Author(s): Morrison, Mark; Hatton MacDonald, Darla; Boyle, Kevin; Rose, John; Roderick, Duncan

Title: Investigating Differences between Internet and Mail Implementation of a Stated-Preference Study While Controlling for Differences in Sample Frames and Self-Selection Effects

Conference: International Choice Modelling Conference

Location: Sydney, Australia **Date:** 3-5 July

Year: 2013 **Pages:** 1-22

Editor:

Publisher: ICMC

Place of Publication: United Kingdom

Abstract: The increasing use of internet surveys for stated preference studies raises questions about the effect of the survey mode on sample representativeness, value estimates and data quality. A number of studies have now been conducted that compare the use of the internet with other survey modes, but all of these comparisons apart from one have confounded testing of mode effects with sample frame effects, and no previous studies have separated these two effects when testing the consequences of using an online panel. In addition, no previous studies have adjusted for survey self-selection across treatments during model estimation which has also confounded testing in previous studies. In this study we seek to identify the effects of using an online panel, and conduct three treatments to decompose model and sample frame effects. Using choice modelling and a case study focusing on riverine health we find that there are some differences in response rates and differences in mean values of the socio-demographics, experience of the river and knowledge. We also use sample weights developed through raking to adjust for potential self-selection effects due to item non-response on key survey questions. We find that self-selection effects are negligible; however in contrast to the previous literature we identify substantive and statistically significant differences in values due to sample frame but especially survey mode, with internet surveys producing values that are on average 30% lower than mail.

URLs:

FT: http://www.icmconference.org.uk/index.php/icmc/ICMC2013/paper/view/591/290

PL:

Investigating Differences between Internet and Mail Implementation of a Stated-Preference Study While Controlling for Differences in Sample Frames and Self-Selection Effects

Mark Morrison, Darla Hatton MacDonald, Kevin Boyle, John Rose and Roderick Duncan

Abstract

The increasing use of internet surveys for stated preference studies raises questions about the effect of the survey mode on sample representativeness, value estimates and data quality. A number of studies have now been conducted that compare the use of the internet with other survey modes, but all of these comparisons apart from one have confounded testing of mode effects with sample frame effects, and no previous studies have separated these two effects when testing the consequences of using an online panel. In addition, no previous studies have adjusted for survey self-selection across treatments during model estimation which has also confounded testing in previous studies. In this study we seek to identify the effects of using an online panel, and conduct three treatments to decompose model and sample frame effects. Using choice modelling and a case study focusing on riverine health we find that there are some differences in response rates and differences in mean values of the socio-demographics, experience of the river and knowledge. We also use sample weights developed through raking to adjust for potential self-selection effects due to item non-response on key survey questions. We find that self-selection effects are negligible; however in contrast to the previous literature we identify substantive and statistically significant differences in values due to sample frame but especially survey mode, with internet surveys producing values that are on average 30% lower than mail.

Key words: choice experiment, internet and mail comparison, error component model, survey mode

1. Introduction

Internet administration of surveys has been growing in popularity, but questions remain about the representativeness of internet samples and the effects that internet administration may have on survey responses (Lindhjem and Navrud, 2011b; Yeager et al., 2011). Lindhjem and Narvud discuss two types of errors that can arise from implementation of internet surveys, representation errors and measurement errors. Any survey can be susceptible to these errors and the key issue is the extent to which internet application of stated-preference surveys are more or less susceptible to these errors than traditional modes of implementation.

Representation has three key elements. The first is the extent to which the sample frame is representative of the targeted population. The second is survey response rate and the third is item non-response to the key questions within the survey. Malhotra and Krosnick (2007) argued that internet surveys using volunteer respondents are less accurate than in-person interviews with a

representative sample frame. To address this issue, survey firms such as Knowledge Networks in the U.S. and Pureprofile in Australia have developed panels for administration of internet surveys where the characteristics of the panels are matched to population characteristics in the U.S. and Australia, respectively. Subjects can be randomly drawn from these panels. Yeager et al. (2011) demonstrated that a probability sample, such as the Knowledge Network's panel, performs better than volunteer internet samples, and while post-stratification with demographics increases the accuracy of volunteer samples, the results are still not comparable to initial random sampling.

Even if the sample frame matches the desired population and subjects are randomly selected, survey non-response can introduce sampling error. Yeager et al. (2011) reported an effective response rate of 15% when non-response allows for initial recruitment to the internet panel through to the final survey response. Lindhjem and Navrud, (2011b) suggested weighting internet-survey data to account for error introduced by survey non-response.

Item non-response and measurement error can be addressed through design features of the internet survey instrument, as with any other survey mode (Dillman, 2007). The outcomes of these design features are dependent on how they facilitate careful consideration by respondents in answering survey questions (Tourangeau, Rips and Rasinski, 2000). Lindhjem and Narvud (2011b) suggested that "satisficing" may result in error from survey respondents not putting sufficient effort into answering survey questions (see also Krosnick 1991).

To our knowledge there have been 13 studies that have compared internet applications of a stated preference survey with application using another mode, e.g., mail, phone and in-person. As will be discussed below, all but one of these studies suffer from sample frames differing between the internet application and the other mode of application. For the one study that is an exception, the researchers selected the sample for their alternative mode of implementation from their internet panel (Lindhjem and Narvud, 2011a). Thus, in most comparisons it is impossible to say whether differences identified in the literature between internet surveys and other modes of implementation are due to survey mode or sample frame effects. In addition, if no differences are observed, this does not mean that there is not a survey mode effect; it is possible that survey mode and survey frame have countervailing effects. Further, no previous studies have allowed for sampling error due to item non-response, introducing another confounding factor into the existing comparisons.

In the research reported here, we report results of a stated-preference survey implemented to investigate survey mode while controlling for difference in sample frame and item non-response. This study is the first stated-preference study to implement an internet survey and another mode (mail) using exactly the same sample frame and to simultaneously consider sample frame effects and non-response effects. The experiment was administered to three independent samples. The internet application (II) was administered to a sample randomly drawn from Pureprofile's internet panel. The mail application (MM) was administered to a geographically stratified random sample of adults from the Australia Post database. The third treatment was a mail survey administered to a sample randomly drawn from the Pureprofile Panel (IM). Comparison of the II and IM samples allows for a comparison of survey mode, holding sampling frame constant, which is comparable to the Lindhjem and Narvud (2011a) study and provides a comparison using a different internet panel. The comparison of the MM and IM samples allows for a comparison of sample frame effects, holding survey mode constant. Then, comparison of the II and MM samples allows sample frame and mode

to vary and is comparable to much of the previous studies in the literature. We repeat all comparisons with weights developed using raking to identify the effect of item non-response. We find some differences between II results and both the IM and MM results, but no difference between IM and MM results. We do not find that our results change substantively when weighting is used to allow for item non-responses. This suggests that there is a survey mode effect with internet administration of a stated-preference survey to a random sample from an established panel.

