The Effect of the Number of Choice Sets on Responses in a Stated Choice Survey

Matthew Heberling[1], Jim Shortle, and Ann Fisher

The Pennsylvania State University
Armsby Building
University Park, PA 16802

Session 4: Monday, July 31

Abstract: An important issue for the stated choice method is the effect of the number of choice sets on responses. Based on a study of this issue in a mailed survey, results indicate that the number of choice sets does not affect survey response rates or item non-response rates.

Paper presented at the 2000 AAEA meeting, Tampa, Florida

Copyright 2000 by Matthew Heberling, Jim Shortle, and Ann Fisher. 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.

---

[1]Email: mth127@psu.edu

Introduction

The stated choice method (SCM), which is similar to conjoint analysis, is drawing considerable interest as a technique for valuing environmental goods with multiple attributes, although only a few studies have valued environmental and natural resource goods (e.g., Adamowicz et al. 1997; Hanley et al. 1998). With this method, preference data are collected by asking respondents to choose between pairs (or among triplets) of alternatives (called choice sets) described by a number of characteristics (called attributes), and their levels (Adamowicz et al. 1997). Some form of price or cost is generally one attribute (e.g., travel cost or increased taxes) (Hanley et al. 1998). Results from the SCM can yield the ranking of attributes, willingness to pay for a unit increase/decrease in attribute levels, and an overall willingness to pay for a specified combination of attributes.

The SCM has not received the scrutiny that contingent valuation has (Arrow et al. 1993), but it has several promising features by comparison. The growing interest among economists lies in some of these features. One is that it allows for attributes of the environmental good to be valued rather than the total good (which is important for benefits transfer (Hanley et al. 1998)). Some problems encountered in dichotomous choice contingent valuation problems (e.g., such as yea-saying) are avoided. Extrapolation problems, common with revealed preference methods, also are avoided by allowing researchers to choose levels outside those currently observed. By design, another promising feature of the SCM is that respondents are required to make tradeoffs for different attributes (a market like decision) rather than overall environmental quality and money. This tradeoff process, however, could be a double-edged sword.

To generate preference data, the researcher must direct respondents to think about many tradeoffs over multiple choice sets. However, the respondents' capacity to ponder those tradeoffs stands in the way of clean preference data. Respondents may lack experience making tradeoff decisions about some environmental or natural resource goods. In fact, the tradeoff process can be quite complicated. Beyond the issues with the tradeoff process, other factors such as fatigue or boredom may affect choices, especially when a large number of choice sets are presented to respondents. Therefore, more research is necessary to design the optimal SCM questionnaire (Hanley et al. 1998).

One design issue that has received little attention is the effects of presenting more choice sets to a respondent. More choice sets per respondent mean more observations, reducing data collection costs. However, the task of answering more choice sets could cause fatigue or boredom. For example, time constraints or fatigue may limit how many choice sets respondents will answer or answer well. Fatigue effects have been found in some SCM studies (Bradley and Daly 1994). The salience of the survey topic, the difficulty of understanding the good, and the SCM task all influence fatigue.

The noise, or unexplained variation, associated with making choices may increase if respondents tire quickly. Swait and Adamowicz (1996) incorporate task complexity into their statistical model to understand choice. One aspect of complexity in their model is prior effort or "cumulative cognitive burden" which is a function of the number of choice sets. They found that the number of choice sets can affect the amount of noise in the data. However, there has been little research on how fatigue (and complexity) affect item and survey non-response (Adamowicz et al. 1998). If the task takes a lot of time or is frustrating, the respondent will be more likely to skip questions or to cease participation. Skipped questions (i.e., item non-response) could cause questionnaires to be unusable. Knowing how the number of choice sets affects item non-response allows one to assess how to maximize the amount of data from individuals. Similarly, too many unreturned questionnaires (i.e., survey non-response) can reduce the credibility of the results.

Individuals are expected to optimize their decision making efforts such that the benefits are greater than the costs. If decision making is based on choices in a questionnaire, respondents can be expected to throw away a questionnaire or skip questions if the decision making costs outweigh the benefits. Hence, more choice sets (i.e., greater costs) would mean more item non-response due to fatigue and boredom, and the group that receives the most choice sets would have the greatest survey non-response.

