Smartphone and tablet effects in contingent valuation web surveys – No reason to worry?

Magnus Aa. Skeie[1,a], Henrik Lindhjem[b,c], Sofie Skjeflo[b] and Ståle Navrud[d, b]

[a] Vista Analysis, Meltzers gate 4, NO-0257 Oslo, Norway.

[b] Menon Centre for Environmental and Resource Economics (MERE)
Sørkedalsveien 10B, NO-0369 Oslo, Norway.

[c] Norwegian Institute for Nature Research (NINA)
Høgskoleringen 9, NO-7034 Trondheim, Norway.

[d] School of Economics and Business, Norwegian University of Life Sciences
Universitetstunet 3, NO-1430 Ås, Norway.

Keywords: Contingent valuation, ecosystem services, mobile device, propensity score matching, survey quality

[1] Corresponding author: magnus.aagaard.skeie@gmail.com.

Skeie, Magnus Aa.; Lindhjem, Henrik; Skjeflo, Sofie Waage; Navrud, Ståle.
Smartphone and tablet effects in contingent valuation web surveys – No reason to worry?. Ecological Economics 2019 ;Volum 165.
DOI10.1016/j.ecolecon.2019.106390

Abstract

Stated preference (SP) web surveys are increasingly completed on mobile devices such as smartphones and tablets instead of computers. Due to differences in technical attributes and response contexts of the devices, this trend may affect the quality of the survey data and elicited welfare measures. Little is known of such device effects in SP research. In the first such study of its kind, we compare willingness to pay (WTP) and response quality between devices in a large, national contingent valuation survey. Propensity score matching is used to distinguish device effects from observed sample composition effects due to self-selection. We find significantly higher WTP for smartphone respondents in the first out of four sequential WTP questions, and no differences for tablets. Concerning data (response) quality, results are mixed, but not consistently lower for smartphones and tablets compared to computers. Measured by indicators of response randomness, shares of don't know and protest zeros, smartphone responses even show signs of higher quality. Only in terms of the extent of internal scope sensitivity, do smartphones and tablets fare somewhat worse than computers. Overall, our results do not indicate substantial loss of response quality or differences in welfare measures for mobile devices.

Skeie, Magnus Aa.; Lindhjem, Henrik; Skjeflo, Sofie Waage; Navrud, Ståle.
Smartphone and tablet effects in contingent valuation web surveys – No reason to worry?. Ecological Economics 2019 ;Volum 165.
DOI10.1016/j.ecolecon.2019.106390

1. Introduction

Stated preference (SP) surveys in environmental economics, i.e. contingent valuation (CV) and choice experiments (CE), are increasingly administered on internet panels (Lindhjem and Navrud 2011a; Menegaki et al. 2016). In fields utilizing such survey data, the share of respondents completing surveys on smartphones and tablets rather than on standard computers has recently been rising fast (Peterson et al. 2017). Due to the differences in attributes of the devices (e.g. screen size and touch screen functions) and the response context (e.g. while commuting), the technical platform may affect the quality of responses and elicited willingness to pay (WTP). If significant differences are found it may jeopardize the validity and trust in web surveys and derived welfare estimates for use in for example cost-benefit analysis (CBA).

Little is still known of such effects in SP research; we have identified only one study that has compared such platform effects, in this case using CE (Liebe et al., 2015). It has been more common to compare survey mode effects more generally, see e.g. the review by Lindhjem and Navrud (2011a) and recent studies comparing web surveys with e.g. mail (Olsen, 2009; Boyle et al., 2016; Campbell et al., 2018), face-to-face (Lindhjem and Navrud, 2011b) and more deliberative settings (Sandorf et al., 2016). The mode effects found so far are small to moderate, but studies are few and results, as judged by the recent guideline on SP, both “mixed and context specific” (Johnston et al. 2017; p340). These results would in any case not be directly transferable to a device effect investigation in internet panels. Hence, while most SP research, at least in high- and middle-income countries, is moving online using such panels of respondents, coupled with rapid changes in mobile phone technologies and use patterns, it is increasingly important to investigate potential platform effects on survey responses and quality. The survey methodology literature is also mobilizing a similar research program for survey research in general (e.g. Callegaro et al., 2015; 2014; Couper et al., 2017).