2. **Previous Research**

The relatively low implementation cost, the opportunity to collect data quickly, the ability to control how people respond to a survey (e.g., prohibit changing responses), and the availability of data on respondent effort (e.g., time to complete individual tasks) makes the use of internet surveys appealing. Within the stated-preference literature, 13 previous studies have compared internet implementation to other implementation modes with the primary goal being to investigate the effects on estimated Hicksian surplus measured as willingness to pay (Table 2). The internet implementation has used volunteer samples (Marta-Pedroso et al., 2007) and carefully developed internet panels (Grandjean, Nelson & Taylor, 2009). The alternative survey modes include mail, telephone, various types of in-person, and drop-off/pick-up modes of implementation.

2.1 Confounding Factors in the Comparison of Survey Modes

The comparison of internet implementation and other modes of survey implementation can be confounded by a number of factors such as different sample frames, different timing of implementation, different response rates, differences in survey presentation, and different sample sizes. Acquiring identical sample frames can be challenging. Only one study used a common sampling frame, common timing of implementation, made the survey instruments as similar as possible and obtained fairly respectable and comparable response rates (Lindhjem and Navrud, 2011).

Lindhjem and Navrud (2011) drew both their internet and mail samples from their internet panel. If this is not done, developing identical samples is challenging. For example, Banzaf et al. (2006) used a Knowledge Networks panel of New York residents for their internet application and random-digit-dial reverse address look-up for their mail survey. The Knowledge Networks panel was initially recruited through random digit dialling, so the sample development potentially started with the same initial frame for both survey implementation modes, but two different confounds occurred. First, the timing for the developement of the knowledge networks panel was different from the timing Banzhaf et al. used for developing their mail sample. Second, different screens were placed on the internet and mail recruitment processes. The Knowledge Networks sample required compliance with joining the panel and the Banzhaf et al. mail sample required that an address could be matched with the phone number. A similar approach was used by Grandjean, Nelson and Taylor (2009), so their mail survey confound would be the same as Banzaf. Their phone survey would not have had a screen, while their Knowledge networks Panel did have the screen for recruitment to the

panel. These studies represent the best in developing comparable samples from an initial sampling frame and one can see that there may still be concerns.

The other end of the continuum in sample frame comparability are the studies by Canavari, Nocella and Scarpa (2005), Marta-Pedroso et al. (2007), van der Heide et al. (2008) and Windle and Rolfe (2011) that do not appear to have made any attempt to develop samples drawn from comparable sample frames for the implementation of their internet and alternative survey modes.

Bell, Huber and Viscusi (2011) implemented their mail survey in 1999-00 and their internet survey at least two years later, 2002-04. Neilsen (2011) had a one year delay, December 2005/January 2006 to January/February 2007. Others researchers cited in Table 1 either reported that the surveys were conducted concurrently or did not report this feature of their experiment. If the timing of the internet survey and alternative mode(s) differ, then results can be confounded by sample frame, survey mode, and survey timing (reliability) (Bowman et al. 2011, Brouwer 2006, Brouwer and Bateman 2005).

Beyond the concern about comparable sample frames, if a survey mode results in a higher response rate, it is is more likely that there will be a different profile of subjects responding to the surveys. The different modes of survey implementation may attract differing proportions of respondents from different socio-demographic classes (e.g., lower income respondents) or differences in attitudinal or behavioural characteristics (e.g., interested in environmental issues). While most studies in Table 1 do not report comparable response rates or do not report response rates at all, two of the best designed studies do present comparable response rates. The results reported by Grandjean, Nelson and Taylor (2009) with 4% internet, 16% telephone and 30% mail survey response rates suggest careful analyses of response data are needed to account for potential differences in sample representativeness introduced by the differing response rates. This would be the same for the Lindhjem and Narvud (2011a) results with an internet response rate of 75% and a mail survey response rate of 60%. Indeed Lindhjem and Navrud (2011a, pp.1631-1632) while they did not correct for this difference recognised that “we cannot rule out that some respondents have been self-selected due to their mode preference” and that “A more comprehensive analysis could have included ... weighting samples by demographics”.

All survey mode comparisons have some differences in presentation and this issue is difficult to address without reviewing survey instruments. The comparison conducted by Covey et al. (2010) may minimize differences because computer assisted personal interviews were used for the in-person application so the internet and in-person subjects would have viewed the same survey content. The in-person surveys still have a potential enumerator effect that is not present in the internet administration. Comparisons with other modes become more complicated because subjects may not be receiving exactly the same survey information. For example, it may not be possible to present pages on a paper on a single screen in an internet survey. Thus, while subjects may see the same survey text between the internet and mail survey modes, they may not see the same blocks of text on survey pages. In addition, an internet survey can restrict subjects’ ability to go back to previous questions and change responses while this is not possible in a mail survey. The concern regarding survey mode is greatest, however, between the internet and telephone applications because the survey information must be read to subjects over the phone and they do not have the ability to visually review the survey information (Grandjean, Nelson and Taylor, 2009; Li

et al., 2004 and 2009). In addition, subjects can complete an internet survey at their pace while a telephone survey may leave respondents feeling that they need to respond quickly, which may engage satisficing behaviour. The interviewer can provide clarifying responses to questions by subjects, which is not possible in a similar manner with an internet survey.

In any empirical study authors must estimate potential response rates to surveys. A quick review of the sample sizes reported in Table 1 suggests that the investigators generally did not try to balance usable observations between their internet and alternative modes of survey implementation (see Grandjean, Nelson & Thomas, 2009; Lindhjem and Narvud, 2011). Some studies , however, report very large differences in sample sizes (see Bell, Huber and Viscusi, 2011; Canavari, Nocella & Scarpa 2005; Neilsen, 2011; Windle and Rolfe, 2011). Differences in samples size, are a concern especially when the sample size for one survey mode treatment can have an effect on results of statistical comparisons, with the likely outcome of failing to reject the null hypothesis if no difference in survey outcomes such as estimates of willingness to pay; a Type 2 error. For example, Bell, Huber and Viscusi report a sample of 88 for their mail survey. Four studies report sample sizes for one or both of their survey modes that are less than 200, which is smaller than is reported for most stated-preference field studies in the literature today, and is suggestive of a lack of statistical power (Marta-Pedroso et al., 2007; Nielson, 2011; Olsen, 2009; Windle and Rolfe, 2011). More will be said regarding this issue below.