In this research, I analyze survey and item non-response rates when the number of choice sets is varied. Performing hypothesis tests, I am able to determine whether the number of choice sets affects survey response rates and whether the number affects the quality of responses.

Studies of Survey Response and Item Non-response Rates

Most research addressing response rates and questionnaire length prior to 1990 is inconsistent (see Bogen 1996; Herzog and Bachman 1981; Dillman et al. 1993). Some researchers discover a negative relationship between response and length (Heberlein and Baumgartner 1978; Childers and Ferrell 1979) while others find a positive relationship between response and length (Champion and Sear 1969).

After 1990, several studies also addressed the relation between length of questionnaire and survey response rates. Dillman et al. (1993) found higher response rates for shorter census surveys than for a longer booklet form. However, the shortest form did not have the highest completion rate, conveying the possibility of a lower limit for improving the response rate. Pennings et al. (1999) examined response rates for a mailed survey of farmers. Their conclusions suggest that farmers spend only 10 minutes on any questionnaire, so the length of the questionnaire negatively affected their response rate. Hoffman et al. (1998), using an epidemiologic study, analyzed the effects of the length, incentives, and follow-up techniques on response rates to a mailed survey. Their analysis found no significant difference in survey response rates between a 16 item, 4 page booklet and a 76 item, 16 page booklet.

Few studies test differences in data quality (i.e., item non-response) related to the length of a questionnaire. Dillman et al. (1974) found larger item non-response rates in the final quarter of a census questionnaire, but felt it was due to irrelevant or difficult questions following the usual socio-economic questions. Herzog and Bachman (1981) examined the effect of questionnaire length on response quality. They found that respondents are more likely to answer the same way for identical response scales near the end of the questionnaire (called straight-lining). Mooney et al. (1993) studied monetary incentives and item non-response. They concluded that a combination of no incentive along with a long questionnaire led to lower response rates. However, data quality was not affected by questionnaire length.

Given the paucity of studies and their lack of consistent results, it is appropriate to examine non-response and task complexity. The following section provides context for the tests conducted to shed light on optimal design for the SCM.

Acid Mine Drainage: A Case Study

Acid mine drainage (AMD), water containing high levels of metals and sulfuric acid, affects some 2400 miles of Pennsylvania streams, mostly in western and central Pennsylvania (Bureau of Abandoned Mine Reclamation 1997). It would be prohibitively expensive to clean up all 2400 miles in any short period, so priorities must be set. Results from a mail survey that explores the benefits of restoring stream damaged by AMD can help evaluate which AMD waterways produce the most benefits after clean up. The results also can help identify the optimal level of restoration.

The AMD questionnaire collects information using four sections: water quality issues, acid mine drainage issues, outdoor recreation participation, and background information. To get respondents thinking about water quality issues, the first section determines respondents' concern about local water pollution sources and their steps taken, if any, to ensure clean drinking water (i.e., purchasing bottled water or using water filters). The next section gathers information about respondents' awareness and concern about AMD. This section also includes the SCM choice sets. Outdoor recreation participation questions follow in the third section. The final section collects socio-economic information such as age, education, and income.

The survey was conducted for a random sample of Pennsylvania residents in three locations known for acid mine drainage problems: Clearfield County, Northumberland County, and parts of Huntingdon and Bedford Counties known as the Broad Top Region. Questionnaires along with a $2 incentive were mailed to a total of 2208 residents evenly distributed among the three regions. The Total Design Method (Dillman 1978) was modified by omitting the registered mail follow-up. Of the 2208 questionnaires mailed, 264 (12%) were returned because of a bad address or deceased. Another 52 questionnaires were sent back blank. In some cases, respondents explained that they were not experienced enough to answer questions about acid mine drainage or not concerned about acid mine drainage. These were considered unuseable. A total of 1171 responses were useable in the survey and item non-response study; thus, the final response rate is 60.2%.