Survey statistics are prone to both errors of representation and measurement; the latter being the gap between the ideal (true) measurement, and the response obtained. If the same respondent provides different answers to questions of the same survey depending on whether a mobile phone, tablet or laptop/stationary computer (PC) is used, a "platform effect" is present. This has its parallel in what is sometimes called the "pure" survey mode effect, where the same respondent would answer differently to equally worded questions across survey modes (Jäckle et al., 2010; Lindhjem and Navrud, 2011a). Two main sources of platform effects have been noted in the literature; namely differences in technical attributes and response context (see e.g. De Bruijne and Oudejans, 2015).

Firstly, the smaller screens and keyboards of tablets and smartphones compared to PCs may induce cognitive fatigue at the hands of the respondent. This in turn may affect response quality in terms of larger acquiescence tendency, [2] more randomness in responses, or through a potentially dampening effect on for instance WTP responses in SP surveys due to less striking visual stimuli on smaller screens (Liebe et al., 2015). Generally, one

[2] Acquiescence is sometimes referred to as "yea-saying", i.e. the tendency to agree with a statement when in doubt.

could expect a higher "satisficing" behavior (Lindhjem and Navrud, 2011a).[3] Some studies find that people handle PCs better technically than they do phones (e.g. Parush and Yuviler-Gavish, 2004) and that smaller screens and keyboards introduce undesirable effects on survey responses from mobile devices, due to scrolling and zooming operations (Peytchev and Hill, 2010). Still, some studies in the general survey literature find that completion on mobile devices need not lead to lower quality or different results, as long as thought is given to design (Antoun et al., 2017; De Bruijne and Wijnant, 2013; Drewes, 2014). Secondly, the typical response context may differ from that of PCs, in that smartphones (and to a lesser extent tablets) more frequently are used away from home, on the move, in the presence of other people or while multitasking (de Brouijne and Oudejans, 2015). The context may influence cognitive processing and concentration/attention levels, and the social context, e.g. the presence of others, may give normative influence on responses (Dillman et al., 2014). Research is still inconclusive and results from the survey literature would in any case not be directly transferable to SP research, as SP surveys are generally more complex and contain more text and visual stimuli than typical population surveys, e.g. where Likert scale type questions often dominate. Hence, SP surveys would be prone to biases observed in the literature when many such elements are present at the same time and the survey is complex. Liebe et al. (2015) use a CE survey to compare response quality from mobile devices (tablets and smartphones) and PCs. They find no differences in scale or in the tendency to choose the status quo option. For mobile devices only, they found a negative correlation

[3] Shortcutting the response process, providing less than optimal effort in answering.

between screen size and interview length and a positive correlation between screen size and acquiescence tendency. Model results for mobile device users indicate a U-shaped relationship between error variance, a measure of survey quality, and screen size. They conclude that using mobile devices seems not significantly to affect survey quality.

The main challenge in studies that investigate survey mode or platform effects is the potential confounding of measurement effects with sample composition effects due to self-selection into one survey mode/platform (Lindhjem and Navrud, 2011a). This is not straightforward to avoid or to control for in practice (Boyle et al., 2016). One could encourage or technically force respondents sampled from the same frame, to answer using mobile or PC and randomize treatment across respondents. This procedure will not avoid self-selection completely, as those who prefer another platform may just not respond or refuse to follow the encouragement (as seen for mobile users in Drewes (2014)). Alternatively, one could, as we do here, follow a more practical approach. We carry out a CV survey using the standard approach survey companies follow to maximize response rates, where the internet panel respondents are free to choose the platform they prefer when invited to the survey. The survey is designed for PC but optimized for answering in the internet browsers of tablets and smartphones. From this, we can first investigate people’s preferences for devices and compare the degree of selection by observable characteristics into the different platforms. Then, we follow the spirit of Liebe et al. (2015) and use propensity score matching to discern likely platform effects. We compare WTP and assess quality of responses based on experience from the survey methodology literature. A

DOI10.1016/j.ecolecon.2019.106390

broader analysis of response quality may help in judging the validity of stated preferences. As basis for the study we use a CV web survey of ecosystem service (ES) damages from accidental coastal oil spills from ships in Norway aiming at producing welfare estimates for CBA of government preventive measures (Navrud et al., 2017). Our study is, to our knowledge, the first to investigate platform effects in CV, and a first step in a continued research program on understanding device effects on response quality and welfare estimates in internet-based SP research.

