The Journal of Choice Modelling 8 (2013) 19–31

Contents lists available at SciVerse ScienceDirect

The Journal of Choice Modelling

journal homepage: www.elsevier.com/locate/jocm

Sequential preference questions factors influencing completion rates and response times using an online panel

Jordan J. Louviere [a,b], Richard T. Carson [a,c,*], Leonie Burgess [d], Deborah Street [d], A.A.J. Marley [a,e]

[a] *Centre for the Study of Choice, University of Technology, Sydney, Australia*
[b] *School of Marketing, University of Technology, Sydney, Australia*
[c] *Department of Economics, University of California, San Diego, USA*
[d] *School of Mathematical Sciences, University of Technology, Sydney, Australia*
[e] *Department of Psychology, University of Victoria, Canada*

A R T I C L E I N F O

Article history:
Received 25 April 2013
Accepted 26 April 2013
Available online 15 June 2013

Keywords:
Discrete choice experiment
Number of choice sets
Web surveys

A B S T R A C T

How many choice sets respondents will answer is a critical issue in how much data a researcher has for analysis. We used 66 separate surveys that ask respondents, from an opt-in web panel, sequences of preference questions about consumer products to study design factors that influence the rate of completing the entire sequence of questions comprising a discrete choice experiment. We do this by systematically varying the number of choice sets, the number of alternatives respondents were asked to consider, the nature of the list of attributes of each alternative and the type of statistical design. Completion rates systematically varied with the factors explored, but perhaps the key finding is that completion rates are reasonably high in all cases. We found that completion rates are relatively unaffected by asking more questions (choice sets), but they decline as one includes more alternatives. Expected time to complete a survey often plays a key role in the cost of web-based panels, so we also look at how the preceding factors impact completion times. Practical implications for applied research using opt-in web panels are discussed.

1. Introduction

Discrete choice experiments (DCEs) involve sequences of preference questions in which respondents choose from repeated sets of alternative choices. DCEs are widely used in marketing and several other fields, such as environmental economics, health policy, and transportation planning.[1] Many published papers contain statements about “task complexity”, “cognitive burden” and the like, and one hears strongly expressed opinions about “too many choice sets” at major conferences. Thus, there are widespread beliefs and much “folk wisdom” that claims that as the number of choice questions in DCEs or the number of alternatives per choice set and/or their attributes increases more respondents fail to complete the entire set of questions. Despite such widely held beliefs, there is surprisingly little available research that actually bears on these claims. The purpose of this paper is to address this gap by systematically examining the influence on DCE completion rates and completion times of four key factors that are associated with task complexity and/or cognitive burden: (1) the number of alternatives in each choice set, (2) the number of attributes that describe the alternatives, (3) the number of distinct levels of

* Corresponding author at: Department of Economics, University of California, San Diego, USA. Tel.: +1 8585343383.
E-mail address: rcarson@ucsd.edu (R.T. Carson).

[1] Louviere et al. (2000) provide an overview of modeling sequences of choice questions with an emphasis on marketing and transportation applications. Hoyos (2010) focuses on environmental applications while Viney et al. (2002) focus on health applications.

1755-5345/$ - see front matter
http://dx.doi.org/10.1016/j.jocm.2013.04.009

each attribute, and (4) the underlying statistical design that generates the alternatives and choice sets. These four factors determine the extent to which a topic can be studied using data collected in a DCE.

We designed, implemented and analyzed 66 separate DCEs using samples from an opt-in online web panel. Consequently, our results should be considered specific to this form of survey administration. Opt-in web panels now are widely used in marketing and other fields due to the high penetration of internet connections in many countries, the large size of many panels that makes it possible to select the demographics of interest for a survey, reduced time needed to collect data, ability to show respondents visual material, and increasing cost and decreasing response rates associated with other modes of survey administration.[2] Online surveys are an ideal way to ask a sequence of choice questions due to the ease with which respondents can be randomly assigned to different treatments in the form of different choice sets.

This paper can be seen as a complement to the recent paper by Bech et al. (2011) which looks at the number of choices, completion rates and completion times using an opt-in web panel. They look at questions involving dental care in Denmark. They do this using 5, 9, and 17 choice sets with 3 alternatives and 6 attributes including cost. We do this for 16 and 32 choice sets varying the number of alternatives, attributes, attribute levels, and statistical designs for two consumer products, pizza and airline flights using an Australian opt-in web panel.

An early review by Carson et al. (1994) noted that the number of sequential choice questions typically asked was in the range of 4–8, but they reported no empirical studies dealing with the impact of numbers of attributes, numbers of choice options and/or numbers of choice sets (scenarios) on completion rates or data quality. Since that review a number of studies have looked at completion rates associated with sets of choice questions in mail and in-person surveys (e.g., Brazell and Louviere, 1998; Johnson and Orme, 1996; Hensher et al., 2001; Hensher, 2004, 2006; Caussade et al., 2005; Carlsson and Martinsson, 2008; Rose et al., 2009).[3] A rough summary of the overall findings from these studies on data quality is as follows: increasing the number of choice tasks typically does not have large effects on key summary statistics like willingness to pay or the value of travel time saved, and when it does, the impact does not seem robust across different populations (e.g., Rose et al., 2009); and there also are often (but not always) some impact of the number of choice sets on the error variances.

There is now a fairly sizeable literature (e.g., Mazotta and Opaluch, 1995; Swait and Adamowicz, 2001; DeShazo and Fermo, 2002; Arentzea et al., 2003; Caussade et al., 2005; Hensher, 2006; Carlsson and Martinsson, 2008; Louviere et al., 2008; Rose et al., 2009) that look at how data quality is influenced by features of the set of choice questions that respondents are asked by focusing on how estimates of preference parameters are influenced by these features.[4] Many of these papers, and particularly more recent ones, used the "design of designs" approach (Hensher, 2004) to look at the relative impacts of multiple aspects of DCEs seen by respondents. The general impression from this work is that the number of attributes may be the most important influence on summary statistics and error variances followed by the number of choice alternatives, with other factors such as number of choice sets, number of attribute levels, and the particular statistical design being of somewhat lesser importance. As before, there are some inconsistencies across studies suggesting that the good and the context in which it is provided can matter. While the work discussed in this paper can be seen to fit into this literature, our primary focus is at a much more basic level, namely how do characteristics of sequences of choice questions influence the number of respondents who drop out by not answering the full set of questions?[5] It may well be that asking someone to answer more choice questions induces various response effects (desirable or undesirable), but the predicate question is whether participants will complete the task(s) that they are asked to do. If, in fact, they will answer a relatively large number of choice questions, the range of models that can be effectively estimated increases considerably.[6] Czajkowski et al. (2012) explicitly make this point by showing that it can be difficult to distinguish several potential features of the data generating process when the number of choice sets is small.

[2] There are many issues related to the use of opt-in web panels versus other modes of survey administration that are beyond the scope of this paper. See Lindhjem and Navrud (2011) for a discussion of issues related to web panels versus other modes of survey administration in the context of non-market valuation.