Designing Choice Sets

Key steps in designing SCM choice sets include identifying and selecting the attributes, choosing the range and increments of levels for each attribute, and deciding the method for combining levels into choice sets (Louviere and Timmermans 1990, Hensher 1994, Adamowicz et al. 1998, and Hanley et al. 1998). For this study, ecologists and engineers provided insight into the possible cleanup levels and their approximate costs. Ecologists and engineers may not provide information that can be transferable to the questionnaire because it does not correspond to how people view the environmental goods. Therefore, focus groups are essential to the questionnaire development. Focus groups, composed of residents from the three regions, provided us with people's perceptions of AMD, stream attributes, and attribute levels. They also expressed concern about the number of attributes and attribute levels they could consider while answering multiple choice sets. Finally, the participants suggested that after six or eight choice sets they became bored and tired.

To estimate the benefits of mitigating AMD, five attributes describe site restoration scenarios (Water Quality, Miles Restored, Travel Time from Home, Easy Access Points, and Household Cost). These attributes form the basis for the choice sets. The levels of Water

Quality include able to swim; able to fish and swim; and able to drink, fish, and swim. Miles Restored indicates the extent of the cleanup. Time from Home explains how long it will take to reach the improved stream. Easy Access Points describes the boat ramps, parking lots and simple trails to the stream or river. Finally, Household Cost, in the form of increased water bill payments for 10 years, provides information to value changes in the attributes. The attributes and their levels appear in Table 1.

The total number of alternatives possible is X^A where X is the number of levels and A is the number of attributes with X levels. This study uses a 3^3 x 2^1 x 6^1 orthogonal main effects design. The main effects design is based on the assumption that all interaction terms between attributes are insignificant. This means the utility function is strictly additive and includes no interaction terms for the attributes (Adamowicz et al. 1998).

To generate alternatives for our choice sets, we utilize a computerized search developed by Zwerina et al. (1996). The search generates a statistically efficient subset of alternatives from the set of all possible alternatives by minimizing the size of the covariance matrix (Zwerina et al. 1996). It allows the incorporation of many attributes and levels; this increases design efficiency, and reduces the number of required choice sets.

Our choice sets comprise of two alternatives and a choice of "neither." A choice of neither (the opt-out choice) can be interpreted as a status quo, or do nothing, option. An example of a choice set is in Table 2. A total of twenty choice sets are generated to examine non-response. Four questionnaire versions (versions 3, 4, 5, and 6) include five choice sets and two versions (versions 1 and 2) consist of ten choice sets. Version 1's choice sets are the same as version 3's and version 4's combined, while version 2's choice sets are the same as versions 5 and

6 combined. Using this research design, we can test differences in response rates. The two versions with ten choice sets were sent to 330 residents each while the four versions with five choice sets were sent to 387 residents each[2].

Estimation of Indirect Utility

Choice set data can be used to estimate the marginal values of watershed restoration attributes. Using a random utility model, we can examine tradeoffs between attributes, and analyze the change in economic welfare. To develop this model, consider an individual faced with the decision of choosing between alternative streams for restoration. Let the individual's utility function have an observable part that includes income, costs of the alternatives, the alternative chosen, and a vector of characteristics for individual i that affects her preferences. It may be expressed as

$$U_{ji} = V_{ji}(Y_i,\ cost_j,\ alt_j,\ char_i)$$

where U_{ji} is the utility of person i choosing stream j and V_{ji} is the observable component of utility (Freeman 1993). Now suppose the researcher cannot observe the utility function perfectly. Let ε_{ji} represent the random, unobservable part that incorporates all researcher and/or respondent error (Adamowicz et al. 1998). This allows a utility function to be estimated even though it is not completely observed.