2. Research questions and hypotheses

The main questions we ask are: (1) Are there systematic differences in stated WTP between mobile device and PC users, and if any, to what extent are these due to platform effects of the device?, (2) Is data quality, assessed using selected quality indicators, from mobile devices different from PCs, and if so, to what extent can this be attributed to platform effects?

Regarding the first question, controlling for (observable) respondent characteristics that influence both platform choice and WTP, there may be residual differences in stated WTP across platforms due to technical attributes and/or response context, as explained above. The main challenge in answering the first question is therefore to control for self-selection.

Regarding the second question, since overall validity of SP surveys is hard to assess (i.e. we do not know the true WTP), general response or survey quality can give an indication of validity (Lindhjem and Navrud 2011b). Response quality can be measured or proxied in several ways. We use four such indicators: shares of “don´t know” and protest zero responses to the WTP questions, response randomness and response inconsistency interpreted as lack of internal scope effect.

The share of “don’t know” and protest zero responses might indicate the extent to which respondents proceed through the survey without carefully considering the questions (so-called satisficing). Stating WTP demands a certain cognitive effort and selecting the “don´t know” or protest zero responses may serve as an easy way out (Krosnick et al. 2002)[4]. In this way we regard (low) share of “don´t know” and protest zero responses as indicative of cognitive efforts in interacting with the survey questions, and the higher is cognitive efforts, the higher is data quality. Regarding response randomness, the results of previous studies are not unanimous with respect to platform differences. The result of Liebe et al. (2015) that choice randomness shows a U-shaped relationship with screen size, means that tablet responses should be associated with less randomness than that of both smartphones and

[4] The SP literature, e.g. Johnston et al. (2017), is not clear about how to interpret “don’t know” responses. It is likely that “don’t know” in practice is a mix of satisficing behaviour and true uncertainty about ones’ preferences (especially if one takes the view advocated by Payne et al. (1999) that preferences are constructed during the valuation exercise, and not the more traditional view that “people know their preferences” (Freeman et al. 2014; p7) presumably without uncertainty). As long as the existence of preference uncertainty does not vary (or vary less) between devices than satisficing behaviour, prevalence of “don’t know” may still be used as an indicator here. We follow Lindhjem and Navrud (2011b) in their approach. “Protest zero” responses are harder to assess (i.e. due to satisficing or some other fault of the survey or the respondent), but the practical implication is that both “don’t knows” and “protest zeros” typically are taken out of the sample leaving a survey with lower information value and quality. We therefore also include protests as an indicator of (low) quality here.

Skeie, Magnus Aa.; Lindhjem, Henrik; Skjeflo, Sofie Waage; Navrud, Ståle.
Smartphone and tablet effects in contingent valuation web surveys – No reason to worry?. Ecological Economics 2019 ;Volum 165.
DOI10.1016/j.ecolecon.2019.106390 CC-BY-NC-ND

PCs. On the other hand, De Bruijne and Oudejans (2015) found that multitasking leads to lower concentration levels, and that both smartphone and tablet users were more likely to report multitasking than PC users. This pulls in the opposite direction with respect to response randomness, and the net effect is an empirical question. For smartphones, the results of De Bruijne and Oudejans (2015) pull in the direction of larger randomness compared to PCs, whereas Liebe et al. (2015) indicate somewhat similar levels between smartphones and PCs. Besides lower attentiveness due to multitasking among smartphone respondents, it might be that smaller screens and the associated scrolling and zooming operations wear respondents out and eventually induce a faster pace through the survey, causing more response randomness with smartphones than with PCs. As a final quality indicator, we investigate response consistency by checking internal scope effect with a definition of inconsistency that demands non-decreasing WTP over increasing ES damage. [5] This leads to the following five hypotheses (H1-H5), after controlling for self-selection into device:

- H1 (level of WTP): The level of WTP differs between mobile device respondents and PC respondents.
- H2 (response quality): The shares of “don’t know” responses are greater for mobile device respondents than for PC respondents.
- H3 (response quality): The shares of protest zero responses are greater for mobile device respondents than for PC respondents.

[5] There is a thriving debate about what should be regarded as “adequate” or “plausible” level of scope (Whitehead 2016). We do not go into this debate here but take as a pragmatic stance concluding that lack of internal scope is a sign of potential inconsistency.

- H4 (response quality): Response randomness, measured as the variance of the unexplained variation in WTP, is greater for mobile device respondents than for PC respondents.
- H5 (response quality): The share of inconsistent responses, indicated by lack of internal scope effect, is greater for mobile device respondents than for PC respondents.

3. Survey design and empirical methods

3.1 Survey design

We use data from a web survey conducted in October 2015 by the survey company Kantar TNS, which maintains an ISO certified, randomly recruited internet panel of respondents. The purpose was to obtain estimates of WTP for preventive measures to avoid oil spills, and associated loss of ecosystem services (ES), from ship accidents in the coastal areas of Norway, to be used in CBA. The dataset was delivered with paradata such as total time used and the type of device used, in addition to a range of background panel variables. Respondents were told in the survey that due to increased traffic along the Norwegian coast, without new safety measures, an oil spill would happen in the next few years. This oil spill could result in four potential levels of dispersion and ES damages illustrated on a map. There were five such sites distributed across the country, each regional population assessing a spill in their home region. The environmental damage and ES loss associated with each oil spill scenario were illustrated in a table, as shown in Figure 1 (from the spill site in the Oslofjord area). This table described the most important damages to seabirds, seals, the ecosystem more generally ("life in the sea") and soiling of the coastal zone. The ES terminology was not used directly, though it is clear that recreation services (incl. consumption of healthy, self-caught seafood) and non-use values associated with protection of coastal ecosystems (incl. biodiversity and specific species) were the most important. Impacts were estimated using oil spill dispersion modelling in combination with a quantitative tool for environmental impact assessment (Jødestøl et al. 2001) in combination with expert knowledge.

Skeie, Magnus Aa.; Lindhjem, Henrik; Skjeflo, Sofie Waage; Navrud, Ståle.
Smartphone and tablet effects in contingent valuation web surveys – No reason to worry?. Ecological Economics 2019 ;Volum 165.
DOI10.1016/j.ecolecon.2019.106390

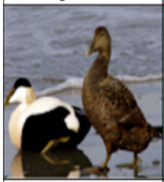
	With new measures	Without new measures			
	Without environmental damage	**Small damage** Equivalent to a spill of 20 tonnes of marine diesel	**Medium damage** Equivalent to a spill of 200 tonnes of bunker oil	**Large damage** Equivalent to a spill of 2000 tonnes of bunker oil	**Very large damage** Equivalent to a spill of 20 000 tonnes of crude oil
Damage to seabirds					
	The area is important to seabirds such as eider, cormorants, goosanders, herring gull, swans and vulnerable seabirds such as velvet scoter, black guillemot and common gull.	**200** dead seabirds **Ignorable** effect on seabird populations	**3000** dead seabirds Populations of ordinary and vulnerable seabirds will restitute after **1 year**	**7000** dead seabirds Populations of ordinary and vulnerable seabirds will restitute after **2 years**	**15 000** dead seabirds Populations of ordinary and vulnerable seabirds will restitute after **3 years**
Damage to seals					
	Parts of the area are important to seals. The seal population is in good condition	**20** dead seals **Ignorable** effect on the seal population	**40** dead seals The population of seals will restitute after **1 year**	**80** dead seals The population of seals will restitute after **2 years**	**120** dead seals The population of seals will restitute after **5 years**
Damage to life in the sea					
	The area is an important breeding and growth area for fish and other life forms in the sea. Feeding grounds for several populations.	**Ignorable** effect on life in the sea	**Small** damage on life in the sea Safe to eat fish and shellfish after **1 year**	**Some** damage on life in the sea, in particular on local populations Safe to eat fish and shellfish after **1-2 years**	**Larger** damage on life in the sea, in particular on local populations Safe to eat fish and shellfish after **1-2 years**
Damage to the coastal zone					
	Very important recreational area	**20 km** of polluted coastline The area can be used as normal in less than **1 year**	**30 km** of polluted coastline The area can be used as normal after **2 years**	**120 km** of polluted coastline The area can be used as normal after **3 years**	**190 km** of polluted coastline The area can be used as normal after **5 years**

