

Valuing Food Safety and Nutrition

EDITED BY
Julie A. Caswell

*Book Originally Published by
Westview Press, Boulder, Colorado, 1995*

PART THREE: A Closer Look at Performing Contingent Valuation

16. Criteria for Evaluating Results Obtained from Contingent Valuation Methods

Richard B. Belzer and Richard P. Theroux

Food Marketing Policy Center
Department of Agricultural and Resource Economics
University of Connecticut

Criteria for Evaluating Results Obtained from Contingent Valuation Methods

Richard B. Belzer

Richard P. Theroux

Keywords: Contingent valuation, evaluation criteria, reliability,
willingness to pay

Copyright © 1997 by Food Marketing Policy Center, University of Connecticut. All rights reserved. Readers may make verbatim copies of this document for non-commercial purposes by any means, provided that this copyright notice appears on all such copies.

16

Criteria for Evaluating Results Obtained from Contingent Valuation Methods

Richard B. Belzer and Richard P. Theroux¹

Recent years have seen rapidly expanding interest in a variety of nonobservational methods of collecting data on economic phenomena. Foremost among these methods is contingent valuation (CV), a methodology that has become prominent through its application to the problem of valuing environmental amenities and similar commodities which are not directly traded in markets (e.g., Cummings et al. 1986, Mitchell and Carson 1989, Environmental Protection Agency 1993). The goal of CV methods is to simulate the same kind of ordered preferences which economic theory argues would be revealed through market behavior if such markets existed (Freeman 1979: 97). Thus, the application of CV methods has been largely limited to public-good commodities that are not traded in markets and for which there are few if any alternative methodologies.

New interest in CV can be attributed primarily to natural resource damage litigation spawned by the Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA), the Clean Water Act (CWA), the Oil Pollution Act of 1990 (OPA), and intensified oversight of federal regulatory activity conducted by the Executive Office of the President. Section 311(f) of CWA and section 107 of CERCLA authorize natural resource trustees to recover compensatory damages for injury to or destruction of natural resources resulting from a discharge of oil into navigable waters or a release of hazardous substances. Section 1006(e) of OPA authorized the National Oceanic and Atmospheric Administration to develop rules for assessing natural resource damages for discharges of oil into navigable waters.

Regulatory oversight is performed by the Office of Information and Regulatory Affairs (OIRA), a statutory office within the Office of Management and Budget (OMB), as directed by Executive Order No. 12866.² Among other things, this Order subjects significant regulatory actions of federal departments

and agencies to a rudimentary benefit-cost test. Included within the realm of regulatory actions are those involving the mandated provision or disclosure of information and requirements for the collection, retention, and submission of data (Office of Management and Budget 1990b). Regulatory actions that fail to offer social benefits in excess of social costs bear a special burden of policy justification, inasmuch as they violate traditional normative welfare economics standards. Much of the interest in applying CV methods to food safety issues has arisen because of these high-level demands for regulatory analysis.

OMB Guidance Concerning the Use of CV in Regulatory Impact Analysis

With respect to federal regulation, OMB has published guidance for agencies to use in performing Regulatory Impact Analyses (RIAs) of major rules (Office of Management and Budget 1990c, Office of Management and Budget 1990d).³ OMB guidance establishes a preference for observational or behavioral data in the development of benefit and cost estimates. Because such data reflect voluntary exchanges they can be presumed to reflect the economic preferences of individuals in the absence of a demonstrated market failure. Where benefits derive from risk assessments and an agency chooses to represent benefits with point estimates, the guidance calls for the use of expected value estimates of risk. Conventional risk assessment methods are not appropriate for use in benefit-cost analyses; embedded conservatism results in highly exaggerated point estimates of both risks and benefits, thus inserting a bias in favor of government intervention to remedy problems whose scope and severity are similarly overstated (Office of Management and Budget 1990a).

OMB's guidance acknowledges the difficulty of estimating individuals' willingness to pay (WTP) for commodities that are not traded in markets and thus are impossible to value using conventional observational methods. However, the guidance also recognizes that nonobservational methods such as CV warrant an additional burden of analytic rigor:

Contingent valuation methods provide the only analytical approaches currently available for estimating the benefits of such untraded goods. The absence of observable and replicable behavior with respect to the benefit in question, combined with the difficulties of avoiding bias in contingent valuation studies, argues for great care and circumspection in the use of such methods. This means, for example, that estimates of willingness to pay must incorporate the variety of alternative means individuals have of expressing value for untraded goods. Moreover, analyses must faithfully capture individuals' budget constraints, which restrict their willingness to pay for untraded as well as traded goods and services. Benefit analyses derived from contingent valuation and similar methods thus require considerable analytic rigor in design and careful

execution. Absent such efforts, analyses based heavily on the benefits of untraded goods and services normally would fail the test of a satisfactory RIA (Office of Management and Budget 1990d: 661).

This flexible performance standard for RIAs was crafted after considering a number of comments from federal agencies. In responding to these comments, OMB acknowledged the need for CV and similar methods for estimating the value of untraded goods and services. Nevertheless, OMB insisted on a very high standard of care in the use of CV, including an extensive effort to test alternative explanatory hypotheses:

Survey estimates may be necessary to estimate certain physical and psychological uses, because relevant behavior is unobservable. However, the problems that arise in the estimation of use value through survey methods are considerably more serious. Great care needs to be taken to ensure that survey designs do not introduce systematic biases by departing from market-based valuation principles. For example, slight changes in the way questions are presented can sometimes result in dramatic changes in responses, because of the hypothetical nature of data derived from survey instruments. This hypothetical character means that survey methods offer considerable opportunities for abuse. Analyses relying on survey instruments to estimate benefits should devote considerable efforts to quality control, data verification, and real-world hypothesis testing. Major departures from market-based principles can lead to serious distortions in the allocation of our Nation's scarce resources (Office of Management and Budget 1990c: 37-38).

Finally, the OMB guidance sounds a warning to federal agencies inclined to employ CV methods in support of federal regulation: CV-based analysis will be judged in accordance with extraordinary standards, particularly if it is relied upon for a large share of the estimated benefits:

Because of the potential for misuse of survey methods, RIAs generally should avoid relying exclusively on value estimates derived from survey approaches... [D]epartments and agencies that develop benefit estimates which rely heavily on the results of survey instruments bear an extraordinary burden to show that estimates obtained are reasonably consistent with observable market behavior and common sense (Office of Management and Budget 1990c: 38).