[3] Johnson and Orme (1996) were perhaps the first to provide evidence suggesting respondents would complete more choice sets than commonly expected by many researchers by summarizing several studies using computer-assisted in-person surveys in marketing. Brazell and Louviere (1998) reported a negative effect on completion rates using a drop-off mail survey due to the numbers of choice sets, but the negative effect occurred only beyond 32 choice sets. Hensher et al (2001) used a mail survey on air travel and reported a 44% response rate for 4 choices sets and an almost constant 30% response rate for several treatments ranging from 6 to 32 choice sets, with a weak suggestion of item non-response in surveys with more choice sets. Carlsson and Martinsson (2008) used separate mail surveys to study issues related to reliability of electricity supplies and reported a 39% response rate for a version with 12 choice sets and 33% for a version with 24 choice sets. More generally, Bogen's (1996) review of the literature suggests that the length of the questionnaire (especially beyond a few pages) seems to influence response rates in mail surveys (but less than many researchers seem to think), as her evidence regarding length of telephone and in-person surveys was mixed.

[4] Much of this literature has been strongly influenced by Heiner (1983) on difficulties that agents may have in optimizing decisions as complexity increases. For a somewhat different perspective more closely rooted in Simon's bounded rationality and satisficing see De Palma et al. (1994) and Conlisk (1996). There is also a long standing psychology literature on information processes and information overload that has been influential in marketing. See Keller and Staelin (1987) for an early effort looking at the issue of task complexity in the context of preference questions and data quality.

[5] It can be argued that completing/not completing a set of choice scenarios is the critical factor due to the availability of fairly well-developed statistical techniques for modeling response behavior that allow for differences in the variability of response errors and various types of strategic and satisficing behavior as respondents proceed through the set of choice questions.

[6] There is a clear tension between collecting a small amount of preference data from a large number of respondents or a large amount of preference data from a small number of respondents. Further, the larger the divergence in the response rate, the more need there is to take sample selection issues related to the particular characteristics of the DCE used into account.

Our DCEs use equivalent random samples from a large opt-in internet panel. This makes them fundamentally different from the two main traditional survey administration modes (in-person and mail) long used to administer sequences of choice questions. For in-person surveys (sometimes computer-assisted) the presence of an interviewer helps ensure respondents answer all of the choice questions (but may lead to other undesirable demand characteristics). In mail surveys potential respondents can observe the number of choice questions; because such choice questions often appear one or two to a page, apparent survey length increases rapidly with the number of these questions. In web-based surveys no interviewer is present to help ensure completion, but respondents also cannot easily determine the actual length of the survey at the outset. Payment for completing surveys in opt-in internet panels provides an incentive to complete the survey that often is not associated with other forms of survey administration and membership in online panels may be associated with factors that predispose members to answer survey questions.

Internet survey use is increasing for academic and commercial purposes (e.g., Couper, 2000; Dillman, 2000); currently there is substantial research on their specific properties (e.g., Smyth et al., 2006), and their properties compared with more traditional modes of survey administration (e.g., Kaplowitz et al., 2004; Fricker et al., 2005; Chang and Krosnick, 2009). The opt-in nature of internet panels like the one we use in this paper is likely to impact completion rates because managers of good opt-in internet panels devote substantial resources to recruiting large samples and collect substantial amounts of background data on panel members. If selection bias is involved, it is likely to be associated with panel members being more inclined to answer surveys. Forecasting from such panels involve several issues, such as weighting (Taylor et al., 2001), that do not concern us in this paper, as we randomly assign respondents to statistically equivalent subsamples drawn from a larger sample chosen to be representative of the general population to test our hypotheses.[7] The completion rates and completion times that are the focus of our research are only one of several criteria that should be used when considering the use of opt-in web panels.

Completion time represents a factor of considerable interest because it is a major driver of the cost to researchers of using an opt-in panel. Thus, our interest differs from studies (e.g., Aaker et al., 1980; Haaijer et al, 2000; Rose and Black, 2006; Otter et al., 2008) that examine the role of response times as it influences choices in various ways. What is of interest in some of these studies is that the marginal time cost of collecting answers to additional choice questions declines. This is driven, in part, by the fact that part of survey administration time, such as the introduction and collection of demographics represent a fixed cost, and in part because respondents take progressively less time to answer each choice set over at least a sizable number of choice sets, after which, choice set response time may stabilize. We look at how survey completion times change as one varies other features of the DCE such as the number of alternatives, the number of attributes, and the number of attribute levels, because these factors influence how much information is collected by the survey. As noted earlier, prior emphasis with respect to these features has been on modeling their influence on summary statistics and error variances. We also look at the impact on completion times of two popular designs. Again, our focus differs from existing literature that has primarily examined the role of experimental designs on efficiency (Bliemer and Rose, 2011) or error variances (Louviere et al., 2008).

2. Research approach

We administered 66 separate DCEs to separate but statistically equivalent samples to which potential respondents were randomly assigned. The 66 choice experiments vary by the number of choice sets respondents received (16 or 32), the number of alternatives in each choice set (3, 4 or 5), the number of attributes used to describe each alternative (6, 8, 10 or 12), the number of attribute levels for some attributes (2 or 4), and the type of statistical design used to construct the choice sets. Table 1 lists the features of the 33 experiments conducted for two product categories (delivered pizzas and cross-country flights) for a total of 66 experiments. Hereafter we refer to them as the study factors: product, number of choice sets, number of alternatives, number of attributes, number of attribute levels, and type of statistical design. The first column indicates the 11 different conditions that categorize each experiment; the second column identifies how each condition is defined by the design. For example, the first design is a $2^3 \times 4^3$, indicating that there are 3 attributes with 2 levels and 3 attributes with 4 levels. The third column contains the number of attributes implied by the design, which repeats the information about the design. The fourth (3 alternatives), sixth (4 alternatives), and eighth (5 alternatives) columns list the number of choice sets associated with different designs for a given number of alternatives.

Columns labeled “Design” indicate which of two experimental design approaches were used. “SB” is a Street and Burgess (2007) design, and “Kuhfeld” is a Kuhfeld (2005) design available as a SAS macro. Both are commonly used design approaches that require researchers to make several assumptions, with statistical efficiency levels associated with the assumptions. Our conjecture is that the most important difference between these two design approaches is the nature of the alternatives respondents “see” in the choice sets. In the SB designs, each attribute level differs across all alternatives seen in each choice set. This is true for many, but not all, Kuhfeld designed choice sets. If each choice set shows all the different attribute levels, it makes it easier for respondents to choose alternatives based on the attributes/levels that matter most to them. Conversely, if all the levels of minor attributes in a choice set are the same, this may simplify a respondent's choice. Whether and how

[7] Strictly speaking, any claim of representativeness cannot extend beyond demographic variables used to construct the sample. There is active debate in the survey sampling community on the importance of selection bias in opt-in panels for particular purposes (e.g., Baker et al., 2010). This issue is routinely ignored in much applied work, and is beyond the scope of this paper because our results assume opt-in web panels are used.

Table 1
33 Experimental conditions in the study.

Condition #	Design	# of attributes	3 Alternatives # choice sets	Design	4 Alternatives # choice sets	Design	5 Alternatives # choice sets	Design
1	$2^3 \times 4^3$	6	16	SB	16	SB	16	SB
2	$2^2 \times 4^4$	6	16	SB	16	SB	16	SB
3	$2^2 \times 4^6$	8	32	SB	32	SB	32	SB
4	$2^6 \times 4^2$	8	16	SB	16	SB	16	SB
5	$2^3 \times 4^7$	10	32	SB	32	SB	32	SB
6	$2^7 \times 4^3$	10	32	SB	32	SB	32	SB
7	$2^6 \times 4^6$	12	32	SB	32	SB	32	SB
8	$2^9 \times 4^3$	12	32	SB	32	SB	32	SB
9	$2^2 \times 4^4$	6	16	Kuhfeld	32	Kuhfeld	32	Kuhfeld
10	$2^3 \times 4^7$	10	16	Kuhfeld	32	Kuhfeld	32	Kuhfeld
11	$2^6 \times 4^6$	12	16	Kuhfeld	32	Kuhfeld	32	Kuhfeld

attributes co-vary may also determine how well choice sets approximate what respondents “see” in real markets and how interesting the choices are to them. Kuhfeld alternatives generally have more correlated attribute levels than SB designs.

