12-2017

Testing the Limits of Temporal Stability: Willingness to Pay Values among Grand Canyon Whitewater Boaters Across Decades

Christopher J. Neher
University of Montana, Missoula

John Duffield
University of Montana, Missoula

Lucas Bair
U.S. Geological Survey

David Patterson
University of Montana, Missoula

Katherine Neher
Bioeconomics, Inc.

Recommended Citation

Neher, C., Duffield, J., Bair, L., Patterson, D., & Neher, K. (2017). Testing the limits of temporal stability: willingness to pay values among grand canyon whitewater boaters across decades. Water Resources Research, 53, 10,108–10,120. https://doi.org/10.1002/2017WR020729

AGU PUBLICATIONS

Water Resources Research

RESEARCH ARTICLE

10.1002/2017WR020729

Key Points:

- Whitewater boater willingness to pay was compared across two studies conducted 30 years apart
- The comparison included estimated trip values for four alternative flow levels
- No significant differences in trip willingness to pay were found between the two studies, although differences in WTP functions were found

Correspondene to:
C. Neher,
neher@montana.com

Citation:
Neher, C., Duffield, J., Bair, L., Patterson, D., & Neher, K. (2017). Testing the limits of temporal stability: willingness to pay values among grand canyon whitewater boaters across decades. *Water Resources Research*, *53*, 10,108–10,120. https://doi.org/10.1002/2017WR020729

Received 9 MAR 2017
Accepted 1 NOV 2017
Accepted article online 7 NOV 2017
Published online 1 DEC 2017

Testing the Limits of Temporal Stability: Willingness to Pay Values among Grand Canyon Whitewater Boaters Across Decades

Chris Neher[1], John Duffield[1], Lucas Bair[2], David Patterson[1], and Katherine Neher[3]

[1]Department of Mathematical Sciences, University of Montana, Missoula, MT, USA, [2]U.S. Geological Survey, Southwest Biological Science Center, Grand Canyon Monitoring and Research Center, Flagstaff, AZ, USA, [3]Bioeconomics, Inc, Missoula, MT, USA

Abstract We directly compare trip willingness to pay (WTP) values between 1985 and 2015 stated preference surveys of private party Grand Canyon boaters using identically designed valuation methods. The temporal gap of 30 years between these two studies is well beyond that of any tests of WTP temporal stability in the literature. Comparisons were made of mean WTP estimates for four hypothetical Colorado River flow level scenarios. WTP values from the 1985 survey were adjusted to 2015 levels using the consumer price index. Mean WTP precision was estimated through simulation. No statistically significant differences were detected between the adjusted Bishop et al. (1987) and the current study mean WTP estimates. Examination of pooled models of the data from the studies suggest that while the estimated WTP values are stable over time, the underlying valuation functions may not be, particularly when the data and models are corrected to account for differing bid structures and possible panel effects.

1. Introduction

The Grand Canyon is an iconic symbol of both the American West and the success of the National Park System. While most visitors view the Colorado River from the vantage point of either the North or South rims of the canyon, far removed from its waters, each year many people recreate on the river as well. One highly prized Colorado River activity is securing a coveted private party permit to boat the world class whitewater from Lees Ferry through the Grand Canyon. Grand Canyon National Park issues both private and commercial permits to boat the Colorado River through the Grand Canyon. Of the approximately 24,000 annual visitors, 7,000 are private party whitewater boaters (Grand Canyon National Park, 2006).

In the 1980s a suite of Colorado River user surveys, hereafter referred to as the Bishop study, was undertaken to estimate how different water flow levels in the Colorado River and different scenarios for releasing water from Glen Canyon Dam impacted the value Grand Canyon whitewater boaters (among other river user groups) placed on their recreational river experience (Bishop et al., 1987). Since the original Bishop study, there have been numerous changes in river flow management and the status of the resource (Grand Canyon Monitoring and Research Center, 2005). Recognizing the length of time since the original Bishop study (nearly 25 years), in 2009 the National Park Service (NPS) funded initial stages of a study design for a replication and extension of the earlier work. However, only preparatory work could be done until 2014, when the Glen Canyon Dam Adaptive Management Program provided the additional funding needed to complete our 2016 Grand Canyon private party whitewater boater study beyond the initial design stage.

While our study was intended to update the nearly 30 years old Bishop study willingness to pay (WTP) estimates, we also recognized the opportunity to test the temporal stability of recreational WTP values from a high-profile, high-quality study over a period much longer than had been reported previously in the literature. Our study design sampled individuals who participated in a private party whitewater trip down the Grand Canyon in 2014 and 2015. Randomly selected private party boaters received a recreational use and value survey containing valuation questions that used a dichotomous choice contingent valuation (DCCV) question format with the same wording as in the earlier Bishop study. The DCCV questions were designed so responses and estimated WTP per trip could be directly compared to parallel results estimated from data collected in the Bishop study 30 years earlier.

2. Literature

In the late 1980s and 1990s, researchers began examining whether estimates of WTP derived from contingent valuation (CV) studies provided stable value estimates over time. The National Oceanic and Atmospheric Administration (NOAA) blue ribbon panel report on CV included a recommended method for assessing temporal stability of CV-based WTP estimates (Arrow et al., 1993). They suggested:

> Time-dependent measurement noise should be reduced by averaging across independently drawn samples taken at different points in time. A clear and substantial time trend in the responses would cast doubt on the 'reliability' of the finding.

Carson et al. (1997) specifically address the NOAA panel suggestion in a test of temporal stability of WTP by examining responses to CV questions related to the Exxon Valdez oil spill over a 2 year period. They find response proportions to alternative bid levels were stable over this period. While many researchers have examined the question of the temporal stability of WTP estimates (see e.g., Jorgensen et al. (2004) or Price et al. (2016) for discussions of this literature), the current study contributes to this body of work by examining CV-based WTP stability over a 30 year period, more than three times longer than the longest period reported in the literature (an 8 year period studied by Price et al. (2016)).

Studies of temporal stability of WTP values have generally fallen into two broad categories. The first is test-retest experiments where the same sample of individuals is surveyed at two points in time (Jorgensen et al. (2004) review a number of test-retest studies). While correlation of individual responses has varied somewhat within this literature, tests of estimated parameters and mean WTP have found no significant difference associated with dichotomous choice (CV) models over periods of 9 months (Loomis, 1990) and 3 years (Cameron, 1997).

The second broad class of WTP temporal stability studies report experiments where the same or very similar surveys are administered to different populations after a period of time. These types of studies have been used to examine the impacts to WTP of dramatic shifts in economic situation (Metcalfe & Baker, 2015) or transformational political events (Kountouris et al., 2012). Studies of WTP stability within less tumultuous settings have found that either controlling for temporal differences in socioeconomics or attitudes results in stable WTP over a 5 year period (Whitehead & Hoban, 1999), and modeling WTP in the absence of covariates (beyond income) results in transferable WTP estimates over a similar 5 year period (Brouwer & Bateman, 2005). Price et al. (2016) report on what they believe to be the longest time span for this type of WTP stability study (8 years, from 2004 to 2012) and find no significant differences between estimated single factor discrete choice mean WTP values across time.