2.2 Comparing Results from Internet and Other Survey Modes

The first result from the stated preference studies comparing internet with other modes is the vast difference in survey response rates reported. Some studies (Grandjean, Nelson & Taylor, 2009) follow the AAPOR (2006) standards for computing and reporting survey response rates while others do not report response rates (Bell, Huber and Viscusi, 2011). Of the studies documented in Table 1, five of the 13 studies do not report response rates for their internet applications. Of the studies that report internet response rates, four report rates of 40% or more and as high as 75 percent, and three report response rates of less than 10%. The high response rates reported likely do not comply with AAPOR guidelines and likely only report rates for compliance within the internet panels and do not factor in consideration of attrition in the creation of the internet panel. When response rates are computed correctly, the internet survey applications generally have lower response rates than the alternative modes of implementing the stated-preference surveys. This is consistent with the results from the wider literature comparing internet and other survey models. Shih and Fan (2009) reported results from a meta-analysis of 39 comparison studies, with all response rates uniformly calculated using AAPOR guidelines, finding that the response rates from internet surveys are on average about 20% lower than mail surveys.

Table 1: Response Rate Differences in Comparisons of Internet and Other Survey Modes

	Response Rate	Sample Source	N[a]	Response Rate	Survey Mode	N
	Internet			***Other Mode***		
Banzhaf et all. (2006)[b]	74%[c]	KN panel[d]	841	24%	mail	293
Bell, Huber and Viscusi (2011)	NR[d]	KN panel	4,244	NR	mail	88
Canavari, Nocella & Scarpa (2005)	6%	e-mail recruitment from Iperbole	306	NR	In-person, at store	1,002
Covey et al. (2010)	NR	YouGov panel	1,957	NR	in-person, at home with computer	1,033
Taylor et al. (2009)	4%	KN panel	1,162	16% 30%	Phone Mail	1,273 904
Li et al. (2004)	24%	KN panel	2162	46%	Phone, RDD	1,699
Li et al. (2009)	NR	Survey Sampling, Inc. Recruitment	1417	53%	Phone, RDD	916
Lindhelm and Navrud (2011)	75%	TNS Gallup panel	385	60%	In-person, at home	300
Marta-Pedroso et al. (2007)	5%	Telepac internet subscribers	192	84%	In-person, at beach	230
Nielson (2011)	40%	GallupForum panel	495	70%	In-person, at home	132
Olsen (2009)	59%	AC-Nielsen panel	161	51%	Mail	181
van der Heide et al. (2008)	NR	Telder B.V. panel)	310	87%	In-person, at nature park	251
Windle and Rolfe (2011)	NR	NR	160	91%	Drop-off/pick-up, at home	179

[a] N → number of competed interviews or completed surveys received.
[b] These authors reported three different internet samples that are all based on Knowledge Networks internet panel. We report the result for the administration of their main survey instrument here.
[c] Based on a response to a Knowledge Networks internet panel and does not reflect attrition in panel recruitment.
[d] KN → Knowledge Networks.
[d] NR → not reported.

Several stated preference studies have compared the socio-demographic and attitudinal characteristics of respondents between internet and other modes, with differences in socio-

demographics between internet and in-person, mail and telephone surveys previously observed (Berrens et al. 2003; Marta-Pedroso et al. 2007; Olsen 2009; Taylor et al. 2009, Nielson 2011). However, fewer studies have examined how closely each of the modes represents the population (Berrens et al. 2003; Olsen 2009, Bell et al. 2011). Berrens et al. (2003) found that both telephone and internet were representative of the population in terms of age and gender. However, they found that respondents to their telephone sample were more likely to be educated and less likely to be Hispanic or African American than the general population. In contrast, the web respondents were more likely to have a lower income than the general population. Olsen (2009) similarly found that there were no differences to the population in terms of gender, but there were some differences to the population for both internet and mail in terms of age and education, and that respondents to the internet sample had a higher income to the population. Bell et al. (2011) found that the internet sample more closely matched population characteristics in terms of age, education, proportion of people from minority backgrounds and income than either phone-mail, mall intercept or sampling in a central location, however they did not test for statistically significant differences. Overall the existing evidence suggests that socio-demographic values do differ between internet and other survey modes, however the direction of effect appears to be context specific, and it is not apparent which mode tends to produce the most representative samples.

Less testing of differences in attitudinal and behavioural characteristics has occurred across modes. Berrens et al. (2003) compared a range of political and environmental variables between a telephone survey and two different surveys using the Harris Interactive panel (a non-RDD panel) and the Knowledge Networks panel (an RDD panel). They found that overall the telephone survey had the most similarity in terms of these variables to the Harris Interactive panel, and that the Knowledge Networks panel had a lower percentage of people eligible to vote, had fewer respondents who were members of an environmental group. All internet samples had a larger proportion of liberal (democrat) voters than with the telephone survey. Banzhaf et al. (2006) also found when comparing a survey using the Knowledge Networks panel to a mail survey that respondents from the panel include fewer environmentalists and fewer conservative voters. Van der Heide et al. (2008) reported that visitation to the Veluwe, the forest area that the study was seeking to value, was much higher amongst those who were interviewed in person (46% visited once per year) than those who responded to the web survey (25% visited once per year). Lastly, Taylor et al. (2009) found substantial behavioural differences across modes, with mail respondents significantly more likely to have visited a national park and undertaken recreational activities than either phone or internet respondents. They also observed attitudinal differences across modes, with phone and mail respondents more satisfied with the national Parks Service, and more in favour of having basic park facilities provided. Overall there is a need for further research on this aspect of mode differences, as attitudinal and behavioural characteristics of those willing to participate in online panels may well be different to those who respond via other modes, and these characteristics may influence mean values.