Let stream j and stream k have observable environmental quality attributes; person i chooses j over stream k. From this observation, person i reveals that $U_j > U_k$. Because ε_{ji} is

[2]The sample sizes were chosen to satisfy the asymptotic properties for maximum likelihood estimation suggested in Adamowicz et al. (1998).

random, the probability that stream j is chosen over all k in set C depends on the probability that $U_j > U_k$. This is given by (Adamowicz et al. 1998):

$$Prob_i(j \mid C) = Prob_i(V_{ji} + \varepsilon_{ji} > V_{ki} + \varepsilon_{ki}, \forall\ k \in C)$$

To estimate this probability, assumptions must be made on the distribution of the differences in errors. If the differences in errors are independent and identically distributed, independent of irrelevant alternatives (IIA), and Gumbel distributed, the probability of choosing option j is given by [3]:

$$Prob(j \mid C) = \frac{e^{\mu V_j}}{\Sigma_{k \in C}\ e^{\mu V_k}}$$

where the scale parameter, μ, is inversely related to the variance of the error term. Typically, it is assumed equal to one (Hanley et al. 1998). This equation can be estimated using the conditional logit regression (Adamowicz et al. 1998; Hanley et al. 1998; Greene 1997).

Results

The variables in this indirect utility model describe the characteristics of the acid mine drainage clean up site (see Table 3). The quality of water (DRINKABLE and FISHABLE) illustrates the possible level of clean up. EASY ACCESS indicates how hard it is to get to the site. MILES RESTORED defines the extent of clean up. TRAVEL TIME indicates how long it takes to get to the site from the respondent's home and COST describes the increase in the water bill for the next 10 years to pay for the clean up. Alternative specific constants, dummy variables

[3] See Adamowicz et al. (1998) or Greene (1997) for details.

for Site A and Site B, are included to measure nonparticipation effects, and allow for unobservable attributes to influence utility (Hanley et al. 1998).

The results for the ten choice set and five choice set versions are estimated using a logit model (see Table 4). Observations for respondents who choose the status quo option for all choice sets are interpreted as "don't know" responses or protest responses, and dropped from the analysis (Adamowicz et al. 1998). This omits 37 respondents from the five choice set versions and 11 respondents from the ten choice set versions.

In the five choice set model, DRINKABLE and FISHABLE are positive and significant, meaning the respondents prefer more clean up to less. EASY ACCESS and MILES RESTORED are positive and significant, indicating that people prefer easy access to limited access and more miles restored to less. The variable TRAVEL TIME is negative and significant revealing the preference for a site close to home. Respondents who receive five choice sets rank the attributes (more important to less) as DRINKABLE, FISHABLE, EASY ACCESS, MILES RESTORED, and TRAVEL TIME. The chi-square statistic rejects the null hypothesis that none of the variables is significantly different from zero.

Two variables are insignificant in the ten choice set model (FISHABLE and EASY ACCESS). Respondents in this group rank the attributes as DRINKABLE, MILES RESTORED, and TRAVEL TIME. The chi-square statistic also rejects the null hypothesis that none of the variables is significantly different from zero.

After estimating an aggregate data model (not shown), we can test whether there are systematic differences in the coefficients (Ben-Akiva and Lerman 1985). The likelihood ratio test is $LR = -2 * [L_R - \sum_{g=1}^{G} L(g)_{UR}]$ where L_R is the log likelihood for the aggregate model, and

$L(g)_{UR}$ is the log likelihood of the model for the g[th] market segment. This test is chi-square distributed test with the degrees of freedom equal to $\sum_{g=1}^{G} R_g - R$ where R_G is the number of coefficients in the g[th] market segment (Ben-Akiva and Lerman 1985). The log likelihood for the aggregate model is -5101.40, and the total sum of the log likelihood for the segments is -5057.12. With 8 degrees of freedom, and the test statistic calculated at 88.56, I can reject the equality of coefficients in the ten choice set and five choice set questionnaire versions. Additional results, such as welfare measures for individual regions, can be found in Heberling (2000).

Survey Response Rates

The next step is to analyze how the number of choice sets affect survey response. This section examines overall response rates and the next section examines item non-response rates. Table 5 presents the response rate for all six versions of the questionnaires; versions 1 and 2 have ten choice sets (18 pages of questions) and versions 3, 4, 5 and 6 have five choice sets (13 pages of questions). The response rate for both version 1 and 2 combined is 59.7% while the response rate for all versions with five choice sets is 60.5%. Table 6 displays response rates for each region.