As previously noted, respondents were members of a large opt-in web panel. We used a random sample drawn to match the census demographic profile of the country. Panelists are compensated with money and points that can be redeemed for prizes. Invitations to participate in online surveys are sent by email from the panel operator. Participants were told that the survey was an academic study looking at how people made choices, and that the choices involved pizzas and flights, as these would be familiar to most people. They also were informed that the sponsor was a university and the research was supported by a government grant. The survey was conducted in October 2005. The acceptance rate was 79.9% of those sent invitations to participate, consistent with the panel operator's prior experience at the time for surveys with a university sponsor.[8]

After introducing the sponsor and nature of the survey, respondents decided whether to participate or not. If they participated, the first page of instructions informed them that they would be answering questions about pizza (or flights) and that “The following pages contain between 16 and 32 scenarios. Each scenario offers you a choice of different pizzas. The features of each option are laid out in a table.” As they progressed through the survey, respondents saw a bar that indicated what fraction of the survey questions had been answered.[9] Because the panel collects basic demographic data for all panelists, we can compare the characteristics of those who started but did not complete the sequence of choice questions with those who started and completed the entire sequence.

Table 2a and b list attributes and associated levels for each of the 11 conditions for pizzas and flights, respectively. Three four-level attributes appeared in all conditions: (a) chain name, price and number of toppings for pizza, and (b) carrier name, fare and total travel time (cross-country flights) for flights. As Table 2a and b indicate, the other attributes were varied at two or four levels. In some conditions, some four-level attributes were varied over only two of their levels and, when this occurred, the two extreme levels were used. Fig. 1 shows a screenshot of one pizza choice set featuring five alternatives. After being asked for their most preferred pizza, respondents were asked for their least preferred pizza and then whether they would purchase any of the offered options.

Task size and complexity (see Table 1) varied from (a) 16 choice sets described by six attributes and three choice options per set (conditions 1, 2 and 9; with 3 choice options) to (b) 32 choice sets described by 12 attributes and 5 choice options per set (conditions 7, 8 and 11; with 5 choice options). Remaining conditions represent intermediate levels of complexity. Conditions 1 to 8 used designs for conditional multinomial logit models constructed using Street and Burgess (2007) designs. Conditions 9, 10 and 11 represent designs constructed using the SAS macros %MktEx, %MktLab and %ChoiceEff described in Kuhfeld (2005). Designs in conditions 9, 10 and 11 differ in statistical efficiency, given the same assumptions, and the differences are nested under the number of choice options.

3. Results

We present two types of results: (1) we analyze the influence of the study factors on completion rates; and (2) we analyze the effects of the study factors on median completion times.[10]

[8] Since 2005, participation rates have fallen as the size of the panel has increased.

[9] Use of such a completion bar is common in on-line surveys. Note that in contrast to a mail survey where the length of the survey can be clearly seen before starting, all respondents start with the same completion time. Respondents in the 32 choice set treatments will see slower progression of the completion bar than do those in the 16 choice set treatments.

[10] We also investigated how well demographic variables could predict completion rates using an unconditional logit model (results available from the authors on request). We examined a large set of demographic variables and found that the resulting model had low explanatory power (pseudo R-square of.02), and only a few parameters (married, no children, live in apartment/condo, and high income categories) were significant at the.10 level or better.

Table 2
List of attributes/levels in the experiments.

a. Cross-country flights

Flight attributes	Level 1	Level 2	Level 3	Level 4
Airline	Qantas	Virgin Blue	JetStar	Oz Jet
Round-trip air fare (exc tax)	$350	$450	$550	$650
Total travel time (Hours)	4	5	6	7
Food[a]	None	Free hot meal	Free snack	Food can be purchased
Audio/video entertainment[a]	Not available	Free	$3	$6
Wait in baggage claim for bags[a]	10 mins	20 mins	30 mins	40 mins
Number of stops	0	1		
% time flight departs on time	100%	80%		
Wine/beer	$6 each	Both free		
Juice/water/soft drinks	Not available	All free		
Frequent flyer club	No	Yes		
Typical wait to check-in	5 mins	20 mins		

b. Delivered pizzas

Pizza attributes	Level1	Level 2	Level 3	Level 4
Brand of delivered pizza	Pizza Hut	Dominos	Eagle Boys	Pizza Haven
Price of a large pizza	$12.00	$14.00	$16.00	$18.00
Number of toppings	1	2	3	4
Type of crust[a]	Regular	Thick	Cheese stuffed	Thin
Average delivery time[a]	10 minutes	20 minutes	30 minutes	40 minutes
Likely range in delivery time[a,b]	10%	20%	30%	40%
How often pizza arrives hot (times/10)[a]	10/10	8/10	6/10	4/10
Free garlic bread/bread sticks	No	Yes		
Free Coke or Pepsi	No	Yes		
Free side salad	No	Yes		
Free hot chicken wings	No	Yes		
Free dessert	No	Yes		

[a] In some conditions treated as 2-level attributes by using the two extreme levels (i.e., levels 1 and 4).
[b] Expressed in range of minutes not percentages (e.g., if delivery time = 20 minutes and range = (40%, the participant would have seen "average delivery time = 20 min" and "likely range in that time = 12–28 min".

	Option A	Option B	Option C	Option D	Option E
Brand of delivered pizza	Pizza Hut	Pizza Haven	Pizza Haven	Dominos	Eagle Boys
Price of a large pizza	$12.00	$14.00	$12.00	$12.00	$16.00
Type of crust	Regular	Regular	Cheese stuffed	Cheese stuffed	Thin
# of toppings	1	3	1	1	2
Average delivery time	10 mins	30 mins	30 mins	20 mins	10 mins
Free garlic bread/bread sticks	No	No	No	No	No
Free Coke or Pepsi	No	Yes	Yes	Yes	No
Free side salad	No	Yes	Yes	No	Yes
Free hot chicken wings	No	No	No	No	Yes
Free dessert	No	Yes	No	No	Yes
Select your most preferred pizza	○	○	○	○	○

Fig. 1. Screenshot of one of the pizza choice sets.

3.1. Completion rates by study factors

Completion rates ranged from 57.14% to 100% across the different treatments for those who started the survey; the overall completion rate was 77.7%.[11] These results are in Appendix Table A1. Appendix Table A2 shows that across the 11 master design conditions in Table 1, average completion rates for those who accepted the invitation to participate ranged from 72.9% to 88.1%. A 2×11 contingency table has a chi-square statistic of 25.51 with $p = .004$. The average completion rate for the 33 treatments ($N = 904$) for flights was lower (76.9%) than the comparable set of treatments ($N = 854$) for pizzas (81.2%); the difference was significant at $p = .027$ on a two-sided t-test. Completion rates for treatments using 32 choice sets were lower (77.3%; $N = 1142$) than for 16 choice sets (82.0%; $N = 617$), and the difference was significant at $p = .021$ using a two-sided t-test.[12]

Now we consider alternatives and attributes. Completion rates differed by the number of alternatives respondents chose among; the rates were 86.6% for three alternatives, 76.3% for four alternatives, and 71.7% for five alternatives; the 3×2 contingency table had a χ^2 statistic of 38.29 ($p < .001$). Completion rates also differed by the number of attributes; they were 82.5% for 6 attributes, 80.6% for 8 attributes, 76.3% for 10 attributes and 73.4% for 12 attributes. The 4×2 contingency table has a χ^2 statistic of 13.78 ($p = .003$). We did not find a consistently significant direction for the number of attribute levels.

