The Australian Journal of
Agricultural and Resource Economics

Australian Journal of Agricultural and Resource Economics, 66, pp. 44–71

Comparing Internet and phone survey mode effects across countries and research contexts*

Jed J. Cohen and Johannes Reichl†

We compare Internet and telephone survey responses across 27 European Union nations and two research contexts: one, a choice experiment of willingness to pay to avoid power outages, and the other, a public acceptance of energy infrastructure question. The various forms of survey mode effects and the challenges of survey mode choice are documented and developed in the context of statistical theory and an application to an economics survey. We find evidence that survey mode effects vary across research contexts, and to a lesser extent, across nations. We suggest that the degree of measurement bias may be varying between research contexts, for example based on the availability of a perceived socially correct response within a given context. Future survey-based research should evaluate the choice of survey mode in a context- and region-specific manner.

Key words: Internet survey, public acceptance, social desirability bias, survey mode, telephone survey, willingness to pay.

1. Introduction

The preferred mode of survey data collection often varies between disciplines and researchers. Popular modes of data collection include face-to-face, mail, telephone and Internet surveys. Survey mode choice is often based on clerical concerns such as budget, accessibility of the target sample and the ease of data collection, as well as the preferences of the researcher. However, this choice may have real consequences for the responses received (Klausch et al., 2015), and ultimately the conclusions of the research (Breunig & McKibbin, 2011), making it critical to better inform the survey mode choices in empirical research. In particular, the chosen survey mode may introduce biases into the study that stem from various behavioural consequences of the survey-respondent ecosystem. Past research in a variety of academic disciplines, contexts and geographies has checked for these forms of bias between different survey modes by comparing obtained responses and corresponding parameter estimates across modes and variants of the same mode (Maslovskaya et al., 2019; Struminskaya et al., 2015). The past research has inconclusive results, as discussed in detail in Section 1.2 below.

*This work was supported by funding from the European Union's Horizon 2020 research and innovation program under the SMARTEES Project [grant agreement #763912].

†Jed J. Cohen (e-mail: jedcohen@vt.edu) is an Integrated Planning Lead at the Salt River Project, 1500 N Mill Ave, Tempe, AZ, USA 85281. Johannes Reichl is a Project Leader at the Energieinstitut at Johannes Kepler University, Altenberger Strasse 69, 4040 Linz, Austria.

doi: 10.1111/1467-8489.12451

A major present-day use of surveys, in combination with advanced statistical methods, comes from stated preference research within economics, wherein no clear consensus regarding a preferred survey mode exists (Dillman et al., 2014; Johnston et al., 2017). This deficiency endures despite the development of a broad and deep literature on survey methodology spurred by the insights of the NOAA stated preference panel in 1993 (Arrow et al., 1993). Stated preference studies investigating survey mode effects in relation to estimated non-market values have mixed and inconsistent findings (Boyle et al., 2016; Ethier et al., 2000; Johnston et al., 2017; Windle & Rolfe, 2011). A recent seminal article on the best practices in survey-based economics research suggests that survey mode effects appear '[. . .] context specific' and that the mode choice should also be made based on context (Johnston et al., 2017, p. 340). However, up until now an explicit investigation of the effects of survey mode across contexts is lacking. We fill this need in the survey research by comparing Internet and telephone survey responses across 27 European Union (EU) nations and two research contexts that were included in a single survey completed by the same respondents. In this paper, we define a 'research context' broadly as the topic and research question(s) under investigation and the tools, both survey and statistical, that are used to investigate the research question(s). The research contexts compared here are a choice experiment of willingness to pay (WTP) to avoid power outages, and a public acceptance of energy infrastructure experiment.

The costs of power outages are a major consideration in power grid planning and have thus been the subject of significant research (Cohen et al., 2018a,b; Leahy & Tol, 2011; Nooij et al., 2007; Praktiknjo et al., 2011; Reichl et al., 2013). Household welfare loss from power outages is typically assessed via survey. Indeed, in 2020 the Agency for the Cooperation of Energy Regulators (ACER) in Europe issued a decision whereby every EU member state must develop non-market surveys for measuring the WTP to avoid power outages (ACER, 2020). Such a wide-scale and high-profile roll-out of these surveys would benefit from a better understanding of their methodological trade-offs, including survey mode choice, as investigated herein. The second research context addresses the issue of public acceptance, often given the unpopular eponym NIMBY (not in my backyard) syndrome, which can hinder de-carbonisation of the electricity grid as the development of important pieces of infrastructure is slowed or blocked by local coalitions (Cohen et al., 2014; Cotton & Devine-Wright, 2013; Devine-Wright, 2007, 2013; Devine-Wright & Batel, 2013; Elliott & Wadley, 2012; Soini et al., 2011; Wuestenhagen et al., 2007). Survey instruments are often used in public acceptance research, making an improved understanding of methodological choices worthwhile. Taken together, these two contexts were chosen for this study as they can be united in a single survey instrument addressing energy topics, but they are distinct enough where the propensity for response biases could plausibly vary between them.

We posit that the inconsistent past findings with respect to the effect of survey mode on collected responses corresponds to the variation in the populations considered and the contexts of previous research. The effects of survey mode on response patterns are tested by estimating choice models that allow for country-specific survey mode effects. We find evidence that survey mode effects vary between research contexts within our study and to a lesser degree across nations. This suggests that the effect of survey mode is not constant, but can instead vary as the survey mode is applied to different regions and research questions, which might explain the varying results of past research of survey mode effects.

The paper proceeds by developing a conceptual framework for the different types of biases related to survey modes and their root causes. The varied results of past literature investigating survey modes is discussed in Section 1.2, with the focus put on research related to stated preferences. We then present our survey instrument in Section 2, followed by the analysis methods and results in Section 3. The final section discusses the findings.

1.1 Typology of survey mode effects

Survey mode effects can be broadly categorised into two types: sampling (i.e. representation) biases arising from different groups of people completing the survey based on the mode employed and measurement (i.e. response) biases, whereby response patterns change across survey mode even when the survey and question formats are held constant (Lindhjem & Navrud, 2011b).

Sampling biases include the following. *Sample-frame bias* where the population that is used to draw the sample is restricted. For example, an Internet-based survey can only sample from the population with Internet access. If this group differs systematically from the target population, the responses can be biased. *Non-response bias* can occur when only a specific, systematically different sub-sample of the population completes, or does not complete, the survey. A specific type of non-response bias is *avidity bias* where only those with an interest in the subject matter participate in the survey (Maguire, 2009).

Measurement biases include *social desirability bias*, where respondents provide answers they deem to be socially desirable when an interviewer is on the phone or present with them, and *satisficing*, where respondents shortcut the cognitive process required to answer the survey question and thus can arrive at a response that does not accurately reflect their beliefs of preferences (Krosnick, 1991). The concepts of sampling and measurement biases are formalised below.

1.1.1 Preliminaries

Consider a survey sample $X = [X_1, X_2, \ldots, X_i, \ldots, X_n]$ of random variables where each respondent $i \in \{1, 2, \ldots, n\}$ contributes a vector of information $\mathbf{X}_i$

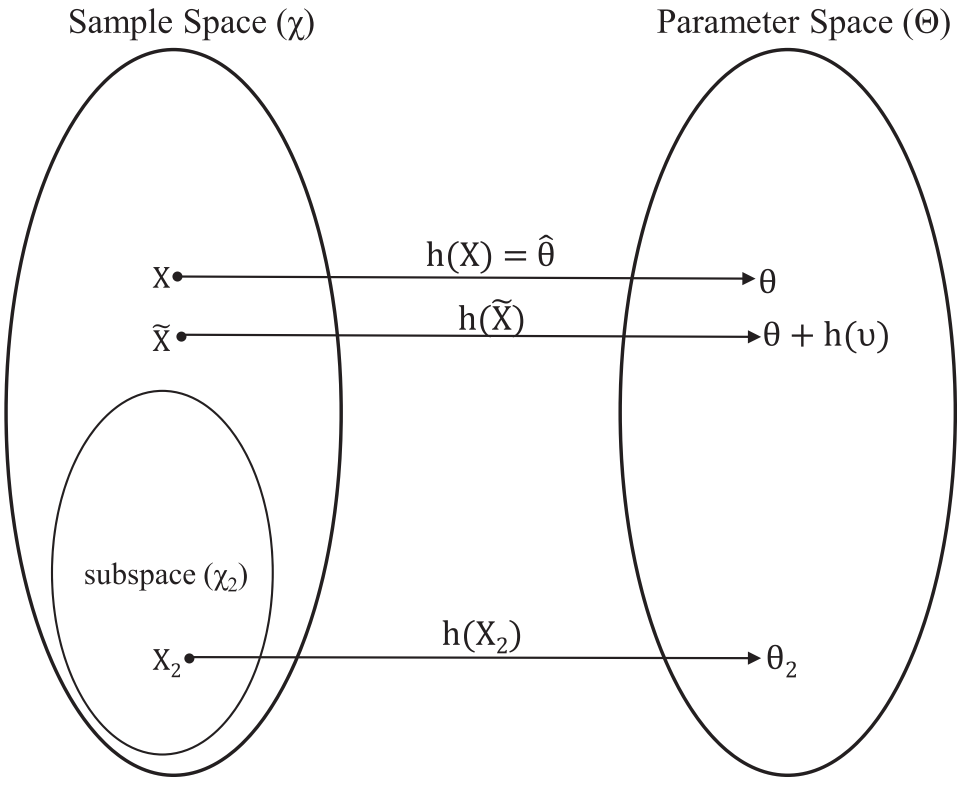

Figure 1 Conceptualisation of survey sample realisations (following the notation of Spanos (1999)).

with probability distribution $f(\mathbf{x}_i|\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ are true parameters relating to the target population (e.g. preferences for an environmental good). The sample $\mathbf{X}$ is assumed to be free of sampling and measurement biases. The researcher's goal is to estimate $\boldsymbol{\theta}$ using an estimator $h(\cdot):\chi\rightarrow\Theta$, which is a real-valued function mapping from the sample space to the parameter space, as shown in Figure 1. A specific value vector of the estimator based on a sample realisation $\mathbf{X}$ is an estimate denoted as $\hat{\theta}$ with $h(\mathbf{X})=\hat{\boldsymbol{\theta}}$. We assume this estimator is well behaved with the usual desirable properties of unbiasedness, consistency and efficiency.