In terms of comparisons of mean values generated using internet surveys with other survey modes, the results from ten previous studies (nineteen comparisons) indicate that the ratio of values of internet and other survey modes is equal on average to 0.96 (see Table 1). However, the average ratio for comparisons with in-person surveys is lower at 0.84, and there is a similar average for those comparisons involving phone or phone-mail (0.83). Comparisons involving mail or drop-off and pick-

up surveys have an average ratio of 1.00. Thus the *prima facie* evidence is that the internet appears to be producing lower value estimates compared to certain modes (ie in-person and phone) and similar value estimates to other modes (ie mail and drop-off and pick-up). However, only in three studies were statistically significant differences identified (Canavari, Nocella & Scarpa 2005, van der Heide et al. 2008, Nielson 2011), and in one of these comparisons the internet mode produced higher value estimates (Canavari et al. 2005). This suggests that even for in-person and phone the value estimates could be interpreted as being relatively similar. However, in interpreting these findings it should be noted, as discussed above, that there are a range of confounding factors that potentially make the comparisons of value estimates across modes presented in almost all of these studies either invalid or inaccurate.

Table 2: Differences in Mean Values for Comparisons between Internet and Other Modes ($US2011)

Study	What is valued	Internet	Other Mode	Significant difference	Ratio of internet/ other mode
Banzhaf et al. (2006)	Adirondack's ecosystems	Not reported separately		No	Not reported
Bell, Huber and Viscusi (2011)[1]	Water quality	$27.5	Phone-Mail $34.3	Not tested	0.88
			Mall Intercept $22.6		1.39
			Central Location $17.1		1.85
Canavari, Nocella & Scarpa (2005)	Organic fruit (O) (per kg) and pesticide ban (P)	O $1.6 P $332.4	In person O $1.2 P $332.4	Yes No	1.38 1.00
Taylor et al. (2009)	Ozone in National Parks	$13.6	Mail $12.4 Phone $15.6	Not tested	1.08 0.86
Li et al. (2004)	Ratification of Kyoto protocol	$810.3	Phone $1095.5	No	0.74
Li et al. (2009)	R&D for	Not reported separately		No	Not reported
Lindhelm and Navrud (2011)	Biodiversity protection	$287.3	In-person $333.8	No	0.86
Marta-Pedroso et al. (2007)	Cereal Steppe, annual tax	$65.0	In-person $153.3	Not tested	0.42
	Cereal Steppe, donation Env. NGO	$118.3	In-person $178.3	Not tested	0.66
Nielson (2011)	Life expectancy, Scenario 1	$107.1	In-person $90.5	Yes, at 10%	1.18
	Life expectancy, Scenario 2	$76.9	In-person $80.0	No	0.96

[1] Bell et al. (2011) also estimated willingness to pay across modes when sociodemographics were held constant. They found with this correction that the Internet Panel ($31.31) produced lower values than Mall Intercept ($42.83) and Phone-Mail ($53.67) and similar values to sampling at a Central Location ($29.36).

Olsen (2009)	Forests	$10.0	Mail $11.5	No	0.87
	Wetlands	$19.9	Mail $19.7	No	1.01
Van der Heide et al. (2008)	Forest restoration, Scenario 1	$95.0	In-person $339.1	Yes	0.40
	Forest restoration, Scenario 2	$235.1	In-person $472.3	No	0.72
Windle and Rolfe (2011)	Reef quality, Scenario 1	$0.90	Drop-off $1.0	No	0.99
	Reef quality, Scenario 2	$2.50	Drop-off $2.90	No	1.07
					Average 0.96

3. Description of the Case Study Area, Survey and the Experiments

The focus of the case study area is the environmental quality of the River Murray and the Coorong in the Murray Darling Basin of Australia. The River Murray, floodplains along the River and the major coastal wetland, the Coorong and Lower Lakes, are in decline due to the loss of high to medium flooding with regulation of the River (Hillman 2008) and the reduced flows due to over-allocation (CSIRO 2008). Major investments in infrastructure and buying back water from irrigators have been made and more are currently underway. It is in this context that a study was undertaken to explore the willingness to pay for improving the quality of the River Murray and the Coorong (see Hatton MacDonald et al. 2011).

3.1 Survey Logistics

The mail-out sample was drawn from a sampling frame from Australia Post which is the largest commercially available mail-out sampling frame. The mail-out sample size was 6000 and was stratified by geographic areas of interest for the objectives of the overall study. More detail of the mail-out survey can be found in Hatton MacDonald et al. (2011). An internet survey was conducted using the sampling frame of one the major internet survey companies in Australia Pureprofile, with participants randomly selected from Pureprofile's sampling frame. Participants logged into their accounts and were asked if they were willing to participate in a survey that may be administered as part of an internet-recruit - mail-back (IM) or the internet recruit – internet response (II) respectively. Willing participants were randomly assigned to one of the split samples.

To ensure the closest possible comparability in survey formats, Pureprofile was provided with the Illustrator pages from the hard copy questionnaire and information sheet. Toggle boxes were used on the internet and tick boxes on the paper copy. Internet respondents were permitted to move backwards and forwards through the questionnaire and bring up the information sheet at anytime to review information.

Small incentives were used with all three samples. Upon completion, the MM and IM samples were provided with a new $1 commemorative coin from the Australian mint and the II sample was provided PureProfile's standard $1.60 credit to use for on-line purchases.

4 Modelling Approach

4.1 Specification of the Utility Function

Let U_{ntj} denote the utility of alternative *j* perceived by respondent *n* in choice situation *t*. U_{ntj} consists of two components, an observed component V_{ntj} and an unobserved component ε_{ntj} such that

$$U_{ntj} = V_{ntj} + \varepsilon_{ntj}. \quad (1)$$

As is common practice, we assume the observed component of utility to be represented as a linear relationship of *k* attributes, *x*, related to each of the *j* alternatives and corresponding parameter weights such that

$$U_{ntj} = \sum_{k=1}^{K} \beta_{nk} x_{ntjk} + \varepsilon_{ntj}, \quad (2)$$

where β_{nk} represents the marginal utility or parameter weight associated with attribute *k* for respondent *n* and the unobserved component, ε_{ntj} is assumed to be independently and identically, extreme value type 1 distributed. As well as containing information on the levels of the attributes, *x* may also contain up to *J*-1 alternative specific constants (ASCs) capturing the residual mean influences of the unobserved effects on choice associated with their respective alternatives; where *x* takes the value 1 for the alternative under consideration or zero otherwise.