Comparing SB statistical designs with Kuhfeld designs shows a 79.2% completion rate for the former and a 74.2% completion rate for the latter. The χ^2 statistic for the 2×2 contingent table is 5.15 ($p = .023$). A more detailed analysis by number of choice alternatives suggests that completion rates for Kuhfeld designs generally are lower than SB designs; however, the fall in completion rates as the number of choice alternatives increases is less steep for Kuhfeld designs.

Table 3 contains logistic regression results for a dependent variable indicating completing the DCE considering all the factors.[13] Flights is an indicator variable for that product rather than pizzas; results indicate higher completion rates for pizzas than flights. We suspect that product- or issue-specific response rates are common; hence, differences in response rates may be due to ease of answering choice questions or differential interest in the products, but this has not yet been systematically investigated.

Turning to numbers of choice sets (16 or 32), the parameter estimate is moderately large in magnitude but does not come close to approaching statistical significance after controlling for other study design factors. This is perhaps the most surprising finding given the widely held belief, seen in opinions expressed in papers and at conferences, that response rates fall rapidly with the number of choice sets.

Regarding the number of attributes, one must control for the number of choice sets as the study design is not balanced for both factors. We created two indicator variables (CS16 for 16 choice sets and CS32 for 32 choice sets) and interacted them with the number of attributes. This allows the effect for the number of attributes to differ for 16 choice sets (the number of attributes generally is smaller in magnitude and range) and 32 choice sets. Increasing the number of attributes substantially decreases completion rates for both choice set sizes substantially. The impact is larger for 16 choice sets treatments but to some extent this may be an artifact of differences between the distributions of the number of attributes for the two choice set sizes examined.

The overall results from our study suggest that increasing complexity of choice sets either in terms of numbers of options or numbers of alternatives decreases completion rates. The results for flights and pizzas also are consistent with this notion because choosing among flight options often involves complicated tradeoffs but choosing between pizza options generally is easier. Our results are less clear on how complexity relates to experimental designs used to construct the particular choice sets shown to respondents, and we think this deserves further study.

We used an indicator variable (SBDesign) for a Street and Burgess design rather than a Kuhfeld SAS design for the multivariate analysis. Our result suggests that a SB Designs had higher response rates than Kuhfeld SAS designs, and this effect is significant.

3.2. Analysis of completion times

Mean/median completion times for DCE conditions to which respondents were randomly assigned are in Appendix Table A1. We focus on median completion times (MEDTIME) by treatment as mean completion times are highly correlated ($r = .84$). MEDTIME varies from just under 11 min to just over 34 min across conditions. Mean times average about 1 min longer

[11] Treating those who did not accept the invitation to participate as non-completers, average completion rates ranged from 45% to 78%, with an overall average completion rate of 66%. An individual's decision to participate or not participate in the survey was independent of the characteristics of the set of choice questions in the treatment to which they were randomly assigned and would have received had they elected to participate. Thus, we only consider the set of individuals who started the questionnaire. An alternative approach is to randomly assign treatments after accepting the invitation to participate. This would have reduced the key source of variability in the number of subjects assigned to each treatment but at the expense of having less control over random assignment to treatments due to temporal patterns that can occur in the number of subjects who decide to accept.

[12] Our study design confounds the number of choice sets and numbers of attributes, as our larger choice sets tend to have more attributes, typical of empirical practice. To sort out these effects, one must use a multivariate approach, such as that in Table 3.

[13] One could directly control for respondent covariates in the analyses, but random assignment to treatments implies such controls are unnecessary other than to potentially increase power, which is not a problem in our study. Further, almost all available covariates were insignificant. A more complex model interacting respondent covariates with treatment indicators is technically possible but the relatively small differences in completion rates across treatments suggests that few will be significant, which was our experience.

Table 3
Logistic regression of completion rates on study factors.

Parameter	Estimate	Robust standard error	p-Value
Constant	4.7297	.8236	.000
Flights	−.2719	.1172	.020
CS32	−.3763	.9702	.698
Alternatives	−.4367	.0720	.000
Attributes*CS16	−.2337	.1151	.042
Attributes*CS32	−.1325	.0530	.012
SBDesign	.2598	.1230	.046
Model LL LR Test (covariates)	−900.95 (χ^2=61.88; df=6)	Observations	1759

(20.51 min for the mean and 11.49 min for the median); but in some conditions the median was larger than the mean. The mean also had a slightly higher standard deviation. Analysis of the means yields similar results to that for MEDTIME.

Table 4 summarizes results for the study factors. Going down the rows of Table 4, it is easy to see that DCEs for flights take longer than pizzas, but the difference is small and not significant at conventional levels. DCEs for 32 choice sets takes significantly longer ($p<.001$ using a two-sided t-test) than those for 16 choice sets, but not twice as long. Due to the short introductory section, it should be the case that 32 choice sets would not take twice as long as the 16 choice sets. Yet, 32 choice sets took only about 50% longer on average than 16 choice sets, suggesting that respondents probably respond more quickly as they proceed through the choice sets. Increasing the number of alternatives increases the average median completion time (p=.012 for an ANOVA with alternatives as a three level factor). Fig. 2 gives median completion times for the various alternatives and choice set combinations.[14] The difference between 16 and 32 choice sets is larger than the difference between different numbers of alternatives, and the increase from 3 to 4 alternatives is smaller than that from 4 to 5.

The results are less straightforward for the number of attributes. MEDTIME increases as the number of attributes increases from 6 to 8 (p=.047). Differences past 8 are not significant, and one sees a slight downturn for 12 attributes. Results for numbers of attribute levels requires one to control for numbers of attributes and numbers of choice sets. We found no significant effects for numbers of attribute levels, so we do not pursue this issue further.

Average median completion time for conditions involving SB Designs took just over half a minute longer than Kuhfeld designs, but this difference is not significant.

Table 5 displays the results of a regression analysis using MEDTIME as the dependent variable. DCEs for flights had an average median completion time almost two minutes longer than pizza DCEs, and the difference is significant (p=.040) with controls for other covariates. Using 32 choice sets increases median completion times a little over 7 min, which is highly significant ($p<.001$). Moving from 3 to 4 alternatives increases MEDTIME a little over 2 min (p=.007), while moving from 3 to 5 alternatives increases MEDTIME over four and half minutes. Neither numbers of alternatives nor type of designs were significant predictors.[15]

4. A simple look at data quality

We do not intend to engage in a comparative modeling exercise, but a natural question to ask is whether the quality of choice responses obtained in the DCEs decreases systematically and rapidly with respect to the factors varied across the 66 DCEs. We address this by focusing on the six extreme conditions, namely 16 choice sets, 6 attributes and 3, 4 or 5 choice options per set versus 32 choice sets, 12 attributes and 3, 4 or 5 choice options per set.[16] We estimated conditional logit models for each of these six extreme conditions and calculated the residuals for each model in each condition. We analyze the squared residuals via an analysis of variance (Table 6), which indicates a very large effect is associated with the number of choice options ($F_{2,1168}=628.2$); there is a smaller, marginally significant effect associated with the number of choice sets/number of attributes ($F_{1,1168}=3.2$), and there is a small, but significant interaction between the number of choice options and the number of choice sets/number of attributes ($F_{2,1168}=3.1$).[17]

We also compare the average squared residuals for 16 and 32 choice sets with t-tests for the 3, 4 and 5 alternative conditions. The results (Table 7) indicate that for 3 and 4 alternatives differences in average squared residuals for 16 and 32

[14] This graph uses only the SBDesign data because it is balanced with respect to the number of alternatives and the alternative levels; hence, it gives a clean comparison of the two characteristics.