Interest in CV methods among federal regulatory agencies is a product of these institutional demands for rigorous policy analysis. It is important to keep foremost in mind that while CV research may serve to advance knowledge and improve economic methodologies generally, agencies' primary interest in CV research lies in its potential capacity to support or expand programmatic responsibilities and achieve organizational objectives.

This creates an unavoidable tension between CV researchers and the agencies that fund their work. Continued funding depends to some extent on the

ability of CV methods to justify regulations and programs—if not now, then in the foreseeable future. Researchers thus face potentially perverse incentives to "see" promise where it is ephemeral and to overinterpret the results of the studies they perform.

The purpose of this chapter is to provide a look at CV methods and applications from the perspective of a "consumer" of benefit-cost analyses prepared in support of governmental programs, policies, and regulations. While some practitioners have celebrated CV's "arrival" (Cummings et al. 1986), the controversy surrounding CV only continues to intensify (Cambridge Economics 1992). Our intent is neither to disparage nor promote the CV method, for we are agnostic as to its ultimate capacity to answer some very difficult empirical questions which the economics profession has largely ignored.

The issues and concerns set forth in this chapter should not be interpreted as comprehensive, nor are the suggested hypothesis tests intended to be either exhaustive or appropriate in every context. Rather, they represent relatively simple issues which have become evident in the course of reviewing government agency CV survey instruments and regulatory analyses relying upon CV methods and results. Further, agencies need encouragement to set high standards for the CV research they sponsor, both to stimulate high quality research and to resist the unavoidable countervailing pressure to utilize flawed results which might nevertheless appear to be attractive.

Throughout this chapter we have made several important simplifying assumptions. First, we presume that a significant market failure has been demonstrated to exist with respect to a specific food safety risk. This means that carefully devised government intervention *may* increase social welfare in accordance with the Kaldor-Hicks criterion. We take no issue at all with private parties applying CV (or any other methods) as long as they alone bear the costs and capture the benefits of acting based on the results of such research. Our concern is solely with efforts to use government outlays or regulations to achieve public purposes.

Second, we assume that researchers are pursuing CV methods because the relevant portion of the demand curve for risk reduction is not observable and there are no market-based data to rely upon for estimating consumers' WTP. Where the relevant segment of the demand curve is observable or can be approximated from other commodity markets in which food safety risk is a significant attribute, we expect that traditional valuation methods based on observations of behavior would be used instead of CV.

Third, it is our firm conviction that governments bear a special burden whenever they propose to intervene in private transactions. Regardless of whether they spend monies collected through taxes or command others to make expenditures through regulation, government officials have a fiduciary duty when it comes to other people's money. Whether one fritters away one's own assets or inheritance is not generally a public concern. However, the power to consume the public purse or command private parties to reallocate the expenditure of

private funds carries with it a special ethical responsibility to be careful and deliberate. Thus, we presume that it is both fair and appropriate that those who would use the financial resources or coercive powers of the government also bear a disproportionately greater burden to demonstrate the wisdom and analytical soundness of their proposals.

Finally, we are not experts in CV, nor are we advocates or critics. Our interest in the CV method and its potential applications arises because we are "consumers" of benefits analyses and regulatory policy initiatives in which CV methods are playing an increasingly important role. Many of the applications we have seen have made us acutely uncomfortable. We see a clear need for comprehensive criteria acceptable to all sides which could guide future public policy applications of CV methods. This chapter represents an effort to begin the dialogue necessary to develop such criteria.

Methodological Concerns Raised by Contingent Valuation Methods

Practitioners of CV continue to struggle with a variety of serious methodological issues. At the risk of oversimplification, these issues may be captured by the following three questions:

1. Do the expressed WTPs provided by CV survey respondents under hypothetical conditions adequately simulate behavior under real-world conditions?
2. Do CV respondents have a clear understanding of the identity and character of the commodity which they have been asked to value?
3. Are the expressed WTPs elicited by CV surveys reliable and valid estimates of respondents' economic preferences?

Each of these questions implicitly suggests the existence of necessary and sufficient conditions for CV estimates to be both valid and reliable. Because the CV method is still relatively new and its capable practitioners few, we seem to be a long way from developing sufficient conditions. Nevertheless, it is clearly necessary that the answers to these three questions must be "yes" before it is appropriate to use CV-based estimates in applied work having important public policy implications.

Hypothetical vs. Real-World Data

Endemic to CV methods is the fact that the transactions examined are hypothetical, and answers to such questions are only hypothetically accurate. The problem, of course, is that absent actual behavior there is no way to ensure

that respondents give "real" answers. Experienced CV practitioners thus devote considerable efforts to establishing realism in an attempt to elicit "real" answers and to develop hypothesis tests which would identify possible biases (e.g., Cummings et al. 1986, Mitchell and Carson 1989). Others have performed experiments in which serious biases have been detected, such as discrepancies between cash and hypothetical payments (Bishop and Heberlein 1986, Neill et al. 1994), noncommitment biases related to budget constraints (Kemp and Maxwell 1992), and the presence of embedded goods and moral satisfaction (Kahneman and Knetsch 1992). Similar concerns were raised recently by a panel of distinguished economists asked by the federal government to evaluate the use of CV methods for valuing natural resource damages (National Oceanic and Atmospheric Administration 1993).

In our view the general lack of effective budget constraints poses the most serious problem with hypothetical data. Public opinion polls routinely ask people to value certain things, but rarely do they ever focus, explicitly or implicitly, on the opportunity costs associated with obtaining these new assets. Absent any clear evidence that respondents understand the concept of opportunity cost and properly incorporate it into their responses, polls offer no useful economic information. Further, CV studies routinely suffer from an absence of effective budget constraints and little has been done in recent years to solve this problem. Kahneman and Knetsch (1992: 59) quote experienced CV practitioners who wrote in a 1983 survey of the state-of-the-art that "the summation of average CV values for public goods thus far available in the literature would *exhaust* the budget of the average individual." One can only imagine how much worse this comparison must be given the large number of CV studies performed since this statement was made.

