additional information beyond the basic description of the change to be valued decreased estimates of central tendency, but the reductions were not statistically significant. The standard errors of the welfare estimates, however, decreased significantly with additional information. This result suggests information also affects the efficiency of value estimates. Bergstrom et al. (1990) investigated the effect of providing information about the services provided by wetlands and found that value estimates were affected by different types of service information. This outcome suggests that specific information on services is important for applications where respondents are not fully aware of how they currently benefit or could benefit in the future from the item being valued. Poe and Bishop (1999) demonstrated that specific information on well-water contamination was required in a study of groundwater protection. These findings are consistent with a more recent study by MacMillan et al. (2006) that found more information is required when people are less familiar with the item being valued. Valuation of a trout fishery by licensed anglers is an example where the study participants would be more familiar with the item being valued, but wetland and groundwater valuation are examples where study participants would likely have less knowledge or experience. These studies clearly indicate that specific information about the item being valued is required in order to elicit credible responses to contingent-valuation questions, and this is done through the qualitative research process to design the survey instrument.

Li et al. (2014) discussed prior knowledge and acquired knowledge people use in answering a survey. Prior knowledge includes knowledge people possess prior to engaging in the survey and any information they seek external to the survey while completing the survey (e.g., doing an Internet search for additional information). Acquired knowledge is the information provided in the survey instrument. Berrens et al. (2004) suggested that when given the opportunity to seek additional

information, respondent use of this information is modest. This suggests that respondent attempts to seek additional prior knowledge could be limited. However, any additional prior information respondents seek introduces potential noise into welfare estimation. Thus, pretesting the information in the survey is needed to establish that survey respondents have the information they need to respond to the value question(s).

While some studies have used pictures and other types of graphics (e.g., maps, graphs, and tables) in valuation scenarios, there do not appear to be any studies that have evaluated whether the use of graphics to display information in the valuation scenario affects value estimates. The use of graphics requires careful pretesting so that the images do not inadvertently introduce unwanted effects into valuation responses. Thus, while pictures and other graphics can be helpful in conveying information in a contingent-valuation survey, they can also generate unwanted effects. The use of multiple modes of portraying information in a contingent-valuation scenario, such as written text, numerical presentation, graphs, or pictures, can facilitate respondent understanding because different people may use different information to understand the change to be valued.

Collectively, the lesson is that respondents to a contingent-valuation survey need to be presented with information that clearly explains the change to be valued and that such information must account for heterogeneity in how respondents use and process information. There is a careful balance between providing too little information such that respondents could misinterpret the valuation question and providing too much information so that respondents do not attend to critical portions of the information provided. Refinement of this information occurs in focus groups, one-on-one interviews, and, if necessary, in small-scale field pretests.

4.1.5.2 Select a Provision Mechanism

In any contingent-valuation study, it is necessary to tell respondents how the change to be valued will be implemented (Step 5.2). Without such information, respondents might not view the change as being credible or might implement a personal assumption that is not appropriate and could inadvertently affect value estimates. In a general sense, the provision mechanism is the "production process" that will accomplish the change respondents are asked to value. Suppose a policy was protection of well water from contamination. One provision mechanism that has been used to provide such protection is to establish protection zones around wellheads, which preclude any activities that might contaminate the groundwater. Sometimes applications have a clear mechanism that is part of the actual action being valued, while in other applications the selection of the provision mechanism is part of the study design.

Choosing the provision mechanism is complicated because the chosen mechanism could affect responses to contingent-valuation questions. For example, consider public concern over chemical residues in fruits and vegetables, genetically modified foods, sweatshop production of clothing, dolphin-free tuna, etc. These

production attributes affect purchase decisions for market goods (Foster and Just 1989; Teisl et al. 2002) and there is no reason why there should not be similar types of provision-mechanism effects in responses to contingent-valuation questions. For example, while sweatshop production might not affect the quality of the clothes in terms of their use by the purchaser, this production process could represent an undesirable externality.

The effect of the selected provision mechanism on welfare estimates has not been formally investigated to my knowledge. At a minimum, careful pretesting in focus groups can identify whether respondents will understand the provision mechanism and if they feel it is credible in producing the change to be valued.

4.1.5.3 Select a Payment Vehicle

A payment vehicle (Step 5.3) is the mechanism by which respondents are told how payments would be made. For example, this might be a tax increase for a public policy or a higher price for a health-enhancing procedure.

This is a design area where the trade-off between credibility and unintended effects has been clearly noted in the literature. Mitchell and Carson (1989) argued that the choice of a payment vehicle requires balancing realism against payment vehicle rejection. That is, as realism increases, the likelihood that the payment vehicle will engender responses that protest the vehicle might also increase. For example, water-use fees are very realistic payment vehicles, but someone who values protecting potable groundwater might still give a valuation response of $0 to protest an increase in water rates. Income tax vehicles can run into problems due to resistance to higher taxes. On the other hand, a donation payment vehicle could yield an underestimate of value because this vehicle might not be incentive compatible for estimating a respondent's full willingness to pay (Wiser 2007).[14]

Failure to provide a realistic payment vehicle can also lead to protest responses. A sales tax would not be credible in an area that does not have sales taxes or when the group providing the item being valued does not have taxing authority. Thus, respondents could reject the valuation scenario even if they value the change because the payment mechanism is not believable. The realism of some vehicles can lead people to give what they think is a reasonable response, not their maximum WTP. For example, where there are only nominal entry fees (e.g., entrance fees at national parks), an increase in an entrance fee could engender responses of what respondents think is a reasonable increase in price rather than statements of maximum WTP (Campos et al. 2007).

Some studies demonstrate that payment vehicles do influence welfare estimates (Rowe et al. 1980; Greenley et al. 1981; Campos et al. 2007), but the line of inquiry

---

[14]Incentive compatibility refers to whether elements of the survey design will motivate people to truthfully reveal their preferences (i.e., statements related to the value they place on the item being valued).

Table 4.2 Payment vehicles used in recent studies

Payment vehicle	Study citation
Higher prices	Desaigues et al. (2011)
Voluntary donation	Garcia-Llorente et al. (2011)
Annual tax	Lindhjem and Navrud (2011)
Network fee	Menges and Beyer (2014)
Water bill	Ramajo-Hernández and del Saz-Salazar (2012)

has not been prominent in recent years.[15] Testing of payment vehicle effects is typically undertaken in survey pretesting to select a payment vehicle that minimizes undesirable effects on value estimates.

A variety of payment vehicles have been used in studies and a sampling of those in the recent literature are presented in Table 4.2. These examples are presented as a general set of payment vehicle examples and each could have relative strengths or weaknesses, as discussed above. While research can generate general insights about payment vehicles (e.g., donations are not incentive compatible and taxes are likely to lead to protest responses), selection of a specific payment vehicle is likely to be study-specific and will always require careful pretesting.

A concern with using prices or taxes as a payment vehicle is that people can adjust the quantity purchased (e.g., take fewer recreational trips or consume less water). This makes the cost to the individual endogenous; the respondent controls cost by adjusting quantity. Further, the choice of a payment vehicle must align with the value to be estimated. For example, using an increase in water rates to estimate passive-use values for groundwater (Eq. (4.1)) might not be logical to respondents and could result in some respondents including use values in their valuation responses. The concerns discussed here highlight how important it is to carefully select a payment vehicle that minimizes unintended effects on value estimates.

4.1.5.4 Select a Decision Rule

The decision rule is the process by which the results of the contingent-valuation study, individual valuation responses or summary statistics on valuation responses, are used to inform the decision as to whether the item valued will be provided (Step 5.4). Such a decision rule might be that the item will be provided if at least 50% of respondents answer "yes" to a dichotomous-choice question.

The choice of a decision rule is closely linked to the payment vehicle. A referendum is clearly applicable when the issue relates to the provision of a public good, such as groundwater protection, and the payment vehicle is an increase in taxes. However, a referendum would not be applicable when dealing with use

---

[15]In a recent conjoint study, Johnston et al. (1999) demonstrated that the choice of a payment mechanism, particularly one that guarantees funding for a change, can influence welfare estimates.

values, such as recreational fishing, and the payment vehicle is an increase in individual trip costs. In this second example, the decision rule mag be if aggregate benefits exceed project costs.

The more complex and perhaps more important cases are the public good cases like the groundwater valuation example. A fundamental goal in the choice of a decision rule is to make a selection that is plausible to respondents and will entice truthful valuation responses. Following Carson and Groves (2007), it is also important that the decision rule is consequential, which means that payment is mandatory if the program is imlemented and there is a non-zero probability that responses to the survey will influence provision of the item being valued.

4.1.5.5 Select a Time Frame of Payment

This step describes the number and frequency of payments respondents will make (Step 5.5). For example, a valuation scenario might be posed where a new filtration system would be installed to remove contaminants from a public water supply. Values could be elicited as a one-time payment now or as annual payments over the lifetime of the system, say 20 years.

While this is another area where there is scant research, Stevens et al. (1997) showed that repeated payments yield statistically different estimates of WTP when compared with a lump-sum payment. More recently, Soliño et al. (2009) found no difference between bimonthly and annual payments when a dichotomous-choice valuation question was employed. These limited results suggest that the choice of a payment time frame must proceed with caution.

There is often a disconnect between the time frame of payment in a contingent-valuation question and the time frame over which survey respondents will enjoy the benefits of the change. Typical changes result in benefits that accrue over a number of years, and the time frame of the payment(s) is often much shorter (e.g., a one-time payment). Thus, survey respondents are asked to undertake personal discounting to answer valuation questions. The research by Stevens et al. (1997) suggested that survey respondents might not do this well.

The time frame of payment varies substantially across studies, from one-time payments to annual payments into perpetuity. The time frame is crucially important because this influences how value estimates are aggregated to compute benefits or costs. This is another design feature that must be carefully addressed in survey pretesting.

4.1.5.6 Substitutes and Budget Constraint Reminders

Substitutes and a budget constraint are fundamental components of economic choices (Step 6). Both the availability of substitutes (Freeman 1993, Chapter 3; Hoehn and Loomis 1993; Flores, Chapter 2 in this book) and income (Flores and Carson 1997) affect the magnitude of welfare estimates.

Though encouraging respondents to consider substitutes and think about their budget constraints when answering contingent-valuation questions is intuitively straightforward, it is difficult to test the effectiveness of these reminders. What might be considered a substitute by one respondent might not be considered a substitute by another respondent. Split-sample studies, where one sample is reminded of substitutes and their budget constraints and another sample is not, reveal that information on substitutes, complements, and budget constraints affect estimates of central tendency and dispersion (Kotchen and Reiling 1999; Loomis et al. 1994; Whitehead and Bloomquist 1995). In a meta-analysis, Schläpfer (2006) found significant income effects for contingent-valuation estimates. Smith (2005) suggested that sensitivity to budget becomes more relevant the higher the cost to the respondent is as a proportion of income. Given the roles that substitutes, complements, and income play in theoretical definitions of economic values, the theoretical component of content validity suggests that respondents should be prompted to consider likely substitutes and complements, and they should be reminded that they could spend their money otherwise.