Figure 1 Table of ES impacts associated with four oil spill damage levels. Source: Lindhjem et al. (2016). Translated from Norwegian. [*Print in colour, 1.5 column fitting with 300 dpi*]

The colour codes matched between the dispersions on the map and the damage levels in the table, to ease the cognitive burden. Respondents were asked to state their WTP to avoid the impacts for each damage level in sequence starting with the small damage, following an advanced disclosure procedure (Bateman et al., 2004). For each scenario, respondents were shown on one screen the table with the green column and one of the four damage columns highlighted (and the other three faded), coupled with the corresponding oil dispersion map (see illustration in Figure 3). The payment vehicle was a one-time tax per household that in its entirety would be used for measures that would avoid the damages with certainty. Respondents were asked to indicate the maximum amount they would be willing to pay to avoid each damage level by sliding a cursor on a payment scale (i.e. a type of payment card) with numbers from zero to NOK 12 000 (Figure 2).[6]

6 Respondents that indicated an amount exceeding 12 000 NOK were asked to specify the exact amount in a follow-up question.

DOI10.1016/j.ecolecon.2019.106390

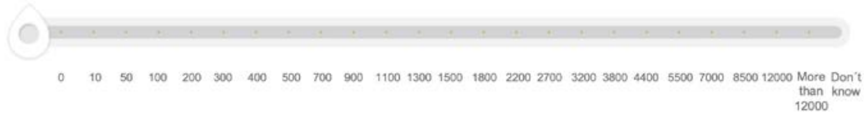


Figure 2 Payment scale with slider cursor to indicate WTP to avoid environmental damage resulting from each oil spill scenario. Source: Lindhjem et al. (2016)
[*1.5 column fitting with 400 dpi*]

Screenshots from a tablet is shown in Figure 3, where the first scenario of "small damage" (yellow colour) is compared with a situation with safety measures achieving "no damage" (green colour). The accompanying map to the right of the table shows the relatively small oil dispersion in this scenario. The look on a PC is similar to a tablet, while on a smartphone zooming and scrolling are required to see all the information and to choose the appropriate WTP. Note that the survey was designed to be relatively robust against differences in web browsers etc. for PC's, so that such differences should be relatively small.