CV studies of environmental amenities have shown that the imposition of a budget constraint can dramatically alter expressed WTPs. In a study of the value of preventing oil spills in Alaska's Prince William Sound, Kemp and Maxwell (1992) obtained WTP estimates about 300 times smaller when they utilized a top-down disaggregation procedure than when they sought direct estimates of value. They concluded that the lower value was more plausible precisely because it required respondents to adapt to a binding budget constraint, but they averred from endorsing it as the "right" value because expressed WTPs tend to decline as the number of disaggregation levels increases and there is no obvious stopping point. Kahneman and Knetsch (1992: 61) observed a 50-fold reduction in median WTP in an experiment involving just two disaggregation steps.

Where CV methods are applied to food safety problems, we believe that similarly rigorous efforts will be needed to develop realistic scenarios and effective budget constraints. It is the responsibility of CV practitioners to devise and test alternative hypotheses concerning possible biases resulting from the hypothetical nature of the transaction in question.

Clear Commodity Definition

Economists asked to value commodities traded in markets look directly at such markets for data. Consumers reveal their preferences in their voluntary transactions. Thus, if the market price for doughnuts is 42 cents each, and there is no particular reason to believe that doughnut markets are rife with either externalities or market power, then the value of a doughnut to the marginal consumer is precisely 42 cents. Most consumers will value doughnuts at more than 42 cents, reflecting the consumers' surplus they reap from not having to pay a price greater than 42 cents despite a willingness to do so, but market price will be set by the marginal consumer.

Most commodities traded in markets have attributes substantially more complex than doughnuts, but consumers still reveal their preferences through voluntary exchange. Consumers need not be able to quantify or even observe every attribute of a product to place an appropriate value on it. Information concerning product attributes is itself a valuable commodity, and one that suffers from certain public good problems, but even where information is available freely it cannot be costlessly internalized and processed. This means that less-than-perfect information is in all cases optimal for individual decision making.

Privately owned environmental amenities also are traded in markets. An ocean view is a valuable real estate attribute, one that raises both its market clearing sales price and the rents which it can command. But the difficulty of quantifying an ocean view does not diminish consumers' capacity to monetize it or the ability of willing buyers and sellers to establish market-clearing prices. Sellers generate and disseminate information to seek out buyers, attempting to persuade them of the unique character, attractiveness, and perhaps exclusivity of these amenities. Potential buyers process this information, collect their own data and comparison data on possible substitutes, and evaluate market prices in the context of their own WTP. Prices rise or fall depending on the usual supply and demand conditions.

Publicly owned amenities do not benefit from these market prices. Indeed, an obvious inefficiency arises insofar as publicly owned environments often have near zero prices. The National Parks suffer congestion externalities, for example, reducing both the quality of the environmental amenity and consumers' WTP to visit. Other valuation approaches, such as the travel cost and hedonic methods, must be used to estimate the value of publicly owned environments due to government's decision not to price these assets efficiently.

Risks to health and safety represent another class of commodities which are clearly traded in markets but for which indirect methods must be used to estimate value. The economics literature on risk valuation is extensive and clearly demonstrates that individuals are willing to pay substantial sums to avoid risk despite the difficulty of identifying analytically rigorous measures for it and quantifying it (Viscusi 1983, Fisher et al. 1989). The task of the researcher is

to statistically disentangle from the gamut of attributes inherent in the product or employment contract the attributes relating to risk.

Alternative methods thus are available for analyzing actual behavioral data and ascertaining the reliability and validity of CV results. Given the range of methodological problems associated with CV and the depth of concern raised by many distinguished economists, researchers should seriously consider whether CV is the most appropriate tool for estimating consumers' WTP to reduce foodborne risks.⁴

For any CV approach to generate potentially useful results, the survey must establish clearly the precise character of the commodity respondents are being asked to value. Thus, if the question is how to value consumers' WTP to avoid *E. coli* infection, then the commodity must be defined as a specified likelihood (e.g., a 1 in 1,000 annual chance) of a well-defined adverse health outcome (e.g., hemorrhagic colitis). Clarity is necessary to minimize the extent to which respondents mistakenly provide values for unspecified alternative commodities (e.g., a 1 in 10,000 chance of debilitating kidney damage from hemolytic uremic syndrome). CV studies which purport to estimate consumers' WTP to reduce ambiguous risks by an unspecified amount do not yield information that is relevant for policy making where government intervention involves restricting consumer choice, mandating the expenditure of private resources, or expending public funds collected through taxation.

The CV literature in the food safety area reveals uncertainty as to whether food safety risks should be portrayed in absolute or relative terms (e.g., see Lin and Milon, Chapter 5). In theory, however, respondents should be indifferent to such framing questions if they have a clear understanding of the commodity they are expected to value and are sufficiently familiar with the commodity of interest. To the extent that framing matters, CV may be an inappropriate method for value elicitation inasmuch as there is no analytic basis for asserting that one particular framing approach is superior to others.

In a similar vein, both the nature of the underlying risk and the commodity to be valued must be believable to survey respondents. If the problem appears to be hypothetical, then respondents may be inclined to give hypothetical answers. CV researchers often encounter zero bids from respondents who believe that the underlying problem does not actually exist or reject the proposed action as a valid response to such a problem. Because CV surveys represent hypothetical market transactions, respondents often fail to recognize their own budget constraints. High bids entail no additional opportunity cost. CV researchers must develop and implement methods to simulate real-world budget constraints. Finally, CV studies must use payment vehicles which would force respondents to bear the full costs of the benefits they receive. Increases in general taxation and similarly broad payment formulae may be acceptable in certain circumstances, but they create powerful incentives for survey respondents to express WTPs that are not founded on economic preferences and hence of limited utility in benefit-cost analysis.

Reliability and Validity in Expressed WTP Values

Because there may be no market transactions providing sure evidence of voluntary exchange, the results of CV studies may be subject to a variety of alternative interpretations. It is not enough to assume that survey respondents interpret valuation questions in precisely the way they were intended. Rather, CV researchers must ensure that plausible alternative explanatory hypotheses can be safely ruled out.