Survey response rates are analyzed by inspecting the proportions of two populations (Anderson et al. 1993). The test statistic is

$$\bar{p} = \frac{n_i p_i + n_j p_j}{n_i + n_j}$$

$$S_{p_i - p_j} = [(\bar{p})(1 - \bar{p})(\frac{1}{n_i} + \frac{1}{n_j})]^{0.5}$$

$$z = \frac{(p_i - p_j)}{S_{p_i - p_j}}$$

where p is the response rate for questionnaire versions i and j, p-bar is the pooled estimator, n is the sample size for versions i and j, S is the point estimator for standard deviation and z is the test statistic.

When comparing V1 to all five choice set versions, no significant differences between response rates are found. Similarly, when comparing V2 against all five choice set versions, no significant differences are found. Even comparing V1's response rate to the response rate of both V3 and V4 is insignificant. Similar results occur when comparing V2 with V5 and V6. In fact, the largest difference in response rate (V4 vs. V6) is also insignificant at the 10% level. Using the same test statistic, Broad Top region consistently has the low response rate which is significantly different from the other two regions at the 10% significance level (see Table 7).

Item Non-response Rates

Data quality is as important as whether people respond, so it is important to address the effects of the number of choice sets on item non-response (one proxy for data quality). For the questionnaire versions with ten choice sets, respondents are asked a total of 72 questions (some questions have multiple parts). Respondents who receive the five choice set versions face a total of 62 questions. The difference between the two sets are the extra choice sets and rating questions. Rather than analyzing non-response for all questions, we drop the follow-up questions to two filter questions. One filter question asks respondents whether they are concerned about the impacts of AMD on stream quality in their area (Q4) and the other asks respondents if they participate in outdoor recreational activities (Q7). Both Q4 and Q7 were included in the item non-response analysis, but their follow-up questions were not (removing 21 questions from the analysis). Numerous respondents did not respond for those recreational activities that they do not participate in and many did not respond for the impacts of AMD that they are not concerned about. Before investigating item non-response, the percent of responses with no missing items are calculated. About a third of all respondents in each version have no items missing (see Table 8).

To test item non-response, individuals' item non-response rates are averaged[4]. Table 9 presents the mean non-response rates for each version along with the rates for comparing V1 with V3 and V4 and V2 with V5 and V6.

Respondents' item non-response also may increase as they progress through the questionnaire. The questionnaire can be divided into three sections: 1) the 1st five questions (i.e., two water quality questions, and three acid mine drainage questions), 2) ranking the attributes and choice sets (i.e., the majority of information about acid mine drainage was presented just before this group of questions), and 3) the outdoor recreation participation questions along with typical background information questions. Table 10 displays the item non-

[4]One observation from version 4 was dropped as an outlier (95% item non-response rate). The respondent only answered age, and type of water system used at home.

response for the different sections of the questionnaire.

A t-test reveals whether the number of choice sets affects average item non-response rates. Having independent populations (or individuals who receive different questionnaires), the significance test statistic is (Ostle and Mennsing 1987):

$$t = \frac{(\overline{X}_1 - \overline{X}_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

where $\overline{X}_i$ is the mean, n_i is the number of observations, and s_i^2 is the variance.

Table 11 presents the mean tests for comparing version 1 and version 2 against the five choice set versions that are combined to create them. The results indicate that version 1 and version 2 do not have a significantly different item non-response than the individual versions combined to create them. Even comparing both version 1 and version 2 to any of the five choice set versions indicates no significant differences.

The final question is whether respondents increase item non-response as they progress through a questionnaire (from Table 10). There is a slight trend of increasing item non-response through the three sections of the questionnaire. However, testing for significant differences reveal that item non-response does not, on average, increase as respondents proceed. Also, comparing the sections across the number of choice sets indicates no significant differences (e.g., 1st five questions in versions 1 and 2 vs. 1st five questions in versions 3, 4, 5 and 6).

Table 12 describes the item non-response tests that were completed, but not displayed. We tested whether any questionnaire versions were significantly different from each other, whether there were regional differences in item non-response rates, and whether there were significant differences for the questions that followed the choice sets. All tests failed to show significant differences.