Different marginal utilities for each attribute are traditionally accounted for in the mixed multinomial logit model by allowing one or more parameters to be specified as

$$\beta_{nk} = \bar{\beta}_k + \eta_k z_{nk}, \quad (3)$$

where $\bar{\beta}_k$ represents the mean of the distribution of marginal utilities held by the sampled population, η_k represents a deviation from the mean and z_{nk} an individual specific (set of) draw(s) from some predefined distribution (e.g., $z_{nk} \sim N(0,1)$)

A variant, the Error Component (EC) model can be rewritten where the utility for alternative j is

$$U_{ntj} = \left(\bar{\beta} + \eta z_{nj}\right) x_{ntj} + \varepsilon_{ntj} \\ = \bar{\beta} x_{ntj} + \eta z_{nj} x_{ntj} + \varepsilon_{ntj}. \quad (4)$$

and rather than associate separate x_{ntj} for each η, the EC model utilise dummy variables to place subsets of alternatives into different 'nests' or 'branches'. This is shown in Equation (5).

$$U_{ntj} = \bar{\beta} x_{ntj} + \eta z_n d_h + \varepsilon_{ntj}, \quad (5)$$

where $d_h = \begin{cases} 1 & \text{if } j \text{ belongs to nest } h \\ 0 & \text{otherwise} \end{cases}$.

The random error term applies only to alternatives where $d_h = 1$ (else $\eta z_n.0 = 0$). It is in this way that the random error term is associated only with subsets (or nests) of selected alternatives and the 'nests' are formed.

The error component model induces covariances amongst the alternatives as shown in Equation (6).

$$Cov(U_{ntj}, U_{nti}) = E\left(\eta z_n d_{hj} + \varepsilon_{ntj}\right)\left(\eta z_n d_{hi} + \varepsilon_{nti}\right) = \sigma_h \quad (6)$$

and with variance of all alternatives in nest *h* equal to

$$Var(\eta_n) = E\left(\eta z_n d_{hj} + \varepsilon_{ntj}\right)^2 = \sigma_h + \frac{\pi^2}{6}. \quad (7)$$

In the current paper, we estimate separate error components for different data sets by allowing different dummy variables (i.e., d_h) to exist for each state dataset, thus approximating the commonly used NL model approach typically used to combine data sets (Train, 2010).

4.2 Experimental Design

An efficient design was employed in these stated preference choice experiments. The generation of an efficient design involves selecting a design that is likely to provide an asymptotic variance-covariance (AVC) matrix containing values which are as small as possible for a given econometric model form. As the asymptotic standard errors obtained from discrete choice models are simply the square roots of the leading diagonal of the AVC matrix of a discrete choice model, the smaller the elements of the AVC matrix (or at least the diagonal elements), the smaller the asymptotic standard errors. Given that dividing the parameter estimates by the asymptotic standard errors results in the asymptotic t-ratios for the model, the smaller the asymptotic standard errors, the larger the asymptotic t-ratios. Designs which attempt to minimise the elements contained within the AVC matrix are referred to as efficient designs. Note that such designs are unlikely to be orthogonal in the attributes (as we are dealing with a nonlinear discrete choice model).

For this study, a D_b-efficient design was used. A D_b-efficient design represents a Bayesian approach to the design generation process, where the analyst assumes a set of prior parameter estimates and

utilises these to infer the best attribute level combinations that will minimise the elements within the expected AVC matrix obtained from models estimated on data collected using the design. Rather than assume precise knowledge of the population parameter estimates, D_b-efficient designs utilise distributions of likely parameter estimates in the design construction process. As is typical of many studies that use such designs, prior parameter estimates were obtained from a pilot survey (for a discussion of the precise design generation procedure, see Bliemer and Rose 2006; Bliemer et al., 2009).

The resulting design had 32 choice sets with a D_b-error of 0.10004. Rather than have each respondent answer all 32 choice sets, the design was further blocked into 8 groups of 4 choice sets each. Given that the final design was not orthogonal in the attributes, the blocking column was constructed so as to minimise the largest correlation (absolute value) of the column with the various other columns that represent the attributes of the design. The largest correlation of the resulting blocking column was with the household cost attribute for the first alternative with a correlation of 0.0413.

5. Empirical Results

5.1 Differences in Response Rates

Three different split-samples[2] are reported in this manuscript: a mail survey collected using the Australia Post (mail) sample frame (MM), a mail survey collected using the Pureprofile (internet) sample frame (IM), and an internet survey collected using the Pureprofile (internet) sample frame (II). The mail-out (MM) sample size was 6000 and a response rate of 54.7% was achieved for this sample using AAPOR response rate 1 (see Table 3). For the internet-mail (IM) survey, from an initial number of internet invitations to participate of 2755 (excluding ineligible respondents), the mail survey was sent to a sample of 837 people and 470 questionnaires were received back, yielding a response rate of 17.1%. For the internet-internet (II) sample, from a sample of 2179 people, a response rate of 23.5% was achieved.

Table 3: Response Rates of Mail-Mail, Internet-Mail and Internet-Internet Treatments

	Mail-Mail	Internet-Mail	Internet-Internet
Number of invitations to participate distributed via the Internet		2755	2179
Number of questionnaires distributed via Mail	6000	837	
Total returned/completed	3158	470	512

[2] Four internet split on-line surveys were undertaken as part of this project, two of which are discussed in this paper.

Response rate	52.5%	17.1%	23.5%

5.2 Differences in Socio-demographics and Behaviours

Table 4 summarises the socio-demographic information and experience of respondents in the three samples. As the mail-out sample is six times larger than the II or IM sample, a random sample of comparable size (n=500) were taken from the MM sample[3] for the purposes of testing for differences in means and proportions.

[3] The mail-out sample was designed to test for differences between States and MDB and Australia. A stratified random sample of 1000 in the Murray-Darling Basin (which incorporates the Australian Capital Territory and parts of New South Wales, Victoria, South Australia, and Queensland), 1400 in NSW, 1200 in Victoria, 1000 South Australia and the 1400 in rest of Australia (Queensland, Western Australia, Tasmania and the Northern Territory). Note that when estimating weighted models, state-based weights were used to adjust for over and under sampling in specific states in this and other treatments.

Table 4: Comparison of Means and Proportions

Variable	Mean Value - MM	Mean Value - II	Mean Value – IM
Income – less than $52,000 per year	52.1%	52.0%	59.9% [1]
Age	46.1 [2,3]	42.8 [1,3]	50.6 [1,2]
Education - year 12 or less	41.2%	35.5%	39.0%
Education - diploma	31.0%	28.1%	30.3%
Gender (Female = 1)	54.0%	52.7%	49.4%
Have been sight-seeing along the River Murray	50.0%	13.1% [1,3]	47.3%
Family Association with Farming	24% [2]	13.1% [1]	14.8%
Test Score	5.94 [2]	6.21 [1,3]	5.97 [2]

[1] statistically different from MM α=5%; [2] statistically different from II α=5%; [3]statistically different from IM α=5%

There is a higher proportion of IM respondents with an income less than $52,000 per year and is significantly different in proportion to the MM sample. The II sample is younger than the MM and IM samples. No statistically significant difference in education is found. The samples are all close to gender balanced. The MM sample has a stronger connection with the land through family members associated with farming than the II. Experience with the River Murray is different among the samples. The internet sample has less experience sightseeing along the River Murray.