This process entails several critical elements. For example, researchers must identify all plausible explanations for the results they obtain and develop tests suitable for testing these alternative hypotheses; they cannot simply assume that the values respondents provide represent economic preferences. Verbal protocols must be employed to ensure that the decision processes used by respondents correspond to the economic model of behavior. Researchers must design CV surveys such that data needed to test alternative hypotheses are collected. Survey instruments must be designed, pre-tested, and refined based on the knowledge gleaned; it is not enough to assert that a particular design has been used effectively elsewhere and thus does not require similar validation. Samples must be large enough to execute the necessary hypothesis tests with sufficient statistical power; multiple alternative scenarios and payment vehicles are probably necessary. Finally, CV results must be compared with other expectations and economic data to ensure consistency between CV results and revealed preferences. Where CV results are intended to be used for policy purposes, they should be replicated by independent researchers.

Suggested Checks for Consistency with the Consumer Choice Axioms of Elementary Economic Theory

Little effort has been devoted to the development and propagation of methods for testing whether CV results are consistent with revealed preferences. In this section we identify a variety of alternative hypotheses which CV researchers ought to be able to reject using generally accepted statistical methods. Each of these alternative hypotheses has been identified in the economics literature as a plausible explanation for observed CV results and as a rationale for rejecting their use in policy contexts.

Calibration of Hypothetical and Actual WTP

The primary use of survey methods involving economic phenomena is in market research, and survey results are routinely calibrated based on prior experience to adjust for the hypothetical nature of the exercise. Nevertheless, many products fail in real-world markets despite successful market research

studies, suggesting that calibration is more art than science. Marketing researchers routinely observe actual (ex post) purchase behavior that falls well short of the quantities consumers state (ex ante) they intend to purchase and have focused their efforts on developing models to explain and predict these shortfalls (Jamieson and Bass 1989).

Recent work by Neill et al. (1994) begins to address this issue in the context of CV. They examined WTP responses in three settings—a CV survey, a hypothetical Vickrey auction, and a real Vickrey auction—to ascertain the extent to which CV methods could replicate a known incentive-compatible instrument and whether either hypothetical exercise could replicate actual purchase behavior. The authors found that the distribution of bids in the two hypothetical markets were not significantly different. However, median WTP in the hypothetical markets exceeded median WTP in the real auction by factors ranging from 6 to 15.

These results have three important implications. First, hypothetical WTP generally exceeds actual WTP—often by substantial amounts—and cannot be assumed to represent actual WTP. Second, the difference between hypothetical and actual WTP may be independent of the nature of the simulated market used to elicit responses. Third, and perhaps most important, the use of an incentive-compatible payment instrument in a hypothetical setting does not assure results consistent with real purchase behavior. More research into the calibration of CV results clearly would be useful. However, this may conflict with the prevailing view among CV advocates that respondents' expressed WTPs properly reflect true economic values and thus do not require calibration at all. This represents an important divergence that deserves further study, because it is difficult to construct a theory under which both CV advocates and market researchers are correct. We believe that this conflict must be resolved before the results of CV studies can be treated as reliable indicators of willingness to pay suitable for public policy applications.

Law of Demand

Subject as always to binding budget constraints, individuals should give consistent answers when they value single or multiple units of a commodity. That is, marginal WTP should decline in response to movement down the demand curve and increase in response to upward movement. As elementary as this notion may seem to be, CV studies do not routinely demonstrate conformity with the Law of Demand.

Suppose that food safety measured as residual risk is the commodity in question. The consumer begins with q_0 units of safety and would pay a price of p_0 for an additional unit. Once this first incremental unit has been acquired, however, the consumer's WTP declines to p' per unit. If entitled only to purchase multiple units of increased safety—that is, travel down the demand

curve—the consumer's total WTP must be less than twice the value of a single unit unless her price elasticity of demand is infinite. The opposite result should be observed for reductions in safety from the baseline. That is, after paying p_0 to avoid losing one unit of safety the consumer places the higher value p'' on preserving the next unit. The consumer's total WTP to avoid a multiple unit loss of safety must exceed twice the value of losing a single unit.

Hypothesis Tests. A proper test of conformity with the Law of Demand requires two demonstrations. First, CV respondents must show that the value they place on a subsequent unit of additional safety is no greater than (and probably less than) the value they place on the first unit. Second, CV respondents must show that the value they place on avoiding a subsequent unit *loss* of safety is no less than (and probably greater than) the value they place on avoiding the initial unit loss of safety. This "marginal value test" makes sense as long as the commodity can be characterized as unidimensional. Food safety risks appear to qualify because they can be readily characterized in quantitative terms based on their probability of occurrence. Goods which clearly have multiple attributes cannot be so easily subjected to this test, but nevertheless should be subjected to an analogous "total value test" in which it is demonstrated that respondents have higher WTPs for an unambiguously superior commodity.⁵

Alternative Hypotheses. Some economists believe that the common absence of CV results consistent with the Law of Demand indicates the presence of "warm glow" effects, which are characterized as the moral satisfaction associated with responding affirmatively when asked to value "good" things (e.g., Diamond et al. 1992, Kahneman and Knetsch 1992). For example, a respondent may feel good about "making a contribution" or doing his "fair share" to acquire or protect a particular environmental amenity. Similarly, respondents may care deeply in an abstract sense about certain *general* food safety risks (e.g., pesticide residues) and embed this concern within their expressed WTP for avoiding *specific* food safety risks (e.g., Alar residues in red apples) or even unrelated risks (e.g., microbiological contamination in ground beef).

Effective tests of upward and downward movement along the demand curve are essential to distinguish between these alternative hypotheses. It is the CV researcher's responsibility to test for the possible presence of "warm glow" effects and similar phenomena unrelated to individual economic preferences, for CV responses must reflect individual economic preferences for the commodity in question before they can be used in benefit-cost analysis (BCA). CV researchers can use split samples to directly test the relevant hypotheses and demonstrate conformity to the Law of Demand.

Casual Benefit-Cost Analysis

Certain payment vehicles such as referenda and generalized tax increases are popular among CV practitioners because they comport with respondents'

personal experiences. Referenda have become so commonplace that voters are now quite familiar with this form of dichotomous choice. Similarly, CV respondents also readily understand how the provision of a certain public good may be paid for through higher taxes.