In the case of boating down the Colorado River through the Grand Canyon, the nature of the recreational activity largely precludes consideration of a test-retest study design. For most Grand Canyon boaters the 225 mile trip, often lasting 18 days, is a "once-in-a-lifetime" experience. Additionally, the long wait periods and difficulty of securing one of the private party trip permits exacerbates the problems with ever identifying a substantial sample of repeat users of the resource.

While a test-retest experiment was not possible for the resource and activity studied, the nature of the Grand Canyon private party trip provides several unusual sample controls over time. First, participants are by necessity highly informed about the resource and the trip due to the substantial time, expenses, and administrative controls associated with securing and completing the trip. Even over a period of several decades, the very nature of the activity ensures that the sample population remains informed and engaged. Further, while there have been changes in flow release protocols (flow levels as well as daily flow fluctuations) from Glen Canyon Dam over the 30 year interval between studies, the primary physical characteristics of the Grand Canyon experience, including the nature of the rapids, side canyon hikes, and the scenic setting of the canyon, have remained generally stable over the period.

3. Methods

In DCCV, used in both the original Bishop study and our study, individuals respond "yes" or "no" as to their WTP a specific cash amount for a specified commodity or service. The advantages of this approach, as compared to open-ended or bidding game questions formats, have been discussed elsewhere (Bowker & Stoll, 1988; Boyle & Bishop, 1987). In DCCV it is assumed that if each individual has a true WTP, then the individual will respond positively to a given bid only if their WTP is greater than the bid (Cameron, 1988; Hanemann, 1984). For example, suppose that an individual is confronted with an offered price (t) for access to a given resource or recreational site. The probability of accepting this offer, $\pi(t)$, given the individual's true (unobserved) valuation, or WTP, is then:

$$\pi(t)=\Pr(WTP > t)=1-F(t) \quad (1)$$

where *F* is a cumulative distribution function (c.d.f.) of the WTP values in the population. In the logit model, *F(.)* is the c.d.f. of a logistic variate, and in the probit model *F(.)* is the c.d.f. of a normal variate. The specification of this model can be briefly illustrated for the case where the WTP values are assumed to have a logistic distribution in the population of interest. The probability of a "yes" response to bid amount *x* is

$$P=\frac{1}{1+e^{-a-bx}} \quad (2)$$

where the parameters to be estimated are α and *b*. This can also be expressed as:

$$L=\ln\left[\frac{P}{1-P}\right]=\alpha+bx \quad (3)$$

where *L* is the "logit" or log of the odds of a "yes" and *p* are observed response proportions. Maximum likelihood estimates of the parameters can be obtained with a logistic regression program. We have utilized SAS Proc Logistic analysis program (SAS Institute, 1988).

Consumer WTP is generally defined as a nonnegative value. Therefore, estimates of the expected value of WTP in this study used integration over the range from 0 to $+\infty$. This estimate is generally referred to as the conditional mean, which is often shown as E(WTP|WTP $\geq$ 0). The expression for the conditional mean consumer surplus for a logit specification that is linear in terms of its parameters, derived by Hanemann (1989) is

$$\mathrm{E}(\mathrm{WTP}|\mathrm{WTP}\geq 0)=\frac{\ln(1+e^{\alpha})}{-b} \quad (4)$$

The standard errors and confidence intervals for these value estimates were estimated using the Krinsky and Robb (1986) simulation method.

4. Data

A sample of potential respondents for the survey was randomly drawn from a list of all September 2014 to August 2015 private party whitewater boaters, provided by the NPS. As was done in the Bishop study, we utilized a repeat contact mail-back survey method to gather data (Dillman, 2007). Our survey procedure included mailing an initial postcard, a full survey information packet, a reminder postcard, and finally a second full survey packet to those who had yet to respond. Our survey had a relatively high final response rate of 65% of deliverable surveys.

The original Bishop study was a fairly early high profile application of DCCV WTP estimation. Administered in an era where response rates to mail surveys were much better than those achieved today (AAPOR, 2016), the Bishop study survey of private whitewater boaters had a final response rate of nearly 85%. The Bishop study utilized a series of DCCV questions asking boaters to value different hypothetical river flow scenarios. While more recent DCCV question designs commonly use randomly assigned bids from a predetermined limited set of specific bid levels for the DCCV questions, this early study used randomly generated bid levels over an interval, resulting in near unique bid amounts presented to each respondent in each question. The observed range of bid levels presented in the four flow scenario DCCV questions by Bishop et al. was from $3 to $2,521 (in 1985 dollars).

One goal of setting DCCV *a priori* bid levels is to ensure that the top bid level captures a significant portion of the upper tail of the WTP distribution. Examination of the responses to the bids presented in the Bishop study showed that for the 108 bid levels exceeding $1,000, only three respondents said they would pay this

Table 1
Characteristics of Surveys and Willingness to Pay Question Design

	Current study	Bishop study
Whitewater trip dates	1 Sep 2014 to 31 Aug 2015	26 Feb to 6 Nov 1985
Flow scenarios modeled (cubic feet per second (cfs))	5,000; 13,000; 22,000; 40,000	5,000; 13,000; 22,000; 40,000
Population	NPS-identified private trip participants	NPS-identified private trip participants
Sample size	404	171
DCCV bid levels presented	$40; $90; $275; $650; $1,200; $2,600 (2015 dollars)	$3 to $2,521 (1985 dollars)[a]
Model specification	Logistic regression-linear specification	Logistic regression-linear specification
Average experienced flow level of survey respondents	12,065 cfs	26,000 cfs

[a]Note: the 1985 data incuded one BID observation $>$ $2,521 ($3,237). This single outlying bid was retained in the analysis for consistency with the original 1985 models.

relatively high amount. Therefore, our study design adopted the same bid range, in the belief that a top bid of $2,600 would also capture much of the upper tail of the WTP distribution. In the current study, amounts asked for each of the DCCV questions were randomly varied between a set of bids ranging from $40 to $2,600 (Table 1 outlines key characteristics of our study and the Bishop study surveys). Examination of the DCCV responses to our study showed that the "percent yes" responses to the top ($2,600) bid level ranged from 5.5% to 20.2% for the range of Colorado River flow scenarios presented to respondents.