4.1.5.7 Summary

There is no cookie cutter or one-size-fits-all set of rules for framing contingent-valuation scenarios, but testing draft information scenarios in focus groups is critically important. Focus group testing provides the opportunity to learn if respondents are using the information, understand and believe the information, and are basing valuation responses on the actual change being valued.

Even seemingly innocuous statements and terms in contingent-valuation scenarios have the potential to affect valuation responses. For example, in a study of preserving agricultural lands we quickly found that open space has entirely different meanings to people involved in land-use policy as compared to the general public. Focus group participants told us that open space conveyed a sense of "outer space," "large foyers," etc.—not undeveloped land.

Pretesting in focus groups and/or one-on-one interviews is the best way to avoid pitfalls that can bias welfare estimates because of incorrect interpretation of information by respondents, the provision of unintended clues to respondents, and information rejection by respondents. This pretesting must be carefully conducted and is not a substitute for more research to understand the effects of each element of the information in a contingent-valuation scenario. This means that careful design must also be accompanied by conceptual and methodological research to refine what and how information should be presented in the survey to guide the design of future empirical studies.

In practice, it is important to recognize that the information set must vary from study to study to fit the issue-specific application and institutions. Further, any information set will not be understood and accepted by everyone in a sample; the design goal is to minimize misinterpretations and scenario rejection to the greatest extent possible. This means that some information in the scenario will be necessary

to satisfy some respondents and other information will satisfy other respondents; the common goal is to elicit consistent value information across all respondents. Little is documented about study pretesting, but reporting of this information in journal articles and other publication will help the valuation community learn about design challenges and successes from prior study designs.

4.1.6   Design the Contingent-Valuation Question

After the information regarding how the change to be valued will be provided, the respondents are asked to reveal information about the value they place on the change described in the valuation scenario. This section provides guidelines and considerations for selecting and designing a contingent-valuation question (Step 6).

4.1.6.1   Select a Response Format

The response format refers to how the contingent-valuation question will be answered (Step 6.1). The three main formats ask respondents to directly provide their maximum willingness to pay (open-ended), choose an amount from a list of possible willingness-to-pay amounts (payment-card), or respond "yes" or "no" to a specified dollar amount (dichotomous-choice). The response format has implications for how the response data are analyzed and interpreted; it is the key characteristic that differentiates the various types of contingent-valuation questions. The information scenario components described above are generally portable from one question format to another with the exception of the decision rule (e.g., a majority vote would work with a dichotomous-choice question but not with an open-ended question).

Early contingent-valuation studies used either an open-ended question (Hammack and Brown 1974) or an iterative-bidding question (Randall et al. 1974). An open-ended question asks respondents how much they would pay for the specified change. An iterative-bidding question starts by asking respondents, "would you pay $SB" for a specified change (SB = starting bid). If respondents answer "yes," then the bid is increased in specified increments (I) until they say "no" and decreased until they answered "yes" if the initial response was "no" ($SB ± $I). The magnitudes of starting bids, magnitudes of bid iterations, and number of iterations varied from study to study. While the open-ended format has persisted, the iterative-bidding format is no longer used because of an anchoring effect where the final bid at the end of the iterations was found to be significantly correlated with the starting bid (i.e., the higher the starting bid, the higher the final bid to which people would answer "yes" (Boyle et al. 1985; Thayer 1981).

If the passage of the proposal would cost you these amounts **every year** for the foreseeable future, what is the highest amount you would pay and still vote for the program? (CIRCLE THE HIGHEST AMOUNT THAT YOU WO ULD STILL VOTE FOR THE PROGRAM)

10¢	50¢	$1	$5	$10	$20
$30	$40	$50	$75	$100	$150
$200		MORE THAN $200			

Fig. 4.1 Example of an unanchored payment card (Welsh and Poe 1998, p. 183)

Open-ended questions are still used in some studies. The following is an example of an open-ended question used by Welsh and Poe (1998):

If passage of the proposal would cost you some amount of money **every year** for the foreseeable future, what is the highest amount that you would pay annually and still vote for the program? (WRITE IN THE HIGHEST DOLLAR AMOUNT AT WHICH YOU WOULD STILL VOTE FOR THE PROGRAM) (p. 183)

Respondents are provided with a blank line where they can write in the maximum they would pay.[16]

In the early 1980s, Mitchell and Carson (1981) introduced the payment card (see also Mitchell and Carson 1993). This was a card with k bid amounts, and it showed respondents how much they pay for selected public services (anchors), which in essence is very general information on substitutes. Respondents were asked to "circle the dollar amount that is the most they would pay" for the change. Current applications of payment cards have proceeded without anchors (Fig. 4.1).

Dichotomous-choice questions, introduced by Bishop and Heberlein (1979), ask respondents, "would you pay $B" for the specified change, which is simply the first round in an iterative-bidding question (Fig. 4.2). The bid amount ($B) is varied over different respondents. The starting-point problems with iterative-bidding questions subsequently led to the adoption of dichotomous-response questions, and the single-shot question is easier to administer than the iterative framework. Some have also posited the heuristic argument that dichotomous-choice questions mimic the take-it-or-leave-it nature of many market purchases. Such a heuristic argument cannot be made for open-ended and payment-card questions.

---

[16]A well-defined contingent-valuation question would also tell the respondents the period over which payments would occur.

Would you vote for this proposal if the proposal would cost you $B **every year** for the foreseeable future? (CIRCLE ONE NUMBER)

1 Yes

2 No

Fig. 4.2 Example of a dichotomous-choice question (Welsh and Poe 1998, p. 183)

This example is a dichotomous-choice response format framed as a referendum. The bid amounts ($B) are entered before the survey is administered.[17] The referendum vote here is the decision rule. A dichotomous-choice question can be framed as a referendum or not. For example, a dichotomous-choice question can be framed as agreeing to pay or not pay an entrance fee to a national park.

A number of researchers have experimented with variations of the dichotomous-choice format. For example, studies have used one-and-one-half bounded questions (Cooper et al. 2002), double-bounded questions (Hanemann et al. 1991), and multiple-bounded questions (Bateman et al. 2001). Each of these questions present follow-up bids to survey respondents. For example, in the one-and-one-half bound, respondents randomly receive an initial bid. If they answer "yes" to the initial bid amount, they receive a higher bid; if they answer "no," they receive a lower bid amount. The multiple-bounded question is a repeated dichotomous choice where a response is required for every bid amount, which is essentially a payment card where respondents indicate their willingness to pay each bid amount, not just the maximum they would pay. These alternative specifications of dichotomous-choice questions were proposed to increase estimation efficiency (Hanemann et al. 1991). Responses to a dichotomous-choice question only reveal if each respondent's value is less than ("no" response) or greater than ("yes" response) the bid amount they received. Adding additional bid amounts reduces the range into which the unobserved values reside.

One of the NOAA Panel recommendations (Arrow et al. 1993) was to allow respondents a "no answer" option in addition to "yes" or "no" when the valuation question was framed as a referendum. This recommendation appeared to logically follow from consideration of undecided voters in predicting election outcomes. While there have been many different interpretations of how "don't know" responses should be elicited and treated in data analyses, at the basic level recommended by the NOAA Panel, it appears that most these respondents would vote "no" in the absence of a "don't know" option (Carson et al. 1998; Groothuis and Whitehead 2002).

---

[17]Welsh and Poe (1998) used nine bid amounts that ranged from $1 to $200.

Payment-card questions and dichotomous-choice questions require an additional design feature of selecting the bid amounts used as the monetary stimuli in the questions. Development of bid values usually follows a three-step process. The first step is to review similar studies in the literature to develop a prior on the distribution of values for the current application. Second, this prior information is used to develop the initial bid amounts used in pretesting the survey instrument, and these bid amounts are adjusted based on what is learned in this survey design process. This should include a field pretest, or pilot, of the full survey instrument and should not be limited to focus group data if possible. Finally, an optimal bid-design approach can be used to select the bids amounts used in the final survey instrument (Alberini 1995a; Dalmau-Matarrodona 2001; Kanninen 1993a, b; Scarpa and Bateman 2000). Alberini (1995a, b), and Kanninen (1993a, b, 1995) have shown that an optimal design has a small number of bids (five to eight), and the bid amounts should span the median WTP, and not placed too close to the median nor in the tails of the distribution. Very low or very high bid amounts may not be credible to respondents. Mis-specification of the bid distribution such that the most bid amounts fall above or below the median seriously compromises the ability to estimate mean WTP. McFadden (1994) proposed a continuous bid design, which might avoid mis-specification when only a small number of bids are employed. These bid-specification issues were empirically investigated by Boyle et al. (1998), with the empirical results supporting the bid-design recommendations of Kanninen (1993a, b), and Alberini (1995a).

The framing of the actual contingent-valuation questions and their respective response formats are quite simple relative to the framing of the valuation scenario that precedes the valuation question. The one exception is that careful selection of bid amounts is crucial in the design of payment-card and dichotomous-choice questions. In the next section, the relative strengths and weaknesses of these question formats are discussed.

4.1.6.2   Relative Strengths and Weaknesses of Response Formats

While dichotomous-choice questions are most commonly used, each of the three main response formats has strengths and weaknesses (Table 4.3). Conceptual arguments by Carson and Groves (2007), Carson et al. (2014), and Hoehn and Randall (1987) suggest that the "take-it-or-leave-it" nature of dichotomous-choice

Table 4.3 Comparison of contingent-valuation response formats

Characteristics	**Open-ended**	**Payment card**	**Dichotomous choice**
Incentive compatible	No	No	Yes
Bid design required	No	Yes	Yes
Responses/statistical efficiency	Continuous	Interval	Greater than or less than a threshold bid
Potential problems	Zero bids	Anchoring	Anchoring

questions, when framed as a referendum vote for a public good, has desirable properties for incentive-compatible revelation of preferences. There is a single bid amount to which respondents respond, and there is no incentive for respondents to pick very high (more than they would pay) or very low (less than they would pay) dollar amounts to purposely misstate their values. This is not the case for open-ended and payment-card questions where respondents can influence the outcome of a study by the value they state or dollar amount they pick. For example, if respondents want to see a change occur, they can state an open-ended value or pick a payment card amount that exceeds their WTP. Alternatively, if they want to send a signal that they want the cost to be low, they might select a value below what they would actually pay. The opportunities for such misstatements of value are not consistent with incentive comparability.

Cummings and Taylor (1998) argued that dichotomous-choice questions must be accompanied by the realism that the referendum vote will be binding (i.e., respondents must believe the change **will** be implemented if more than 50% of respondents vote "yes"). This concept has been more formally developed by Carson and Groves (2007) and Vossler et al. (2012). In contrast to Cummings and Taylor (1998), Carson et al. (2014) argued that it is not necessary that the referendum is binding but that the results of the survey referendum will have a nonzero probability of being used in the decision-making process to provide the item being valued. This provides a strong incentive for dichotomous-choice questions framed as a referendum as the preferred framing of a contingent-valuation question.