5.2 Econometric Model Results

Both unweighted and weighted[4] panel multinomial logit error-components (EC) models were estimated using NLOGIT 5.0 to identify whether survey mode and sample frame are associated with different responses. The weights were developed using both sociodemographic and behavioural variables as previous research has demonstrated that respondents answering internet surveys can differ in their sociodemograhic profile as well as in their usage of the natural asset being valued (eg Van der Heide et al. 2008, Taylor et al. 2009). The weights were developed using raking or sample balancing of the data using the STATA add-in *survwgt* from Winter (2008).

Given the use of socio-demographic and behavioural variables for weighting the data, sociodemographic and behavioural variables are not also included as regressors in the estimated

[4] Three different weighted models were estimated. In the first model, the weights were derived using raking and were based on four socio-demographic variables: gender, age, income and education. In the second model, the weights were based on three behavioural variables: family association with farming, have been sightseeing on the Murray, and have been boating along the Murray. In the third model, the weights were derived using the full set of socio-demographic and behavioural variables. As the results did not differ substantively across the three weighted models, only the results for the third model are reported.

models. Rather, the regressors in the utility function consist of the choice set attributes, plus two alternative specific constants (ASCs): one associated with the status quo alternative and the second with the first choice set alternative. The second ASC is included to account for the possibility of left to right bias (i.e., to account for respondents potentially being more likely to select alternatives presented closer to the left in the experiment).

Table 5 reports the results of the unweighted EC model, where model IM for the internet recruit-mail back survey, the II model for the internet recruit-internet return survey and model MM was estimated on the mail recruited- mail back survey sample. Examining the ASCs, all constants associated with the status quo alternative, as well as left-right bias are statistically significant. Further, the SQ ASC parameters are all negative while the SP1 ASC parameters are all positive, implying that a*ll else being equal*, respondents have a propensity to choose an alternative other than the SQ alternative, but in doing so, tend to select the first SP alternative over the second. For the design attributes, all parameters have the expected signs and are statistically significant at conventional levels. Table 6 reports the results of the weighted model. In terms of the sign and significance of the various coefficients in the model, the results for this model are not substantively different to the unweighted model.

Table 5: Panel Error Components Unweighted Model Results

	Mail/Mail		**Internet/Internet**		**Internet/Mail**	
Variable	**Coefficient**	**t-ratio**	**Coefficient**	**t-ratio**	**Coefficient**	**t-ratio**
Constants						
SP1 constant	0.1327***	2.88	0.1848***	3.79	0.1714***	3.44
SQ constant	-6.8614***	-6.37	-4.0681***	-6.54	-3.2723***	-6.93
Design attributes (all alternatives)						
Bird Breeding along the Murray	-0.1058***	-8.31	-0.1133***	-9.00	-0.0848***	-6.69
Native Fish along the Murray	0.0139***	3.55	0.0166***	4.26	0.0134***	3.81
Healthy Vegetation along the Murray	0.0231***	7.57	0.0260***	8.95	0.0161***	5.49
Healthy Water Bird Habitat in the Coorong	1.1066***	10.06	1.0715***	9.68	1.0659***	10.37
Household cost	-0.0061***	-9.12	-0.0072***	-10.44	-0.0077***	-12.26
Error components – SP Alts only						
Error component (MM)	-6.9234***	-7.79				
Error component (II)	4.4634***	9.15				
Error component (IM)	4.8170***	10.23				
Model statistics: Number of observations = 5924, Log likelihood function = -4771.13, McFadden Pseudo R-square = 0.63						

Table 6: Panel Error Components Model – Weighted by Socio-demographics and Experience of the River

	Mail/Mail		Internet/Internet		Internet/Mail	
Variable	**Coefficient**	**t-ratio**	**Coefficient**	**t-ratio**	**Coefficient**	**t-ratio**
Constants						
SP1 constant	0.1387***	3.03	0.1868***	3.29	0.1335***	2.71
SQ constant	-6.4201***	-6.35	-4.2789***	-6.84	-3.6818***	-7.05
Design attributes (all alternatives)						
Bird Breeding along the Murray	-0.1167***	-9.17	-0.1158***	-9.46	-0.0789***	-6.12
Native Fish along the Murray	0.0154***	4.00	0.0145***	3.80	0.0140***	3.91
Healthy Vegetation along the Murray	0.0239***	7.75	0.0244***	9.23	0.0164***	5.68
Healthy Water Bird Habitat in the Coorong	1.1154***	10.25	1.0963***	10.17	0.9947***	9.70
Household cost	-0.0063***	-9.46	-0.0071***	-10.61	-0.0071***	-11.51
Error components - SP alts only						
Error component (MM)	6.6589***	7.94				
Error component (II)	5.1601***	9.54				
Error component (IM)	4.9547***	9.79				
Model statistics: Number of observations = 5924, Log likelihood function = -4874.38, McFadden Pseudo R-square = 0.63						

For both the unweighted and weighted error component models, likelihood ratio tests were used to identify significant differences in the design variables across the three treatments. This testing demonstrated that for the unweighted models there were significant differences[5] in only one of the model coefficients for the choice set attributes at α=0.10 in each of the treatment comparisons. When weights for state of residence and demographics were added, two out of five coefficients (breeding and vegetation) were different in the comparison between Mail-Mail and Internet-Internet, and one out of five coefficients (vegetation) were different in the comparison between Internet-Mail and Internet-Internet. There were no significant differences in the comparison between Mail-Mail and Internet-Mail. When weights for state of residence and activities were used, all five attribute coefficients were different at α=0.05 between the Mail-Mail and Internet-Internet treatments, and four out of five coefficients (excluding household cost) were different at α=0.05 in the comparison between Internet-Mail and Internet-Internet. Again there were no significant differences in the comparison between Mail-Mail and Internet-Mail. Finally, when all weights were used (state, demographics and activities), all attribute coefficients were different at α=0.05 for each of the treatments, apart from two cases where the models did not converge and testing was not

[5] Note that because the models were jointly estimated they have a common scale. Hence it is appropriate to test equivalence of model parameters.

possible. Overall these results suggest the presence of both sample mode and sample frame effects, but that the sample mode effects are more dominant.