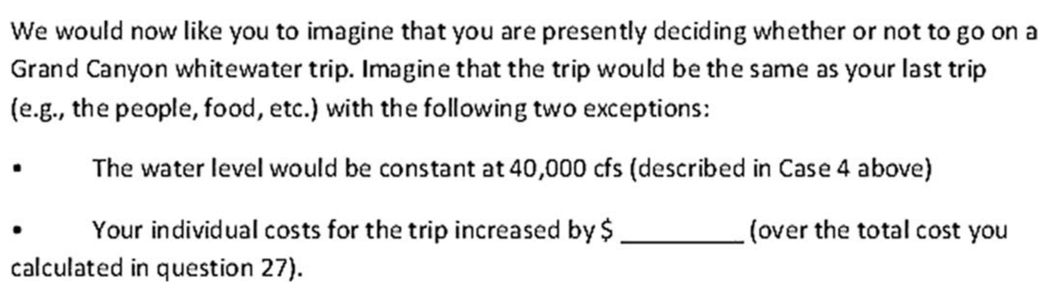
The format of the DCCV questions in our survey was designed to present the exact information and use the exact question wording as was employed in the Bishop study. As is standard in the literature, the DCCV question format asked boaters whether they would be willing to pay $XXX more than they actually did for their trip (Figure 1). Contingent valuation questions were asked for valuation of four hypothetical constant flow scenarios (5,000, 13,000, 22,000, and 40,000 cubic feet per second (cfs)).

5. Results

The 30 year gap between the original Bishop study and our study made comparison of logistic regression models and WTP estimates challenging. In order to estimate identical functional forms and mean WTP values, the original Bishop study data were obtained. Using the Bishop study data, we verified that the estimated conditional mean WTP values using a bivariate logistic regression model and a linear specification were nearly identical to those reported by Bishop et al. (1987), which included an additional covariate for trip expenditures. These replicated Bishop study results were compared to our models that were identically specified and estimated.

5.1. Comparison of Boater Preferences and Perceptions of Flow

Private boaters on the Grand Canyon are an extremely specialized and engaged population. For our survey responses, 64% reported having primary responsibility for operating a boat on their trip. Further, the average length of trip was 18 days. The common thread of necessary experience and commitment required by


We would now like you to imagine that you are presently deciding whether or not to go on a Grand Canyon whitewater trip. Imagine that the trip would be the same as your last trip (e.g., the people, food, etc.) with the following two exceptions:

- The water level would be constant at 40,000 cfs (described in Case 4 above)
- Your individual costs for the trip increased by $ _________ (over the total cost you calculated in question 27).

Figure 1. Structure of dichotomous choice contingent valuation survey question.

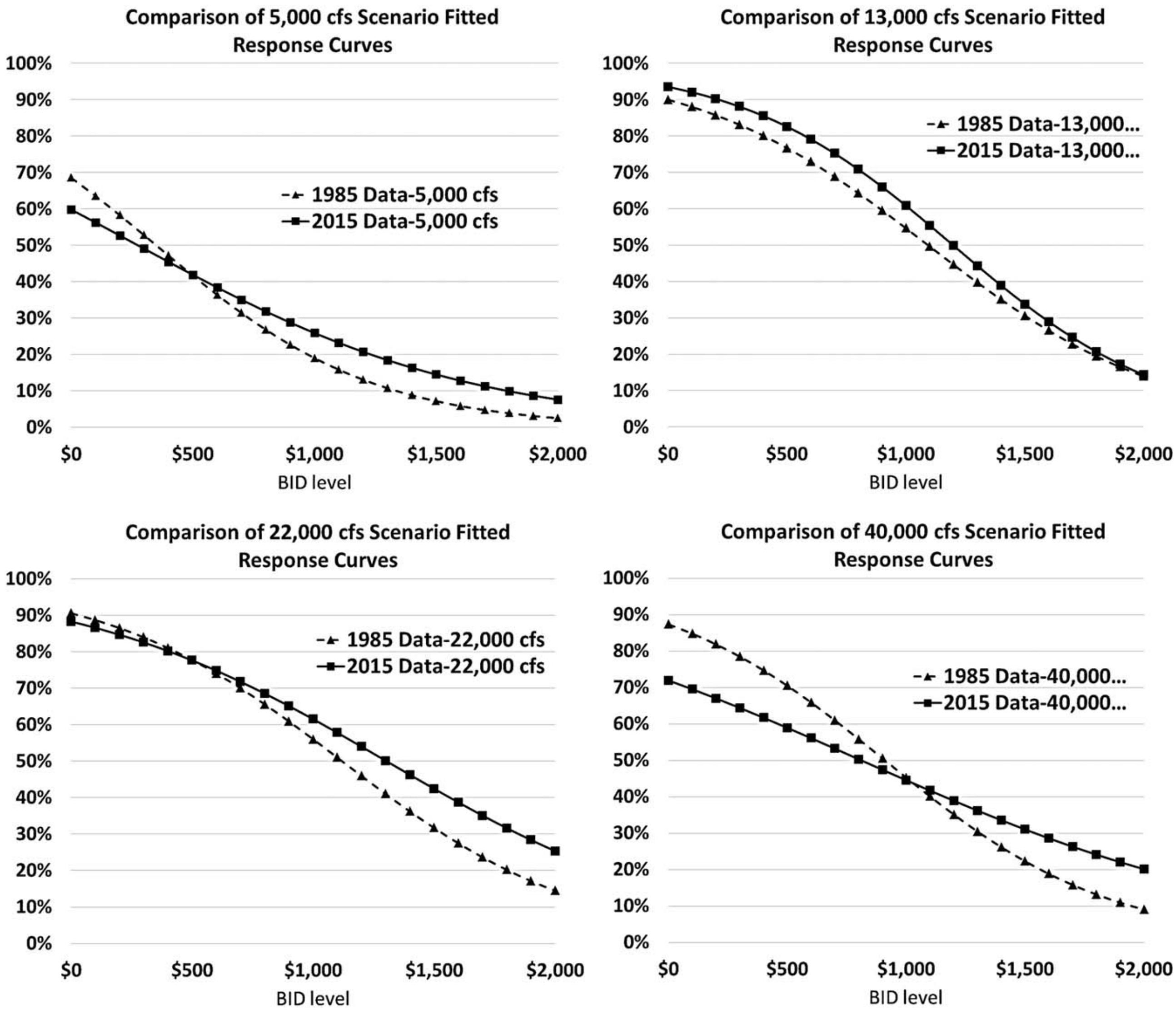


Figure 2. Comparison of 1985 and 2015 logistic probability response curve, by hypothetical flow scenario (note: response probability for 1985 data scaled to 2015 price levels).

boating the Grand Canyon on a private trip also is shown in remarkable stability of preferences voiced by boaters across time. Both the suite of Grand Canyon user surveys by Bishop et al. (1987) and our survey asked boaters to rate the importance of 24 separate features of a Grand Canyon whitewater trip. It is of note that both the choice and ranking of the top five trip features in the Bishop study and our study were identical. They were: (1) being in the unique natural setting of Grand Canyon, (2) being in a natural setting, (3) stopping at side canyons or creeks, (4) hiking the side canyons, and (5) observing flora, fauna, and geology. Further, Stewart et al. (2000) compared the Bishop study ratings for these trip features to a 1998–1999 survey of private whitewater trip leaders (which used the same question). They also found strong consistency in ratings of the importance of trip features between the Bishop study and their survey.