[15] An analysis using mean completion times, which averages a little over a minute longer and is a bit noisier, gave similar results to the median analysis.

[16] The confound in our design noted earlier effectively makes it impossible to independently sort out the full range of choice set numbers and attribute numbers so we have taken the extreme case of 16 choices sets with 6 attributes versus 32 choice sets with 12 attributes to look at the effect on error variances. There would need to be some type of offsetting between the number of choice sets and attributes to drive individually significant effects. The fairly weak interaction between the number of choice sets and the number of alternatives suggests that this is unlikely even though it cannot be ruled out.

[17] Conditional logit models estimate “scaled” parameters, and scale seems to differ across treatments. So, one should not directly compare estimated parameters for different models. Calculations that use estimated parameters like marginal willingness to pay for an attribute may cancel out scale factor differences, allowing one to compare such statistics across models. We focus on the squared residuals, as that is where we expect most of the differences between conditions to manifest itself.

Table 4
Median completion time by study design factors.

Factor	Average of median time in minutes	Standard error	Treatments
Product			
Flights	20.02	5.11	33
Pizzas	18.27	5.15	33
No. of choice sets			
16	14.65	3.55	24
32	21.71	4.09	42
No. of alternatives			
3.00	16.91	3.94	22
4.00	19.06	5.38	22
5.00	21.46	5.22	22
No. of attributes			
6.00	14.75	3.63	18
8.00	18.35	5.86	12
10.00	22.21	4.84	18
12.00	21.00	2.96	18
No. of 2-level attributes			
2.00	16.78	5.43	18
3.00	20.32	5.11	18
6.00	18.18	3.82	18
7.00	21.91	6.45	6
9.00	22.51	4.06	6
No. of 4-level attributes			
2.00	14.36	3.58	6
3.00	20.22	5.63	18
4.00	14.00	3.01	12
6.00	20.94	3.01	18
7.00	22.36	4.15	12
Statistical design			
SB_Design	19.35	5.15	48
Kuhfeld_SAS Design	18.57	5.32	18

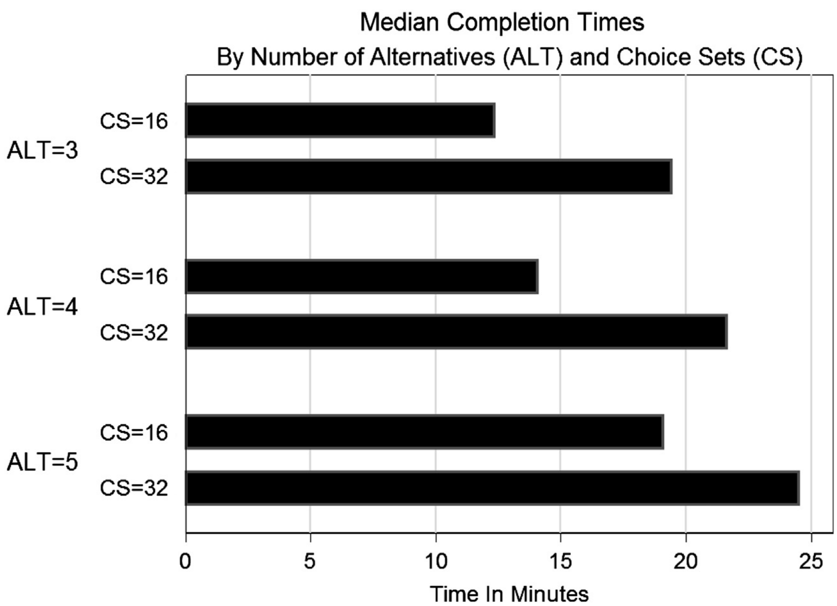

Fig. 2. Survey completion time.

Table 5
Regression of median completion time on study factors.

Parameter	Estimate	Robust standard error	p-Value
Constant	11.5493	.7072	.000
Flights	1.7481	.8319	.040
ChS32	7.0584	.7707	.000
Alternatives=4	2.1480	.7690	.007
Alternatives=5	4.5534	1.0442	.000
R^2	.599	Observations	66

Table 6
ANOVA results for model squared residuals.

Source	Type III sum of squares	df	Mean square	F	P(F)
Corrected model	1721.900[a]	5	344.380	295.268	.000
Intercept	5448.266	1	5448.266	4671.285	.000
16 versus 32 choice sets	3.766	1	3.766	3.229	.072
3,4,5 options per set	1465.485	2	732.743	628.246	.000
16vs32 × 3,4,5 options	7.157	2	3.578	3.068	.047
Error	13597.089	11658	1.166		
Total	23090.890	11664			
Corrected total	15318.989	11663			

[a] R^2=.112.

Table 7
t-Tests for difference in model squared residuals.[a]

# Alternatives	Average squared residual 16 sets	Average squared residuals 32 sets	*t*-Statistic	*P*-Value
3 options	.275	.296	−.88	.38
4 options	.720	.709	.35	.73
5 options	1.160	1.267	−3.11	.00

[a] Number of observations by row [(3200, 1600), (2560, 1280), (2016, 1008)].

choice sets are small and insignificant. For 5 alternatives, the average squared residual is significantly smaller for 32 choices sets than for 16 choice sets. In all three cases this result is opposite to what one should obtain if data quality was rapidly declining with numbers of choice sets. There are significant differences ($p < .001$) between 3, 4 and 5 alternatives that are large; their magnitudes overshadow differences between 16 and 32 choice sets.

Fig. 3 shows that error variability increases approximately linearly in numbers of choice alternatives per set, but decreases slightly with more choice sets or attributes. In turn, this suggests that adding more choice sets is not deleterious to data quality. Error variance increases quite markedly with the number of options that participants are asked to process.

Another simple way to gauge quality across the 16 and 32 choice set treatments is to compare the last 8 choice sets answered for both numbers of sets with respect to some common indicators of satisficing. Slightly fewer respondents who received the 32 choice sets treatments indicated that they would not purchase any of the offered alternatives (less than 1% fewer for pizzas and less than 2% fewer for flights). Neither difference is statistically significant. The main observable difference is that in moving from 4 to 5 options, respondents become significantly less likely to choose the fifth option than the 20% that one would have expected given the experiment designs used. That is, for pizzas, 12.1% chose option 5 in the 16 choice set/5 alternative treatments versus 12.9% in the 32 choice set/5 alternative treatments. For flights, the comparable numbers are 17.6% and 14.4%. Neither difference is significant at the 5% level and the estimated differences for pizzas and flights are in different directions.