1.1.2 Sampling bias

Imagine a sample space χ with a distinct subspace χ_2, as in Figure 1; if we sample exclusively from χ_2, we obtain sample realisation $\mathbf{X}_2\in\chi_2$, where respondent i's realisation is subject to probability distribution $f(\mathbf{X}_i|\boldsymbol{\theta}_2)$. For example, χ_2 may represent the subset of the population with Internet access, and the data generated by sampling exclusively from the Internet population may be subject to different underlying preference parameters if $\boldsymbol{\theta}_2 \neq \boldsymbol{\theta}$.[1] The effect of sampling bias on the estimator using the $\mathbf{X}_2$ sample realisation can be expressed as in (1).

$$Bias(h(\mathbf{X}_2)) = E\left[\hat{\boldsymbol{\theta}}|\mathbf{X}_2\right] - E[\boldsymbol{\theta}] = E[E[h(\mathbf{X}_2)]] - E[\boldsymbol{\theta}] = \boldsymbol{\theta}_2 - \boldsymbol{\theta} \tag{1}$$

where the third equality in (1) follows from the unbiasedness assumptions of $h(\cdot)$ in estimating the underlying parameters of the data. However, while $h(\mathbf{X}_2)$ gives an unbiased estimate of $\boldsymbol{\theta}_2$, it gives a biased estimate of the true population parameter if $\boldsymbol{\theta}_2 \neq \boldsymbol{\theta}$. This result is derived for concrete families of estimators in Bethlehem (2010).

1.1.3 Measurement bias

Now consider a third survey sample denoted $\tilde{X} = \left[\underset{1}{X}, \underset{2}{X}, \ldots, \underset{i}{X}, \ldots, \underset{n}{X}\right]$, where for each respondent $i \in \{1, 2, \ldots, n\}$, the data may be subject to measurement bias expressed as $E[\tilde{X}_i] = E[X_i + \nu_i] = X_i + \nu_i$, where $\mathbf{v}_i$ is a random variable; we write this succinctly for all i as $\tilde{\mathbf{X}} = \mathbf{X} + \mathbf{v}$.[2] For simplicity, we assume that $h(\mathbf{X} + \mathbf{v}) = h(\mathbf{X}) + h(\mathbf{v})$, as would be the case in the classic linear regression model (Greene, 2012, Section 8.5.1). Also, all $\tilde{\mathbf{X}}_i \in \tilde{\mathbf{X}}$ are subject to the probability distribution $f(\mathbf{x}_i|\boldsymbol{\theta})$ based on the true population parameters $\boldsymbol{\theta}$. Thus, the $\tilde{\mathbf{X}}$ sample is free from sampling biases. However, under these assumptions the bias of the estimator when using the $\tilde{\mathbf{X}}$ sample realisation can be expressed as in (2).

$$\begin{aligned} Bias\big(h(\tilde{\mathbf{X}})\big) &= E[\hat{\boldsymbol{\theta}}|\tilde{\mathbf{X}}] - E[\boldsymbol{\theta}] = E\big[E\big[h(\tilde{\mathbf{X}})\big]\big] - E[\boldsymbol{\theta}] \\ &= E[E[h(\mathbf{X}+\mathbf{v})]] - E(\boldsymbol{\theta}) = E[h(\mathbf{X})] + E[h(\mathbf{v})] - E[\boldsymbol{\theta}] \\ &= \boldsymbol{\theta} + E[h(\mathbf{v})] - \theta = E[h(\mathbf{v})] \end{aligned} \tag{2}$$

1.1.4 Outcomes of interest

Under the framework described above, we are interested in explaining observed outcomes in survey studies that compare parameter estimates across survey modes, as such differences drive the concern over survey mode choice.

[1] One can always argue that different sub-populations are subject to different underlying parameters. As shown in Bethlehem (2010) and Biemer (2001), sampling proportionally from different sub-populations allows for unbiased estimation of the population parameters. This is also a key idea behind stratified random sampling methods (e.g. Muneer et al., 2017).

[2] The measurement bias is here formulated as classical measurement error, as in Hyslop and Imbens (2001). Other forms of measurement error from Hyslop and Imbens (2001) could be considered within this framework, but this is beyond the scope of the present work. The classic measurement error assumes that $E[\mathbf{v}_i|\mathbf{x}_i] = 0$, no other distributional assumptions are made for $\mathbf{v}_i$, but we note that even if $E[\mathbf{v}_i] = 0$ the $E[h(\mathbf{v}_i)]$ can be non-zero, as shown in Greene (2012, Section 8.5.1) for the classic linear regression model.

From above, we have three samples under consideration. A sample $\mathbf{X}$ that is a random sample such that each $\mathbf{x}_i \in \mathbf{X}$ is subject to the probability distribution $f(\mathbf{x}_i|\boldsymbol{\theta})$ and the data are obtained without measurement bias. A second random sample $\mathbf{X}_2$ may be drawn from a systematically different sample of respondents where each $\mathbf{x}_i \in \mathbf{X}_2$ is subject to the probability distribution $f(\mathbf{x}_i|\boldsymbol{\theta}_2)$. A third random sample $\tilde{\mathbf{X}}$ may be subject to measurement bias such that $\tilde{\mathbf{X}} = \mathbf{X} + \mathbf{v}$, but is free of sampling bias where each $\tilde{\mathbf{x}}_i \in \tilde{\mathbf{X}}$ is subject to the probability distribution $f(\tilde{\mathbf{x}}_i|\boldsymbol{\theta})$.

Two hypotheses of interest to survey mode research are defined below. Hypothesis 1.1 tests for the presence of sampling biases in the $\mathbf{X}_2$ sample.

Hypothesis 1.1:

$$H_0 : E[\hat{\boldsymbol{\theta}}|\mathbf{X}] - E[\hat{\boldsymbol{\theta}}|\mathbf{X}_2] = 0 \Rightarrow \textit{no sampling bias in } X_2.$$

In relation to survey mode, this is most relevant to a situation where $\mathbf{X}_2$ is sampled using a Web-based survey and the $\mathbf{X}$ sample is obtained via mail or telephone. Survey mode literature generally views Web-based surveys as more prone to sampling biases (e.g. Johnston et al., 2017; Lindhjem & Navrud, 2011a, 2011b; Taylor et al., 2009), especially from sample frame bias, as most Internet panels from which survey respondents are drawn are voluntary opt-in panels and omit the subset of the population that does not access or regularly use the Internet (Baker et al., 2013).

Hypothesis 1.2:

$$H_0 : E[\hat{\boldsymbol{\theta}}|\mathbf{X}] - E[\hat{\boldsymbol{\theta}}|\tilde{\mathbf{X}}] = 0 \Rightarrow \textit{no measurement bias in } X$$

To test Hypothesis 1.2, researchers would need to ensure that both samples are subject to the same underlying distribution, that is $f(\mathbf{x}_i|\boldsymbol{\theta})$, and inform the same underlying population parameters. In other words, both samples need to be free of sampling bias and be random samples of the same target population. Studies that accomplish this are noted in the Section 1.2. Furthermore, one sample would need to minimise the chance of measurement biases in response patterns. Web-based surveys are at lower risk for social desirability bias than survey modes that use an interviewer, such as telephone and face-to-face data collections (Chang & Krosnick, 2010; Heerwegh, 2009; Holbrook & Krosnick, 2009; Kreuter et al., 2009; Lindhjem & Navrud, 2011a). With respect to satisficing bias, any survey mode can be at risk. Some of the limited literature on the issue suggests that Web-based surveys are subject to lower risk (Borkan, 2010; Chang & Krosnick, 2010; Schwappach & Strasmann, 2006), although others find that Internet surveys are subject to equal or higher satisficing bias (Heerwegh, 2009; Malhotra & Krosnick, 2007). As such, it is difficult to conclude that any one mode has higher risk of satisficing bias (Lindhjem & Navrud, 2011a), and more research is needed into the specific mechanisms of satisficing in non-market valuation, as in Box all et al. (2009).

We derive the conceptual form of the observed bias for studies that compare parameter estimates across samples collected via different survey modes in the Propositions below. Proposition 1.1 shows that the estimated parameters across samples (e.g. Internet and phone samples) will differ systematically based on the differences in the underlying parameter vectors that define these samples.

Proposition 1.1:

$$E[\hat{\boldsymbol{\theta}}|\mathbf{X}] - E[\hat{\boldsymbol{\theta}}|\mathbf{X}_2] = E[\boldsymbol{\theta}] - E[\boldsymbol{\theta}_2] = \boldsymbol{\theta} - \boldsymbol{\theta}_2$$

Proposition 1.2 shows that the difference in observed estimates based on samples from two survey modes where one mode is subject to measurement bias is a function of the measurement bias and specifically of how this bias manifests within the estimator function $h(\cdot)$.

Proposition 1.2:

$$\begin{aligned} E[\hat{\boldsymbol{\theta}}|\tilde{\mathbf{X}}] - E[\hat{\boldsymbol{\theta}}|\mathbf{X}] &= E[h(\tilde{\mathbf{X}})] - E[h(\mathbf{X})] = E[h(\mathbf{X})] + E[h(\mathbf{v})] - E[h(\mathbf{X})] \\ &= \boldsymbol{\theta} - \boldsymbol{\theta} + E[h(\boldsymbol{v})] = E[h(\boldsymbol{v})] \end{aligned}$$

Proposition 1.3 shows that under sampling bias in one sample ($\mathbf{X}_2$), and measurement bias in the other ($\tilde{\mathbf{X}}$), the differences in the parameter estimates are a function of both forms of bias. In general, there is no reason to expect that these origins of bias will cancel out, although biemer2001a derives a way to isolate the bias origins under specific conditions.