Respondents to both the Bishop study and our study were presented with four alternative hypothetical flow scenarios and asked a series of questions about each. The Bishop study and our surveys included DCCV questions on trip attribute scenarios not examined in this comparison. The current study compares only the four primary alternative constant flow level scenarios presented in both surveys, which were 5,000, 13,000, 22,000, and 40,000 cfs. After a description of each scenario, respondents were asked whether that scenario would be better or worse than conditions they experienced on their own recent Grand Canyon

Table 2
Comparison of Whitewater Boater Perceptions of Alternative Flow Levels

Constant river flow level	Better		About the same		Worse	
	Bishop study	Current study	Bishop study	Current study	Bishop study	Current study
5,000 cfs	4%	1%	3%	4%	93%	95%
13,000 cfs	25%	17%	36%	75%	39%	7%
22,000 cfs	30%	50%	66%	40%	1%	10%
40,000 cfs	8%	24%	36%	13%	56%	64%

whitewater trip. Table 2 shows a comparison of responses to these questions between the Bishop study and our study.

At a hypothetical flow of 5,000 cfs, the Bishop study and our results are very similar, with 93% and 95% of respondents, respectively, indicating those levels would be worse than the ones they experienced. More pronounced differences begin to arise at 13,000 cfs—75% of the participants in our study said these levels would be "about the same" as the ones they experienced, compared to 36% in the Bishop study. This disparity is consistent the change in river flows in the past three decades. Where those sampled in the Bishop study experienced an average flow of 26,000 cfs (thus prompting 66% to say 22,000 cfs would be about the same as their experience), those sampled in our 2015 study experienced an average flow of 12,065 cfs and appropriately 75% said 13,000 cfs was "about the same" as they experienced. Moreover, the average high flow experienced by respondents in the Bishop study was 29,200 cfs. Among those sampled in our study, on the other hand, only six respondents experienced flows of 20,000 cfs or more. These circumstances likely also account for the differences in boater perceptions of a hypothetical flow of 40,000 cfs; those surveyed in our study, not having experienced flows even close to that level on their trips, indicated such a scenario would be better than their experience 24% of the time (compared to 8% in the Bishop study).

To a great extent, the actual baseline flows experienced by the 1985 and 2015 floaters fell within the range of the two middle hypothetical flow levels (13,000 and 22,000 cfs). Despite the large differences in the baseline experienced flows from the 1985 and the 2015 survey responses, for the remaining two hypothetical flows, which fell largely outside of the baseline actual experienced flows, preferences between the studies showed consistency. Both studies reflected the strong belief that a flow levels of both 5,000 and 40,000 cfs would be "worse" than what respondents actually experienced.

5.2. Comparisons of Models of WTP at Alternative Hypothetical Flow Levels

As noted, while the 2015 survey presented respondents with a set of six possible bid levels ($0 to $2,600) in the DCCV questions, the earlier 1985 survey (with a few exceptions) used a unique bid level for each respondent in each question presented. This difference across bid structures complicates the comparison of inspection of response proportions across bid levels. Table 3 shows a generalized comparison of response proportions between the two studies. In order to more appropriately compare the studies, a two-step process was used. First, the six bid levels in the 2015 survey were deflated to 1985 dollars using the CPI-U. Thus a range of $40 to $2,600 became $18 to $1,172. Next the bid levels from the 1985 survey were aggregated into ranges (0–18, 19–40, 41–124, 125–293, 294–541, 542–1,172) and response proportions (% yes responses) to bids within these ranges were calculated.

The fact that the original 1985 bid structure used nearly unique bid levels for each respondent combined with a relatively small sample size (171 individuals) leads to challenges in comparing responses from the two studies. Table 3 shows that 9 of 24 bid ranges in the table for the Bishop data have five or fewer "yes" responses. Therefore, these cells of the table are very sensitive to the shift of even one answer within the range. Additionally, in Table 3 nearly all cell comparisons between the studies show the Bishop et al. data with higher percentages of "yes" responses then the 2015 data at individual bid levels. This is as would be expected with the Bishop data being aggregated across a number of bids for each bid range. At a given bid level in the distribution, the threshold bid faced by respondents to the 2015 survey was (for instance) $293, while the aggregated responses from the Bishop study faced a range of threshold bids from the lowest bid in the range ($125 in this case) up to $293. It is therefore expected that the aggregated Bishop et al.

Table 3
Comparison of DCCV Response Proportions, by Study Year and Bid Range

	5,000 cfs		13,000 cfs		22,000 cfs		40,000 cfs	
Bid level	Bishop study % yes (n = yes)	Current study (2015)	Bishop study % yes (n = yes)	Current study (2015)	Bishop study % yes (n = yes)	Current study (2015)	Bishop study % yes (n = yes)	Current study (2015)
$18	75%	64%		96%		96%	100%	71%
	(3)	(45)		(69)		(54)	(3)	(49)
$40	86%	59%	100%	96%		95%	100%	75%
	(6)	(38)	(2)	(75)		(59)	(2)	(51)
$124	45%	43%	100%	91%	100%	80%	85%	77%
	(5)	(26)	(12)	(70)	(6)	(56)	(11)	(53)
$293	52%	35%	85%	64%	94%	64%	67%	47%
	(15)	(23)	(28)	(39)	(17)	(49)	(14)	(30)
$541	19%	18%	44%	48%	50%	48%	47%	21%
	(9)	(14)	(18)	(26)	(19)	(39)	(17)	(15)
$1,172	6%	5%	27%	7%	23%	18%	17%	20%
	(3)	(4)	(18)	(5)	(22)	(12)	(12)	(14)

percentages would tend to be higher than the 2015 percentages, since the average effective bid faced for each bid range was somewhat lower.

Comparison of the DCCV response proportions across studies and bid levels/bid ranges, while constrained by significant differences in bid structure between the studies as well as limited sample size in the 1985 survey, still shows consistency with theory in that all four flow levels in both studies show a pattern of declining bid acceptance as bid levels increase.

All of the estimated models of WTP—both in the Bishop study and our study—showed the expected parameter signs (a negative sign on bid level), and all but one estimated parameter were statistically significant at the 99% level of confidence (the remaining parameter was significant at the 95% level) (Table 4).