Unfortunately, the advantage of familiarity is offset by the fact that respondents may be inclined to base their responses on casual BCA rather than personal preferences. Despite clear instructions to report personal valuations, some respondents apparently interpret their task as one of making a broad societal judgment about the desirability of a program or policy. This behavior may reflect widespread familiarity with voting and public opinion polls in which some element of choosing for a group seems to be implied. When such situations arise, respondents are unlikely to report the same preferences they would if the transaction actually arose in a conventional private-goods market (Buchanan 1954).

A common practice among experienced survey researchers is to test survey instruments using verbal protocols. Respondents are asked to think aloud as they fill in the questionnaire. Researchers then review audio or video tapes of these sessions to learn more about the mental processes people use in deriving answers to critical valuation questions. Learning about these mental processes often is more important for effective market research than ascertaining precise value estimates.

Casual BCA can be readily observed in these verbal protocols. Sensitivity to the payments of others reflects concerns about fairness and other values which are perfectly legitimate but cannot be incorporated into the benefit-cost paradigm. For example, excerpts of verbal responses in a recent CV study of ground water valuation clearly reveal that respondents provided values based on casual BCA rather than their own economic preferences (McClelland et al. 1992). Some respondents indicated WTP values which were either conditional upon others paying their "fair share" or mentally derived from considering how much revenue would be generated if everyone contributed a similar amount.

The fundamental problem with casual BCA is that lay estimates of societal welfare are not equivalent to the underlying economic preferences of individuals. Results from CV studies tainted by casual BCA thus offer little insight into the underlying valuation question. Further, if BCA is to be used for societal decision making, expert analysis performed by trained economists is more likely to pass tests of validity and reliability than casual analysis performed on the spot by untrained survey respondents with limited information.

Hypothesis Tests. Perhaps the best way to detect casual BCA is through a verbal protocol study executed as part of or parallel with a CV valuation study. Considerable effort must be devoted to develop hypothesis tests that are compatible with each particular CV survey, for no off-the-shelf test exists. One possible test of the hypothesis that survey responses reflect casual BCA rather than individual preferences would be to employ otherwise identical survey

instruments with radically different payment vehicles. Suppose that one payment vehicle involves a general tax increase that would be paid by all. The alternative payment vehicle could be a targeted tax or user fee aimed directly and solely at the respondent. A necessary condition for respondents to be providing values based on economic preferences is that the median WTP be the same across the two surveys.

As a final note, careful efforts to define the commodity of interest or to provide extensive contextual information are generally worthwhile because they reduce sample variance resulting from uncertainty. They cannot overcome the problem of casual BCA, however. At the same time that additional information reduces respondents' uncertainty and gives them greater capacity to articulate meaningful answers, variance across answers may widen if respondents' preferences truly differ. Thus, reduced variance resulting from enhanced contextual information may reflect improved casual BCA rather than better articulation of individual economic preferences.

Altruism and Vengeance

Standard welfare economics, the foundation for BCA, treats value as a personal matter independent of the gains and payments of others or the process by which decisions are made. Altruism, the expression of WTP derived from the enhanced utility of others, is a perfectly acceptable and economically rational expression of individual economic preferences with respect to private goods where the benefactor bears the full cost. However, it is not a legitimate quantity for BCA of public goods because it results in the double-counting of the benefits which are enjoyed by certain (but not all) members of the community (Milgrom 1992). That is, the incorporation of altruistic values in BCA would explicitly provide for the welfare of some individuals to be weighted more highly than others. Most economists are uncomfortable with such an approach because there are an infinite number of outcomes associated with weighting individual utilities, and economics offers no special expertise or legitimacy in the assignment of such weights.

Another reason for leaving altruism out of BCA is that there is no ethical way to distinguish it from its antithesis, a construct we call "vengeance value." This is the utility an individual obtains from imposing costs on (or denying benefits to) others. Vengeance clearly motivates behavior in private markets, and while it may be subject to opprobrium or other social sanctions when discovered it cannot be readily detected or deterred, especially when individuals quite sensibly keep their ulterior motives to themselves. Like altruism, vengeance value cannot be a legitimate element of BCA because it would implicitly reduce and could even eliminate the worth of some individuals merely due to the enmity of others.

The possibility of vengeance value is particularly troublesome in CV surveys where the payment vehicle encourages respondents to think about the payments

which would be made by others. An individual who actually places a low value on a specified commodity may overstate his true economic preference if he believes that everyone else in the community would be compelled to pay as well and he places a high value on imposing sacrificial expenditures on others.

Hypothesis Tests. A variety of tests may be devised to identify altruism (or vengeance) as an individual's underlying motive, but it is difficult to construct devices for filtering it out. One way to identify altruism may be to test otherwise identical surveys having different representation of the benefit. In one survey, only the respondent could enjoy the benefits of the program or policy of interest; in the other survey the same benefits would be distributed broadly. In the absence of altruism the median WTP of each survey sample should be the same. However, if the median WTP of the sample valuing a broadly distributed good exceeds the median WTP of the sample valuing an individually targeted good, then the former WTP estimate is tainted by altruism and should not be used for BCA. Unfortunately, we know of no way to tell whether altruism is embedded in the latter WTP estimate.⁶

As before, verbal protocols may be extremely useful for detecting altruism (or vengeance). CV respondents often reveal quite legitimate and salutary concerns for their friends and neighbors. (They are unlikely to reveal enmity even if they are highly motivated by it.) Further refinement of a survey instrument may decrease the prevalence of this phenomenon, but we do not know of any way to eliminate it. CV practitioners must devote extraordinary efforts to control the problem in cases where altruism is particularly likely to arise.⁷

"Absolution Value"

Except for problems such as altruism (and now vengeance), economists generally do not dwell on the motives behind individuals' preferences. One such motive, which to our knowledge has received no attention in the benefits valuation literature, involves an individual's WTP to be absolved of sin. The sin in question may be personal (e.g., alcohol or drug abuse) or societal (e.g., rain forest destruction). Environmental restoration projects, such as the cleanup of the federal government's nuclear weapons testing and manufacturing facilities, appear to be strongly motivated by such "absolution values." A verbal protocol CV survey might well reveal that many respondents place a relatively high value on cleaning up these facilities because of the shame they feel for the nation's nuclear heritage rather than the demonstrable human health or environmental risks posed by such sites.