Proposition 1.3:

$$\begin{aligned} E[\hat{\boldsymbol{\theta}}|\tilde{\mathbf{X}}] - E[\hat{\boldsymbol{\theta}}|\mathbf{X}_2] &= E[h(\tilde{\mathbf{X}})] - E[h(\mathbf{X}_2)] = E[h(\mathbf{X})] + E[h(\mathbf{v})] - E[h(\mathbf{X}_2)] \\ &= \boldsymbol{\theta} - \boldsymbol{\theta}_2 + E[h(\boldsymbol{v})] \end{aligned}$$

The conceptual framework is used to understand and categorise the survey mode results from past literature in the next section and to contextualise the results from the empirical exercise in Section 3.

1.2 Survey mode studies in non-market valuation

The past literature that is most closely related to our effort here is that which has investigated the impact of data collection mode on responses to economic stated preference surveys. These studies are summarised in Table 1. The specific findings and methods of key studies are discussed in more detail below. However, most critical is the variety of results obtained with respect to the effect of survey mode on the estimated non-market values. Five of the fifteen studies surveyed show little to no effect from survey mode choice, and two show smaller effects that do not substantially impact the non-market valuation and resulting policy recommendations, while six of the fifteen studies find substantial effects, including the more recent studies of Boyle et al. (2016) and Sandorf et al. (2017). Most studies report a bias of the form in Proposition 1.3, since their survey mode samples may be subject to differences in both measurement and sampling biases. However, some studies carefully control for sampling biases and thus are able to test Hypothesis 1.2 and any remaining observed bias is of the form in Proposition 1.2. Studies typically test Hypothesis 1.1 via comparing descriptive statistics across

Table 1 Economic stated preference studies comparing survey modes

Article	Modes compared	Geography	Context	Result
Loomis and King (1994)	Phone, mail	California	Wildlife habitat	Lower WTP for phone
Whittaker et al. (1998)	Phone, mail	Colorado	State park fees	Higher WTP for phone
Ethier et al. (2000)	Phone, mail	Buffalo, NY	Green electricity	No difference
Marta-Pedroso et al. (2007)	Web, face to face	Portugal	Steppe bird habitat conservation	Lower WTP for Web
Olsen (2009)	Web, mail	Denmark	Landscape conservation	Small effects
Fleming and Bowden (2009)	Web, mail	Australia	Valuing access to an island	No effects
Maguire (2009)	Phone, mail, face to face	USA	Support Nature Conservancy	No effects
Bell et al. (2011)	Web, mail, face to face	USA	Inland water quality	Significant differences across all modes
Nielsen (2011)	Web, face to face	Denmark	Air pollution	Small effects
Windle and Rolfe (2011)	Web, mail	Australia	Condition of Great Barrier Reef	No effects
Scasny and Alberini (2012)	Face to face, Web	Czech Republic	Premature death from climate change	Small effects
Lindhjem and Navrud (2011a)	Web, face to face	Norway	Biodiversity	No effects
Boyle et al. (2016)	Web, mail	Brisbane, Australia	River and floodplains	Lower values for Web
Sandorf et al. (2017)	Face to face, Web	Norway	Protection of cold water coral	Lower values for Web
Campbell et al. (2018)	Web, mail, mixed	MT, CO, AZ (USA)	Biomass energy generation	No sig. differences in WTP estimates

samples and testing for the equality of means for response rate and socio-demographic indicators.

Contradictory findings of survey mode effects began in non-market valuation with Loomis and King (1994) and Whittaker et al. (1998). Loomis and King (1994) found that Californian respondents to the phone survey had lower WTP for wildlife habitat improvements, while Whittaker et al. (1998) found that phone respondents had higher WTP for state park fees in Colorado. In an attempt to clear up this contradiction, Ethier et al. (2000) carry out a careful comparison of phone and mail survey methods and find no evidence that survey mode effects responses. The authors ensure the same sampling frame between respondents receiving the mail survey and those being surveyed by phone by selecting potential respondents from a pool of customers of a specific utility company in Buffalo, New York, and randomly assigning them a survey mode, thus constituting a test of Hypothesis 1.2. Their study is limited to a highly specific population and does not consider Web-based surveys, which may be the most distinct survey mode in terms of its low potential for social desirability bias. Also, we note that Ethier et al.'s study again differs from previous work in the population sampled, and also in context, WTP for green electricity options in this case, which could be driving the difference in results. The European study from Lindhjem and Navrud (2011a) compares face-to-face and Internet surveys in the context of WTP for protecting biodiversity in Norway. The authors are able to sample from the same sampling frame to minimise sampling biases and find no evidence that survey mode is affecting responses.

A recent study by Boyle et al. (2016) also managed to test Hypothesis 1.2 by taking particular care to ensure that the comparison of Web- and mail-based surveys did not suffer from sampling frame bias by drawing both samples from an Internet panel and randomly selecting a group of respondents that would receive a hard copy of the survey via post. In their valuation of implicit attribute prices for aspects of a riparian ecosystem in eastern Australia, the survey mode significantly affected attribute prices in most cases, with mail-based respondents attributing higher values to the various attributes tested. The paper concludes by calling attention to the lack of definitive evidence regarding survey mode effects, and the need to consider *why* respondents may respond differently across modes.

Other studies outside of economics have directly compared telephone and Internet survey media (Beck et al., 2009; Booth-Kewley et al., 1992; Dillman et al., 2009; Knapp & Kirk, 2003; Szolnoki & Hoffmann, 2013; Zhang et al., 2017). These studies differ in context and in the populations sampled, and perhaps correspondingly give varied conclusions with respect to the degree and types of survey mode-induced biases. For instance, Booth-Kewley et al. (1992) in their survey of attitudes, and Knapp and Kirk (2003) in their survey of personal sensitivity find little difference between paper and Web-based surveys. Meanwhile, Dillman et al. (2009) and Szolnoki and Hoffmann (2013) find evidence of sampling bias and changing response patterns across these survey media. A broad social science literature investigates the related issues of survey response rates based on respondent incentives and the cost-effectiveness of data collection (Cook et al., 2000; Frederiks et al., 2020).

Thus, we note that past literature, both inside environmental economics and in the broader survey research community often find conflicting results with respect to the effects of survey mode on survey responses. While many studies have noted this inconsistency in the literature (e.g. Ethier et al., 2000; Maguire, 2009; Zhang et al., 2017), none have investigated its aetiology across research contexts. We note that even in economics studies that have carefully controlled for sampling effects across the tested survey modes, one finds significant mode effects (Boyle et al., 2016), while the other two do not (Ethier et al., 2000; Lindhjem & Navrud, 2011a).

We posit that the inconsistent conclusions with regard to survey mode effects are due to survey mode effects varying across sampled populations, regions and the contexts of the research. Populations across regions and nations may have differing degrees of susceptibility to non-response, avidity and social desirability bias, among others, if socio-cultural proclivities for responding to surveys and appealing to those around you vary by nation, and also relating to how the context of the study is generally perceived in the nation. Indeed, psychologists and political scientists have noted that the propensity for social desirability bias can vary across nations and cultures in various contexts; however, this insight has not yet been applied to survey analyses (Dunn & Shome, 2009; Karp & Brockington, 2005; Kim & Kim, 2016).

2. Survey methodology

The survey examined herein was conducted for the SESAME (Securing the European Electricity Supply Against Malicious and Accidental Threats) EU-FP7 research project during the second half of 2012 and first quarter of 2013 in all 27 nations that comprised the EU at that time.[3] The goals of the survey were to understand the public's experience with, and perception of, power outages, quantify their WTP to avoid power outages, and understand their acceptance of new energy infrastructure projects in their area. The survey had two main experiments included: a choice experiment designed to elicit the respondent's WTP to avoid power outages, and an ordinal question designed to assess the respondent's acceptance of a new hypothetical power line being sited near their home.

Administration of the survey proceeded in two distinct phases. The first phase consisted of a recruitment and a screening step. Respondents were recruited into the sample through initial telephone or email contact. Contact information was obtained from general panel lists maintained for survey purposes. The lists primarily came from two companies: one for telephone numbers and one for email addresses, and we note that the quality and exact composition of the lists

[3] See Gutierrez et al. (2013) for a detailed account of survey methodology and the English version of the full questionnaire.

Table 2 Average respondent characteristics by survey mode

Variable	Phone respondents	Internet respondents
Male	0.47 (0.5)	0.53 (0.5)
Age	53.76 (13.29)	39.57 (12.68)
Urban	0.26 (0.44)	0.38 (0.49)
College	0.42 (0.49)	0.39 (0.49)
hhsize	2.67 (1.25)	2.62 (1.23)
Children	0.31 (0.46)	0.39 (0.49)
Income (1000's EUR)	19.422 (14.802)	15.209 (11.685)

Note: N = 7659; Differences in means between Internet and phone subsamples are significant at the 5% level for all variables, based on between-group mean-comparison t-tests. Variables defined in Table 5.

varied across countries, so we must assume that the sample frame differed across survey mode. At this stage, potential participants were given a screening questionnaire which asked for respondents' socio-demographic information. Based on the answers to this questionnaire, respondents were either offered to take the full survey or thanked and dismissed in order to ensure that the final national samples are representative of each nation's population in the dimensions of gender, age, working status, income and rural residents.[4] The varied sampling frames between survey and phone respondents led to small differences in sample composition, where Web-based respondents were more likely to be male, younger and urban residents, as shown in Table 2.

The second phase was the administration of the actual survey. Phone respondents were sent a cover letter and information booklet, provided in their language, which they were asked to find and reference when the survey administrator made contact via telephone. Internet respondents were sent an email with the cover letter information contained within, and a link to the online version of the survey. Respondents to the Web-based survey were incentivised to participate in a small way by the company that maintains the email list. In order to counteract the possibility that this incentive would cause Internet respondents to speed through the survey, the online questionnaire required that respondents spend at least 3 min reading the important text sections of the survey, including the texts introducing the choice experiment and acceptance question. Furthermore, if a respondent took less than 10 min to complete the survey, they were not included in the final sample. Even so, survey duration varied between survey mode, with the telephone survey lasting an average of 30 min and the online survey lasting an average of 16 min. This is unsurprising given that most online respondents can read faster to themselves than an interviewer is able to read aloud.