Responses to open-ended questions result in a continuous distribution of responses on the interval $[0, +\infty)$, while payment-card responses reveal whether the respondents' values reside within a $k + 1$ interval where k is the number of bid amounts ($B) on the payment card [$B_L \leq WTP \leq \$B_U$), where WTP is willingness to pay, B_L is the bid chosen by the respondent, and B_U is the next bid higher than the chosen bid. Responses to dichotomous-choice questions indicate only whether each respondent's values lie below [$0, $B) or above [$B, +\infty$) the bid threshold. Assuming truthful revelation of responses, a person with a value of $15 would respond in the following manner to each of the three basic contingent-valuation response formats:

– The response to an open-ended question would be "$15."
– The response to a payment-card question with bids of $1, $10, $20, and $30 would be "$10."
– The response to a dichotomous-choice question with a bid amount of $10 would be "yes."

Thus, the dichotomous-choice response format reveals if a respondent's value lies in the interval [$10, +\infty$); for the payment-card response format, the respondent's value resides in a narrower interval [$10, $20); and for an open-ended response format, the value of $15 is observed. Therefore, in terms of estimating

central tendency, the open-ended format provides the most efficient estimates, while the dichotomous-choice format provides the least efficient estimates ($se_{oe} <$ $se_{pc} < se_{dc}$, where se is the standard error of the estimated mean).[18] This relationship assumes that all three response formats incentivize respondents to truthfully reveal their preferences.

However, each of the response formats can have unique impacts on respondents' answers to a contingent-valuation question. For example, open-ended questions are believed to yield an unusually high percentage of responses of $0 in that some people might hold a value, but answer $0. It is also argued that people have difficulty coming up with a maximum willingness to pay amount for policies they are not familiar with. A manifestation of this issue is that the empirical distributions of responses to open-ended questions are not smooth and tend to have spikes at $5 increments. This rounding to the nearest $5 further attests to the difficulty respondents might have giving a precise dollar value. For examples of issues with open-ended questions, see Bohara et al. (1998), Donaldson et al. (1997) and Welsh and Poe (1998). All in all, very few applications use open-ended questions today.

Payment cards appear to avoid the issues of a spike of zero values and respondents having to provide a specific dollar value. However, Rowe et al. (1996) found that the bid amounts on the payment card can influence value responses. With careful framing, a payment card question can be posed in the context of a referendum and presented as consequential. Only a few studies still use payment cards (Covey et al. 2007; Ryan and Watson 2009), but despite this low usage, payment-card questions might be the best alternative to dichotomous-choice questions.

While dichotomous-choice questions gained popularity to avoid the anchoring in iterative-bidding questions, dichotomous-choice questions are not free from anchoring problems (Boyle et al. 1997, 1998; Green et al. 1998). That is, respondents have a propensity to say they would pay bid amounts that likely exceed their true values and to say they would not pay low bid amounts below their true values. The issue seems to be most problematic with high bids, which would serve to inflate value estimates (Boyle et al. 1998). Prices and quality are often perceived as being correlated in market goods, and this market intuition could lead respondents to interpret single bids as implicit signals of quality that lead to anchoring (Gabor and Granger 1966; Shapiro 1968).

Further, concerns about the effects of bid amounts on responses extend to one-and-one-half-bound, double-bound, and multiple-bound questions (Bateman et al. 2001, 2009; Herriges and Shogren 1996; Roach et al. 2002; Watson and Ryan 2007). Thus, while dichotomous-choice questions are theoretically incentive compatible, research suggests that value estimates might not be robust to manipulations in the bid design. An interesting question is whether a highly consequential survey would be less susceptible to bid effects.

---

[18]These denote the standard error (se) of the mean value estimated using the open-ended, payment-card and dichotomous-choice response formats.

Dichotomous-choice questions, posed as a referendum vote, are the safe approach to frame contingent-valuation questions, and the extensive use of this approach in the peer-reviewed literature supports this endorsement. However, the referendum framing of a dichotomous-choice question is not practical in some contexts such as recreation use values. Following are some additional considerations in the framing of contingent-valuation questions.

4.1.6.3   Allowing for Values of $0

Some people include the issue of zero bidders under the general heading of protest responses, but there are two issues here (Step 6.2). The first relates to people who give a response of $0 because they reject some component of the contingent-valuation scenario; these are protest responses that will be dealt with in the next section. This section considers those who truly hold values of $0. It is quite possible that a change might not be utility-increasing for some segment of the sampled population, and respondents need a way to indicate such a lack of value.

With an open-ended question, a respondent can simply enter a response of $0, and a payment card can include a value of $0 for respondents to circle. The more problematic case is a dichotomous-choice question where respondents can answer "no" to the bid but do not have the opportunity to express a value of $0. In these cases, we know only whether respondents' values lie within the interval $(-\infty, \$B)$. We do not know if there is a spike in the probability distribution at $0, and it is necessary to have a separate question to identify respondents whose values are $0.[19] This $0-value screen question has been implemented by posing it before the contingent-valuation question and then administering the valuation question only to those who answer "yes" to this screen. Alternatively, this question could probe respondents after they have answered the contingent-valuation question by asking respondents who answer "no" to the bid if they would "pay anything" for the change.

For example, Ahearn et al. (2003) used the following question that preceded the contingent-valuation question: "Would you vote for the proposal if passage of the proposal would increase your household's 1998 income tax?" Respondents who answered "no" were not asked the contingent-valuation question.

A related issue is that policies might actually give some people disutility, which would imply that their values would be strictly negative. While it appears that most studies treat people with negative values as $0s or that such outcomes are artifacts of the statistical distributions assumed in econometric estimation (Haab and McConnell 1997), others have attempted to investigate the plausibility of negative values (Berrens et al. 1998; Bohara et al. 2001).

---

[19]A similar question is required for a payment card if a bid amount of $0 is not included.

4.1.6.4 Protests and Other Types of Misleading Responses

There are at least three types of potential response categories under the heading of protests, all based on a presumption that these are respondents who do not report their true values (Step 6.3). It is important to note that these can be overtly misleading responses or misleading responses that occur inadvertently. Inadvertent misstatements can occur because someone does not fully understand the valuation scenario or because of experimentally induced errors.

The first category includes people who protest some component of the contingent-valuation scenario. These respondents might answer "$0" even though they hold a value for the item, which biases the estimate of central tendency downward, or they might choose not to complete the survey, leaving the effect on central tendency dependent on how these respondents are treated in the analysis of the contingent-valuation data.

The second category includes people who do not understand what they are being asked to value and who answer the valuation question anyway. The effect of this misunderstanding might not introduce a bias into estimates of central tendency, but it most likely will increase noise in the data that will increase the standard error of the mean.

The third category is people who behave strategically in an attempt to influence survey results and ultimately the decision. If everyone who is behaving strategically acts in a similar manner, the effect will be to introduce a bias into the estimate of central tendency. However, some people could have incentives to understate values, and others could have incentives to overstate the values, leaving the overall effect on estimates of central tendency indeterminate.

Within the contingent-valuation literature, two types of misleading responses have received specific attention: warm glow and social desirability. "Warm glow" arises from the utility that people receive from stating a willingness to pay and not for the change actually being valued (Andreoni 1989). Some have suggested that warm glow confounds estimation of WTP, but perhaps the effects can be removed from estimates (Nunes and Schokkaert 2003). However, the warm glow literature has largely been developed for philanthropy and donations (Harbaugh 1998), and the extension to contingent-valuation estimation of WTP is not fully explored. Donations are not a desirable payment vehicle for contingent-valuation questions because the goal is to estimate willingness to pay at the point of indifference, which is shown by the equality in Eqs. (4.1) and (4.2). This is what some people refer to as "maximum willingness to pay." It is not a donation toward provision of the item being valued, but the measurement of a specific economic concept.

Social desirability bias arises when respondents answer questions in a manner to please another person such as the interviewer in a personal interview. While some have detected social desirability (Leggett et al. 2003) in contingent-valuation estimates, it is likely that this effect is limited to interview formats where there is a clear incentive to please, which is not the general case for contingent-valuation studies. For example, the Leggett study was conducted in person on a boat as people were returning from visiting a military historical site in an area people visit because of the

military history. Even in cases where social desirability might arise, it can be addressed through survey implementation procedures (Kreuter et al. 2008).

In terms of identifying misleading responses, empirical applications have used a variety of approaches to identify anomalies in responses to contingent-valuation questions. Some have included questions in the survey to probe respondents' understanding and motivations when answering the contingent-valuation question (Ajzen et al. 1996; Berrens, Bohara et al. 1998; Blamey et al. 1999; Stevens et al. 1994). Others have trimmed the upper values if they are greater than a certain percentage (e.g., 10%) of a respondent's income (Mitchell and Carson 1989, pp. 226-227). Others have used statistical routines as described in Belsey et al. (1980) to identify responses that have undue influence on estimation results (Desvousges et al. 1987).

While most acknowledge that there are potentially some misleading responses in contingent-valuation data, there is no established procedure with a sound conceptual basis for excluding responses. The reasons for this are varied. What if a person gives one response that suggests a protest, but provides another response that indicates they are providing a valid response? Which response is correct or more meaningful? Or what constitutes a sufficient lack of understanding such that a respondent's valuation response should be excluded from statistical analyses? For example, it would not be appropriate to exclude respondents with low levels of education because they still have preferences and hold values. They might not be able to understand the valuation scenario as well as other respondents, but they make market decisions with a similar ability on a daily basis. Questioning people after they answered the valuation question is problematic; people who are behaving strategically would be unlikely to tell you that they are doing this. People who do not understand the valuation question also might not understand the follow-up question. In addition, these responses to follow-up questions cannot be assumed to be exogenous to responses to the valuation question.

Another approach, trimming the tails of the distribution by deleting outliers, must be done with care (e.g., people with high values might be those who have the most to lose). For example, some people give up income to live in areas that are near desirable resources; for these people, welfare losses could be quite large relative to income.

Another question deals with how much of an effect misleading responses actually have on estimates of central tendency. The famous Marwell and Ames (1981) study found that "economists free ride," but others do not. This suggests that strategic behavior might be relegated to a small segment of any sample. Thus, those who behave strategically or who more generally provide protest responses might be a small segment of the sampled population and might not behave in a way that influences sample statistics. The first contingent-valuation study that I conducted was undertaken with personal interviews of people while they were recreating on-site. An environmental group came through one of the survey locations; they talked among each other and encouraged each other to behave strategically by giving high-value responses to influence a desirable environmental outcome. We marked all of the surveys to identify these individuals in data analyses. Despite

behaving strategically, none of their responses were statistical outliers and most responses were quite close to the sample mean; their strategic behavior was not effective at manipulating sample statistics. Thus, while strategic behavior could occur, it is possible that it is not sufficiently pervasive or of a magnitude to have an effect on welfare estimates.

Recent studies have looked at segmenting true and protest zero responses. Jones et al. (2008), based on an open-ended question, segmented true and protest zero responses and then analyzed the data using a Tobit model after removing protest zero responses. Strazzera et al. (2003) provided a more sophisticated econometric approach that accounts for both true and protest zero responses without any ex ante segmenting. Meyerhoff and Liebe (2006) looked at protest more generally throughout their data and found that a variable that scales the level of protest response resulted in a higher level of protest, decreasing the likelihood that someone would hold a nonzero value and decreased estimated willingness to pay.