Conclusions

How the difficulty of a questionnaire task affects survey response rates and item non-response rates is assessed in a stated choice survey to value the benefits of cleaning acid mine drainage. The difficulty of the task is measured by the number of choice sets. Two questionnaire versions include ten choice sets and four versions include five choice sets.

Results show that survey responses do not differ across the number of choice sets. This rejects the hypothesis that a more difficult questionnaire will cause survey non-response, at least for the difference between choosing for ten choice sets rather than five choice sets. However, survey response rates differ across regions. Broad Top region's response rate is significantly lower than Clearfield or Northumberland Counties in Pennsylvania.

Mean item non-response rates also do not appear to be affected by the number of choice sets. This contradicts the hypothesis that a more difficult questionnaire will cause respondents to skip questions. All tests failed to show any significant differences between mean item non-response rates.

Unfortunately, this information alone does not reveal the whole picture. Learning or fatigue effects are possible factors for the differences in indirect utility models. However, survey non-response and item non-response may not be the reason the models differ.

Further Research

A more complete study would compare two more questionnaire versions: one that includes no choice sets and another with all the choice sets generated (in this case, 20 choice sets). Although differences from an extra 10 questions were not found, an extra 40 questions may cause significant differences. Tests between no choice sets and some choice sets could have significant differences between both response rates and item non-response rates. The expense of a larger sample size and two more questionnaire versions made it infeasible to include them in this research.

Table 1: Attributes Describing Site Restoration

Attributes	**Levels**
Water Quality	Able to Swim Able to Fish and Swim Able to Drink, Fish and Swim
Miles Restored	5 miles 20 miles 50 miles
Time from Home	10 minutes 30 minutes 2 hours
Easy Access Points	Limited Excellent
Household Cost per Year	$5 $30 $100 $250 $500 $750

Table 2: Example of Choice Set

Features	Site A	Site B	Neither
Water Quality *(cleaned to this level of use)*	Able to Fish and Swim	Able to Drink, Fish, and Swim	I Prefer Cleaning Neither Site A Nor Site B
Miles Restored *(length of stream cleaned)*	50 miles	20 miles	
Time from Home *(how close to home)*	2 hours	10 minutes	
Easy Access Points *(how easy to get to the stream)*	Excellent	Limited	
Household Cost per year *(for next 10 years)*	$30	$750	
Check the site you prefer the most	☐ I Prefer Site A	☐ I Prefer Site B	☐ I Prefer Neither

Table 3: Variable Names

Names	Description
ASC_1, ASC_2	Alternative specific constants
DRINKABLE	Clean water so you can drink with filter system
FISHABLE	Clean water so you can fish
MILES RESTORED	Number of miles restored to water quality level
TRAVEL TIME	Travel time from home
EASY ACCESS	How easy it is to get to restored stream
COST	Household cost for the next ten years

Table 4: Logit Estimates

Variables	**Ten Choice Sets**	**Five Choice Sets**
ASC_1	.724** (7.90)	1.19** (13.14)
ASC_2	.853** (9.17)	1.40** (15.36)
DRINKABLE[a]	.710** (17.87)	.606** (16.78)
FISHABLE[a]	.037 (0.97)	.079** (2.36)
EASY ACCESS	.049 (0.99)	.079* (1.79)
MILES RESTORED	.006** (3.44)	.007** (5.00)
TRAVEL TIME	-.005** (-8.89)	-.005** (-10.00)
COST	-.003** (-22.30)	-.002** (-20.54)
N	2859	3236
Log Likelihood (max)	-2447.25	-2609.87
Log Likelihood (0)	-3140.93	-3555.11
Chi-square	1019.45	910.66
R^2=1-[LogL(max) /LogL(0)]	.221	.266
% Correct Predictions	50%	52%

[a]Effects coded variables
Numbers in parentheses are t values
**significant at the 5% level
*significant at the 10% level