Table 3 shows the response rates to the survey, broken down by survey mode. Overall, 4.75% of contact attempts resulted in a completed survey with

[4] The specific quota targets were balanced genders with up to 70% male as males are often the heads of households in many countries, 70% urban residents (live in towns of over 10,000 people), 25–30% of respondents in any one of the four age brackets, no more than 20% students/unemployed respondents who are often easier to recruit and a representative sample from the top and bottom 20% of the income spectrum.

Table 3 Response rates by survey mode and country

	Total contacts	Completed surveys	Total response rate	Internet contacts	Completed Internet surveys	Internetresponse rate	Phone contacts	Completed phone surveys	Phone response rate
Total	176072	8359	4.75%	28270	3652	12.92%	147802	4707	3.18%
Austria	7445	305	4.10%	834	148	17.75%	6611	157	2.37%
Belgium	5168	276	5.34%	601	90	14.98%	4567	186	4.07%
Bulgaria	5866	306	5.22%	519	126	24.28%	5347	180	3.37%
Cyprus	5481	266	4.85%	288	75	26.04%	5193	191	3.68%
Czech Republic	8577	334	3.89%	1693	166	9.81%	6884	168	2.44%
Denmark	8209	294	3.58%	1067	133	12.46%	7142	161	2.25%
Estonia	3092	299	9.67%	688	121	17.59%	2404	178	7.40%
Finland	9250	306	3.31%	933	138	14.79%	8317	168	2.02%
France	4108	311	7.57%	1108	131	11.82%	3000	180	6.00%
Germany	8330	344	4.13%	989	195	19.72%	7341	149	2.03%
Greece	7462	317	4.25%	1096	131	11.95%	6366	186	2.92%
Hungary	5194	336	6.47%	2056	146	7.10%	3138	190	6.05%
Ireland	6265	310	4.95%	796	140	17.59%	5469	170	3.11%
Italy	5605	308	5.50%	850	135	15.88%	4755	173	3.64%
Latvia	5280	314	5.95%	597	120	20.10%	4683	194	4.14%
Lithuania	5715	313	5.48%	627	132	21.05%	5088	181	3.56%
Luxembourg	10088	303	3.00%	757	140	18.49%	9331	163	1.75%
Malta	5763	278	4.82%	433	65	15.01%	5330	213	4.00%
Netherlands	4289	304	7.09%	860	128	14.88%	3429	176	5.13%
Poland	8750	335	3.83%	2281	156	6.84%	6469	179	2.77%
Portugal	8896	322	3.62%	2005	155	7.73%	6891	167	2.42%
Romania	5316	294	5.53%	682	131	19.21%	4634	163	3.52%
Slovakia	4906	289	5.89%	1122	99	8.82%	3784	190	5.02%
Slovenia	6745	300	4.45%	1075	136	12.65%	5670	164	2.89%
Spain	8448	328	3.88%	1517	169	11.14%	6931	159	2.29%
Sweden	5337	313	5.86%	1182	156	13.20%	4155	157	3.78%
UK	6487	354	5.46%	1614	190	11.77%	4873	164	3.37%

Note: A ‘completed’ survey in this table means the respondent answered all questions. Some respondents were dropped from our estimation sample of 7659 due to responses of ‘don’t know’ to key questions.

13% of emails resulting in a completed Internet-based survey and only 3.2% of phone calls resulting in a completed phone survey. These response rates are not directly comparable as many incorrect phone numbers were dialled. 40 144 or 27% of the total 147 802 numbers dialled, and to the adherence to quotas that disqualified 27 841 respondents, or 16% of the total contact attempts. Nevertheless, the lower response rates to the phone survey suggest potentially larger non-response bias in the phone-based sample. The noted differences between the phone and Web-based samples in our study suggest that our data reject Hypothesis 1.1, namely that sampling biases are present between survey mode samples.

2.1 The choice experiment

The first experiment included in the survey is a non-market valuation exercise with the goal of estimating respondents WTP to avoid power outages. The choice experiment asked respondents to imagine a power outage with a specified duration, start and end time, season and area (residential street or whole country). A visual depiction of one of the eight scenarios shown to respondents is reproduced in Figure 2. We use 1, 4, 12 and 24 h as durations for the outage scenarios, generally reflecting the durations found in the literature (Baarsma & Hop, 2009; Carlsson & Martinsson, 2007, 2008; Layton & Moeltner, 2004; Reichl et al., 2013). Since the number of choice tasks is limited to eight due to length and budgetary constraints, we revealed two random scenarios of each duration to every respondent.

After seeing a depiction of a power outage scenario, respondents were offered the option to pay a specified amount of money in their native currency to avoid experiencing the outage. This is referred to as the 'bid price' and varies for each respondent, across countries and across outage scenarios. The bid price design is based on a previous, similar WTP study conducted by Reichl et al. (2013) for the nation of Austria, which used the D-efficiency

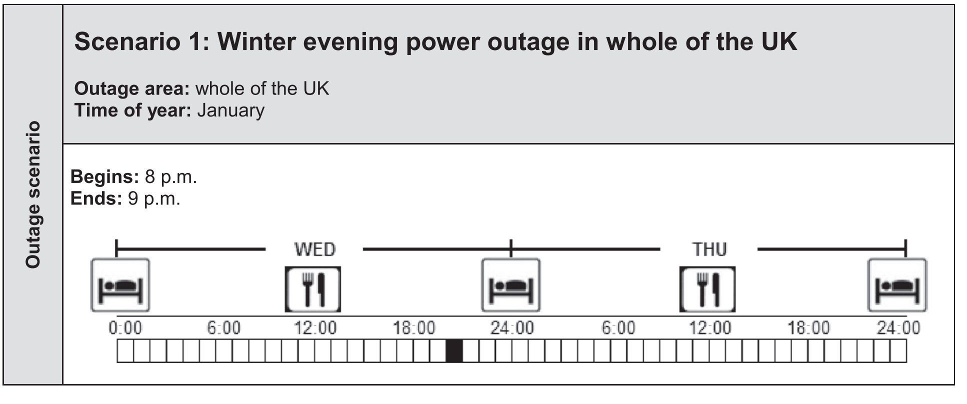

Figure 2 Example of blackout scenario from the English/UK version of the survey.

Table 4 Proportion of respondents who accepted the offered bid by scenario

	Scenario			
	1	2	3	4
Lowest bid	0.65	0.81	0.60	0.55
Low bid	0.48	0.68	0.47	0.41
High bid	0.34	0.44	0.25	0.34
Highest bid	0.29	0.33	0.20	0.18

Note: $N = 7741$.

criterion with balanced utilities to set the bids (Burgess & Street, 2005; Huber & Zwerina, 1996; Scarpa & Rose, 2008). Two of the four bids of each scenario used in the Austrian study are adopted here with a correction for the difference in income distribution between Austria and every other nation. The other two bids of each scenario are held constant between countries to enable cross-country comparison. Bids in the Austrian version ranged from €0.25 to €70 with 16 values included in total. As expected, the proportion of respondents accepting the option to pay is decreasing in bid price for all four scenarios as shown in Table 4.[5] The survey text stipulated that by refusing to pay the bid price the respondent would experience the full extent of the outage depicted by the scenario.

2.2 Public acceptance question

The second major experiment included in the survey presented respondents with a hypothetical scenario where a new transmission power line with pylons was to be built 250 metres from their home. This was followed by an ordinal question designed to assess the level of acceptance that respondents would have to this scenario. In some scenarios, respondents were given one of three 'treatment' scripts, which espoused a potential benefit from the power line in the preface of the question.

Next, respondents were asked: 'How do you think YOU would react to the announcement of this power infrastructure program?' With the ability to choose between four possible reactions: 'definitely not accept without opposition' (DNA), 'probably not accept without opposition' (PNA), 'probably accept without opposition' (PYA) and 'definitely accept without opposition' (DYA).[6]

An initial look at the survey responses from the full sample in Figure 3 illustrates the challenge of public acceptance across Europe, with high proportions of DNA responses in many EU nations, and heterogeneous

[5] For a full analysis of this choice experiment and further discussion of the WTP elicitation, please see Cohen et al. (2018a) and Cohen et al. (2018b).

[6] For more details on the structure and results of the public acceptance experiment, please see Cohen et al. (2016).

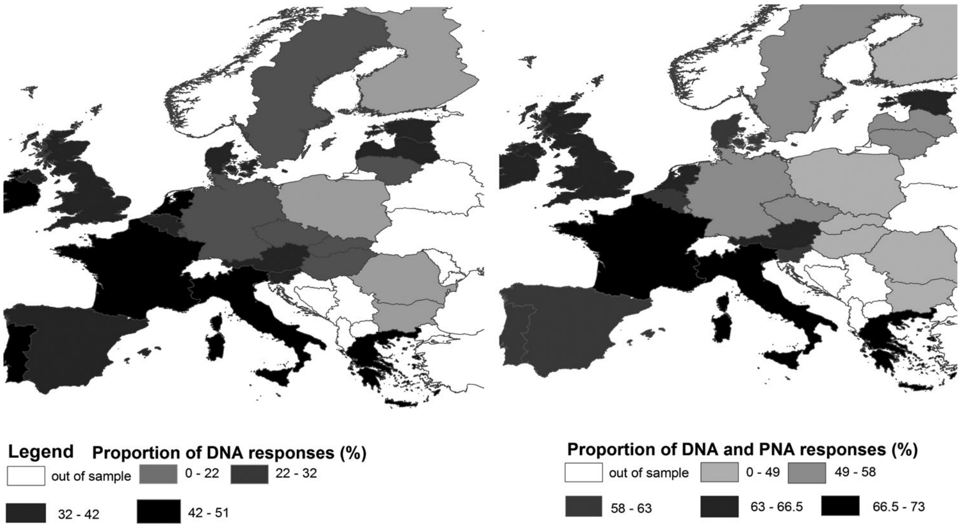

Figure 3 Proportion of opposition responses to acceptance question in full sample.

response trends between nations. Overall, 34% of respondents indicated that they would definitely not accept the new project without opposition, while only 12% said that they would definitely accept. Furthermore, as shown by the second panel of Figure 3, for all but 6 of the 27 EU nations the proportion of responses of either definitely not or probably not accept is greater than half.