Table 4
Estimated Bivariate Models of Whitewater Boater Willingness to Pay, by Flow Scenario

	5,000 cfs	13,000 cfs	22,000 cfs	40,000 cfs
Current study				
Intercept	0.3958	2.6686	2.0180	0.9394
(Standard Error (S.E.))	(0.1498)	(0.2258)	(0.1892)	(0.1474)
Bid	−0.00145	−0.00223	−0.00155	−0.00116
(S.E.)	(0.000205)	(0.000231)	(0.000167)	(0.000152)
Sample Size	412	410	412	411
Conditional Mean WTP[a]	$628	$1,227	$1,382	$1,094
(S.E. of Mean)	($72)[b]	($87)	($102)	($112)
Bishop study				
Intercept	0.7768[c]	2.1939	2.2592	1.9277
(S.E.)	(0.3595)	(0.4053)	(0.4811)	(0.4151)
Bid	−0.00494	−0.00445	−0.00447	−0.00470
(S.E.)	(0.000992)	(0.000733)	(0.000773)	(0.000774)
Sample Size	171	170	171	170
Conditional Mean WTP[a]	$234	$517	$528	$439
(S.E. of Mean)	($32)	($43)	($39)	($39)

[a]Current study results in 2015 dollars and Bishop study results in 1985 dollars.
[b]Standard errors were simulated with 10,000 random draws using the method by Krinsky and Robb (1986).
[c]Significant at 95% level of confidence. All other estimated parameters significant at 99% level of confidence.

The conditional mean WTP values in Table 4 reflect only the original bid and response levels with no adjustments for changes over time. Much has changed between the original survey years of 1985 and 2015 in terms of income levels, price levels, and economic conditions. Additionally, tastes and preferences for outdoor recreation have evolved as well during this period. While there are different approaches to indexing WTP across time, we have adopted the approach suggested by Hensher et al. (2012) of adjusting WTP estimates *ex post* using changes in the Consumer Price Index for all urban consumers (CPI) (U.S. BLS at https://data.bls.gov/cgi-bin/surveymost?cu) between 1985 and 2015.

In addition to indexing for changes in price levels, changes in income between the 1985 and 2015 surveys were also considered. Income was included as a covariate in the eight WTP models in order to examine the stability of income elasticity between the two studies. While the coefficient on income had the expected sign (+) in all models, it was only statistically significant in one of the eight models, leading to very large levels of uncertainty in estimated elasticities. Lacking a strong empirical basis for adjusting for income over time, only an adjustment for changing price levels (CPI-U) was used.

In order to visually compare the estimated logistic BID response curves implied by the Table 4 models, the 1985 study BID coefficients

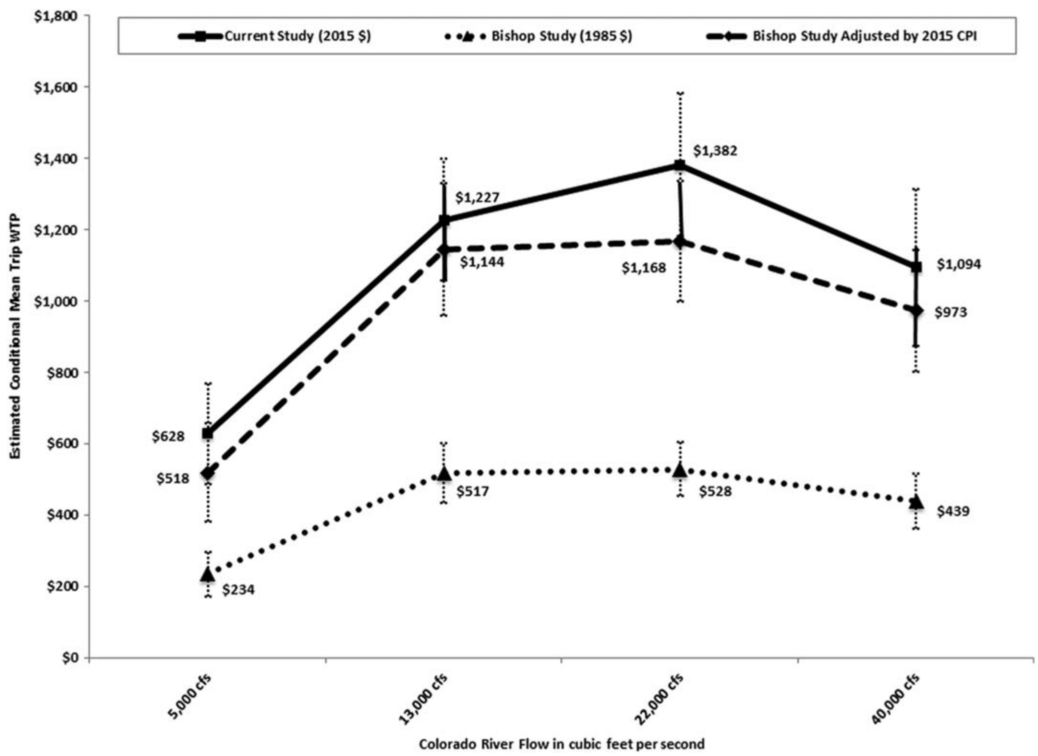


Figure 3. Comparison of current study and bishop study dichotomous choice contingent valuation willingness to pay estimates with 95% confidence intervals. (note: solid portions of error bars indicate area of overlap).

were scaled to reflect the change in the CPI-U between the 1985 and 2015 study years. Figure 2 shows comparisons of the estimated probability of a "YES" response to alternative BID levels between the 1985 and 2015 data models for each of the four hypothetical flow level scenarios. Generally, the predicted probabilities from the two studies show consistency and clearly highlight the relative attractiveness of alternative flow scenarios.

The plot of both original and adjusted WTP from the Bishop study in comparison to our study shows two things of note (Figure 3). First the shape of the traced WTP "curves" follow the a priori expectation that WTP for alternative flow levels follows a generally quadratic form, with lower WTP at both very low and very high flow levels, and higher WTP at intermediate flows (Loomis & McTernan, 2014). At very low river flows whitewater boaters spend more time on the river and have less time at camp and for hiking. Also, at low flows river navigation can become more difficult due to exposed hazards. At high flows river navigation can become challenging due to the size of rapids and campsite beaches become more limited. A scope test of differences between mean values for alternative flow scenarios using the conditional mean WTP and associated standard errors from Table 4 also provides support for the inverted-U relationship of value to flow level. A two-tailed test of differences in mean values shows WTP for the three higher flows (13,000, 22,000, and 40,000 cfs) significantly higher ($p = 0.001$) than WTP for the 5,000 cfs level for both studies. Furthermore, a test between $WTP_{(22,000)}$ and $WTP_{(40,000)}$ shows a decline in WTP over this interval that is significant at the 90% level of confidence in both the 1985 data ($p = 0.09$) and the 2015 data ($p = 0.10$).There are no significant differences between WTP estimates for 13,000 and 22,000 or 13,000 and 40,000 for either data set. A second finding is that adjusted WTP from the Bishop study closely traces the WTP values from our study.

A statistical comparison of the inflation/income-adjusted mean WTP values is shown in Table 5. The null hypothesis is:

$$\boldsymbol{H}_0 : \; WTP_F^{Current \;\; Study} - WTP_F^{Bishop \;\; Study} = 0 \tag{5}$$

Where F denotes the specific flow level scenario being compared. The alternative hypothesis is one of no equality between the two means.