In contrast to treating protest responses as faulty data, Garcia-Llorente et al. (2011) suggested that protest responses can be useful in identifying specific design elements of environmental programs that will engender greater public support. This changes the perspective and dimensionality of a contingent-valuation study from simply estimating a value to support decision-making to providing richer information to decision-makers regarding program design.

Despite the issues discussed above, contingent-valuation study designs should consider including questions to differentiate true $0 responses from potential protest $0s, investigate the presence of data outliers, and include questions to probe respondent acceptance of the change valued, the provision mechanism, the payment vehicle, the decision rule, the time frame of payment, and the actual payment amount in payment-card and dichotomous-choice questions. While the issue of misleading responses to contingent-valuation questions deserves consideration, it is a tough conceptual and empirical issue that deserves greater consideration at a conceptual level of what actually constitutes a protest response such that the observation should not be included in data analyses to compute value estimates. Further, additional robustness studies are needed to determine if there is a systematic effect on estimates of central tendency and dispersion.

4.1.7 Develop Auxiliary Questions for Statistical Analyses of Contingent-Valuation Responses

Step 7 calls for development of auxiliary questions designed to collect data to be used in the analyses of responses to the contingent-valuation questions, and auxiliary questions can provide information to support decision-making beyond what is provided by the valuation component of the study. These are in addition to probing for potential protest responses, which was discussed in the preceding section. The

most obvious candidates are income (Schläpfer 2006) and other variables that might influence value estimates (e.g., Dupont 2004; Farreras et al. 2005).

Questions such as income, sex, and other demographic questions that have objective responses are commonly placed at the end of a questionnaire following standard survey practice. Opinion questions, whose responses are to be used in analyzing responses to the valuation question, might include a question asking respondents to rate the current condition of the item being valued. Such a question should be placed before the valuation question(s) in a survey. If such questions are placed after respondents have answered the valuation question, their responses cannot be assumed to be exogenous to responses to the valuation question. Data on these types of variables are commonly used in the estimation of econometric models of payment-card and dichotomous-choice data.

Secondary data may also be incorporated in econometric analyses. For example, in a study of lost value from contamination of an aquifer, Poe and Bishop (1999) argued that well-specific water quality data are needed in the valuation question. Spatial data can also be useful to develop variables that describe proximity of respondents' households to wells with known levels of contamination or proximity to the source of the contamination. It is important to have questions in the survey that will help in matching the respondent's answers to valuation questions to the auxiliary data that will be merged with the survey response data.

Finally, it is important to consider whether existing surveys (e.g., the U.S. Census, NORC General Social Survey, etc.) have similar questions. Using the same framing of questions as existing surveys allows for direct comparisons of the data for assessing whether sample selection has occurred, merging data sets for richer statistical analyses, and filling in missing data due to item nonresponse. Cameron et al. (1999) used such an approach to address nonresponse bias in a mail survey.

4.1.8 Pretest and Implement the Survey

Chapter 3 provides details of survey design and administration (Step 8). This section will briefly discuss the use of cognitive interviews, focus groups, and pilot surveys in the design of a contingent-valuation survey.

Surveys are pretested through an iterative process of one-on-one interviews, focus groups, and/or field trials (pilot surveys). Here again, there is no one-size-fits-all approach. Some studies might use one focus group, while other studies use multiple focus groups. Some studies might use focus groups and cognitive interviews, while other studies only use focus groups. These choices can depend on the available budget and the complexity of the valuation topic. For studies with a large budget and a complex valuation topic, the design process could be the following:

- Focus groups to learn about potential respondent's knowledge and beliefs about the item being valued.
- Focus groups to test the valuation scenario.
- Cognitive interviews to learn what people think about the valuation scenario in the absence of group effects and to learn more deeply about how potential respondents are reacting to the survey information.
- Focus groups to pretest a complete draft of the survey instrument.
- A field pilot to develop information on the survey response rate and item responses to individual questions in the survey with particular interest in responses to the valuation question.

How qualitative research tools and field pilots are used in the survey design process varies from study to study, but there is a common feature. The design process typically starts with a simple structure to learn from potential respondents, and the complexity of the survey instrument is built through what is learned in each focus group, cognitive interview, and pilot. While this qualitative research is common to survey design, the unique component here is the design of the contingent-valuation scenario.

The give and take of the focus group process can be useful by learning what is important to one person in terms of information about the item being valued is not very important to the rest of the group. Learning how different people in the focus groups respond to different features of the valuation scenario is a crucial feature of this survey design research. Cognitive interviews allow for more in-depth probing of potential respondents' reactions to the surveys materials in the absence of the group dynamics. The pilot survey is particularly useful for developing information on potential values to be used to develop bids for payment-card and dichotomous-choice questions.

This qualitative pretesting of the survey instrument and administration process ensures that survey questions are understandable to respondents and are actually eliciting the information they are designed to elicit, and that the survey will yield adequate responses to support statistical analyses of the resulting data.

Following the pretesting, the survey should be implemented using best practices for survey administration as discussed in Chap. 3.

4.1.9 Data Analysis

The process of data analysis (Step 9) varies with the response format used for the contingent-valuation question. We start with responses to open-ended questions, move to payment-card responses, and close with dichotomous-choice responses, thereby covering the three main types of contingent-valuation response formats.

Responses to open-ended questions are mathematically the easiest to analyze in terms of computing the arithmetic mean:

$$\overline{\text{WTP}} = \sum_{i=1}^{N} \frac{\text{WTP}_i}{N}, \tag{4.4}$$

where WTP_i is the open-ended response for the ith respondent, and N is the number of observations (complete surveys). The responses to an open-ended question (WTP_i) are individual statements of value, WTP or op in Eqs. (4.1) and (4.2), respectively. If open-ended responses are analyzed as a function of variables that explain WTP, a theoretical specification would be based on a definition of the value as in Eqs. (4.1) and (4.2), which would be solved for WTP or op as the dependent variable.

Analyses of payment-card and dichotomous-choice data require econometric analyses and the estimated equations are used to derive estimates of WTP. These analyses start with the specification of a function based on theoretical definitions of value, like the examples presented in Eqs. (4.1) and (4.2). Willingness to pay can be expressed as

$$\log(\text{WTP}_i) = x_i'\alpha + e_i, \tag{4.5}$$

where x_i' is a vector of arguments that affect the magnitude of individual's WTP, α is a vector of preference coefficients to be estimates, and e_i is the random error that might be assumed to be distributed normally with mean zero and standard deviation σ. The function $x'\alpha$ is specified as the solution to equations such as (4.1) or (4.2) that define the value being estimated. The vector x_i logically includes variables that describe the change valued. As with most econometric analyses, the explanatory variables are often chosen based on economic theory and previous research, and are hypothesized to affect the magnitudes of respondents' WTP.

Analysis of payment-card data proceeds by modeling the interval where respondents have revealed that their values reside. These intervals are bounded by the bid amounts each respondent circled and the next highest amount on the payment card. Following Cameron and Huppert (1989), respondents' true values $\left(\text{WTP}_i^t\right)$ reside in the interval ($\$B_{1i}$, $\$B_{ui}$], where "1" denotes the lower bid the respondent circled and "u" denotes the next highest bid amount on the payment card.[20]coefficients. Since WTP_i is not actually observed, the probability that WTP_i falls into the chosen interval on the payment card is modeled as

$$\Pr(\text{WTP}_i \in (\$B_{1i}, \$B_{ui}]) = \Pr\left(\frac{\log \$B_{1i} - x_i'\alpha}{\sigma} < t_i < \frac{\log \$B_{ui} - x_i'\alpha}{\sigma}\right), \tag{4.6}$$

where t_i is a standard normal variable. Using the estimates coefficients ($\hat{\alpha}$), willingness to pay can be derived as[21]

$$E(\log(\text{WTP})) = x'\hat{\alpha}, \tag{4.7a}$$

---

[20]True values are as theoretically defined in the examples provided in Eqs. (4.1) and (4.2).

[21]Wooldridge (2012, pp. 212-213) noted that the log-normal estimator is consistent but could be biased.

or

$$E(\text{WTP}) = \exp(x'\hat{\alpha})\exp\left(\frac{\hat{\sigma}^2}{2}\right). \tag{4.7b}$$

The point estimate of WTP that results from Eq. (4.7b) is a function of chosen levels of the x vector, and several computation approaches are generally followed. The first is to insert mean values for each element of x to predict WTP. The second is to insert the unique values for the elements of x for each person in the sample, predict the individual-specific WTP, and then compute mean WTP from the predictions. To compute WTP for a specific change in quality (or quantity), the variable(s) in x representing the change need to be set to the appropriate level(s).

Again, following Cameron and Huppert (1989)[22] and continuing with the notation established for the analysis of payment-card data, analysis of dichotomous-choice data proceeds as follows:

$$\begin{aligned}
\Pr(\text{yes}_i) &= \Pr(\log(\text{WTP}_i) > \log \$\text{B}_i) \\
&= \Pr\left(\frac{e_i}{\sigma} > \frac{\log \$\text{B}_i - x'\alpha}{\sigma}\right) \\
&= 1 - \emptyset\left(\frac{\log \$\text{B}_i - x'\alpha}{\sigma}\right).
\end{aligned} \tag{4.8}$$

The computation of mean and median values proceeds as described above for payment-card data.

The issue of a spike in the probability distribution at \$0 for people who do not value the change has received limited attention in the literature. In order to address zero values in the analysis of contingent-valuation data, this issue needs to be considered early in the study design to identify people who hold values of \$0. With this information, open-ended responses can be analyzed using a Tobit model (Jones et al. 2008; Strazzera et al. 2003), and alternative specifications have been considered for payment-card and dichotomous-choice data (Kriström 1997; Bohara et al. 2001).

In addition to the parametric approach described above, semiparametric and nonparametric approaches have been used to analyze dichotomous-choice data (Araña and León 2005; Creel and Loomis 1997; Crooker and Herriges 2004; Fernández et al. 2004; Haab and McConnell 1997; Kriström 1990; Li 1996). A nonparametric approach reduces the distribution and functional form assumptions that must be imposed on the data when estimating WTP. A lower-bound nonparametric estimator is often used to compute WTP and this estimator is specified as

---

[22]Hanemann (1984) proposed an alternative random-utility approach to analyzing dichotomous-choice data, and McConnell (1990) showed the Hanemann (1984), and Cameron and Huppert (1989) approaches to be the duals of each other.

$$\overline{\text{WTP}_l} = \sum_{i=1}^{k} (\$B_i - \$B_{i-1}) * p(\$B_i), \qquad (4.9)$$

where $\overline{\text{WTP}_l}$ is a lower-bound estimator of WTP, k is the number of bids used in the study design, and $p(\bullet)$ is the percentage of respondents that answered "yes" to bid amount $\$B_i$. Here it is assumed that $\$B_{i-1} = 0$ for $i = 1$.[23] This is a lower-bound estimator because anyone who might hold a value greater than the highest bid amount is assigned a value of \$0, and the percentage of people who answered "yes" in each bid interval are assigned the lower percentage of the higher bid. For discussion of the econometric analysis of this nonparametric lower-bound estimator, see Lewbel et al. (2011) and Wantanabe (2010).