3. Analysis and results

The survey effort yielded 7741 questionnaires where all the necessary questions had been completed without ‘don’t know’ answers, leaving an average of 286 survey respondents per nation. We analyse the survey data separately for each of the two contexts discussed above. The variables pulled used in one or both of these estimations are defined in Table 5.

3.1 Willingness to pay model

The choice experiment, explained in Section 2.1, is designed to elicit each respondent’s WTP to avoid power outages. Each respondent completed eight choice tasks where they could either experience a hypothetical power outage of specified characteristics or pay the offered bid price P_{is} to avoid the outage. Thus, the data are in panel format with eight observations per respondent. We model the data using a probit model that allows for respondent-specific random effects. As is becoming standard in environmental economics, identification of the probit model is achieved by dividing the respondent *i*’s indirect utility related to scenario *s* by the respondent marginal utility of

Table 5 Descriptions of explanatory variables used in econometric analysis

Variable name	Description
Respondent-specific variables included in both models	
web	Took the Web-based survey
urban	Lives in urban area
male	Is male
age35t45	Between ages 35 and 45
age46t60	between ages 46 and 60
over60	Over age 60
hhsize	Number of household members
college	College degree
out1t4	Experienced 1- to 4-h outage
out4t8	Experienced 4- to 8-h outage
out8t24	Experienced 8- to 24-h outage
over24	Experienced over 24-h outage
numoutages	Number of outages in past year
Variables relevant only to the WTP choice experiment	
rotation	Dummy variable for the ordering of the choice scenarios
summer	Scenario specified summer power outage
wholecountry	Scenario specified power outage across entire nation
Variables relevant only to the social acceptance experiment	
income	Self-reported net income (1000's EUR)
T1	Was told new power lines will have economic benefits
T2	Was told new power lines will have environmental benefits
T3	Was told new power lines will have community benefits
posutil	Respondent has positive view of their energy provider
negutil	Respondent has negative view of their energy provider
satisfied	Respondent is satisfied with their supply security
yearsinhome	Number of years respondent has lived at current home
needsgrids	Believes grid expansion is necessary

income, resulting in a price coefficient that is fixed at 1 for all respondents with corresponding unit-level variances. This practice is denoted as the 'WTP space' method in contrast to modelling respondents' choices in 'utility-space', where the scale of the utility is fixed for all respondents (Scarpa et al., 2008; Sonnier et al., 2007; Train & Weeks, 2005). Further details on the WTP space model are given online in Cohen and Reichl (2021).

The model of each respondent *i*'s indirect utility in the WTP space for outage scenario *s* can thus be written as:

$$\begin{gathered} \nu_{is} = d_s(X'_{is}\beta) - P_{is} + d_s(\alpha_i + \varepsilon_{is}) \\ \alpha_i \sim N(0, \delta^2_\alpha)\ \varepsilon_{is} \sim N(0, \delta^2_{i,\varepsilon}) \end{gathered} \tag{3}$$

where d_s is the duration of outage s in hours,[7] P_{is} is the bid value shown to respondent i for scenario s, $\mathbf{X}$ are the explanatory variables defined in

[7] Assuming a linear effect of duration is a simplifying assumption for our purpose here of estimating the mode effects. See Cohen et al. (2018b) for the derivation of a full simultaneous equation model.

Table 5, α_i is the random effect term for respondent i, and ε_{is} is a stochastic unobservable component of indirect utility. We will observe a positive response to the survey question when $v_{is} > 0$, indicating that the respondent increases their utility by paying the bid price rather than experiencing the outage. The error components $(\alpha_i + \varepsilon_{is})$ are both assumed to be normally distributed with population variances, δ^2_α the variance between respondents, and $\delta^2_{i,s}$ the variance within respondents. Both error terms are scaled by the duration of the outage to account for the fact that longer outages can have a wider range of possible effects. Finally, β is the vector of slope coefficients associated with the explanatory variables. The elements of β are interpreted as the marginal effects of a one unit increase in the associated variable on the WTP to avoid an hour of power outage.

We estimate two runs of the model in (3) that vary based on the survey mode variables included. Both models contain fixed effects[8] at the country level to control for unobserved heterogeneity and a group of respondent-specific and scenario-specific control variables. The included variables are defined in Table 5. Since both models are estimated in the Bayesian fashion, the output is a posterior distribution for each parameter. We summarise each posterior distribution by its mean, the probability to the right of zero and the 95% credible interval, which here is defined as the highest posterior density (HPD) interval that 95% of the posterior probability lies between.

Using the output from the second model, which contains country-specific indicators for survey mode, we calculate the average WTP to avoid an hour of power outage in each nation; following Campbell (2007), we include the random effects term in this calculation. The mean WTP figures are shown in Table 6 for the full sample and for the subsamples of Web-based and phone survey respondents. Mean WTP is shown to be highly heterogeneous across nations and higher for phone respondents than Web-based respondents, with an average value across the sample of €0.74, an average value of €0.13 for Internet respondents and an average value of €1.18 for phone respondents. The results confirm the implicit hypothesis of this choice experiment, that citizens have a positive WTP to avoid power outages, on average.

The first model accounts for survey mode with a simple indicator variable that takes a value of 1 if the respondent used the Internet survey medium, we call this variable '*web*'. The estimates of the $\boldsymbol{\beta}$ parameters from this model are summarised in Table 7. The results show that survey mode has an effect on response patterns in this context, where Web-based surveys decrease the WTP to avoid an hour of power outage by €0.73, on average.

A second model is estimated, which includes interaction terms between the *web* variable and country indicator terms. From this model, we can estimate the effect of survey mode on WTP in each nation. The results relating to survey mode from this model are summarised in Table 8. We find that Web-

[8] Specifically, we specify independent priors over each of the country indicator variables as discussed in, for example, Rossi et al. (2005) and Rendon (2013).

Table 6 Mean WTP to avoid 1 h of power outage in each EU-27 nation overall and by survey mode (€)

Country	Full sample		Web-based respondents		Phone respondents	
	Mean WTP	SD of WTP	Mean WTP	SD of WTP	Mean WTP	SD of WTP
France	–0.242	0.085	–0.199	0.127	–0.274	0.115
Germany	0.704	0.073	0.248	0.096	1.295	0.110
Italy	0.743	0.076	–0.135	0.126	1.390	0.097
UK	0.668	0.065	0.375	0.092	0.950	0.091
Austria	0.701	0.076	0.191	0.106	1.145	0.108
Belgium	0.497	0.082	0.161	0.141	0.659	0.099
Denmark	1.424	0.101	0.355	0.121	2.254	0.148
Finland	1.096	0.077	0.007	0.119	1.907	0.101
Netherlands	0.432	0.076	0.122	0.120	0.647	0.094
Spain	0.680	0.075	0.107	0.095	1.249	0.118
Sweden	0.578	0.072	0.019	0.110	1.088	0.098
Portugal	0.466	0.070	0.000	0.098	0.937	0.097
Ireland	0.944	0.065	0.391	0.098	1.390	0.090
Luxembourg	1.484	0.103	0.157	0.134	2.519	0.153
Bulgaria	0.686	0.081	0.255	0.100	0.988	0.118
Czech Republic	0.104	0.081	–0.588	0.139	0.719	0.091
Estonia	0.481	0.074	–0.386	0.137	1.002	0.082
Hungary	0.788	0.061	0.357	0.088	1.099	0.081
Latvia	0.363	0.065	0.054	0.111	0.529	0.082
Lithuania	0.621	0.060	0.278	0.092	0.854	0.081
Poland	0.962	0.069	0.157	0.099	1.618	0.095
Romania	0.908	0.074	0.624	0.106	1.126	0.105
Slovakia	0.612	0.070	0.046	0.113	0.905	0.091
Slovenia	1.036	0.079	0.086	0.106	1.780	0.110
Greece	0.902	0.071	0.414	0.109	1.237	0.094
Cyprus	1.214	0.097	–0.317	0.210	1.655	0.110
Malta	1.069	0.077	0.600	0.146	1.199	0.087
Average	0.738	0.076	0.125	0.116	1.180	0.102

Note: Mean WTP figures are aggregated across respondents and outage scenarios; SD (Standard Deviation) is reported. The 95% credible intervals for avg. Web and phone WTP do not overlap, suggesting a significant survey mode effect.

based survey mode reduces estimated WTP in 26 of the 27 countries in our sample. While the sign of the estimate is the same across these 26 nations, there is notable heterogeneity in the magnitude of the estimated effects of survey mode across nations. Some nations exhibit mean effects much greater in magnitude than the average effect of –0.73, including Denmark, Finland, Luxembourg, Slovenia and Cyprus. The lone outlier is France, which shows no effect from survey mode; however, this may be due to the low WTP exhibited by France overall.

3.2 Public acceptance model

The dependent variable for the public acceptance model is the ordinal response given by the respondent to the question regarding a new

Table 7 Marginal effects on WTP to avoid an hour of power outage (€)

Variable	95% credible interval		95% credible interval	
	Lower bound	Mean	Upper bound	$p > 0$
web	−0.79	−0.73	−0.68	0.00
rotation	0.06	0.10	0.15	1.00
urban	0.08	0.13	0.19	1.00
male	−0.13	−0.09	−0.04	0.00
age35t45	−0.04	0.03	0.10	0.81
age46t60	−0.05	0.02	0.09	0.71
over60	−0.01	0.07	0.15	0.96
hhsize	0.04	0.06	0.08	1.00
college	0.14	0.19	0.23	1.00
out_1t4	−0.01	0.05	0.10	0.95
out_4t8	−0.09	−0.02	0.06	0.33
out_8t24	−0.16	−0.06	0.04	0.11
out_over24	−0.25	−0.13	−0.02	0.01
numoutages	−0.01	0.00	0.00	0.08
summer	−0.82	−0.78	−0.73	0.00
wholecountry	0.56	0.60	0.65	1.00

Note: $N = 7741$. Model also includes country fixed-effects Credible interval defined by the highest posterior density method.

transmission line in their area. Respondents chose one of four possible responses ranging from DNA (definitely not accept) to DYA (definitely accept), and DNA responses were coded as 1 so that higher values of the dependent variable reflect higher levels of acceptance. This ordinal response structure lends itself to econometric modelling using an ordered probit approach. Respondents were also able to respond ‘don’t know’ to this question, a response observed in 4% of respondents who were dropped from the sample.