While the differences in model coefficients suggest that survey mode effects have been identified, of greater practical interest is the effect that survey mode or sample frame has on value estimates. Furthermore, it is possible that while effects on coefficients have been identified, these effects may not translate into significant effects on value estimates. Therefore in Table 7, implicit prices are reported for each of the three treatments. Compared to the Mail-Mail treatment, across the four implicit prices, the Internet-Mail treatment produced implicit prices that were 8% (unweighted) and 13% (weighted) lower, while the Internet-Internet treatment produced implicit prices that were on average 32% lower in the unweighted model and 30% lower in the weighted model.

Table 7: Implicit Prices for the Unweighted and Weighted Models

	Unweighted			**Weighted – State, Socio-Demographic & Experience**		
	MM	**II**	**IM**	**MM**	**II**	**IM**
Water Bird Breeding along the Murray	$17.46	$11.06 (-36.7%)*	$15.84 (-9.3%)	$18.61	$11.07 (-40.5%)	$16.33 (12.2%)
Native Fish Populations along the Murray	$2.30	$1.75 (-23.9%)	$2.32 (0.8%)	$2.46	$1.97 (-20.0%)	$2.05 (16.9%)
Healthy Vegetation along the Murray	$3.82	$2.10 (-44.9%)	$3.64 (-4.7%)	$3.81	$2.30 (-39.6%)	$3.43 (10.0%)
Waterbird Habitat in the Coorong	$182.61	$138.98 (-32.3%)	$149.85 (-17.9%)	$177.89	$139.51 (-21.6%)	$154.62 (13.1%)
Average Percentage Difference to MM		-32.3%	-7.8%		-30.4%	-13.0%

* Percentage difference to MM treatment in brackets

The test recommended by Poe, Giraud and Loomis (2005) was used to identify whether any of the implicit prices are significantly different across treatments (Table 8). Across weighted and unweighted models, significant differences were found for two of the four attributes – waterbird breeding and vegetation for the comparison between internet-mail and internet-internet (at α=0.05). For the comparison between mail-mail and internet-internet (both weighted and unweighted models) there were significant differences for three attributes (at α=0.05), breeding, fish and habitat. For the comparison between Mail-Mail and Internet-Mail only one statistically significant difference was identified, for habitat. This was at α=0.05 in the unweighted model and at α=0.10 in the weighted model.

Table 8: Tests of Differences Between Implicit Prices (p-values listed)

	Unweighted			Weighted		
	MM vs IM	MM vs II	IM vs II	MM vs IM	MM vs II	IM vs II
Water Bird Breeding along the Murray	0.27	0.00	0.01	0.18	0.00	0.01
Native Fish Populations	0.49	0.21	0.18	0.26	0.22	0.46
Healthy Vegetation	0.39	0.00	0.00	0.28	0.01	0.02
Waterbird Habitat in the Coorong	0.03	0.00	0.22	0.07	0.01	0.16

6. Conclusions

The practical advantages of using internet surveys have lead to many recent non-market valuation studies comparing results from internet surveys with other modes. While the majority of these studies have produced results that are supportive of the use of the Internet for non-market valuation, the results of all previous studies have been confounded by either not disentangling sample frame and survey mode effects, not adjusting for item non-response and consequent sample selection effects, or by other problems with experimental control (eg time differences across treatments, low sample sizes etc).

In this study we have sought to overcome these problems with existing studies and selected treatments to separately allow for both sample frame and survey mode effects. We have also used a series of weights derived using raking to determine whether non-response and self-selection influences the comparability of results. We have also sought to deal with several other issues of experimental control that may have influenced the results of other studies. For example, the treatments were conducted at a similar time period, sample sizes were of a reasonable size, and a similar size across treatments.

Consistent with the existing literature we found that the response rates are not the same across the samples with the internet-internet sample response rate about 30 percentage points lower than the internet-internet treatments. There are also differences in two the key socio-demographic variables with the internet-internet sample being younger than either of the other samples, and the Internet-Mail sample having a lower income. Consistent with previous studies (eg Van der Heide et al. 2008, Taylor et al. 2009), we also found that the internet-internet sample also has a lower mean family association with farming and has less recreational experience with the River Murray. However, the internet-internet sample had a higher test score for understanding information contained in the information sheet and questionnaire than either the mail-out MM or the internet-recruit IM sample.

Of most interest to practitioners and policy makers is the effect that choice of survey mode has on value estimates. For previous internet-mail comparisons, the existing literature on average suggests that the values produced are almost identical. However, as shown in this study, this is not necessarily the case. In terms of percentage differences, across all four implicit prices our results show that use an internet panel produces values that are about one-third lower than the value estimates derived using a mail survey. There is also evidence of statistically significant differences, with three out of four of the implicit prices being significantly different across these two treatments.

When comparing sample frames, but holding survey mode constant, the implicit prices were found to be 15% lower when using the internet sample frame for the mail survey. However, neither the model coefficients nor the implicit prices were statistically different between the mail-mail and internet-mail treatments. This suggests overall that sample frame effects are likely to be modest, if they occur at all.

Thirdly, weighting was used to correct for non-response and sample selection effects. While a sophisticated form of weighting was used, and three different weighted models were estimated, weighting was found to have little effect on the final results. This suggests that non-response and sample selection effects have had a minimal role in our comparison between treatments.

Thus our results suggest that sample frame and non-response and self-selection have a limited effect on the comparability of non-market values, but sample mode can have a substantive effect. This raises the question of what might be driving this effect? One possibility is that people are primed differently when participating in internet surveys. Because their motivation for agreeing to complete surveys via internet panels is to gain a financial reward, they may be less open to indicating support for improvements in environmental quality that are going to cost them money. Another possibility is that internet surveys may be less consequential (Carson and Groves 2007) if respondents if use of this mode implies to respondents that government are investing less in a survey and consequently are less likely to implement its findings. A third possibility is that respondents think and behave differently on the internet. There is evidence that for some groups of people the internet stimulates different parts of a person's brain than other forms of interaction, such as would occur when a respondent reads a mail survey. Small et al., (2009) compared brain activation of internet savvy and non-internet savvy people when completing an internet search and reading text (as in reading a book). They found little difference in brain activation for a non-internet savvy group, but for an internet savvy group they found significant increases in signal intensity in regions controlling decision making and complex reasoning. Thus the use of the internet may change the way this groups makes decisions or their capacity to do so. Further experimentation would, however, be needed to test if any of these reasons are the cause for the survey mode effects identified in this study.