Table 5
Comparisons of Conditional Mean Willingness to Pay (WTP) Values Across Studies and Value Adjustments

	5,000 cfs	13,000 cfs	22,000 cfs	40,000 cfs
Conditional mean WTP				
Current Study	\$628	\$1,227	\$1,382	\$1,094
(Standard Error (S.E.))	(\$72)	(\$87)	(\$102)	(\$112)
Bishop Study Indexed by CPI	\$518	\$1,144	\$1,168	\$973
(S.E.)	(\$71)	(\$95)	(\$86)	(\$86)
Wald test (significance of difference in mean WTP)				
Current Study WTP—Bishop Study CPI-Adjusted WTP	1.18 ($p = 0.28$)	0.41 ($p = 0.52$)	2.59 ($p = 0.11$)	0.73 ($p = 0.48$)

$$H_A: \; WTP_F^{Current \; Study} - WTP_F^{Bishop \; Study} \neq 0 \qquad (6)$$

Using the Wald test for differences between the estimated mean values showed no significant difference between the adjusted Bishop study means (using either indexing method) and the means from our study. Additionally, conditional mean WTP values are fairly precisely estimated. The simulated standard errors (S.E.) of the WTP means range from 7% to 13% of the mean values. The inflation/income-adjusted S.E./mean ratios for the Bishop study are as precisely estimated as those from our study (Table 5). With simulated 95% confidence intervals ranging from +/–14% to +/V25% across the models, we conclude that the lack of significance in the difference between the means from the Bishop study to our study is not due simply to imprecise parameter estimation.

A more robust comparison of mean WTP values is simply to compare the respondent rankings of the alternative flow scenarios (as reflected in the estimated WTP). The global ranking of the flow scenarios is the same between the studies, separated by 30 years' time.

6. Sensitivity of WTP Functions to Model Specification

The preceding comparison of WTP estimates from the parallel 1985 and 2015 surveys of Grand Canyon whitewater boaters utilizes the linear functional form of individual flow scenario models to consistently mirror the estimation methods and results reported by Bishop et al. for the four constant flow scenario DCCV estimates. Comparison of the stability of WTP estimates helps to inform the decision of whether to employ "value transfer" in a specific application of benefits transfer. It is also of interest to examine, given the constraints of the data, the degree to which the estimated WTP functions are stable over time. An alternative to estimating eight separate models (four flow scenarios for each of the two studies) is to estimate one pooled model incorporating all the data (Table 6). In this pooled model, BID levels from the 2015 study are indexed to 1985 price levels for consistency. Model 1 includes indicator variables for three of the flour flow levels and for the SAMPLE (0 = 1985 study; 1 = 2015 study) as well as an interaction term between SAMPLE and BID. Model 1 showed highly significant coefficients for both the BID and BID*SAMPLE interaction term. Additionally, two of the three flow scenario indicator variables were also significant at the 99% level of confidence.

One clear lack of comparability between the 1985 and 2015 data lies in the range of the BID levels. While generally consistent in nominal terms, when the bids are adjusted for inflation (both scaled to 1985 price levels) the 1985 bids range much higher than the 2015. To test the sensitivity of the model results to this bid structure, we eliminated all observations in the 1985 data (N = 57) greater than \$1175 (a value consistent with the upper bid level in the 2015 study), Model 2 shows the results of this truncated data analysis. The results of Model 2 are generally consistent with those shown in Model 1. A comparison of the Log Likelihood statistics from Models 1 and 2 show a better fit in the case of the truncated data model (Model 2).

The bivariate models presented in Table 4 assume independence between the four DCCV responses from each survey respondent. It is reasonable, however, to expect these responses for a given individual to be correlated. This is a common issue related surveys where respondents are asked a series of DCCV questions. The Bishop et al. (1987) private party whitewater DCCV data was analyzed for this effect by Poe et al. (1997). Using the Bishop data, Poe et al. found that statistically correcting for cross-question correlation "has only a very small effect on distributions of estimated mean WTP. . . Confidence ranges change only slightly, if at all, between the independent and joint models." However, Poe et al. further cautioned that in cases of high correlation between WTP responses correcting for this correlation may lead to slight efficiency gains in parameter estimation. They note that in cases where the significance in the difference in two means being compared falls near a critical value (i.e., 0.05 in our case) this increase in estimation efficiency may force a difference to cross a significance level. Based on the Poe et al. results the pooled model was re-estimated to account for within respondent correlation.

Model 3 in Table 6 shows the pooled truncated data logistic regression model (Model 2) with a random effect specified for unique survey respondents. In our estimated random effects model, we assume

Table 6
Estimated Pooled Model of Both 1985 and 2015 DCCV Trip/Flow Valuation Responses

	Model 1	Model 2	Model 3	Model 4
Parameter	Pooled model	Pooled model with truncated upper BID	Pooled random effects model with truncated upper BID	Pooled random effects model with flow and sample interactions and truncated upper BID
Intercept	2.416***	2.4629***	2.5032***	2.2754***
	0.2244	0.2293	0.2337	0.3765
BID	−0.0046***	−0.00471***	−0.0049***	−0.0047***
	0.00038	0.000397	0.0005	0.0008
BID*SAMPLE	0.00114***	0.00124***	0.0014***	−0.0003
	0.000422	0.000436	0.0005	0.0010
F5000	−1.6610***	−1.6681***	−1.6741***	−1.5377***
	0.1497	0.1501	0.1247	0.5084
F22000	−0.0602	−0.0691	−0.0811	0.2614
	0.1492	0.1496	0.1181	0.4951
F40000	−0.8062***	−0.8091***	−0.7980***	−0.2016
	0.1480	0.1484	0.1304	0.5589
SAMPLE	−0.333	−0.3750*	−0.4767**	0.3378
	0.2160	0.2205	0.2368	0.4293
F5000*SAMPLE				−0.7177
				0.5609
F22000*SAMPLE				−0.9145*
				0.5524
F40000*SAMPLE				−1.5084**
				0.6072
F5000*BID				−0.0003
				0.0012
F22000*BID				−0.0005
				0.0009
F40000*BID				−0.0004
				0.0011
BID*F5000*SAMPLE				0.0020
				0.0014
BID*F22000*SAMPLE				0.0020*
				0.0011
BID*F40000*SAMPLE				0.0028**
				0.0013
Sample size	2327	2269	2269	2269 obs; 587 clusters
AIC	2320.44	2309.216	QICu = 2311.06	QICu = 2307.5402
−2(Log L)	2306.44	2295.216		

Note. SAMPLE coding: 0 = 1985 data; 1 = 2015 data. Flow indicators are coded 1/0. Random Effect in Models 3 & 4 is respondent ID.
*Significant at the 90% C.I; **Significant at the 95% C.I; ***Significant at the 99% C.I.

heterogeneity across survey respondents and that this heterogeneity can be modeled by a probability distribution. By using this approach, the correlation between DCCV questions from the same respondent arises from the specific but unobserved properties of those respondents which are constant across the four DCCV questions. A random effects logistic regression model is given by:

$$L_{ij} = \ln\left[\frac{P}{1-P}\right] = \alpha + b * BID_{ij} + u_i \quad (7)$$

where BID_{ij} is equal to the bid level for respondent i to DCCV question j, with $u_i \sim N(0,\nu 2)$.