It is also important to consider the variability of estimated WTP for studies to support decision-making and for experiments. Variance can be a function of heterogeneity in values across the sample and can be affected by the study design. That is, not everyone in a sample holds the same value, and larger within-sample variability will lead to larger standard errors of the estimated mean. In addition, the quality of the contingent-valuation scenario (and survey) can influence the estimated standard error; a poorer scenario design might result in a larger standard error of estimated WTP than a better scenario design. This can arise because there is less or greater understanding of the description of the item being valued. Considering variability in estimated willingness to pay is important to support decision-making and for experiments. Larger variances can lead to failure to reject the null hypotheses when the hypotheses should be rejected in statistical tests, and they can reduce the credibility of WTP estimates to support decision-making. Thus, it is important to compute the standard error of estimated WTP and consider design approaches that minimize undue variability in valuation responses through careful pretesting of valuation scenarios.

A number of approaches have been used to develop confidence intervals for WTP estimates (Cooper 1994). A quick review of recent articles suggests that two approaches are used the most commonly. The first is the well-known Krinsky-Robb approach that Park et al. (1991) introduced to the nonmarket valuation literature. The second is the convolutions approach introduced by Poe et al. (1994). See Cooper (1994), and Poe et al. (2005) for comparisons of the alternative approaches.

There are other approaches that have been used to analyze payment-card and dichotomous-choice response data, and I have simply presented the approachs here for illustrative purposes. Readers seeking more information on the econometric analysis of contingent-valuation data should see Haab and McConnell (2002).

---

[23]Haab and McConnell (1997) did not discuss this normalization in their presentation of the nonparametric estimator.

4.1.10 Report Study

The singular most important element in reporting a contingent-valuation study that supports decision-making is to present an aggregate welfare estimate (aggWTP). This calculation, in simplest terms, is

$$\text{aggWTP} = \overline{\text{WTP}} * N, \tag{4.10}$$

where $\overline{\text{WTP}}$ is the estimated mean willingness to pay from the sample and N is the size of the affected population. But this calculation is not simple at all, and a number of questions need to be addressed:

- Will mean or median willingness to pay be used?
- Will weighting be used to bring respondent sample characteristics in line with the affected population characteristics?
- Should individuals who are affected by the change valued and eligible to participate in the study, but are precluded for some reason (e.g., a gated community not allowing access to interviewers) be included in N?
- How should people who resfuse to complete the survey or answer the valuation question be treated?

There are a variety of other questions in this line that might need to be addressed, and addressing these issues is critically important because they can substantially affect the magnitude of the aggregate value estimate.

Reporting of study results (Step 10) requires an understanding that the findings serve multiple purposes—the current decision analysis, transfers to new decision applications at the same application site or to an entirely new application, and advancing the literature by helping future investigations understand the study design, estimation, and testing. Even methodological studies, while not designed to support decision-making, could still be used in benefit transfers (Chap. 11). Thus, clear and detailed study documentation within the limits of the publication outlet is crucial for study credibility and advancing the contingent valuation literature. The steps (Table 4.1) in this chapter provide a guide to assist in this documentation, such as

- The study application.
- The theoretical definition of the estimated value that includes the current condition or baseline and the change valued.
- The steps in the design process and discussions of why items in the final design worked well, what design items were tried and discarded, and why they were discarded.
- The sample frame.

- Survey mode and response rates reported according to the American Association for Public Opinion Research guidelines (www.aapor.org/Standards-Ethics/Best-Practices.aspx).
- The verbatim commodity description and valuation scenario from the survey.
- The contingent-valuation format used, including the question wording.
- Respondents' demographic characteristics, and use or preferences for the item valued.
- Methods of data analysis, including treatment of $0 values and protest responses, and econometric analysis.
- Estimates of central tendency and dispersion, methods used to calculate these sample statistics, and any robustness checks.

This information allows readers to evaluate the content validity of value estimates for the current application and to evaluate the transferability of value estimates to new applications. With limitations on what can be reported in journal articles, best practices should include providing online appendices that fully document studies.

4.2 Reliability and Validity

A reliability study investigates the variance in value estimates. Tests of validity ask whether a contingent-valuation study accurately measures the value concept it is designed to estimate (see Chap. 12). The consideration of reliability and validity collectively constitute what is referred to as a credible or accurate contingent-valuation study.

4.2.1 Reliability

Reliability investigates the variance in value estimates. The common investigation approach is test–retest reliability, where a contingent-valuation survey is repeated at two different points in time. This can be a replication with the same respondents or a between-subjects design where the two samples are drawn identically from the same sample frame.

The consensus in the literature appears to support a conclusion that contingent-valuation estimates are reliable (Brown et al. 2008; Carson et al. 1997; Kealy et al. 1988, 1990; Loomis 1989, 1990; Onwujekwe and Fox-Rushby 2005; Reiling et al. 1990; Stevens et al. 1994; Teisl et al. 1995). Thus, reliability of contingent-valuation estimates is not an issue of concern.

It is also important to recognize that values can and should change over time. Thus, failure to establish statistical equivalence in values over time does not refute reliability in instances where values have legitimately changed.

4.2.2 Validity

Three types of validity are commonly investigated: content, construct and criterion (Carmines and Zeller 1979). Criterion validity compares contingent-valuation estimates to a measurement that is external to the contingent-valuation study and is a presumed measure of the true value. The cash transactions in the Bishop and Heberlein (1979) study provided such a criterion upon which the parallel contingent-valuation estimate is validated. Convergent validity, which is a specific type of construct validity, investigates the consistency of contingent-valuation estimates with estimates provided by another nonmarket valuation method. This is what was done when Bishop and Heberlein compared the contingent-valuation estimate with those derived from a travel-cost model. Content validity asks whether the elements in the design of the contingent-valuation survey and data analyses are consistent with economic theory, established practice, and the valuation objective.

Validity assessments are only as good as the counterfactual that they are compared against. The criterion must be a true representation of the economic construct being valued. Convergent validity only establishes whether two estimation approaches are comparable; comparability can hold when both approaches have the same bias. Or, failure to establish convergent validity only implies that one or both estimates are biased. Two types of criterion validity studies are discussed: comparisons with cash experiments and comparisons with outcomes of actual referendum votes. Content validity requires a collective protocol for what constitutes best practices; the material provided in this chapter is one such documentation of best practices.

Researchers have conducted meta-analyses of studies using experiments to investigate differences in outcomes between contingent-valuation (or stated preference) treatments and control treatments where a simulated market is used that involves cash transactions (List and Gallet 2001; Little and Berrens 2004; Murphy et al. 2005). List and Gallet reviewed 29 studies that provided 174 validity comparisons.[24] Note, the comparisons are not all based on contingent-valuation studies, but experiments where stated preference value estimates are compared to "parallel" values estimated using cash transactions, and the comparisons may be for private goods and not public good applications. List and Gallet report a calibration factor for each study, which is the stated-preference value divided by the cash value. They conclude that the "calibration factors for the most prevalent type of study are 1.26 (minimum), 1.28 (median), and 1.30 (maximum), which suggests that the most common type of (WTP) study will tend to produce a slightly (upward) biased estimate of the actual value" (p. 250). This difference (calibration factor) is what is known as hypothetical bias. However, the credence of this conclusion crucially depends on two features: first, the experiment is similar to what would be conducted

---

[24]Little and Berrens (2004), and Murphy et al. (2005) analyzed essentially the same data. Little and Berrens only considered the percentage choosing to buy and not the welfare estimates, and Murphy et al. is a reanalysis of List and Gallet (2001).

in a contingent-valuation experiment, and second, the cash value measures the same underlying Hicksian concept of value as contingent valuation.

Some researchers have compared contingent-valuation study results to those from a parallel referendum vote. While many have asked how contingent-valuation estimates would compare to values if a market existed, comparisons to referendum votes make logical sense. Many contingent-valuation applications value public goods where a majority vote would be a logical decision rule. Further, the preferred question framing is as a dichotomous choice on a referendum vote. Vossler and Kerkvliet (2003) found that "survey responses match the actual voting outcome and WTP estimates based on the two (survey and voting) are not statistically different" (p. 631). Johnston (2006) similarly found no statistical difference between the proportion of "yes" votes in a stated preference study and an actual referendum at the expected program cost per household. Vossler et al. (2003) found a similar result when undecided votes were excluded or treated as "no" responses in the contingent-valuation portion of the experiment. These results indicate that responses to contingent-valuation questions, framed as votes to a referendum, mimic how people vote when they go to polls to vote on environmental issues. While there have been far fewer of these studies than of the comparisons with cash transactions discussed above, these studies indicate that comparing contingent-valuation outcomes with referendum votes is an important and promising line of inquiry.[25]

Finally, the discussion of validity has been extended to investigations of whether contingent-valuation results follow an expected pattern, which has been termed a "test of scope." A test of scope asks the simple question, "Are people willing to pay more for a larger change in the item being valued than for a smaller change?" but the conduct of these tests is varied (Desvousges et al. 1993). A meta-analysis of scope experiments suggests that contingent-valuation studies do pass a scope test (Ojea and Loureiro 2011).

Some, however, have argued that a scope test should address more than the simple question of whether more is valued more highly than less of the item being valued (e.g., the adding-up test; Diamond 1996; Desvousges et al. 2012). However, for these additional tests to be applied, assumptions about the structure of preferences must be imposed (Haab et al. 2013). Thus, if an empirical test fails, it is impossible to differentiate failure due to a problem in the structure of a contingent-valuation study or failure due to the wrong assumptions about preferences being imposed. Thus, while a scope test can provide insight about the credibility of a contingent-valuation estimate of value, it is a weak test of validity at best (see Heberlein et al. 2005).

More will be said about these points and other considerations of validity in Chap. 12.

---

[25]Schläpfer et al. (2004) also compared contingent-valuation results with voting behavior and found differences that indicate contingent valuation leads to overestimates of value, but confounds in this study limit the generalizability of the result.

4.2.3 Enhancing Validity

Based on the presumption that contingent valuation leads to overestimates of what people would actually pay (List and Gallet 2001), a number of protocols have been proposed to remove overestimation. These include the lower-bound nonparametric estimator of WTP discussed above (Haab and McConnell 1997), insertion of cheap talk into a survey instrument (Cummings and Taylor 1999), use of self-reported uncertainty scales after respondents have answered the contingent-valuation question (Champ et al. 1997), and consequentiality (Carson and Groves 2007). The latter of these approaches holds the most promise.

Each of these approaches leaves questions to be answered. Is it desirable to provide a lower-bound estimate of WTP that might actually underestimate true willingness to pay? Does cheap talk provide the correct incentive to provide truthful answers to a contingent-valuation question or just challenge people to lower their value responses? What is the basis for the correct interpretation of the uncertainty scale, and does it vary over specific valuation applications? Consequentiality, which requires a binding payment and a non-zero probability that the value estimates will inflence decision making, holds the most promise (Carson et al. 2014; Vossler et al. 2012).