We estimate two ordered probit models for this analysis, which are distinguished by the included survey mode variables. Both models contain country-level fixed-effects variables to control for unobserved heterogeneity and a suite of respondent-specific control variables. Included variables are defined in Table 5. Since we include *income* directly in this model, we lose 82 respondents due to improperly entered income information, leaving a final sample size of 7,659.

The first ordered probit model accounts for survey mode using only the *web* variable. The marginal effect of the *web* variable in this case shows the expected change in the predicted probability that the respondent gives a DNA response when the respondent took the Web-based survey as opposed to the telephone survey. The results from this model are shown in Table 9. The results show that, on average, taking the Web-based survey is associated with a more accepting response and decreases the probability that a respondent chooses DNA by 1.9%. This result is not statistically significant at the 5% level, but is at the 10% level. While our results show that survey mode may

Table 8 Country-specific effects of Web-based survey mode on WTP to avoid power outages (EUR)

Variable	95% credible interval		95% credible interval	
	Lower bound	Mean	Upper bound	$p > 0$
France	−0.24	0.04	0.31	0.60
Germany	−0.99	−0.76	−0.52	0.00
Italy	−1.32	−1.07	−0.82	0.00
UK	−0.60	−0.40	−0.20	0.00
Austria	−0.91	−0.66	−0.43	0.00
Belgium	−0.58	−0.30	−0.04	0.01
Denmark	−1.80	−1.47	−1.15	0.00
Finland	−1.60	−1.35	−1.10	0.00
Netherlands	−0.64	−0.41	−0.17	0.00
Spain	−1.00	−0.77	−0.53	0.00
Sweden	−0.97	−0.74	−0.50	0.00
Portugal	−0.90	−0.68	−0.47	0.00
Ireland	−0.95	−0.73	−0.52	0.00
Luxembourg	−2.22	−1.87	−1.52	0.00
Bulgaria	−0.74	−0.49	−0.26	0.00
Czech Republic	−1.22	−0.90	−0.61	0.00
Estonia	−1.25	−0.98	−0.72	0.00
Hungary	−0.67	−0.48	−0.30	0.00
Latvia	−0.58	−0.36	−0.15	0.00
Lithuania	−0.62	−0.43	−0.24	0.00
Poland	−1.28	−1.05	−0.83	0.00
Romania	−0.64	−0.41	−0.17	0.00
Slovakia	−0.86	−0.63	−0.41	0.00
Slovenia	−1.46	−1.22	−0.97	0.00
Greece	−0.82	−0.59	−0.36	0.00
Cyprus	−1.98	−1.57	−1.20	0.00
Malta	−0.70	−0.44	−0.18	0.00

Note: $N = 7741$; Model also includes country fixed-effects and respondent-specific variables (excluding *web*) shown in Table 7; variables are constructed as country indicators multiplied by *web*; credible interval defined by the highest posterior density method.

drive responses, the effect of survey mode on response probability is of low magnitude and is on the border of statistical significance. In comparison, changing the question preface script, by stipulating economic or environmental benefits from the transmission line, has a much larger effect, decreasing the probability of a DNA response by 10.2% and 10.5%, respectively. The results confirm the implicit hypothesis of this survey experiment that citizens are more willing to accept new energy infrastructure in their area if the infrastructure project brings environmental, economic or community-related benefits.

We use the second ordered probit model to dig deeper into the effect of survey mode across nations. We estimate the model with the interactions between the *web* survey mode indicator and all 27 country indicators as explanatory variables. This allows us to estimate the effect of survey mode at a country-specific level. The results of the survey mode variables in this model

Table 9 Public acceptance probit model results for respondent-specific variables

Variable	Coeff	SE	Marg. Eff.† prob(DNA)	SE prob(DNA)
web	0.0573	0.0298	−0.0197	0.0102
T1	0.296**	0.0341	−0.102**	0.0116
T2	0.305**	0.0346	−0.105**	0.0118
T3	0.0902**	0.0340	−0.0310**	0.0117
income	−0.00420**	0.00150	0.00144**	0.000515
urban	0.0458	0.0287	−0.0157	0.00985
male	0.218**	0.0255	−0.0751**	0.00867
age35t45	−0.119**	0.0350	0.0409**	0.0120
age46t60	−0.174**	0.0379	0.0598**	0.0130
over60	−0.147**	0.0475	0.0505**	0.0163
hhsize	−0.0119	0.0116	0.00409	0.00398
college	−0.0587*	0.0275	0.0202*	0.00944
posutil	0.0644*	0.0262	−0.0221*	0.00901
negutil	−0.182**	0.0597	0.0626**	0.0205
supsecsatis	0.0779	0.0474	−0.0268	0.0163
yearsinhome	−0.00153	0.00104	0.000524	0.000357
needgrids	0.260**	0.0263	−0.0895**	0.00893

Note: $N = 7659$; model also includes country fixed-effects; variables are defined in Table 5; * denotes statistical significance at 5% level and ** at 1% level. †Marginal effect of a one unit increase in variable on probability respondent gives a DNA (definitely not accept) response.

are shown in Table 10, given as the marginal effect of Web-based survey mode on the predicted probability of each response category. The results show that the positive overall effect of Internet survey mode on acceptance is driven by the responses in two nations, Greece and Cyprus. In these two nations, we estimate that Web surveys decreases the probability of a DNA response by 13.4% and 14.9%, respectively. In other nations, the marginal effect of survey mode is not statistically significant, except for Lithuania, where the Web-based survey mode is associated with decreased acceptance. Thus, the overall positive effect of Web surveys on acceptance is mostly driven by the strong effect of this survey mode exhibited in Greece and Cyprus, whereas the survey mode is not affecting responses to the acceptance question to a statistically significant degree in 24 of the 27 sample nations.

4. Discussion and conclusions

In the context of a choice experiment assessing WTP to avoid power outages, we find that Internet survey respondents state lower WTP than their phone-surveyed counterparts. This result holds over 26 of the 27 nations in the sample, suggesting survey mode effects exist across the majority of sample nations in this research context. However, our results show that the magnitude of survey mode bias can, and likely will, vary across nations. European nations, despite their geographic proximity, are highly heterogeneous in terms of their historical, cultural, and institutional backgrounds. All

Table 10 Country-specific marginal effects of Web-based survey mode on the probability of choosing a given response category

Variable	P(DNA)		P(PNA)		P(PYA)		P(DYA)	
	Marg. Eff.	SE	Marg. Eff.	SE	Marg. Eff.	SE	Marg. Eff.	SE
France	−0.0489	0.0423	−0.00318	0.00278	0.0252	0.0219	0.0268	0.0232
Germany	−0.0681	0.0424	−0.00443	0.00279	0.0351	0.0219	0.0374	0.0232
Italy	−0.0135	0.0448	−0.000875	0.00291	0.00694	0.0231	0.00739	0.0246
UK	−0.0581	0.0434	−0.00378	0.00284	0.0300	0.0224	0.0319	0.0238
Austria	−0.0445	0.0437	−0.00289	0.00286	0.0230	0.0225	0.0244	0.0240
Belgium	−0.00539	0.0434	−0.000350	0.00282	0.00278	0.0224	0.00296	0.0238
Denmark	−0.0385	0.0466	−0.00250	0.00304	0.0198	0.0241	0.0211	0.0256
Finland	−0.0104	0.0366	−0.000676	0.00238	0.00537	0.0189	0.00571	0.0201
Netherlands	−0.0440	0.0453	−0.00286	0.00295	0.0227	0.0234	0.0242	0.0249
Spain	−0.00943	0.0420	−0.000613	0.00273	0.00487	0.0217	0.00518	0.0230
Sweden	−0.0222	0.0400	−0.00144	0.00260	0.0115	0.0207	0.0122	0.0220
Portugal	−0.0300	0.0482	−0.00195	0.00313	0.0155	0.0249	0.0165	0.0265
Ireland	−0.0610	0.0461	−0.00397	0.00303	0.0315	0.0238	0.0335	0.0253
Luxembourg	0.00184	0.0417	0.000119	0.00271	−0.000948	0.0215	−0.00101	0.0229
Bulgaria	0.0150	0.0434	0.000972	0.00283	−0.00771	0.0224	−0.00821	0.0238
Czech Republic	0.0218	0.0428	0.00142	0.00280	−0.0113	0.0221	−0.0120	0.0235
Estonia	0.0306	0.0460	0.00199	0.00300	−0.0158	0.0237	−0.0168	0.0253
Hungary	−0.0512	0.0394	−0.00333	0.00258	0.0264	0.0203	0.0281	0.0216
Latvia	0.0100	0.0466	0.000651	0.00303	−0.00517	0.0241	−0.00550	0.0256
Lithuania	0.0972*	0.0441	0.00632*	0.00300	−0.0502*	0.0227	−0.0534*	0.0243
Poland	−0.0451	0.0394	−0.00293	0.00257	0.0233	0.0203	0.0248	0.0216
Romania	0.0521	0.0475	0.00339	0.00312	−0.0269	0.0245	−0.0286	0.0261
Slovakia	0.0663	0.0471	0.00431	0.00313	−0.0342	0.0242	−0.0364	0.0259
Greece	−0.134**	0.0457	−0.00872**	0.00317	0.0692**	0.0236	0.0737**	0.0251
Cyprus	−0.149*	0.0587	−0.00970*	0.00401	0.0769*	0.0303	0.0819*	0.0323
Slovenia	−0.0491	0.0477	−0.00319	0.00311	0.0253	0.0246	0.0270	0.0262
Malta	0.0213	0.0531	0.00138	0.00346	−0.0110	0.0274	−0.0117	0.0291

Note: $N = 7659$; model also includes country fixed-effects and respondent-specific variables (excluding *web*) shown in Table 9; * denotes statistical significance at 5% level and ** at 1% level.

of these factors, along with sampling effects, likely play a role in driving the magnitude of the observed survey mode bias across nations, which makes unravelling the causes of this heterogeneity difficult. However, this does not rule out the possibility that more homogenous, but distinct regions, such as U.S. states, could also have varying responses to survey mode. Such a dynamic would explain some of the conflicting findings of earlier literature (e.g. Whittaker et al. (1998) vs. Loomis and King (1994)).