5. Discussion

Our most useful finding for empirical researchers is simply that completion rates for DCE surveys remain relatively high even when numbers of choice sets are larger than in typical applications. We found most of the study factors significantly impacted completion rates and completion times, but none of the effects was large. In particular, completion rates for differences in 16 and 32 choice sets fell less than 5%, while the time needed to complete 32 choice sets was only 50% larger than for 16 sets.

The impact of the number of alternatives offered in each choice set was much larger, with a fall of 10% between three and four alternatives, and a fall of an additional 5% between four and five alternatives. The number of effective choices in a DCE

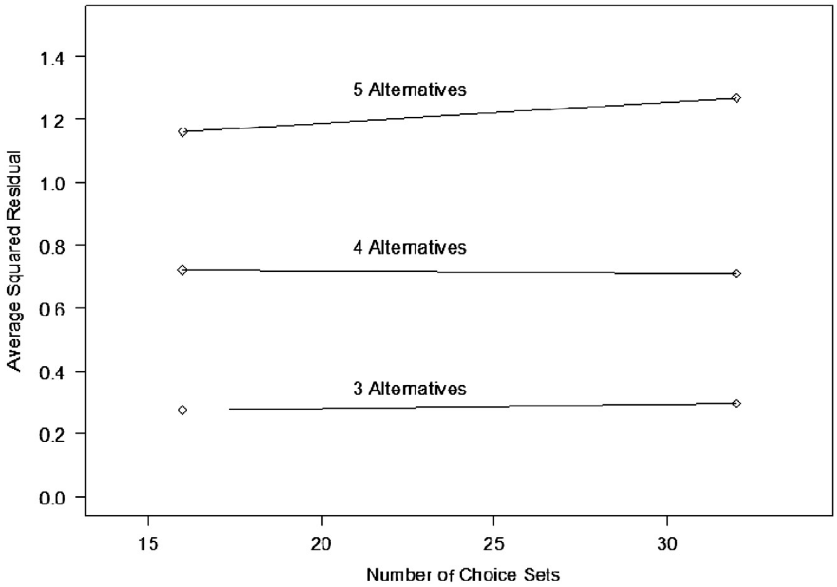

Fig. 3. Average squared residuals by condition.

equals the number of implied paired comparisons in a DCE (Chapman and Staelin, 1982; Horsky and Rao, 1984). This number increases rapidly with the number of alternatives presented in choice sets. The fall in completion rates with increasing numbers of alternatives would be troubling if those who fail to complete are systematically different than those who do. Our analysis of completer/non-completer demographic differences suggests few differences, but our sample size, although large ($N = 1759$), may be too small to find subtle differences given that the non-completion rise is not that large.

The number of attributes negatively impacted the completion rate, which fell by 9% as the attributes increased from 6 to 12. While this is not a trivial difference it is somewhat smaller than many researchers (ourselves included) might have expected. A multivariate analysis that controls for choice set size suggests that the fall is larger between 6 and 8 than it is between 8 and 10 attributes. More attributes allow tradeoffs between more complex (and often realistic) programs to be examined, so these results should be useful for applied researchers who want to understand the tradeoff between numbers of attributes and completion rates. The effect of numbers of attributes on completion times were small and statistically insignificant, suggesting either (a) that our respondents could process up to 12 attributes per alternative without spending substantially more time, or (b) more troubling, that they ignored some attributes as the number of attributes increased (see, e.g., Hensher et al., 2005). Clearly, this issue deserves further study, as does the influence of other commonly used statistical design strategies on completion rates and response times.

In contrast, while the number of alternatives had a large impact on completion times, we found no substantive effects for numbers of attribute levels. This result is not surprising since the number of levels of an attribute in a DCE is not directly relevant for answering any particular DCE question. Thus, this finding should be particularly relevant to researchers who want to trace out potentially non-linear responses to changes in attribute levels.

One should be concerned about increasing the number of choice sets if it reduces data quality. We used a very simple approach to show that this did not happen. That is, for three or four alternatives we found no relationship between error variability and the number of choice sets. We found a drop in the average error variability for five alternatives and 32 choices sets relative to 16 choice sets. However, it is worth noting that the error variability for 5 alternatives was more than four times larger than that for 3 alternatives, and it was 60% larger than for 4 alternatives. Taken together, this suggests that other DCE design decisions are likely to be much more important than the number of choice sets in determining error variability, a finding which is consistent with earlier papers (e.g., Caussade et al., 2005) that examined this issue in considerably more detail.

Each DCE survey cost us approximately $10 (US), and we can use that estimate to calculate the cost of statistical information. For example, a DCE with 16 choice sets, 6 attributes and 3 choice options per set can be viewed as 32 data records, whereas a DCE with 16 choice sets, 6 attributes and 5 choice options per set can be viewed as 64 data records.[18] The average completion rate associated with the former is 89.8%, and the average completion rate associated with the latter is 80.9%, or a difference of 9%. Therefore, if 100 individuals attempt the two DCEs, then, on average, 91 would complete the first and 74 would complete the second. However, these would generate 2874 and 5178 data records, respectively, with the larger design returning about 80% more data from which to estimate models. The cost is 19 cents per data record for the larger

[18] For three alternatives {A, B, C} and a respondent arbitrarily picking (say) A, there are two binary comparisons: A is preferred to B and A is preferred to C. For five alternatives {A, B, C, D, E} and a respondent again picking A, there are four binary comparisons: A is preferred to B, C, D, and E.

Table A1
Summary statistics including aggregate completion rates and times for all conditions.