4.3 Conclusions

It is important to recognize that the design of a contingent-valuation study involves constraints and trade-offs. The constraints could come from the actual change being valued, where the realism challenges the design (e.g., a tax will be used to fund the provision of the item to be valued, but pretesting reveals people value the item and oppose a tax increase). Trade-offs could arise in an attempt to develop a clear scenario while also seeking brevity. These decisions are part of the "art" of contingent valuation and have provided opportunities for critics to question the credibility of contingent-valuation estimates. This criticism is not all bad as it has led to better-designed studies and more-focused validity research. The outcome has been improved contingent-valuation studies that are in common use to support public and private decision-making.

It is important also to realize that the influence the survey design has over the outcome of any contingent-valuation study is **not** any different from any other line of research or empirical analysis. Simply put, contingent valuation—like any other empirical method—requires a high degree of skill and considerable art that must be combined with careful pretesting and validity checks.

Finally, the next chapter (Chap. 5) discusses choice experiments, which are a close cousin to contingent valuation in the family of stated preference valuation methods. There are some unique differences between these valuation approaches:

- Contingent-valuation scenarios define the change to be valued in a written scenario, while choice-experiment scenarios define the change using specific levels of attributes.
- Contingent valuation, using a dichotomous-choice question, asks respondents to choose between the change and a status quo condition, while choice experiments typically ask respondents to choose between two or more changes (alternatives) and a status quo condition.
- Contingent-valuation studies often include one valuation question, while choice experiments typically include multiple valuation questions.

The key distinction between contingent valuation and choice experiments is the scenario presentation; in a choice experiment respondents are presented with alternatives to choose among (usually three or more) that are described in terms of attributes (again, usually three or more). Whether contingent valuation or a choice experiment should be employed is not clear-cut. However, if multiple alternatives are not plausible and the item being valued cannot be logically divided into attributes from a design, policy, or consumer-preference perspective, contingent valuation is the appropriate choice. If multiple alternatives and attributes are plausible, the policy question seeks marginal values for individual attributes and such segmentation into attributes is plausible to respondents, a choice experiment would be the appropriate method. A choice experiment with one alternative, two attributes (including cost), and one valuation question only varies from a dichotomous-choice question in the presentation of the change to be valued, and they are conceptually and analytically equivalent. Thus, many of the design insights discussed in this chapter also relate to choice experiments and vice versa for much of the design feature discussion in Chap. 5.

References

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.

Ahearn, M. C., Boyle, K. J. & Hellerstein, D. R. (2003). Designing a contingent-valuation study to estimate the benefits of the conservation reserve program on grassland bird populations. In J. Kahn, D. Bjornstad & A. Alberini (Eds.), The contingent-valuation handbook. Cheltenham, United Kingdom: Edward Elgar.

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

Alberini, A. (1995a). Optimal designs for discrete choice contingent valuation surveys: Single-bound, double-bound, and bivariate models. Journal of Environmental Economics and Management, 28, 287-306.

Alberini, A. (1995b). Willingness-to-pay models of discrete choice contingent valuation survey data. Land Economics, 71, 83-95.

Alberini, A., Boyle, K. & Welsh, M. (2003). Analysis of contingent valuation data with multiple bids and response options allowing respondents to express uncertainty. Journal of Environmental Economics and Management, 45, 40-62.

Alberini, A., Rosato, P., Longo, A. & Zanatta, V. (2005). Information and willingness to pay in a contingent valuation study: The value of S. Erasmo in the lagoon of Venice. Journal of Environmental Planning and Management, 48, 155-175.

Andreoni, J. (1989). Giving with impure altruism: Applications to charity and Richardian equivalence. Journal of Political Economy, 97, 1447-1458.

Araña, J. E. & León, C. J. (2005). Flexible mixture distribution modeling of dichotomous choice contingent valuation with heterogeneity. Journal of Environmental Economics and Management, 50, 170-188.

Arrow, K., Solow, R., Portney, P. R., Leamer, E. E., Radner, R. & Schuman, H. (1993). Natural resource damage assessments under the Oil Pollution Act of 1990. Federal Register, 58, 4601-4614.

Bateman, I. J., Day, B. H., Dupont, D. P. & Georgiou, S. (2009). Procedural invariance testing of the one-and-one-half-bound dichotomous choice elicitation method. The Review of Economics and Statistics, 91, 806-820.

Bateman, I. J., Day, B. H., Georgiou, S. & Lake, I. (2006). The aggregation of environmental benefit values: Welfare measures, distance decay and total WTP. Ecological Economics, 60, 450-460.

Bateman, I. J., Jones, A. P., Lovett, A. A., Lake, I. R. & Day, B. H. (2002). Applying geographic information systems (GIS) to environmental and resource economics. Environmental and Resource Economics, 22, 219-269.

Bateman, I. J., Langford, I. H., Jones, A. P. & Kerr, G. N. (2001). Bound and path effects in double and triple bounded dichotomous choice contingent valuation. Resource and Energy Economics, 23, 191-213.

Bateman, I. J., Langford, I. H., Turner, R. K., Willis, K. G. & Garrod, G. D. (1995). Elicitation and truncation effects in contingent valuation studies. Ecological Economics, 12, 161-180.

Bateman, I. J. & Munro, A. (2009). Household versus individual valuation: What's the difference? Environmental and Resource Economics, 43, 119-135.

Becker, G. S. (1981). Treatise on the family. Cambridge, MA: Howard University Press.

Belsey, D. A., Kuh, E. & Welsch, R. E. (1980). Regression diagnostics. New York: Wiley.

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

Berrens, R. P., Bohara, A. K., Jenkins-Smith, H. C., Silva, C. L. & Weimer, D. L. (2004). Information and effort in contingent valuation surveys: Application to global climate change using national internet samples. Journal of Environmental Economics and Management, 47, 331-363.

Berrens, R. P., Bohara, A. K., Jenkins-Smith, H., Silva, C. L., Ganderton, P. & Brookshire, D. (1998). Joint investigation of public support and public values: Case of instream flows in New Mexico. Ecological Economics, 27, 189-203.

Berrens, R. P., Brookshire, D., Ganderton, P. & McKee, M. (1998). Exploring nonmarket values for the social impacts of environmental policy change. Resource and Energy Economics, 20, 117-137.

Bishop, R. C. (1982). Option value: An exposition and extension. Land Economics, 58, 1-15.

Bishop, R. C. & Heberlein, T. A. (1979). Measuring values of extra-market goods: Are indirect measures biased? American Journal of Agricultural Economics, 61, 926-930.

Blamey, R. K., Bennett, J. W. & Morrison, M. D. (1999). Yea-saying in contingent valuation surveys. Land Economics, 75, 126-141.

Bohara, A. K., Kerkvliet, J. & Berrens, R. P. (2001). Addressing negative willingness to pay in dichotomous choice contingent valuation: A Monte Carlo simulation. Environmental and Resource Economics, 20, 173-195.

Bohara, A. K., McKee, M., Berrens, R. P., Jenkins-Smith, H., Silva, C. L. & Brookshire, D. S. (1998). Effect of total cost and group-size information on willingness to pay responses: Open ended vs. dichotomous choice. Journal of Environmental Economics and Management, 35, 142-163.

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

Boyle, K. J. & Bishop, R. C. (1988). Welfare measurements using contingent valuation: A comparison of techniques. American Journal of Agricultural Economics, 70, 20-28.

Boyle, K. J., Bishop, R. C. & Welsh, M. P. (1985). Starting point bias in contingent valuation bidding games. Land Economics, 61,188-196.

Boyle, K. J., Desvousges, W. H., Johnson, F. R., Dunford, R. W. & Hudson, S. P. (1994). An investigation of part-whole biases in contingent-valuation studies. Journal of Environmental Economics and Management, 27, 64-83.

Boyle, K. J., Johnson, F. R. & McCollum, D. W. (1997). Anchoring and adjustment in single-bounded, contingent-valuation questions. American Journal of Agricultural Economics, 79, 1495-1500.

Boyle, K. J., MacDonald, H. F., Cheng, H. & McCollum, D. W. (1998). Bid design and yea saying in single-bounded, dichotomous-choice questions. Land Economics, 74, 49-64.

Boyle, K. J., Morrison, M., MacDonald, D. H., Duncan, R. & Rose, J. (2016). Investigating Internet and mail implementation of stated-preference surveys while controlling for differences in sample frames. Environmental and Resource Economics, 64, 401–419.

Brown, T. C., Kingsley, D., Peterson, G. L., Flores, N. E., Clarke, A. & Birjulin, A. (2008). Reliability of individual valuations of public environmental goods: Choice consistency, response time, and preference refinement. Journal of Public Economics, 92, 1595-1606.

Cameron, T. A. & Huppert, D. D. (1989). OLS versus ML estimation of non-market resource values with payment card interval data. Journal of Environmental Economics and Management, 17, 230-246.

Cameron, T. A., Shaw, W. D. & Ragland, S. R. (1999). Nonresponse bias in Maine survey data: Salience vs. endogenous survey complexity. In J. A. Herriges & K. L. Kling (Eds.), Valuing recreation and the environment: Revealed preference methods in theory and practice (pp. 217–251). Cheltenham, United Kingdom: Edward Elgar.

Carmines, E. G. & Zeller, R. A. (1979). Reliability and validity assessment. Beverly Hills, CA: Sage.

Campos, P., Caparrós, A. & Oviedo, J. L. (2007). Comparing payment-vehicle effects in contingent valuation studies for recreational use in two protected Spanish forests. Journal of Leisure Research, 39, 60-85.

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

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.

Carson, R. T., Hanemann, W. M., Kopp, R., Krosnick, J. A., Mitchell, R. C., Presser, S., Ruud, P. A. & Smith, V. K. (1998). Referendum design and contingent valuation: The NOAA panel's no-vote recommendation. Review of Economics and Statistics, 80 (3), 484-487.

Carson, R. T., Hanemann, W. M., Kopp, R., Krosnick, J. A., Mitchell, R. C., Presser, S., Ruud, P. A., Smith, V. K., Conaway, M. & Martin, K. (1997). Temporal reliability of estimates from contingent valuation. Land Economics, 73, 151-163.

Champ, P. A., Bishop, R. C., Brown, T. C. & McCollum, D. W. (1997). Using donation mechanisms to value nonuse benefits from public goods. Journal of Environmental Economics and Management, 33, 151-162.

Chiappori, P. A. & Ekeland, I. (2009). The microeconomics of efficient group behavior: Identification. Econometrica 67, 763-799.

Cicchetti, C. J. & Smith, V. K. (1973). Congestion, quality deterioration, and optimal use: Wilderness recreation in the Spanish peaks primitive area. Social Science Research, 2, 15-30.

Cooper, J. C. (1994). A comparison of approaches to calculating confidence intervals for benefit measures from dichotomous-choice contingent valuation surveys. Land Economics, 70, 111-122.