In the research context of public acceptance to energy infrastructure, we find a different result. While the overall effect of Internet survey mode suggests this mode may increase stated acceptance, the magnitude of this effect is small, and the coefficient is only significant at the 10% level. Looking at the country-specific effects, we find that two nations are mainly driving this result, Greece and Cyprus, where the effect of survey mode is having a strong impact on responses to the acceptance question. The survey mode does not show a statistically significant effect in the other 24 nations. It is interesting to note that Greece and Cyprus share a language, and to some extent have similar culture and history, which could possibly account for their similarity in our study.

Comparing the results with respect to survey mode across contexts, we find evidence that survey mode effects can vary based on the research question posed. Even though both of our research contexts deal broadly with energy, and both questions were included in the same survey, responses to the questions were not equally affected by survey mode. Specifically, the responses to the choice experiment eliciting WTP to avoid power outages showed much stronger and more consistent survey mode effects than the responses to the public acceptance of energy infrastructure question.

To attempt to understand why survey mode effects may differ across research contexts, we consider how the different types of survey-induced biases may vary across contexts. We note the conceptual findings in Section 1.1 and specifically Proposition 1.3, which shows the effects of a situation where sampling biases and measurement biases are both present. The high likelihood of sampling biases in between the Web- and phone-based samples in this study is noted in Section 2 and shown by sample statistics such as response rates and socio-demographic composition of the samples. This is due to the different sampling frames from the Internet and phone respondents within a given nation, and different propensities for non-response and avidity bias since the Web participants were incentivised slightly for their cooperation. However, these sampling effects are expected to be present in both research contexts, as the same individuals responded to both experimental scenarios. Nevertheless, we find that one research context is affected by survey mode and the other context is generally unaffected by survey mode. This may imply that the remaining bias type, measurement bias, may be varying across research contexts. This supposition is upheld by the recent work of Boyle et al. (2016) who took great care to control for sampling biases, but still find significant differences between values elicited via web and mail surveys.

There are many types of measurement bias that may could plausibly drive this result. However, we speculate that social desirability bias may be the most likely candidate. With social desirability bias the respondent wants to give the socially appropriate answer to the survey question when on the phone, or present, with a survey administrator. In the context of power outages, we argue that the socially appropriate behaviour is clear. Since power outages have a negative impact on the majority of people affected, and can even be dangerous if crucial systems go down, when given an opportunity to avoid this bad outcome by paying a small fee, the socially responsible thing to do is to pay the fee. Phone survey respondents may feel this social pressure more acutely due to the presence of an interviewer on the other end of the line and may be more likely to accept the bid price. This dynamic would explain our survey mode results in the context of the WTP to avoid power outages.

Conversely, in the context of public acceptance of energy infrastructure, we argue that the socially appropriate response is not clear. On one hand, the

NIMBY response is often stigmatised as hypocritical (Meyer, 2010) such that a phone respondent may feel more social pressure to accept the nearby infrastructure, which is consistent with the result we see in Cyprus and Greece. On the other hand, defending one's locale from the perceived detriments of a new transmission line may be viewed as the appropriate response (Oreskes, 2014). This sentiment is evidenced by the numerous local organisations that have sprung up to fight against energy infrastructure projects across Europe. A respondent of this mindset may feel social pressure to oppose the infrastructure project when on the phone with an interviewer, which would lead to the response bias exhibited by Lithuanian participants in Table 10.

In conclusion, our results show differences in the magnitude of survey mode bias across research contexts. This suggests that the inconsistent findings of past literature with respect to survey mode effects may be due to the varying contexts under which these studies were performed. Researchers should think critically about their specific research context and the potential causes of measurement bias within this context when planning a survey. While bias in survey responses remains a critical issue for the applied social science community moving forward, this paper has taken a first step in illuminating the heterogeneous effects of survey mode across research contexts. Further research is needed in this area to unveil the optimal survey strategy for specific contexts and geographies.

Data availability statement

The data and code that support the findings of this study are freely available alongside this publication on the journal website.

References

ACER (2020) Methodology for calculating the value of lost load, the cost of new entry and the reliability standard. Agency for the Cooperation of Energy Regulators: Regulatory decision.

Arrow, K., Solow, R., Portney, P., Leamer, E., Radner, R. & Schuman, H. (1993) Report of the NOAA panel on contingent valuation. Technical report, National Oceanic and Atmospheric Administration.

Baarsma, B.E. & Hop, J.P. (2009) Pricing power outages in the Netherlands. *Energy*, 34, 1378–1386.

Baker, R., Brick, J.M., Bates, N.A., Battaglia, M., Couper, M.P., Dever, J.A. et al. (2013) Summary report of the AAPOR task force on non-probability sampling. *Journal of Survey Statistics and Methodology*, 1, 90–143.

Beck, K.H., Yan, A.F. & Wang, M.Q. (2009) A comparison of web-based and telephone surveys for assessing traffic safety concerns, beliefs, and behaviors. *Journal of Safety Research*, 40, 377–381.

Bell, J., Huber, J. & Viscusi, K. (2011) Survey mode effects on valuation of environmental goods. *International Journal of Environmental Research and Public Health*, 8.

Bethlehem, J. (2010) Selection bias in web surveys. *International Statistical Review*, 78, 161–188.

Biemer, P. (2001) Nonresponse bias and measurement bias in a comparison of face to face and telephone interviewing. *Journal of Official Statistics*, 17, 295–320.
Booth-Kewley, S., Edwards, J. & Rosenfeld, P. (1992) Impression management, social desirability, and computer administration of attitude questionnaires: Does the computer make a difference? *Journal of Applied Psychology*, 77, 562–566.
Borkan, B. (2010) The mode effect in mixed-mode surveys: mail and web surveys. *Social Science Computer Review*, 28, 371–380.
Boxall, P., Adamowicz, W.L.V. & Moon, A. (2009) Complexity in choice experiments: choice of the status quo alternative and implications for welfare measurement*. *Australian Journal of Agricultural and Resource Economics*, 53, 503–519.
Boyle, K.J., Morrison, M., MacDonald, D.H., Duncan, R. & Rose, J. (2016) Investigating internet and mail implementation of stated-preference surveys while controlling for differences in sample frames. *Environmental and Resource Economics*, 64, 401–419.
Breunig, R. & McKibbin, R. (2011) The effect of survey design on household reporting of financial difficulty. *Journal of the Royal Statistical Society: Series A*, 174, 991–1005.
Burgess, L. & Street, D.J. (2005) Optimal designs for choice experiments with asymmetric attributes. *Journal of Statistical Planning and Inference*, 134, 288–301.
Campbell, D. (2007) Willingness to pay for rural landscape improvements: combining mixed logit and random-effects models. *Journal of Agricultural Economics*, 58, 467–483.
Campbell, R.M., Venn, T.J. & Anderson, N.M. (2018) Cost and performance tradeoffs between mail and internet survey modes in a nonmarket valuation study. *Journal of Environmental Management*, 210, 316–327.
Carlsson, F. & Martinsson, P. (2007) Willingness to pay among Swedish households to avoid power outages: a random parameter tobit model approach. *Energy Journal*, 28, 75–89.
Carlsson, F. & Martinsson, P. (2008) Does it matter when a power outage occurs? A choice experiment study on the willingness to pay to avoid power outages. *Energy Economics*, 30, 1232–1245.
Chang, L. & Krosnick, J.A. (2010) Comparing oral interviewing with self-administered computerized questionnaires an experiment. *Public Opinion Quarterly*, 74, 154–167.
Cohen, J., Moeltner, K., Reichl, J. & Schmidthaler, M. (2016) An empirical analysis of local opposition to new transmission lines across the EU-27. *The Energy Journal*, 37, 59–82.
Cohen, J., Moeltner, K., Reichl, J. & Schmidthaler, M. (2018a) Effect of global warming on willingness to pay for uninterrupted electricity supply in European nations. *Nature Energy*, 3, 37–45.
Cohen, J., Moeltner, K., Reichl, J. & Schmidthaler, M. (2018b) Valuing electricity-dependent infrastructure: an essential-input approach. *Energy Economics*, 73, 258–273.
Cohen, J., Reichl, J. & Schmidthaler, M. (2014) Re-focussing research efforts on the public acceptance of energy infrastructure: a critical review. *Energy*, 76, 4–9.
Cohen, J.J. & Reichl, J. (2021) Supplementary materials for the article 'Comparing internet and phone survey mode effects across countries and research contexts'. Zenodo database.
Cook, C., Heath, F. & Thompson, R.L. (2000) A meta-analysis of response rates in web- or internet-based surveys. *Educational and Psychological Measurement*, 60, 821–836.
Cotton, M. & Devine-Wright, P. (2013) Putting pylons into place: a UK case study of public perspectives on the impacts of high voltage overhead transmission lines. *Journal of Environmental Planning and Management*, 56, 1225–1245.
Devine-Wright, P. (2007) Reconsidering public acceptance of renewable energy technologies: a critical review. In: Jamas, T., Grubb, M. & Pollitt, M. (Eds.) *Delivering a low carbon electricity system: Technologies, economics and policy*. Cambridge University Press.
Devine-Wright, P. (2013) Explaining NIMBY objections to a power line: the role of personal, place attachment and project-related factors. *Environment and Behavior*, 45, 761–781.
Devine-Wright, P. & Batel, S. (2013) Explaining public preferences for high voltage pylon designs: an empirical study of perceived fit in a rural landscape. *Land Use Policy*, 31, 640–649.