Design	Prod	No of options	No. of sets	Design	No. of 2-levels	No. of 4-levels	No. of attributes	Comps	Non-comps	Total	Comp (%)	Median comp time	Mean comp time
1	Flights	3	16	SB	3	3	6	21	1	22	95.45	15.50	20.56
2	Flights	3	16	SB	2	4	6	22	4	26	84.62	11.84	11.10
3	Flights	3	32	SB	2	6	8	21	3	24	87.50	22.46	23.18
4	Flights	3	16	SB	6	2	8	21	5	26	80.77	12.52	11.46
5	Flights	3	32	SB	3	7	10	21	2	23	91.30	19.99	21.42
6	Flights	3	32	SB	7	3	10	22	7	29	75.86	19.35	19.18
7	Flights	3	32	SB	6	6	12	20	5	25	80.00	19.94	19.54
8	Flights	3	32	SB	9	3	12	20	8	28	71.43	21.23	17.10
9	Flights	3	16	Kuhfeld	2	4	6	20	3	23	86.96	13.20	16.40
10	Flights	3	32	Kuhfeld	3	7	10	20	3	23	86.96	22.05	24.78
11	Flights	3	32	Kuhfeld	6	6	12	22	4	26	84.62	17.58	25.30
1	Flights	4	16	SB	3	3	6	21	3	24	87.50	18.47	17.48
2	Flights	4	16	SB	2	4	6	21	3	24	87.50	18.11	16.86
3	Flights	4	32	SB	2	6	8	21	4	25	84.00	22.36	25.35
4	Flights	4	16	SB	6	2	8	21	15	36	58.33	10.88	10.39
5	Flights	4	32	SB	3	7	10	21	5	26	80.77	24.15	24.08
6	Flights	4	32	SB	7	3	10	20	5	25	80.00	22.76	21.07
7	Flights	4	32	SB	6	6	12	20	8	28	71.43	24.21	28.03
8	Flights	4	32	SB	9	3	12	20	9	29	68.97	19.20	20.92
9	Flights	4	16	Kuhfeld	2	4	6	20	10	30	66.67	11.38	14.60
10	Flights	4	32	Kuhfeld	3	7	10	22	9	31	70.97	32.81	30.70
11	Flights	4	32	Kuhfeld	6	6	12	20	10	30	66.67	19.37	26.70
1	Flights	5	16	SB	3	3	6	20	2	22	90.91	23.35	24.39
2	Flights	5	16	SB	2	4	6	20	6	26	76.92	18.41	21.16
3	Flights	5	32	SB	2	6	8	22	4	26	84.62	31.80	31.96
4	Flights	5	16	SB	6	2	8	21	6	27	77.78	20.43	19.68
5	Flights	5	32	SB	3	7	10	21	10	31	67.74	20.34	21.45
6	Flights	5	32	SB	7	3	10	20	15	35	57.14	16.39	21.54
7	Flights	5	32	SB	6	6	12	20	8	28	71.43	21.94	27.04
8	Flights	5	32	SB	9	3	12	21	8	29	72.41	28.17	26.14
9	Flights	5	16	Kuhfeld	2	4	6	20	8	28	71.43	17.19	14.71
10	Flights	5	32	Kuhfeld	3	7	10	20	15	35	57.14	22.78	22.61
11	Flights	5	32	Kuhfeld	6	6	12	21	13	34	61.76	20.43	25.19
1	Pizzas	3	16	SB	3	3	6	21	4	25	84.00	10.93	12.43
2	Pizzas	3	16	SB	2	4	6	21	0	21	100.00	10.82	12.49
3	Pizzas	3	32	SB	2	6	8	21	1	22	95.45	18.12	22.47
4	Pizzas	3	16	SB	6	2	8	21	3	24	87.50	12.71	15.42
5	Pizzas	3	32	SB	3	7	10	21	4	25	84.00	14.90	17.81
6	Pizzas	3	32	SB	7	3	10	22	2	24	91.67	20.14	22.56
7	Pizzas	3	32	SB	6	6	12	22	3	25	88.00	20.58	19.98
8	Pizzas	3	32	SB	9	3	12	22	2	24	91.67	17.36	19.64
9	Pizzas	3	16	Kuhfeld	2	4	6	21	3	24	87.50	11.82	11.79
10	Pizzas	3	32	Kuhfeld	3	7	10	21	2	23	91.30	21.13	23.87
11	Pizzas	3	32	Kuhfeld	6	6	12	21	3	24	87.50	17.84	26.10
1	Pizzas	4	16	SB	3	3	6	20	5	25	80.00	11.85	12.59
2	Pizzas	4	16	SB	2	4	6	20	9	29	68.97	12.35	10.42
3	Pizzas	4	32	SB	2	6	8	20	4	24	83.33	20.38	21.25
4	Pizzas	4	16	SB	6	2	8	21	5	26	80.77	12.74	14.95
5	Pizzas	4	32	SB	3	7	10	21	4	25	84.00	20.52	22.06
6	Pizzas	4	32	SB	7	3	10	20	6	26	76.92	18.45	17.00
7	Pizzas	4	32	SB	6	6	12	20	4	24	83.33	21.29	22.82
8	Pizzas	4	32	SB	9	3	12	20	7	27	74.07	23.28	23.50
9	Pizzas	4	16	Kuhfeld	2	4	6	21	5	26	80.77	12.69	14.23
10	Pizzas	4	32	Kuhfeld	3	7	10	21	5	26	80.77	22.27	24.44
11	Pizzas	4	32	Kuhfeld	6	6	12	20	5	25	80.00	19.75	27.71
1	Pizzas	5	16	SB	3	3	6	20	2	22	90.91	17.37	16.95
2	Pizzas	5	16	SB	2	4	6	22	2	24	91.67	18.22	19.57
3	Pizzas	5	32	SB	2	6	8	21	6	27	77.78	18.94	22.38
4	Pizzas	5	16	SB	6	2	8	22	5	27	81.48	16.88	17.98
5	Pizzas	5	32	SB	3	7	10	20	8	28	71.43	24.92	21.93
6	Pizzas	5	32	SB	7	3	10	20	7	27	74.07	34.38	28.31
7	Pizzas	5	32	SB	6	6	12	20	10	30	66.67	22.40	22.74
8	Pizzas	5	32	SB	9	3	12	21	13	34	61.76	25.81	23.51
9	Pizzas	5	16	Kuhfeld	2	4	6	21	9	30	70.00	12.00	12.18
10	Pizzas	5	32	Kuhfeld	3	7	10	20	7	27	74.07	22.46	23.63
11	Pizzas	5	32	Kuhfeld	6	6	12	20	15	35	57.14	17.59	26.57

Table A2
Completion rates by master experimental condition.

Mean percent completing												
Exp. cond.	1	2	3	4	5	6	7	8	9	10	11	Total
Mean	.881	.849	.854	.778	.799	.759	.768	.734	.772	.769	.729	.790
Std. error	.023	.045	.024	.041	.036	.046	.034	.041	.037	.050	.052	.013
Treatments	6	6	6	6	6	6	6	6	6	6	6	66

design versus 35 cents for the smaller design; hence, the larger design is 46% less expensive than the smaller one in terms of the number of effective data records collected for the same cost.

The gain in moving from 16 to 32 choice sets is much larger partly because the fall in response rates is small relative to doubling the number of effective choices.[19] Taking into account a 50% increase in time would change this comparison somewhat from a cost perspective if survey completion costs increase rapidly in terms of expected completion times, but this typically does not happen in opt-in web panels. For the study presented here, a 50% increase in time translates into about a 20% increase in cost.[20]

It would be useful to replicate our study with different types of alternatives including some that are more complex than the ones we studied. However, there is no reason to think that the general pattern of results will differ. Administering DCEs via web surveys to opt-in panels is attractive to choice modelers because the technology can randomly assign respondents to (potentially many) distinct versions of a questionnaire. Our results suggest that the amount of information that can be collected from respondents via this mode of survey administration is larger than many researchers believe.

Appendix A

See Tables A1 and A2.