Cooper, J. C., Hanemann, M. & Signorello, G. (2002). One-and-one-half-bound dichotomous choice contingent valuation. The Review of Economics and Statistics, 84, 742-750.

Covey, J., Loomes, G. & Bateman, I. J. (2007). Valuing risk reductions: Testing for range biases in payment card and random card sorting methods. Journal of Environmental Planning and Management, 50, 467-482.

Creel, M. & Loomis, J. (1997). Semi-nonparametric distribution free dichotomous choice CV. Journal of Environmental Economics and Management, 32, 341-358.

Crooker, J. R. & Herriges, J. A. (2004). Parametric and semi-nonparametric estimation of willingness-to-pay in the dichotomous choice contingent valuation framework. Environmental and Resource Economics, 27, 451-480.

Cummings, R. G., Brookshire, D. S. & Schulze, W. D. (Eds.). (1986). Valuing environmental goods: An assessment of the contingent valuation method. Totowa, NJ: Rowman & Allanheld.

Cummings, R. G. & Taylor, L. O. (1998). Does realism matter in contingent valuation surveys? Land Economics, 74, 203-215.

Cummings, R. G. & Taylor, L. O. (1999). Unbiased value estimates for environmental goods: A cheap talk design for the contingent valuation method. American Economic Review, 89, 649-665.

Dalmau-Matarrodona, E. (2001). Alternative approaches to obtain optimal bid values in contingent valuation studies and to model protest zeros: Estimating the determinants of individuals' willingness to pay for home care services in day case surgery. Health Economics, 10, 101-118.

Davis, R. K. (1963). Recreation planning as an economic problem. Natural Resources Journal, 3, 239-249.

Davis, D. D. & Holt, C. A. (1993). Experimental economics. Princeton, NJ: Princeton University Press.

Desaigues, B., Ami, D., Bartczak, A., Braun-Kohlová, M., Chilton, S., Czajkowski, M., Farreras, V., Hunt, A., Hutchinson, M., Jeanrenaud, C., Kaderjak, P., Máca, V., Markiewicz, O., Markowska, A., Metcalf, H., Navrud, S., Nielsen, J. S., Ortiz, R., Pellegrini, S., Rabl, A., Riera, R., Scasny, M., Stoedckel, M. E., Szánto, R. & Urban, J. (2011). Economics valuation of air pollution mortality: A 9-country contingent valuation survey of value of a life year (VOLY). Ecological Indicators, 11, 902-910.

Desvousges, W. H., Johnson, F. R., Dunford, R. W., Hudson, S. P., Wilson, K. N. & Boyle, K. J. (1993). Measuring natural resource damages with contingent valuation. In J. A. Hausman (Ed.), Contingent valuation: A critical assessment (Contributions to economic analysis, vol. 220, pp. 91–164). Amsterdam, The Netherlands: Emerald Group.

Desvousges, W. H., Mathews, H. K. & Train, K. (2012). Adequate responsiveness to scope in contingent valuation. Ecological Economics, 84, 121-128.

Desvousges, W. H., Smith, V. K. & Fisher, A. (1987). Option price estimates for water quality improvements: A contingent valuation study for the Monongahela River. Journal of Environmental Economics and Management, 14, 248-267.

Diamond, P. (1996). Testing the internal consistency of contingent valuation surveys. Journal of Environmental Economics and Management, 30, 337-347.

Dillman, D. A. (2007). Mail and Internet surveys: The tailored design method. Hoboken, NJ: John Wiley & Sons.

Donaldson, C., Thomas, R. & Torgerson, D. J. (1997). Validity of open-ended and payment scale approaches to eliciting willingness to pay. Applied Economics, 29, 79-84.

Dupont, D. P. (2004). Do children matter? An examination of gender differences in environmental valuation. Ecological Economics, 49, 273-286.

Farreras, V., Riera, P. & Mogas, J. (2005). Does gender matter in valuation studies? Evidence from three forestry applications. Forestry, 78, 239-248.

Fernández, C., León, C. J., Steel, M. F. J. & Vázquez-Polo, F. J. (2004). Bayesian analysis of interval data contingent valuation models and pricing policies. Journal of Business Economics and Statistics, 22, 431-442.

Flores, N. E. & Carson, R. T. (1997). The relationship between the income elasticities of demand and willingness to pay. Journal of Environmental Economics and Management, 33, 287-295.

Foster, W. & Just, R. E. (1989). Measuring welfare effects of product contamination with consumer uncertainty. Journal of Environmental Economics and Management, 17, 266-283.

Freeman, A. M., III. (1993). The measurement of environmental and resource values: Theory and methods. Washington, DC: Resources for the Future.

Gabor, A. & Granger, C. W. J. (1966). Price as an indicator of quality. Economica, 33, 43-70.

Garcia-Llorente, M., Martin-López, B. & Montes, C. (2011). Exploring the motivations of protestors in contingent valuation: Insights for conservation policies. Environmental Science & Policy, 14, 76-88.

Green, D., Jacowitz, K., Kahneman, D. & McFadden, D. (1998). Referendum contingent valuation, anchoring and willingness to pay for public goods. Resource and Energy Economics, 20, 85-116.

Greenley, D. A., Walsh, R. G. & Young, R. A. (1981). Option value: Empirical evidence from a case study of recreation and water quality. Quarterly Journal of Economics, 96, 657-672.

Groothuis, P. A. & Whitehead, J. C. (2002). Does don't know mean no? Analysis of 'don't know' responses in dichotomous choice contingent valuation questions. Applied Economics, 34, 1935-1940.

Groves, R. M., Brick, J. M., Couper, M., Kalsbeek, W., Harris-Kojetin, B., Krueter, F., Pennell, B.-E., Raghunathan, T., Schouten, B., Smith, T., Tourangeau, R., Bowers, A., Jans, M., Kennedy, C., Levenstein, R., Olson, K., Peytcheve, E., Ziniel, S. & Wager, J. (2008). Issues facing the field: Alternative practical measures of representativeness of survey respondent pools. Survey Practice, 1 (3), 1-6.

Groves, R. M. & Peytcheva, E. (2008). The impact of nonresponse rates on nonresponse bias: A meta-analysis. Public Opinion Quarterly, 72, 167-189.

Haab, T. C., Interis, M. G., Petrolia, D. R. & Whitehead, J. C. (2013). From hopeless to curious? Thoughts on Hausman's 'dubious to hopeless' critique of contingent valuation. Applied Economic Perspectives and Policy, 35, 593-612.

Haab, T. C. & McConnell, K. E. (1997). Referendum models and negative willingness to pay: Alternative solutions. Journal of Environmental Economics and Management, 32, 251-270.

Haab, T. C. & McConnell, K. E. (2002). Valuing environmental and natural resources: The econometrics of non-market valuation. Cheltenham, United Kingdom: Edward Elgar.

Hammack, J. & Brown, G. M., Jr. (1974). Waterfowl and wetlands: Toward bioeconomic analysis. Baltimore, MD: Johns Hopkins University Press.

Hanemann, W. M. (1984). Welfare evaluations in contingent valuation experiments with discrete responses. American Journal of Agricultural Economics, 66, 332-341.

Hanemann, W. M., Loomis, J. & Kanninen, B. (1991). Statistical efficiency of double-bounded dichotomous choice contingent valuation. American Journal of Agricultural Economics, 73, 1255-1263.

Hanley, N., Schläpfer, F. & Spurgeon, J. (2003). Aggregating the benefits of environmental improvements: Distance-decay functions for use and nonuse values. Journal of Environmental Management, 68, 297-304.

Harbaugh, W. T. (1998). What do donations buy? A model of philanthropy based on prestige and warm glow. Journal of Public Economics, 67, 269-284.

Hausman, J. A. (Ed.). (1993). Contingent valuation: A critical assessment. Amsterdam, The Netherlands: Elsevier Science.

Hausman, J. A. (2012). Contingent valuation: From dubious to hopeless. Journal of Economic Perspectives, 25 (4), 43-56.

Heberlein, T. A., Wilson, M. A., Bishop, R. C. & Schaeffer, N. C. (2005). Rethinking the scope test as a criterion for validity in contingent valuation. Journal of Environmental Economics and Management, 50, 1-22.

Herriges, J. A. & Shogren, J. F. (1996). Starting point bias in dichotomous choice valuation with follow-up questioning. Journal of Environmental Economics and Management, 30, 112-131.

Hoehn, J. P. & Loomis, J. B. (1993). Substitution effects in the valuation of multiple environmental programs. Journal of Environmental Economics and Management, 25, 56-75.

Hoehn, J. P. & Randall, A. (1987). A satisfactory benefit-cost indicator from contingent valuation. Journal of Environmental Economics and Management, 14, 226-247.

Johnston, R. J. (2006). Is hypothetical bias universal? Validating contingent valuation responses using a binding referendum. Journal of Environmental Economics and Management, 52, 469-481.

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., and Vossler, C.A. (2017). Contemporary Guidance for Stated Preference Studies. Journal of the Association of Environmental and Resource Economists (forthcoming).

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

Johnston, R. J., Swallow, S. K. & Weaver, T. F. (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, 97-120.

Jones, N., Sophoulis, C. M. & Malesios, C. (2008). Economic valuation of coastal water quality and protest responses: A case study in Mitilini, Greece. The Journal of Socio-Economics, 37, 2478-2491.

Kanninen, B. J. (1993a). Design of sequential experiments for contingent valuation studies. Journal of Environmental Economics and Management, 25, S1-S11.

Kanninen, B. J. (1993b). Optimal experimental design for double-bounded dichotomous choice contingent valuation. Land Economics, 69,138-146.

Kanninen, B. J. (1995). Bias in discrete response contingent valuation. Journal of Environmental Economics and Management, 28, 114-125.

Kealy, M. J., Dovidio, J. F. & Rockel, M. L. (1988). Accuracy in valuation is a matter of degree. Land Economics, 64, 158-171.

Kealy, M. J., Montgomery, M. & Dovidio, J. F. (1990). Reliability and predictive validity of contingent values: Does the nature of the good matter? Journal of Environmental Economics and Management, 19, 244-263.

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, 125-148.

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, 25 (4), 3-26.

Kotchen, M. J. & Reiling, S. D. (1999). Do reminders of substitutes and budget constrains influence contingent valuation estimates? Another comment. Land Economics, 75, 478-482.

Kreuter, F., Presser, S. & Tourangeau, R. (2008). Social desirability bias in CATI, IVR, and Web surveys: The effects of mode and question sensitivity. Public Opinion Quarterly, 72, 847-865.

Kriström, B. (1990). A non-parametric approach to the estimation of welfare measures in discrete response valuation studies. Land Economics, 66, 135-139.

Kriström, B. (1997). Spike models in contingent valuation. American Journal of Agricultural Economics, 79, 1013-1023.

Leggett, C. G., Kleckner, N. S., Boyle, K. J., Duffield, J. W. & Mitchell, R. C. (2003). Social desirability bias in contingent valuation surveys administered through in-person interviews. Land Economics, 79, 561-575.