While the reasons for the mode effect identified in this study remain ambiguous, its existence and extent remain clear. The results suggest that the mode effect from use of the internet is potentially quite large, and substantial enough to influence the outcomes of a benefit-cost analysis. Hence, our results imply that internet surveys should be used with caution for non-market valuation, particularly when used for major policy decisions. If they are to be used, then it should be recognised that they are likely to produce conservative value estimates up to one third less than use of a mail survey.

References

American Association for Public Opinion Research. (2006). Standard definitions: Final dispositions of case codes and outcome rates for surveys. Retrieved June 18, 2007, from www.aapor.org/uploads/standarddefs_4.pdf.

Banzhaf, S. Burtraw, D., Evans, D. and Krupnick, A. (2006). Valuation of Natural Resource Improvements in the Adirondacks," Land Economics 82: 445-464.

Berrens, R. P., A. K. Bohara, H. Jenkins-Smith, C. Silva and Weimer, D. (2003). The advent of Internet surveys for political research: A comparison of telephone and Internet samples. Political Analysis 11: 1-22.

Bliemer, M. and Rose, J. (2006). Designing stated choice experiments: The state of the art, 11th International Conference on Travel Behaviour Research, Kyoto, Japan, 20th August 2006.

Bliemer, M., Rose, J., and Hensher, D. (2009). Constructing efficient stated choice experiments allowing for differences in error variances across subsets of alternatives, Transportation Research Part B, 43, 19-35.

Bowman, M., Mattsson, L., Ericsson, G. and Kriström (2011). Moose Hunting Values in Sweden Now and Two Decades Ago: The Swedish Hunters Revisted. Environmental and Resource Economics, 50: 515-530.

Boyle, K.J. (2003). Contingent Valuation in Practice, in Champ, P.A., Boyle, K.J., Brown, T.C. (2003), A Primer on Nonmarket Valuation. Kluwer Academic Publishers, The Netherlands.

Brouwer, R. (2006). Do stated preference methods stand the test of time? A test of the stability of contingent values and models for health risks when facing an extreme event, Ecological Economics, 60(2): 399-406.

Brouwer, R. and I. J. Bateman (2005), Temporal stability and transferability of models of willingness to pay for flood control and wetland conservation, *Water Resources Research*, 41:

Canavari, M., G. Nocella, and R. Scarpa (2005). Stated willingness-to-pay for organic fruit and pesticide ban, Journal of Food Products Marketing. 11: 107-134.

Carlsson, F., Frykblom, P., Lagerkvist, C. (2005). Using cheap talk as a test of validity in choice experiments, Economics Letters 89,147-152.

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

CSIRO (2008). Water availability in the Murray-Darling Basin. A report to the Australian Government from the CSIRO Murray-Darling Basin Sustainable Yields Project. CSIRO, Australia.

Dillman, D.A. 2007. Mail and internet surveys: the tailored design method. John Wiley and Sons.

Hillman, T. (2008). Ecological Requirements: Creating a Working River in the Murray-Darling Basin, in Crase, L. (ed.), Water Policy in Australia: the impact of Change and Uncertainty. Resources for the Future, Washington, DC.

Li, H., Berrens, R., Bohara, A., Jenkins-Smith, H., Silva, C. and Weimer, D. (2004). Telephone versus Internet samples for a national advisory referendum: are the underlying stated preferences the same? Applied Economics Letters 11: 173 - 176

Lindhjem, H. and Navrud, S. (2011a). Internet CV surveys – a cheap, fast way to get large samples of biased values? Ecological Economics, 70: 1628-1637.

Krosnick, J. (1991), 'Response strategies for coping with the cognitive demands of attitude measures in surveys', Applied Cognitive Psychology 5: 213-36.

Lindhjem, H., and S. Navrud. 2011b. Using Internet in Stated Preference Surveys: A review and Comparison of Survey Modes" International Review of Environmental and Resource Economics, 5: 309-351.

List, J. Sinha, P., Taylor, M. (2006). Using Choice Experiments to Value Non-Market Goods and Services: Evidence from Field Experiments. Advances in Economic Analysis & Policy 6, Article 2.

Malhotra, N., and J.A. Krosnick. 2007. The Effect of Survey Mode and Sampling on Inferences about Political Attitudes and Behavior: Comparing the 2000 and 2004 ANES to Internet Surveys with Nonprobability Samples. Political Analysis, Vol. 15:286–323

Marta-Pedroso, C., Freitas, H., and Domingos, T. (2007). Testing for the survey mode effect on contingent valuation data quality: A case study of web based versus in-person interviews. Ecological Economics 62: 388-398.

Morrison, M.D., Blamey, R.K. and Bennett, J.W. (2000). Minimising Payment Vehicle Bias in Contingent Valuation Studies. Environmental and Resource Economics. 16(4): 407-422.

Morrison, M., Brown, T. (2009). Testing the Effectiveness of Certainty Scales, Cheap Talk, and Dissonance-Minimization in Reducing Hypothetical Bias in Contingent Valuation Studies. Environmental Resource Economics 44, 307-326

Nielsen, J. (2011). Use of the Internet for willingness-to-pay surveys: A comparison of face-to-face and web-based interviews. Resource and Energy Economics 33(1), 119-129

Olsen, S. (2009). Choosing Between Internet and Mail Survey Modes for Choice Experiment Surveys Considering Non-Market Goods. Environmental and Resource Economics 44: 591-610

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

Shih, T-H and Fan, X (2009). Comparing response rates in e-mail and paper surveys: A meta-analysis. Education Research Review 4: 26-40.

Tourangeau, R., L.J. Ripps and K.A. Rasinski. 2000. *The Psychology of Survey Responses*. Cambridge University Press.

van der Heide, M. van den Bergh, van Ierland, E. and Nunes, P. (2008) Economic valuation of habitat defragmentation: A study of the Veluwe, the Netherlands. Ecological Economics 67: 205-216

Windle, J. and Rolfe, J. (2011). Comparing Responses from Internet and Paper- Based Collection Methods in more Complex Stated Preference Environmental Valuation Surveys. Economic Analysis & Policy, 41(1): 83-97.

Winter, N. (2008) SURVWGT: Stata module to create and manipulate survey weights" downloaded from http://econpapers.repec.org/software/bocbocode/s427503.htm.

Yeager, D.S., J.A. Krosnick, LC. Chang, H.S. Javitz, M.S. Levendusky, A. Simpser and R. Wang. 2011. "Comparing the Accuracy of RDD Telephone Surveys and Internet Surveys Conducted with Probability and Non-Probability Samples." Unpublished paper, Stanford University.