Table 5: Response Rates by Questionnaire Version

Version (sent)	Deceased or Bad Address	Useable or Incomplete	Response Rate
1 (330 sent)	41	173	59.9%
2 (330 sent)	36	175	59.5%
3 (387 sent)	40	214	61.7%
4 (387 sent)	45	200	58.5%
5 (387 sent)	59	196	59.8%
6 (387 sent)	43	213	61.9%
3 and 4 (774 sent)	85	414	60.1%
5 and 6 (774 sent)	102	409	60.9%

Table 6: Response Rates by Region

Region (sent)	Deceased or Bad Address	Useable or Incomplete	Response Rate
Clearfield County (736 sent)	105	394	62.4%
Northumberland County (736 sent)	63	414	61.5%
Broad Top Region (736 sent)	96	363	56.7%

Table 7: Differences in Response Rates (region to region)

Compare	p (1-p)	z-statistic
Clearfield vs. Northumberland	0.236	0.355
Clearfield vs. Broad Top	0.241	2.23
Northumberland vs. Broad Top	0.242	1.87

Table 8: Percent of Responses with No Items Missing

Version	Questions	N	N with no Missing	Percent with no missing items
1	51	173	62	36%
2	51	175	57	33%
3	41	214	68	32%
4	41	200	72	36%
5	41	196	65	33%
6	41	213	74	35%

Table 9: Mean Item Non-response Rate by Questionnaire Version

Version	Questions	N	Mean	Stand Dev.	Max
1	51	173	08.6%	0.160	84.3%
2	51	175	08.7%	0.141	70.6%
3	41	214	08.4%	0.140	87.8%
4	41	199	09.7%	0.176	87.8%
5	41	196	08.7%	0.131	75.6%
6	41	213	09.3%	0.155	85.4%
3 and 4	41	413	09.1%	0.158	87.8%
5 and 6	41	409	09.0%	0.144	85.4%

Table 10: Mean Item Non-response: Divide Questionnaire Into Three Sections

Versions	Divisions	Questions	N	Mean	Stand Dev.
1 and 2	**1st five questions**	16	348	08.1%	0.216
	Choice sets plus ranking	25	348	08.9%	0.206
	Background Information	10	348	09.0%	0.195
3,4,5, and 6	**1st five questions**	16	822	08.4%	0.220
	Choice sets plus ranking	15	822	09.2%	0.209
	Background Information	10	822	09.7%	0.192

Table 11: Testing Means (ten choice sets vs. five choice sets)

Versions	$(X_1\text{-}X_2)$	$\{(s_1^2/n_1)+(s_2^2/n_2)\}^{0.5}$	T-statistic
V1 vs. V3 and V4	-0.005	0.014	-0.33
V2 vs. V5 and V6	-0.003	0.013	-0.24

Table 12: Other Item Non-Response Tests

Tests	Statistic	Significant Differences
Testing mean item non-response for all questionnaire versions	T-statistic	No significant differences between any questionnaire versions
Testing mean item non-response across regions	T-statistic	No significant differences between regions
Testing mean item non-response for questions following choice sets	T-statistic	No significant differences for ten choice sets versus five choice sets

REFERENCES

Adamowicz, W.; J. Swait; P. Boxall; J. Louviere. "Perceptions Versus Objective Measures of Environmental Quality in Combined Revealed and Stated Preference Models of Environmental Valuation." Journal of Environmental Economics and Management. 32: 65-84. 1997.

Adamowicz, W.; J. Louviere; J. Swait. "Introduction to Attribute-Based Stated Choice Methods." Report submitted to Resource Valuation Branch, Damage Assessment Center, NOAA. January 1998.

Anderson, D.; D. Sweeney; T. Williams. Statistics for Business and Economics. NY: West Publishing Co., 1993. Pp. 360-361.

Arrow, Kenneth, Robert Solow, Paul R. Portney, Edward E. Leamer, Roy Radner, Howard Schuman. "Report of the NOAA Panel on Contingent Valuation." Federal Register. 58(10): 4602-4614, January 15, 1993.

Ben-Akiva, M.; S. Lerman. Discrete Choice Analysis: Theory and Application to Travel Demand. Massachusetts: MIT Press, 1985.