Absolution value is a perfectly legitimate component of BCA under certain restrictive conditions. In particular, each individual's WTP must reflect *only* her own absolution. Once the absolution of others or society at large enters into the analysis, the same problems associated with altruism (and vengeance) arise.

An analogous situation could be observed in the food safety arena. Suppose that there are both naturally occurring and anthropogenic sources of a health risk of specified probability and magnitude. Suppose further that the median expressed WTP among survey respondents is significantly greater when the risk is anthropogenic than when it occurs naturally. Clearly, other factors besides risk *per se* have entered into the valuation exercise. It may be that respondents truly are willing to pay more to avoid risks if they are anthropogenic. However, it may also be the case that some respondents have embedded a societal absolution value into their expressed WTPs that is unrelated to their true economic preferences. It is the researchers' responsibility to ascertain these factors to ensure that if absolution value is present only the respondent and no others enjoy absolution.⁸

Hypothesis Tests. Two otherwise identical surveys could be administered in which the specific food safety risk in question has either a natural or anthropogenic source. If CV respondents are expressing WTPs related to the probability and magnitude of the underlying risk without concern for its origin, then the median expressed WTP will not be significantly different between these two surveys. If they are significantly different, however, then the origin of the risk matters and further work is necessary to explore the respondents' valuation processes. Other evidence of absolution value may be obtained from survey questions or verbal protocols. Special ethical or moral concerns about past individual or collective actions may be revealed.

Strategic Behavior Such as Free-Riding

A long-standing concern among economists is the possibility that survey respondents will give strategic answers to CV surveys. In particular, concern has focused on the effect of free-riding on expressed WTPs. Respondents may underbid if, for example, they fear that they will actually have to pay more than their true WTP. Alternatively, they may overbid if they expect the community average WTP will be substantially less than their own WTP.

The problem of strategic behavior in general (and free-riding in particular) has been addressed at length in the CV literature. Mitchell and Carson (1989: 133), for example, reject the traditional Samuelsonian formulation that free-riding is highly rational and inevitable in favor of an alternative view that it is both irrational and avoidable. They find support for this view in the experimental economics literature and in the development of a variety of incentive-compatible demand revelation devices, such as voting schemes and auctions.⁹

Hypothesis Tests. A broad review of the problems strategic behavior poses for CV analyses is beyond the scope of this chapter. However, Mitchell and Carson (1989: 165-168) suggest three types of hypothesis tests which represent minimum efforts to detect strategic behavior in CV surveys. Further, they

recommend procedures to be applied to deal with outliers symptomatic of strategic overbidding, such as the use of "trimmed means" (p. 369). The problem with all such devices, of course, is that researchers do not know *a priori* the precise level of expressed WTP that is "too high" to be plausible. Nor is there any acceptable methodology for upwardly adjusting zero (or just very low) values to account for possibly strategic underbids. Any technique used to remove implausible data and thus reduce bias runs the risk of introducing new biases by mistakenly altering strategic bids incorrectly, as well as altering bids which appeared suspicious but in fact reflected true WTP.

Payment Vehicles and Value Elicitation Formats

CV surveys use a variety of payment vehicles to obtain respondents' expressed WTP, including bidding games, payment cards, open-ended elicitation, and dichotomous choice referenda. Bidding games use an iterative process like an auction to ascertain respondents' highest WTP. In payment card surveys, respondents are asked to select the value on the card which most closely approximates their maximum WTP. Open-ended elicitation formats ask respondents to provide their maximum WTPs without such prior cues. Finally, referenda approaches ask respondents to vote for or against a single proposal based on an explicit dollar value price.

Many experienced CV practitioners have clear preferences for one or another of these formats, but there is no consensus in the CV literature in favor of any of them. Bidding games require the provision of a starting point which tends to impart a downward (upward) bias to the valuation exercise among respondents whose true WTP is above (below) the starting point. Open-ended value elicitation suffers from high nonresponse rates believed to arise from respondents' inability to specify prices without prior cues. The payment card approach helps solve this problem by providing such cues, but it also may cause starting-point bias similar to that observed in bidding games. The dichotomous choice referendum approach has the advantage of better approximating markets that consumers are familiar with in which prices appear to be set by the seller and not generally negotiable. It also reduces the likelihood of strategic behavior because respondents have less capacity to exaggerate their expressed WTPs. Unfortunately, dichotomous choice referenda also suffer from starting-point bias, they require a substantial increase in the number of surveys which must be administered, and assumptions must be made to parameterize the data to obtain estimates of the median WTP.

CV researchers also must confront the possibility that the expressions of WTP provided by survey respondents represent wild guesses or random values. This problem is exacerbated in CV designs in which respondents are pressured to provide answers. When a substantial percentage of respondents prefers not to assign a value—perhaps because the commodity in question is

unfamiliar—then further efforts to coax value elicitation increases the chance that the values provided will be random numbers, guesses, or otherwise poorly related to their true WTP.

Hypothesis Tests. At a minimum, CV researchers should test for the types of biases which can be expected to arise based on the value elicitation method they have employed. For example, if payment cards are used then it is necessary to show that the values included on these cards have not unwittingly created anchors. Two otherwise identical payment-card instruments could be administered where the value ranges on alternative payment cards overlap asymmetrically. Where both medians fall in the overlapping range, the median expressed WTPs should not be different when subjected to an appropriate statistical test.¹⁰

Further, CV researchers should search for the presence of guesses and random numbers. A relatively weak test would involve comparing the median expressed WTP from a payment card instrument lacking a "don't know" option with another instrument in which this alternative is not available. A significant difference in expressed WTP suggests a systematic difference in the way these different groups of respondents processed survey information and formulated their answers.

Framing Effects

Many psychologists and decision scientists have extensively criticized economists' reliance on theories of strict rationality. The fundamental dispute has been expressed recently as a debate between the perspectives of "articulated" versus "basic" values (Fischhoff 1991). The articulated values perspective assumes that individual responses to survey questions are rational and it attributes apparent contradictions to (sometimes imperceptible or unknown) differences in problem or question formulation. In contrast, the basic values perspective holds that individuals lack well-differentiated values except for the most familiar and noncontroversial evaluation efforts. CV methods are based on the articulated values paradigm; unsurprisingly, many of its critics subscribe to the basic values perspective.