Dillman, D.A., Phelps, G., Tortora, R., Swift, K., Kohrell, J., Berck, J. et al. (2009) Response rate and measurement differences in mixed-mode surveys using mail, telephone, interactive voice response (IVR) and the internet. *Social Science Research*, 38, 1–18.

Dillman, D.A., Smyth, J.D. & Christian, L.M. (2014) *Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method*. Wiley.

Dunn, P. & Shome, A. (2009) Cultural crossvergence and social desirability bias: ethical evaluations by Chinese and Canadian business students. *Journal of Business Ethics*, 85, 527–543.

Elliott, P. & Wadley, D. (2012) Coming to terms with power lines. *International Planning Studies*, 17, 179–201.

Ethier, R.G., Poe, G.L., Schulze, W.D. & Clark, J. (2000) A comparison of hypothetical phone and mail contingent valuation responses for green pricing electricity programs. *Land Economics*, 76, 54–67.

Fleming, C.M. & Bowden, M. (2009) Web-based surveys as an alternative to traditional mail methods. *Journal of Environmental Management*, 90, 284–292.

Frederiks, E.R., Romanach, L.M., Berry, A. & Toscas, P. (2020) Making energy surveys more impactful: testing material and non-monetary response strategies. *Energy Research & Social Science*, 63, 101409.

Greene, W. (2012) *Econometric Analysis*, 7th edition. Pearson.

Gutierrez, F.G., Schmidthaler, M., Reichl, J., Voronca, S. & Roman, T.E. (2013) Public effects knowledge base - deliverable d2.2 of the project SESAME (securing the European electricity supply against malicious and accidental threats). Technical report, Transelectrica and Politecnico di Torino.

Heerwegh, D. (2009) Mode Differences between face-to-face and web surveys: an experimental investigation of data quality and social desirability effects. *International Journal of Public Opinion Research*, 21, 111–121.

Holbrook, A.L. & Krosnick, J.A. (2009) Social desirability bias in voter turnout reports: tests using the item count technique. *Public Opinion Quarterly*, 74, 37–67.

Huber, J. & Zwerina, K. (1996) The importance of utility balance in efficient choice designs. *Journal of Marketing Research*, 39, 214–227.

Hyslop, D.R. & Imbens, G.W. (2001) Bias from classical and other forms of measurement error. *Journal of Business & Economic Statistics*, 19, 475–481.

Johnston, R.J., Boyle, K.J., Adamowicz, W.V., Bennett, J., Brouwer, R., Cameron, T.A. et al. (2017) Contemporary guidance for stated preference studies. *Journal of the Association of Environmental and Resource Economists*, 4, 319–405.

Karp, J.A. & Brockington, D. (2005) August). Social desirability and response validity: a comparative analysis of overreporting voter turnout in five countries. *Journal of Politics*, 67, 825–840.

Kim, S. & Kim, S. (2016) National culture and social desirability bias in measuring public service motivation. *Administration & Society*, 48, 444–476.

Klausch, T., Hox, J. & Schouten, B. (2015) Selection error in single- and mixed mode surveys of the Dutch general population. *Journal of the Royal Statistical Society: Series A*, 178, 945–961.

Knapp, H. & Kirk, S.A. (2003) Using pencil and paper, internet and touch-tone phones for self-administered surveys: does methodology matter? *Computers in Human Behavior*, 19, 117–134.

Kreuter, F., Presser, S. & Tourangeau, R. (2009) Social desirability bias in CATI, IVR, and web surveys: the effects of mode and question sensitivity. *Public Opinion Quarterly*, 72, 847–865.

Krosnick, J.A. (1991) Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, 5, 213–236.

Layton, D.F. & Moeltner, K. (2004) The cost of power outages to heterogenous households an application of the mixed gamma-lognormal distribution. In: Alberini, A. & Scarpa, R. (Eds.)

Applications of Simulation Methods in Environmental and Resource Economics. New York: Springer, pp. 35–54.

Leahy, E. & Tol, R. (2011) An estimate of the value of lost load for Ireland. *Energy Policy*, 39, 1514–1520.

Lindhjem, H. & Navrud, S. (2011a) Are internet surveys an alternative to face-to-face interviews in contingent valuation? *Ecological Economics*, 70, 1628–1637.

Lindhjem, H. & Navrud, S. (2011b) Using internet in stated preference surveys: a review and comparison of survey modes. *International Review of Environmental and Resource Economics*, 5, 309–351.

Loomis, J. & King, M. (1994) Comparison of mail and telephone-mail contingent valuation surveys. *Journal of Environmental Management*, 41, 309–324.

Maguire, K.B. (2009) Does mode matter? A comparison of telephone, mail, and in-person treatments in contingent valuation surveys. *Journal of Environmental Management*, 90, 3528–3533.

Malhotra, N. & Krosnick, J.A. (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*, 15, 286–323.

Marta-Pedroso, C., Freitas, H. & 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.

Maslovskaya, O., Durrant, G.B., Smith, P.W., Hanson, T. & Villar, A. (2019) What are the characteristics of respondents using different devices in mixed-device online surveys? Evidence from six UK surveys. *International Statistical Review*, 87, 326–346.

Meyer, J.M. (2010) Hypocrisy, NIMBY, and the politics of everybody's backyard. *Ethics, Place & Environment*, 13, 325–327.

Muneer, S., Shabbir, J. & Khalil, A. (2017) Estimation of finite population mean in simple random sampling and stratified random sampling using two auxiliary variables. *Communications in Statistics - Theory and Methods*, 46, 2181–2192.

Nielsen, J.S. (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, 119–129.

Nooij, M., Koopmans, C. & Bijvoet, C. (2007) The value of supply security: the costs of power interruptions: economic input for damage reduction and investment. *Energy Economics*, 29, 277–295.

Olsen, S.B. (2009) Choosing between internet and mail survey modes for choice experiment surveys considering non-market goods. *Environmental and Resource Economics*, 44, 591–610.

Oreskes, N. (2014) *In defense of NIMBYism*. Web Resource.

Praktiknjo, A., Hahnel, A. & Erdmann, G. (2011) Assessing energy supply security: outage costs in private households. *Energy Policy*, 39, 7825–7833.

Reichl, J., Schmidthaler, M. & Schneider, F. (2013) The value of supply security: the costs of power outages to Austrian households, firms and the public sector. *Energy Economics*, 36, 256–261.

Rendon, S. (2013) Fixed and random effects in Classical and Bayesian regression. *Oxford Bulletin of Economics and Statistics*, 3, 460–476.

Rossi, P.E., Allenby, G.M. & McCulloch, R. (2005) *Bayesian Statistics and Marketing*. Wiley.

Sandorf, E.D., Campbell, D. & Hanley, N. (2017) Disentangling the influence of knowledge on attribute non-attendance. *Journal of Choice Modelling*, 24, 36–50.

Scarpa, R. & Rose, J.M. (2008) Design efficiency for non-market valuation with choice modelling: how to measure it, what to report and why. *Australian Journal of Agricultural and Resource Economics*, 52, 253–282.

Scarpa, R., Thiene, M. & Train, K. (2008) Utility in willingness to pay space: a tool to address confounding random scale effects in destination choice to the Alps. *American Journal of Agricultural Economics*, 90, 994–1010.

Scasny, M. & Alberini, A. (2012) Valuation of mortality risk attributable to climate change: investigating the effect of survey administration modes on a VSL. *International Journal of Environmental Research and Public Health*, 9, 4760–4781.

Schwappach, D.L. & Strasmann, T.J. (2006) "Quick and dirty numbers"?: The reliability of a stated-preference technique for the measurement of preferences for resource allocation. *Journal of Health Economics*, 25, 432–448.

Soini, K., Pouta, E., Salmiovirta, M., Uusitalo, M. & Kivinen, T. (2011) Local residents' perceptions of energy landscape: the case of transmission lines. *Land Use Policy*, 28, 294–305.

Sonnier, G., Ainslie, A. & Otter, T. (2007) Heterogeneity distributions of willingness-to-pay in choice models. *Quantitative Marketing & Economics*, 5, 313–331.

Spanos, A. (1999) *Probability theory and statistical inference: Econometric modelling with observational data*. Cambridge University Press.

Struminskaya, B., Weyandt, K. & Bosnjak, M. (2015) The effects of questionnaire completion using mobile devices on data quality. Evidence from a probability-based general population panel. methods, data, analyses 9.

Szolnoki, G. & Hoffmann, D. (2013) Online, face-to-face and telephone surveys: comparing different sampling methods in wine consumer research. *Wine Economics and Policy*, 2, 57–66.

Taylor, P., Nelson, N., Grandjean, B., Anatchkova, B. & Aadland, D. (2009) Mode effects and other potential biases in panel-based internet surveys: Final report NCEE economic reports environmental economics US EPA. Technical report, Wyoming Survey and Analysis Center.

Train, K. & Weeks, M. (2005) Discrete choice models in preference space and willingness-to-pay space. *Applications of Simulation Methods in Environmental and Resource Economics, The Economics of Non-Market Goods and Resources*. Dordrecht: Springer, pp. 1–16. https://doi.org/10.1007/1-4020-3684-1-1

Whittaker, D., Vaske, J., Donnelly, M. & DeRuiter, D. (1998) Mail versus telephone surveys: potential biases in expenditure and willingness-to-pay data. *Journal of Park and Recreation Administration*, 16, 15–30.

Windle, J. & Rolfe, J. (2011) Comparing responses from internet and paper-based collection methods in more complex stated preference environmental valuation surveys. *Economic Analysis and Policy*, 41, 83–97.

Wuestenhagen, R., Wolsink, M. & Bürer, M.J. (2007) Social acceptance of renewable energy innovation: An introduction to the concept. *Energy Policy*, 35, 2683–2691.

Zhang, X., Kuchinke, L., Woud, M.L., Velten, J. & Margraf, J. (2017) Survey method matters: Online/offline questionnaires and face-to-face or telephone interviews differ. *Computers in Human Behavior*, 71, 172–180.