As was found by Poe et al. (1997), the inclusion of a random effect for individual respondents in Model 4 led to some slightly improved efficiency in parameter estimates.

Models 1–3, through the BID*SAMPLE variable allows BID to vary between the two samples, but constrains BID across flow levels. A more complete examination of the data is shown in Model 4. This last model includes interactions between flow scenarios, bid level, and SAMPLE, as well as the required underlying constituent variables. The estimated coefficients on the three-way interaction terms included in Model 4 show statistical significance at the 90% level for BID*F22000*SAMPLE, and at the 95% level for BID*F40000*-SAMPLE, suggesting SAMPLE has differing effects on BID response across flow scenarios.

A final examination of functional form was a comparison of the individual bivariate models (Table 4) to the parallel set of models using a log transformation of the BID variable. Comparisons of the results from the two functional forms shows all coefficients significant at the 95% level in all models with the expected signs on the coefficients. Comparison of the AIC statistics for the model pairs shows the linear specification produces a superior model in five of the eight models, with the log transformation marginally better in three models. Based on the close correspondence between the specifications and the marginally better fits of the linear models, the linear specification was employed in this analysis.

One additional issue related to the design of both the 1985 and the 2015 surveys bears mentioning. When presenting a series of DCCV questions within a survey, the impact of question order on survey responses can be of concern (Cummings et al., 1986; Mitchell & Carson, 1989). In designing the original 1985 survey, Bishop et al. varied the DCCV question order among respondents. They reported no statistical difference in WTP between differing question orders. This result was confirmed by Boyle et al. (1993) in a more detailed analysis of the 1985 data. The later paper reported that for the private party floater sample (which was the sample replicated in the 2015 study) no effects associated with question order were present. Based on these findings, the 2015 survey was designed without randomly ordered survey questions.

Models 1 through 3 presented in Table 6 suggest that after correcting the models for differing bid structures (Model 2) and possible panel effects (Model 3), we could conclude that while the estimated WTP values between the two studies are reliable over time, the underlying WTP functions may not be. The results from Model 4 are consistent with this finding and further suggest statistically significant variation among the three-way interactions which include FLOW, SAMPLE, and BID.

7. Conclusions and Limitations

This paper compares the conditional mean WTP per trip between Grand Canyon whitewater boaters in the Bishop study and our recent study. We find no significant difference between the mean WTP per trip over four hypothetical river flow scenarios assessed in both studies. The 30 year interval between the Bishop study and our recent study certainly tests the concept of temporal stability of WTP as reported in the current literature. While the nature and quality of the original study and the sample population is somewhat unique, the comparisons of WTP estimates show that given this type of controlled population and careful study design, estimated WTP values can show temporal stability over a period of several decades. This result should bolster the confidence of decision makers in relying on value estimates from somewhat dated high-quality WTP studies. It must be noted, however, that the results of a pooled model of the data also suggests possible differences in the underlying WTP functions between the two studies, particularly when the data and model are adjusted to correct for differing bid structures and possible panel effects.

The distinct nature of the Grand Canyon whitewater experience and the financial, experiential, and institutional requirements for undertaking this trip all serve to preserve a very stable population of resource users over time in terms of recreational experience and knowledge of the resource. However, both the original Bishop study, and by reference our study, benefited from extensive and careful qualitative research by Bishop et al. (1987) during their study and survey design. The Bishop study research team undertook numerous pilot surveys and attribute surveys in order to inform the questions asked in their final DCCV survey. The comparison is further supported by the fact that the Grand Canyon whitewater boating experience is defined by a limited number of important trip attributes (e.g., rapids, side-canyon hiking) that are strongly influenced by river flow scenarios. Other aspects of the whitewater experience (e.g., geological history, remote wilderness setting) are very stable over time. These qualities strengthen our confidence in the results of the comparison of conditional mean WTP across studies.

Natural extensions of this analysis could include testing the stability of WTP estimates beyond the case of bivariate WTP modeling. Specifically, comparable covariates from the two studies could be added to the WTP models and examined as to how these additions impact WTP. This extension would be particularly relevant given the finding of instability of the WTP functions over time. Estimating the impact of comparable covariates across studies could provide insight into how demand for trip attributes is changing over time. This is important when resource managers are consistently faced with balancing tradeoffs in managing resources (e.g., hydropower, recreation, ecosystem services) below Glen Canyon Dam in the Grand Canyon.

While the consistency of the estimated WTP values between the 1985 and 2015 surveys provides some comfort in the stability of WTP estimates across time, care should still be taken in application of these findings to other settings. First, temporal reliability in WTP estimates is achieved only if there is no or little change in the primary characteristics of a resource or recreational experience and also no change in WTP across time. Alternatively, temporal reliability may be achieved if there are changes in the resource or experience and WTP changes accordingly. Therefore, stability of WTP by itself is not evidence of temporal stability in the case of valuing Grand Canyon whitewater trips. However, as discussed above, there is strong evidence that the attributes of a Grand Canyon float trip cited as most important to boaters have remained consistent between the 1985 and 2015 surveys. Additionally, those top-rated five trip characteristics involve resource and trip attributes that have remained virtually unchanged between the studies (characteristics involving the unique setting of the canyon, use of side canyons, and observing flora, fauna, and geology). A further source of consistency across time concerns the flow scenarios presented. As described in the surveys, and in actuality, the effect of the presented flow levels on the whitewater experience has remained unchanged over time as well. Finally, the excess demand for private party Grand Canyon float permits that existed in 1985 still is evident today, even though the process of permit allocation has been modified. These factors provide evidence that the consistency between the estimated WTP values from the two studies truly reflects temporal stability and reliability of the Bishop et al. (1987) estimates.

A corollary to the previous point is that the unique nature of the Grand Canyon float experience and the protections and preservation of the resource itself combine to make it an ideal candidate for a long-term test of temporal stability of WTP. However, care should be taken in applying these findings to other settings without carefully examining potential changes in the resource, demand for its use, or user preferences related to that use. In the case of Grand Canyon whitewater floating, not only was the original study a carefully conducted high-quality DCCV application, but additional analyses based on the 1985 data have been published to bolster the confidence in the original study. This would likely not be the case in many other settings where recreational experiences are not as high-profile or the resource not as world-renowned. Therefore, as in any other case of benefits transfer, researchers should take care to verify the stability of the underlying recreational experience and user preferences for that experience before applying the conclusion of temporal stability found in the case of Grand Canyon floaters to other distant and likely quite different settings.