This debate is well beyond the scope of this chapter. Nevertheless, the debate has revealed important issues surrounding survey methods such as CV which deserve to be addressed before CV estimates make the giant leap toward inclusion in BCA. For example, if the values respondents provide to early survey questions influence how they interpret and respond to later questions, individual respondents may force their answers to be internally consistent but demonstrate incoherence across respondents (Diamond 1992). More importantly, it is then unclear which values are the appropriate ones to rely upon.

Problems Using CV Results in Federal Policy Making

It is important to reiterate here that, given the breadth of methodological problems involved, CV surveys should not be used where there are market transactions which could provide the data necessary to perform traditional economic analysis. Proper CV survey design involves multiple alternative scenarios, multiple payment vehicles, verbal protocols, and a host of additional niggles likely to make the enterprise very expensive. Even if CV researchers can successfully solve problems such as those described briefly above, another set of issues arises regarding how CV results can be transferred to public policy applications.

Replication

Because of the range of methodological problems involved, CV studies that pass the necessary tests should be replicated, preferably by different researchers. The purpose of replication is to increase the likelihood that the initial study was not merely serendipitous. Government guidelines concerning quantitative risk assessment methodologies assert similar replication requirements as a means of deterring reliance on chance results (Office of Science and Technology Policy 1985, Environmental Protection Agency 1986).

At a minimum, results obtained in the replication should be consistent with those obtained initially. If the median expressed WTP across essentially identical studies is significantly different, then it is unclear which study should be deemed authoritative. Researchers may be tempted to use the study yielding the lowest median WTP as a "conservative" estimate of potential benefits. This temptation should be avoided, for it creates a variety of perverse incentives among both researchers and government officials. The objective always should be to develop unbiased estimates of baseline conditions and the costs and benefits of various intervention alternatives (Office of Management and Budget 1990a, 1990d).

CV results must be exceptionally robust where they could influence important public policy interventions. They must be able to withstand significant design perturbations without substantial deviation in median expressed WTPs. Again, if CV results are highly sensitive to survey design there is no clear basis for distinguishing among alternative designs.

Discrepancies Between Commodities in CV Studies and Proposed Government Intervention

A carefully constructed CV study may involve important simplifications and abstractions which are inconsistent with the manner in which government would

actually implement a food safety program, policy, or regulation. However, it is not obvious how benefit estimates derived from a successful but narrowly constructed CV study can be properly extrapolated to a government program or regulation whose characteristics are substantially different from those analyzed. For example, a CV study in which respondents are asked to value reduced risks from eating raw oysters may not be applicable to a government program of general seafood inspection.

Perceived Versus Expert Estimates of Risk

Where individuals make their own decisions concerning risk-taking and risk-avoidance, it is perfectly acceptable to rely on risks that are perceived even if they are substantially different from expert risk estimates. Individuals can expend resources to obtain better information, update prior risk estimates, and make different decisions. Ultimately, however, they alone bear the cost of their own errors.

Where governments make these decisions, however, they bear an obligation to use expert estimates of risk consistent with their fiduciary responsibility to properly care for other people's money. Failing to do so will waste resources and result in a suboptimal risk management portfolio: fewer health risks will be prevented for any given expenditure of societal resources. Further, once the government signals its intent to base policy decisions on perceived risks, it rewards those who strategically exaggerate their risk perceptions. Ignorance and intransigence become potential assets rather than liabilities.

Conclusion

Contingent valuation represents an important technique for attempting to estimate the value of commodities which are not traded in markets. Generally, there are market transactions with risk attributes available to study. Thus, CV may be useful where segments of the market demand curve do not exist and cannot be approximated by examining substitutes.

A decision to use CV opens up a large set of new methodological problems. Experienced CV practitioners believe that these problems can be overcome. CV critics argue that the problems are insurmountable and the method should be abandoned. Food safety researchers interested in CV should be wary of trying to finesse this conflict.

OMB's guidance concerning the use of CV-based estimates for benefit estimation is largely performance oriented. While it avoids taking any position on CV, it does impose an extraordinary burden on any methodology that does not rely on real-world behavioral data. Food safety researchers and government agencies interested in CV thus are well advised to explore this method with considerable care.

We do not believe that we are imposing an unfair or discriminatory burden on CV studies relative to traditional methods based on observed behavior. In particular, we expect that economic studies based on observed market behavior also demonstrate that consumers clearly understand the commodities they are purchasing; that their preferences conform to the Law of Demand; that they do not display casual benefit-cost analysis apart from that which is consistent with individual utility maximization; that if they display altruism or vengeance they bear as individual consumers the full cost of these motives; that they cannot free-ride; and that no calibration is necessary to render observed transactions consistent with elementary axioms of economic theory. Unless their express purpose is to demonstrate the existence of a significant market failure, we would reject for purposes of benefit-cost analysis any study based on market observations that failed any one of these requirements. By elucidating criteria and suggested hypothesis tests, it is our hope that practitioners of the new art of contingent valuation will be able to achieve the same level of legitimacy and respect commonly accorded to economic analysis derived from market observations.

Notes

1. The views expressed in this paper are those of the authors and do not necessarily represent the views of the Office of Management and Budget.

2. Executive Order No. 12866 was signed by President Clinton on September 30, 1993 (58 *Federal Register* 51735-51744). It superseded Executive Order No. 12291 signed by President Reagan on February 17, 1981 (46 *Federal Register* 13193-13198).

3. This document has been reprinted in each annual *Regulatory Program* published since 1990.

4. In the remainder of this chapter we assume that a decision to proceed with a CV approach has already been made despite these warnings.

5. CV practitioners may be tempted to purposefully characterize the food safety commodity in question as multidimensional because of the difficulty of satisfying both elements of the marginal value test. This temptation should be resisted. The presence of multiple dimensions increases the likelihood that respondents will fail to adequately understand the commodity, thus rendering the valuation exercise meaningless. In addition, reviewers and "consumers" of CV studies are increasingly likely to question a survey that has been designed in such a fashion that consistency checks are infeasible.