Lewbel, A., McFadden, D. & Linton, O. (2011). Estimating features of a distribution from binomial data. Journal of Econometrics, 162, 170-188.

Li, C. Z. (1996). Semiparametric estimation of the binary choice model for contingent valuation. Land Economics, 72, 462-473.

Li, X., Boyle, K. J., Holmes, T. P. & Pullis-LaRouche, G. (2014). The effect of on-site forest experience on stated preferences for low impact timber harvesting programs. Journal of Forest Economics, 20, 348-362.

Lindhjem, H. & Navrud, S. (2011). Are Internet surveys an alternative to face-to-face interviews in contingent valuations? Ecological Economics, 70, 1628-1637.

Lindhjem, H. & Navrud, S. (2012). Asking for individual or household willingness to pay for environmental goods? Environmental and Resource Economics, 43, 11-29.

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.

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. B. (1989). Test-retest reliability of the contingent valuation method: A comparison of general population and visitor responses. American Journal of Agricultural Economics, 71, 76-84.

Loomis, J. B. (1990). Comparative reliability of the dichotomous choice and open-ended contingent valuation techniques. Journal of Environmental Economics and Management, 18, 78-85.

Loomis, J., Gonzalez-Caban, A. & Gregory, R. (1994). Do reminders of substitutes and budget constraints influence contingent valuation estimates? Land Economics, 70, 499-506.

MacMillan, D., Hanley, N. & Leinhoop, N. (2006). Contingent valuation: Environmental polling or preference engine? Ecological Economics, 60, 299-307.

Madureira, L., Nunes, L. C., Borges, J. G. & Falcão, A. O. (2011). Assessing forest management strategies using a contingent valuation approach and advanced visualisation techniques: A Portuguese case study. Journal of Forest Economics, 17, 399-414.

Maguire, K. B., Taylor, L. O. & Gurmu, S. (2003). Do students behave like adults? Evidence from valuation experiments. Applied Economics Letters, 10, 753-756.

Manfreda, K. L., Bosnjak, M., Berzelak, J., Haas, I. & Vehovar, V. (2008). Web surveys versus other survey modes: A meta-analysis comparing response rates. Journal of the Market Research Society, 50, 79.

Mansfield, C., Van Houtven, G., Henderschott, A., Chen, P., Potter, J., Nourani, V. & Kilambi, V. (2012). Klamath River basin restoration nonuse value survey. Final report to the U.S. Bureau of Reclamation, RTI International, RTI Project Number 0212485.001.010.

Marwell, G. & Ames, R. E. (1981). Economists free ride, does anyone else? Experiments on the provision of public goods. Journal of Public Economics, 15 (3), 295-310.

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

McFadden, D. (1994). Contingent valuation and social choice. American Journal of Agricultural Economics, 76 (4), 689-708.

Menges, R. & Beyer, G. (2014). Underground cables versus overhead lines: Do cables increase social acceptance of grid development? Results of a contingent valuation survey in Germany. International Journal of Sustainable Energy Planning and Management, 3, 33-48.

Messonnier, M. L., Bergstrom, J. C., Cornwell, C. M., Teasley, R. J. & Cordell, H. K. (2000). Survey response-related biases in contingent valuation: Concepts, remedies, and empirical application to valuing aquatic plant management. American Journal of Agricultural Economics, 82 (2), 438-450.

Meyerhoff, J. & Liebe, U. (2006). Protest beliefs in contingent valuation: Explaining their motivation. Ecological Economics, 57 (4), 583-594.

Mitchell, R. C. & Carson, R. T. (1981). An experiment in determining willingness to pay for national water quality improvements. Unpublished report. Washington, DC: Resources for the Future.

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

Mitchell, R. C. & Carson, R. T. (1993). The value of clean water: The public's willingness-to-pay for boatable, fishable, and swimmable quality water. Water Resources Research, 29 (7), 2445-2454.

Munro, A. (2005). Household willingness to pay equals individual willingness to pay if and only if the household income pools. Economics Letters, 88, 227-230.

Murphy, J. J., Stevens, T. H. & Yadav, L. (2010). A comparison of induced value and home-grown value experiments to test for hypothetical bias in contingent valuation. Environmental and Resource Economics, 47, 111-123.

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.

Nunes, P. A. & Schokkaert, E. (2003). Identifying the warm glow effect in contingent valuation. Journal of Environmental Economics and Management, 45, 231-245.

Ojea, E. & Loureiro, M. L. (2011). Identifying the scope effect in a meta-analysis of biodiversity valuation studies. Resource and Energy Economics, 33, 706-724.

Onwujekwe, O. & Fox-Rushby, J. (2005). Inter-rater and test-retest reliability of three contingent valuation question formats in south-east Nigeria. Health Economics, 14, 529-536.

Park, T., Loomis, J. & Creel, M. (1991). Confidence intervals for evaluating benefit estimates from dichotomous choice contingent valuation studies. Land Economics, 67, 64-73.

Poe, G. L. & Bishop, R. C. (1999). Valuing the incremental benefits of groundwater protection when exposure levels are known. Environmental and Resource Economics, 13, 347-373.

Poe, G. L. & Bishop, R. C. (2001). Information and the valuation of nitrates in ground water, Portage County, Wisconsin. In J. C. Bergstrom, K. J. Boyle & G. L. Poe (Eds.), The economic value of water quality. Cheltenham, United Kingdom: Edward Elgar.

Poe, G. L., Clark, J. E., Rondeau, D. & Schulze, W. D. (2002). Provision point mechanisms and field validity tests of contingent valuation. Environmental and Resource Economics, 23, 105-131.

Poe, G. L., Giraud, K. L. & Loomis, J. B. (2005). Computational methods for measuring the difference in empirical distributions. American Journal of Agricultural Economics, 87, 353-365.

Poe, G. L., Severance, E. K. & Welsh, M. P. (1994). Measuring the difference (X-Y) of simulated distributions: A convolutions approach. American Journal of Agricultural Economics, 76, 904-915.

Quiggin, J. (1998). Individual and household willingness to pay for public goods. American Journal of Agricultural Economics, 80, 58-63.

Ramajo-Hernández, J. & del Saz-Salazar, S. (2012). Estimating the non-market benefits of water quality improvement for a case study in Spain: A contingent valuation approach. Environmental Science & Policy, 22, 47-59.

Randall, A., Ives, B. & Eastman, C. (1974). Bidding games for evaluation of aesthetic environmental improvements. Journal of Environmental Economics and Management, 1, 132-149.

Reiling, S. D., Boyle, K. J., Phillips, M. L. & Anderson, M. W. (1990). Temporal reliability of contingent values. Land Economics, 66, 128-134.

Roach, B., Boyle, K. J. & Welsh, M. P. (2002). Testing bid design effects in multiple bounded contingent valuation. Land Economics, 78, 121-131.

Rowe, R. D., d'Arge, R. C. & Brookshire, D. S. (1980). An experiment on the economic value of visibility. Journal of Environmental Economics and Management, 7, 1-19.

Rowe, R. D., Schulze, W. D. & Breffle, W. S. (1996). A test for payment card biases. Journal of Environmental Economics and Management, 31, 178-185.

Ryan, M. & Watson, V. (2009). Comparing welfare estimates from payment card contingent valuation and discrete choice experiments. Health Economics, 18, 389-401.

Salant, P. & Dillman, D. A. (1994). How to conduct your own survey. New York, NY: John Wiley & Sons.

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

Scarpa, R. & Bateman, I. (2000). Efficiency gains afforded by improved bid design versus follow-up valuation questions in discrete-choice cv studies. Land Economics, 76, 299-311.

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

Schläpfer, F., Roschewitz, A. & Hanley, N. (2004). Validation of stated preferences for public goods: A comparison of contingent valuation survey response and voting behavior. Ecological Economics, 51, 1-16.

Schneemann, M. (1997). A meta-analysis of response rates to contingent valuation surveys conducted by mail. Unpublished master's thesis. University of Maine.

Schouten, B., Cobben, F. & Bethlehem, J. (2009). Indicators for the representativeness of survey response. Survey Methodology, 35, 101-113.

Scott, A. (1965). The valuation of game resources: Some theoretical aspects. Canadian Fisheries Report, iv, Ottawa, Ontario, Canada: Department of Fisheries of Canada.

Shapiro, B. P. (1968). The psychology of pricing. Harvard Business Review, 46 (7), 14-25.

Smith, R. D. (2005). Sensitivity to scale in contingent valuation: The importance of the budget constraint. Journal of Health Economics, 24, 515-529.

Smith, V. L., Suchanek, G. L. & Williams, A. W. (1988). Bubbles, crashes and endogenous expectations in experimental spot asset markets. Econometrica, 56, 1119-1151.

Soliño, M., Vázquez, M. X. & Prada, A. (2009). Social demand for electricity from forest biomass in Spain: Does payment periodicity affect the willingness to pay? Energy Policy, 37, 531-540.

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

Stevens, T. H., More, T. A. & Glass, R. J. (1994). Interpretation and temporal stability of CV bids for wildlife existence: A panel study. Land Economics, 70, 355-363.

Strazzera, E., Scarpa, R., Calia, P., Garrod, G. D. & Willis, K. G. (2003). Modeling zero values and protest responses in contingent valuation surveys. Applied Economics, 35, 133-138.

Teisl, M. F., Boyle, K. J., McCollum., D. W. & Reiling, S. D. (1995). Test-retest reliability of contingent valuation with independent sample pretestPretest and post-test control groups. American Journal of Agricultural Economics, 77, 613-619.

Teisl, M. F., Roe, B. & Hicks, R. D. (2002). Can eco-labels tuna market? Journal of Environmental Economics and Management, 43, 339-359.

Thayer, M. (1981). Contingent valuation techniques for assessing environmental impacts: Further evidence. Journal of Environmental Economics and Management, 8, 27-44.

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.

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.

Vossler, C. A., Kerkvliet, J., Polasky, S. & Gainutdinova, O. (2003). Externally validating contingent valuation: An open-space survey and referendum in Corvallis, Oregon. Journal of Economic Behavior & Organization, 51, 261-277.

Wantanabe, M. (2010). Nonparametric estimation of mean willingness to pay from discrete response valuation data. American Journal of Agricultural Economics, 92, 1114-1135.

Ward, K. L. & Duffield, J. W. (1992). Natural resource damages: Law and economics. New York: John Wiley & Sons.

Watson, V. & Ryan, M. (2007). Exploring preference anomalies in double bounded contingent valuation. Journal of Health Economics, 26, 463-482.

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

Whitehead, J. & Bloomquist, G. (1995). Do reminders of substitutes and budget constraints influence contingent valuation estimates? Comment. Land Economics, 71, 541-543.

Whittington, D. & Pagiola, S. (2012). Using contingent valuation in the design of payments for environmental services mechanisms: A review and assessment. World Bank Research Observer, 27, 261-287.

Wiser, R. H. (2007). Using contingent valuation to explore willingness to pay for renewable energy: A comparison of collective and voluntary payment vehicles. Ecological Economics, 62, 419-432.

Wooldridge, J. (2012). Introductory econometrics: A modern approach. Boston, MA: Cengage Learning.