Acknowledgements
The authors would like to thank the three anonymous reviewers for their thoughtful and informed questions and suggestions. Modeling and analysis suggestions provided by the reviewers have substantially improved the analysis. Funding for the research underlying this paper was provided by the National Park Service Social Sciences Division and the Glen Canyon Dam Adaptive Management Program. Substantial assistance in the early stages of this work was provided by Bruce Peacock. Grand Canyon National Park personnel efficiently provided needed listings of Grand Canyon boaters. Data underlying this analysis is archived as a U.S. Geological Survey Data Release (Neher et al. 2017). Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government. All errors and omissions are the responsibility of the authors.

References

AAPOR (2016, March 3). *Response rates: An overview*. Retrieved from http://www.aapor.org

Arrow, K., Solow, R., Leamer, E., Portney, P. R., Radner, R., & Schuman, H. (1993, January 15). Report of National Oceanic and Atmospheric Administration panel on the reliability of natural resource damage estimates derived from contingent valuation, *Federal Register*, *58*, 4601–4614.

Bishop, R. C., Boyle, K. J., Welsh, M. P., Baumgartner, R. M., & Rathbun, P. R. (1987). *Glen Canyon Dam releases and downstream recreation: An analysis of user preferences and economic values*. (Glen Canyon Environ. Stud. Rep. GCES/27/87). Springfield, VA: Huberlein-Baumgartner Research Service. (Supplemental note: (NTIS Rep. PB88–182546/AS). Madison, WI: National Technical Information Service.)

Bowker, J. M., & Stoll, J. R. (1988). Use of dichotomous choice methods to value the whooping crane resource. *American Journal of Agricultural Economics*, *70*, 372–381.

Boyle, K. J., & Bishop, R. C. (1987). Valuing wildlife in cost-benefit analysis: A case study involving endangered species. *Water Resources Research*, *23*(5), 943–950.

Boyle, K. J., Welsh, M. P., & Bishop, R. C. (1993). The role of question order and respondent experience in contingent-valuation studies. *Journal of Environmental Economics and Management*, *25*(1), S80–S99.

Brouwer, R., & Bateman, I. J. (2005). Benefits transfer of willingness to pay estimates and functions for health-risk reductions: A cross-country study. *Journal of Health Economics*, *24*(3), 591–611.

Cameron, I. I. (1997). Applying Socio-ecological economics: A case study of contingent valuation and integrated catchment management. *Ecological Economics*, *23*, 155–165.

Cameron, T. A. (1988). A new paradigm for valuing non-market goods using referendum data: Maximum likelihood estimation by censored logistic regression. *Journal of Environmental Economics and Management*, *15*, 355–379.

Carson, R. T., Hanemann, W. M., Kopp, R. J., Krosnick, J. A., Mitchell, R. C., Presser, S., . . . Martin, K. (1997). Temporal reliability of estimates from contingent valuation. *Land Economics*, *3*, 151–163.

Cummings, R. G., Brookshire, D. S., & Schulze, W. D. (1986). *Valuing environmental goods: An assessment of the contingent-valuation method*. Totowa, NJ: Rowan and Allenheld.

Dillman, D. (2007). *Mail and internet surveys: The tailored design* (2nd ed.). New York, NY: Wiley and Sons, Inc.
Grand Canyon Monitoring and Research Center (2005). *SCORE (Status of the Colorado River Ecosystem) report*. Flagstaff, AZ: Grand Canyon Monitoring and Research Center.
Grand Canyon National Park (2006). *Colorado river management plan*. Grand Canyon, AZ: Office of Planning and Compliance, Grand Canyon National Park, Department of Interior.
Hanemann, W. M. (1984). Welfare evaluation in contingent valuation experiments with discrete responses. *American Journal of Agricultural Economics*, *66*, 332–341.
Hanemann, W. M. (1989). Welfare evaluation in contingent valuation with discrete response: Reply. *American Journal of Agricultural Economics*, *66*, 332–341.
Hensher, D., Rose, J., & Li, Z. (2012). Does the choice model method and/or the data matter? *Transportation*, *39*, 351–385.
Jorgensen, B. S., Syme, G. J., Smith, L. M., & Bishop, B. J. (2004). Random error in willingness to pay measurement: A multiple indicators, latent variable approach to the reliability of contingent values. *Journal of Economic Psychology*, *25*, 41–59.
Kountouris, Y., Nakic, Z., & Sauer, J. (2012). *Is the valuation of water quality sensitive to external shocks? Evidence from political instability in Croatia*. Paper prepared for presentation at the Agricultural and Applied Economics Association's 2012 AAEA Annual Meeting, Seattle, Washington.
Krinsky, I., & Robb, A. L. (1986). On approximating the statistical properties of elasticities. *Review of Economics and Statistics*, *68*(4), 715–719.
Loomis, J., & McTernan, J. (2014). Economic value of instream flow for non-commercial whitewater boating using recreation demand and contingent valuation methods. *Environmental Management*, *53*, 510–519.
Loomis, J. B. (1990). Comparative reliability of the dichotomous choice and open-ended contingent valuation techniques. *Journal of Environmnental Economics and Management*, *18*, 78–85.
Metcalfe, P. J., & Baker, W. (2015). The sensitivity of willingness to pay to an economic downturn. *Journal of Environmental Economics and Policy*, *4*, 105–121.
Mitchell, R. C., & Carson, R. T. (1989). *Using surveys to value public goods: The contingent valuation method*.. Washington, DC: Resources for the Future.
Neher, C., Duffield, J., Bair, L., Patterson, D., & Neher K. (2017). *Grand canyon whitewater boater data, temporal stability of willingness to pay values*. Reston, VA: U.S. Geological Survey. https://doi.org/10.5066/F7M044C9
Poe, G. L., Welsh, M. P., & Champ, P. A. (1997). Measuring the difference in mean willingness to pay when dichotomous choice contingent valuation responses are not independent. *Land Economics*, *73*(2), 255– 267.
Price, J., Dupont, D., & Adamowicz, W. (2016). As time goes by: Examination of temporal stability across stated preference question formats. *Environmental and Resource Economics*, 1–20. https://doi.org/10.1007/s10640-016-0039-x
SAS Institute (1988). *SAS/STAT users guide. Release 6.04 edition*. Cary, NC: SAS Institute.
Stewart, W., Larkin, K., Orland, B., Anderson, D., Manning, R., Cole, D., . . . Tomar, N. (2000, April). *Preferences of recreation user groups of the Colorado river in grand canyon* (57 pp.). Flagstaff, AZ: Grand Canyon Monitoring and Research Center.
Whitehead, J., & Hoban, T. (1999). Testing for temporal reliability in contingent valuation with time for changes in factors affecting demand. *Land Economics*, *75*(3), 453–465.