Bogen, Karen. "The Effect of Questionnaire Length on Response Rates–A Review of the Literature." Proceedings of the Section on Survey Research Methods, Alexandria, VA: American Statistical Association. 1020-1025. 1996.

Bradley, Mark; Andrew Daly. "Use of the Logit Scaling Approach to Test for Rank-Order and Fatigue Effects in Stated Preference Data." Transportation. 21: 167-184. 1994.

Bureau of Abandoned Mine Reclamation. "Pennsylvania's Comprehensive Plan for Abandoned Mine Reclamation." 5400-BK-DEP2144. June 1997.

Champion, D.; A. Sear. "Questionnaire Response Rate: A Methodological Analysis." Social Forces. 47: 335-339. 1969.

Childers, Terry; O. Ferrell. "Response Rates and Perceived Questionnaire Length in Mail Survey." Journal of Marketing Research. 16(3): 429-31. August 1979.

Dillman, Don; James Christenson; Edwin Carpenter; Ralph Brooks. "Increasing Mail Questionnaire Response: A Four State Comparison." American Sociological Review. 39(5): 744-756. October 1974.

Dillman, Don. Mail and Telephone Surveys, the Total Design Method. New York: Wiley, 1978.

Dillman, Don; M. D. Sinclair; J.R. Clark. "Effects of Questionnaire Length, Respondent-Friendly Design, and a Difficult Question on Response Rates for Occupant-Addressed Census Mail Surveys." Public Opinion Quarterly. 52: 289-304. Autumn 1993.

Freeman, A. Myrick. The Measurement of Environmental and Resource values. Washington, D. C.: Resources for the Future. 1993.

Greene, William. Econometric Analysis. New Jersey: Prentice Hall, 1997.

Hanley, Nick; Robert Wright; Vic Adamowicz. "Using Choice Experiments to Value the Environment." Environmental and Resource Economics. 11(3-4): 413-428. 1998.

Heberlein, Thomas; R. Baumgartner. "Factors Affecting Response Rates to Mailed Questionnaires: A Quantitative Analysis of the Published Literature." American Sociological Review. 43(4): 447-462. 1978.

Heberling, Matthew. Valuing Public Goods Using the Stated Choice Method. University Park: The Pennsylvania State University. Dissertation. December 2000.

Hensher, David. "Overview of Stated Preference in Travel Choices." Transportation. 21: 107-133. May 1994.

Herzog, Regula; Jerald Bachman. "Effects of Questionnaire Length on Response Quality." Public Opinion Quarterly. 45(4): 549-559. Winter 1981.

Hoffman, Sandra; Alyce E. Burke; Kathy J. Helzlsouer; George W. Comstock. "Controlled Trial of the Effect of Length, Incentives, and Follow-up Techniques on Response to a Mailed Questionnaire." American Journal of Epidemiology. 148: 1007-1011. 1998.

Louviere, Jordan; H. Timmermans. "Stated Preference and Choice Models Applied to Recreation Research: A Review." Leisure Sciences. 12(1): 9-32. 1990.

Mooney, Geraldine; L. Giesbrecht; C. Shettle. "To Pay or Not to Pay; That is the Question." Paper presented at the 48th Annual Conference of the American Association for Public Opinion Research, St. Charles, Illinois, May 20-23, 1993.

Ostle, Bernard; Richard W. Mennsing. Statistics in Research. Ames: The Iowa State University Press, 1987. Pp. 120-121.

Pennings, Joost; Scott Irwin; Darrel Good. "Surveying Farmers: A Research Note." University of Illinois at Urbana-Champaign, Department of Agricultural and Consumer Economics. AgMAS Project Research Report 1999-04. October 1999.

Swait, Joffre; Wiktor Adamowicz. "The Effect of Choice Environment and Task Demand on Consumer Behavior: Discriminating Between Contribution and Confusion." Working Paper. University of Alberta: Department of Rural Economy. Staff Paper 96-09. 1996.

Zwerina, Klaus; Joel Huber; W. Kuhfeld. "A General Method for Constructing Efficient Choice Designs." Working Paper at the Fuqua School of Business, Duke University. September 1996.