References

Aaker, D.A., Bagozzi, R.P., Carman, J.M., MacLachlan, J.M., 1980. On using response latency to measure preference. Journal of Marketing Research 17, 237–244.
Arentzea, T., Borgersa, A., Timmermans, H., Del Mistrob, R., 2003. Transport stated choice responses: effects of task complexity, presentation format and literacy. Transportation Research Part E: Logistics and Transportation Review 39, 229–244.
Bech, M.T., Kjaer, Lauridsen, J., 2011. Does the number of choice sets matter?: results from a web survey applying a discrete choice experiment. Health Economics 20, 273–286.
Baker, R., Blumberg, S.J., Brick, J.M., Couper, M.P., Courtright, M., Dennis, J.M., Dillman, D., Frankel, M.R., Garland, P., Groves, R.M., Kennedy, C., Krosnick, J., Lavrakas, P.J., Lee, S., Link, M., Piekarski, L., Rao, K., Thomas, R.K., Zahs, D., 2010. Research synthesis: AAPOR report on online panels. Public Opinion Quarterly 74, 711–781.
Bliemer, M.C.J., Rose, J.M., 2011. Experimental design influences on stated choice outputs: an empirical study in air travel choice. Transportation Research Part A: Policy and Practice 45, 63–79.
Bogen, K., 1996. The Effect of Questionnaire Length on Response Rates: A Review of the Literature. In: Proceedings of the Section on Survey Research Methodology, Alexandria, VA: American Statistical Association, pp. 1020–1025.
Brazell, J.D., Louviere, J.J., 1998, Length Effects in Conjoint Choice Experiments and Surveys: An Explanation Based on Cumulative Cognitive Burden, Working Paper, Department of Marketing, University of Sydney.
Carlsson, F., Martinsson, P., 2008. How much is too much: an investigation of the effect of the number of choice set, starting point, the choice of bid vector in choice experiments. Environmental and Resource Economics 40, 165–176.
Carson, R.T., Louviere, J.J., Anderson, D.A., Arabie, P., Bunch, D.S., Hensher, D., Johnson, R.M., Kuhfeld, W.F., Steinberg, D., Swait, J., Timmermans, H., Wiley, J. B., 1994. Experimental analysis of choice. Marketing Letters 5, 351–367.
Caussade, S., de Dios Ortúzara, J., Rizzia, L.I., Hensher, D.A., 2005. Assessing the influence of design dimensions on stated choice experiment estimates. Transportation Research Part B: Methodological 39, 621–640.
Chang, L., Krosnick, J., 2009. National surveys via rdd telephone interviewing versus the internet comparing sample representativeness and response quality. Public Opinion Quarterly 73, 641–678.
Chapman, R.G., Staelin, R., 1982. Exploiting rank ordered choice data within the stochastic utility model. Journal of Marketing Research 19, 288–301.
Conlisk, J., 1996. Why bounded rationality? Journal of Economic Literature 34, 669–700.
Couper, M.P., 2000. Web surveys: a review of issues and approaches. Public Opinion Quarterly 64, 464–494.
Czajkowski, M., Giergiczny, M., Greene, W.H., 2012, Learning and Fatigue Effects Revised: The Impact of Accounting for Unobserved Preference and Scale Heterogeneity on Perceived Ordering Effects in Multiple Choice Task Discrete Choice Experiments, Faculty of Economics Working Paper 8/2012, University of Warsaw.
De Palma, A., Meyers, G.M., Papageorgiou, Y.Y., 1994. Rational choice under an imperfect ability to choose. American Economic Review 84, 419–440.
DeShazo, J.R., Fermo, G., 2002. Designing choice sets for stated preference methods: the effect of complexity on choice consistency. Journal of Environmental Economics and Management 44, 123–147.

[19] It is worth noting that having twice as many respondents is better from a statistical standpoint than having twice as many choice sets for the same respondents due to likely correlations across any one respondent's choices. However, more choices per respondent can improve the fit of models with individual fixed effects and estimation of models with correlated error terms, as well, as facilitating the estimation of individual level models.

[20] Our estimates of the percentage increase in completion times is likely to be on the high side, as the introductory portion of the survey was short and we relied on the panel provider for demographics. Increasing the length of these proportions of the survey increases the overall survey length but makes the percentage change to an increase in the number of choice sets smaller.

Dillman, D.A., 2000. Navigating the rapids of change: some observations on survey methodology in the early twenty-first century. Public Opinion Quarterly 66, 473–494.
Fricker, S., Galesic, M., Tourangeau, R., Yan, T., 2005. An experimental comparison of web and telephone surveys. Public Opinion Quarterly 69, 370–392.
Haaijer, R., Kamakura, W., Wedel, M., 2000. Response latencies in the analysis of conjoint choice experiments. Journal of Marketing Research 37, 376–382.
Heiner, R.A., 1983. The origin of predictable behaviour. American Economic Review 73, 560–595.
Hensher, D.A., 2004. Identifying the influence of stated choice design dimensionality on willingness-to-pay for travel time savings. Journal of Transport Economics and Policy 38, 425–446.
Hensher, D.A., 2006. How do respondents process stated choice experiments?: attribute consideration under varying information load. Journal of Applied Econometrics 21, 861–878.
Hensher, D.A., Rose, J., Greene, W.H., 2005. The implication on willingness to pay of respondents ignoring specific attributes. Transportation 32, 203–222.
Hensher, D.A., Stopher, P.R., Louviere, J.J., 2001. An exploratory analysis of the effects of numbers of choice sets in designed choice experiments: an airline choice application. Journal of Air Transport Management 7, 373–379.
Horsky, D., Rao, M.R., 1984. Estimation of attribute weights from preference comparisons. Management Science 30, 801–822.
Hoyos, D., 2010. The state of the art of environmental valuation with discrete choice experiments. Ecological Economics 69, 1595–1603.
Johnson, R.M., Orme, B.K., 1996, How Many Questions Should You Ask In Choice-Based Conjoint Studies?, Paper presented at the American Marketing Association Advanced Research Techniques Forum, Beaver Creek, CO.
Kaplowitz, M.D., Hadlock, T.D., Levine, R., 2004. A comparison between web and mail survey response rates. Public Opinion Quarterly 68, 94–101.
Keller, K.L., Staelin, R., 1987. Effects of quality and quantity of information on decision effectiveness. Journal of Consumer Research 14, 200–213.
Kuhfeld, W.F., 2005, Marketing Research Methods in SAS: Experimental Design, Choices, Conjoint and Graphical Techniques. 〈http:/support.sas.com/techsup/tnote/tnote_stat.html〉.
Lindhjem, H., Navrud, S., 2011. Using internet in stated preference surveys: a review and comparison of survey modes. International Review of Environmental and Resource Economics 5, 309–351.
Louviere, J.J., Hensher, D., Swait, J., 2000. Stated Choice Methods: Analysis and Application. Cambridge University Press, New York.
Louviere, J.J., Islam, T., Wasi, N., Street, D., Burgess, L., 2008. Designing discrete choice experiments: do optimal designs come at a price? Journal of Consumer Research 35, 360–375.
Mazotta, M.J., Opaluch, J.J., 1995. Decision making when choices are complex: a test of Heiner's hypothesis. Land Economics 71, 500–515.
Otter, T., Allenby, G.M., van Zandt, T., 2008. An integrated model of discrete choice and response time. Journal of Marketing Research 55, 593–607.
Rose, J., Black, I., 2006. Means matter, but variance matter too: decomposing response latency influences on variance heterogeneity in stated preference experiments. Marketing Letters 17, 295–310.
Rose, J.M., Hensher, D.A., Caussade, S., de D. Ortúzar, J., Jou, R.-C., 2009. Identifying differences in willingness to pay due to dimensionality in stated choice experiments: a cross country analysis. Journal of Transport Geography 17, 21–29.
Smyth, J.D., Dillman, D.A., Christian, L.M., Stern, M.J., 2006. Comparing check-all and forced-choice question formats in web surveys. Public Opinion Quarterly 70, 66–77.
Street, D.J., Burgess, L., 2007. The Construction of Optimal Stated Choice Experiments. John Wiley, New York.
Swait, J., Adamowicz, V., 2001. The influence of task complexity on consumer choice: a latent class model of decision strategy switching. Journal of Consumer Research 28, 135–148.
Taylor, H., Bremer, J., Overmeyer, C., Siegel, J.W., Terhanian, G., 2001. The record of internet-based opinion polls in predicting the results of 72 races in the November 2000 US elections. International Journal of Market Research 43, 127–136.
Viney, R., Lanscar, E., Louviere, J., 2002. Discrete choice experiments to measure consumer preferences for health and healthcare. Expert Review of Pharmaco-economics and Outcomes Research 2, 319–326.