6. A similar experiment could be performed to test for the presence of vengeance value by altering the payment vehicle in analogous ways.

7. Examples may include situations where programs or policies are intended to benefit children, the elderly, the infirm, or the disabled.

8. Note that if absolute value is present the nature of the commodity is radically changed. Absent both private markets for absolute value and governmental mechanisms to provide it, estimates of WTP infected by absolute value are unlikely to have any practical use.

9. As indicated earlier in the discussion of calibration, the use of an incentive-compatible voting scheme or auction is not sufficient to overcome the hypothetical nature of the CV exercise.

10. If one of the medians is not in the overlapping range, then the median WTP for the other survey should be close to the relevant extreme value in its range.

References

- Bishop, Richard C. and Thomas A. Heberlein. 1986. Does Contingent Valuation Work? In *Valuing Environmental Goods: An Assessment of the Contingent Valuation Method*, ed. Cummings et al., Chapter 9. Savage, MD: Rowman & Littlefield.
- Buchanan, James. 1954. Individual Choice in Voting and the Market. *Journal of Political Economy* 62:334-343.
- Cambridge Economics, Inc. 1992. *Contingent Valuation: A Critical Assessment*. Cambridge, MA: Cambridge Economics.
- Cummings, Ronald G., David S. Brookshire, and William D. Schultz, eds. 1986. *Valuing Environmental Goods: An Assessment of the Contingent Valuation Method*. Savage, MD: Rowman & Littlefield.
- Diamond, Peter. 1992. Comment on "Methods of Measuring Non-Use Values: A Contingent Valuation Study of Groundwater Cleanup," by Gary H. McClelland, William D. Schultz, Jeffrey K. Lazo, Donald M. Waldman, James K. Doyle, Steven R. Elliot, and Julie R. Irwin. Unpublished comment submitted to Environmental Economics Advisory Committee of the Science Advisory Board, U.S. Environmental Protection Agency (December 14).
- Diamond, Peter A., Jerry A. Hausman, Gregory K. Leonard, and Mike A. Denning. 1992. Does Contingent Valuation Measure Preferences? Experimental Evidence. In *Contingent Valuation: A Critical Assessment*, Tab 3. Cambridge, MA: Cambridge Economics.
- Environmental Protection Agency. 1986. Guidelines for Carcinogen Risk Assessment. *Federal Register* 51(46):33991ff (September 24).
- Environmental Protection Agency. 1993. *Clean Water and the American Economy, October 19-21, 1992; Proceedings: Ground Water (Volume 2)*. Washington, D.C.: U.S. EPA, Office of Water (EPA 800-R-93-001b).
- Fischhoff, Baruch. 1991. Value Elicitation: Is There Anything in There? *American Psychologist* 46(8):835-847.
- Fisher, Ann, Lauraine G. Chestnut, and Daniel M. Violette. 1989. The Value of Reducing Risks of Death: A Note on New Evidence. *Journal of Policy Analysis and Management* 8(1):88-100.
- Freeman, A. Myrick. 1979. *The Benefits of Environmental Improvement*. Baltimore, MD: Johns Hopkins Press.
- Jamieson, Linda F. and Frank M. Bass. 1989. Adjusting Stated Intention Measures to Predict Trial Purchase of New Products: A Comparison of Models and Methods. *Journal of Marketing Research* 26:336-345.

- Kahneman, Daniel and Jack L. Knetsch. 1992. Valuing Public Goods: The Purchase of Moral Satisfaction. *Journal of Environmental Economics and Management* 22:57-70.
- Kemp, Michael A. and Christopher Maxwell. 1992. Exploring a Budget Constraint for Contingent Valuation Estimates. In *Contingent Valuation: A Critical Assessment*, Tab 4. Cambridge, MA: Cambridge Economics.
- McClelland, Gary H., William D. Schultze, Jeffrey K. Lazo, Donald M. Waldman, James K. Doyle, Steven R. Elliott, and Julie R. Irwin. 1992. *Methods for Measuring Non-Use Values: A Contingent Valuation Study of Groundwater Cleanup*. Boulder, CO: Center for Economic Analysis.
- Milgrom, Paul. 1992. Is Sympathy an Economic Value? Philosophy, Economics, and the Contingent Value Method. In *Contingent Valuation: A Critical Assessment*, Tab 11. Cambridge, MA: Cambridge Economics.
- Mitchell, Robert C. and Richard T. Carson. 1989. *Using Surveys to Value Public Goods*. Washington, D.C.: Resources for the Future.
- National Oceanic and Atmospheric Administration. 1993. Natural Resource Damage Assessments Under the Oil Pollution Act of 1990. *Federal Register* 58(10):4601-4614 (January 15).
- Neill, Helen R., Ronald G. Cummings, Philip T. Ganderton, Glenn W. Harrison, and Thomas McGuckin. 1994. Hypothetical Surveys and Real Economic Commitments. *Land Economics* 70(2):145-154.
- Office of Management and Budget. 1990a. Current Regulatory Issues in Risk Assessment and Risk Management. In *Regulatory Program of the United States Government: April 1, 1990-March 31, 1991*, 13-26. Washington D.C.: U.S. Government Printing Office.
- Office of Management and Budget. 1990b. Information as an Alternative Regulatory Strategy. In *Regulatory Program of the United States Government: April 1, 1990-March 31, 1991*, 26-32. Washington D.C.: U.S. Government Printing Office.
- Office of Management and Budget. 1990c. Regulatory Impact Analysis Guidance: Discussion of Comments. In *Regulatory Program of the United States Government: April 1, 1990-March 31, 1991*, 32-41. Washington D.C.: U.S. Government Printing Office.
- Office of Management and Budget. 1990d. Regulatory Impact Analysis Guidance. In *Regulatory Program of the United States Government: April 1, 1990-March 31, 1991*, 653-666 (Appendix V). Washington, D.C.: U.S. Government Printing Office.
- Office of Science and Technology Policy. 1985. Chemical Carcinogens: A Review of the Science and Its Associated Principles. *Federal Register* 50(50):10374ff (March 14).
- Viscusi, W. Kip. 1983. *Risk by Choice*. Cambridge, MA: Harvard University Press.