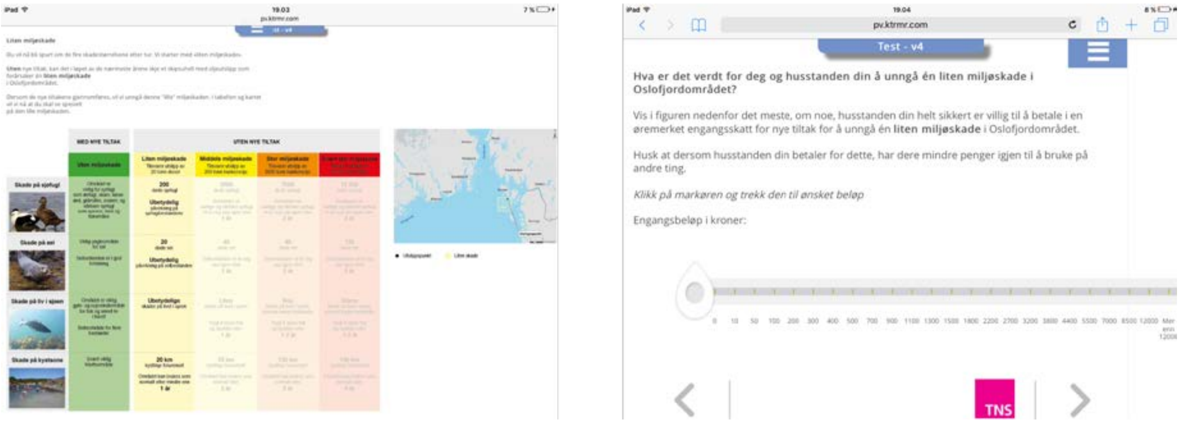


Figure 3 The damage scenario table and payment slider as seen by respondents in two screens using a large tablet, prior to scrolling and zooming. Screenshots from an iPad 2 (9.7-inch diagonal screen size). [*Print in colour, 2 column fitting with 300 dpi*]

The survey instrument was developed and thoroughly tested over several years in pilots, focus groups and in personal interviews with survey respondents (see e.g. Navrud et al., 2017).

3.2 Empirical methods

We use linear regression models and propensity score matching models as the main methods in our analysis. Respondents choose their maximum WTP with lower limit WTP_{li} and upper limit WTP_{ui} for i intervals. In our main models, we use interval midpoints as estimates of WTP for each amount WTP_{li} chosen on the payment scale[7]. We use propensity score matching (PSM) to control for observable characteristics that may be correlated with both platform choice and WTP response (see e.g. Liebe et al., 2015). PSM involves two steps: First, estimating the probability of being in the “treated” group (in our case, smartphone respondents and tablet respondents, respectively), as a function of observable characteristics that are correlated with platform choice and WTP, by using a Logit model. In the second step, we match observations on the predicted probability of treatment (the propensity score) from the first step. We use a nearest neighbour approach, where observations in the treatment group are matched with the observation in the untreated group with the most similar propensity score. This allows us to estimate the average treatment effect, i.e. the average effect of smartphone and tablet response, respectively, on stated WTP.

4. Results

4.1 Descriptive statistics

[7] When there are many amounts in the payment card the difference between this and using an interval estimation approach should not be large (Mahieu et al. 2012).

The final sample consisted of 5535 respondents and is close to representative of the Norwegian population, except a slight underrepresentation of those below 44 years of age, and an overrepresentation of those above 60 years of age.[8] The total response rate was 54 per cent[9], which is high for this kind of survey. The majority of respondents answered on a PC (68 per cent), while 21 per cent and 11 per cent used tablet and smartphone, respectively. Columns (1) to (3) of Table 1 shows mean values for a number of respondent characteristics for the three groups choosing different devices, with standard errors in parentheses. The final three columns show the p-values of t-tests for the null hypothesis of no difference in the means between groups.

[8] As noted, the data were collected in five regions of Norway for five oil spill sites. The survey is otherwise identical across regions, and we have pooled the data without using sampling weights.

[9] Unfortunately, the survey company did not supply the response rate by device.

Table 1 Descriptive statistics across platforms; mean values with standard errors in parenthesis, and P-values of t-tests of the difference in mean across platforms

	PC[a] (1)	Tablet[b] (2)	Smartphone[c] (3)	(1) vs. (2), p-value	(1) vs. (3), p-value	(2) vs. (3), p-value
Age of respondent	52.422 (0.269)	51.617 (0.464)	35.789 (0.535)	0.140	0.000	0.000
Dummy for female	0.440 (0.008)	0.576 (0.014)	0.666 (0.019)	0.000	0.000	0.000
Dummy for higher educ. (MA or PhD)	0.138 (0.006)	0.094 (0.008)	0.135 (0.014)	0.000	0.844	0.009
Household gross annual income, thousand NOK	701.952 (5.553)	713.835 (9.035)	759.253 (15.445)	0.285	0.000	0.007
Response time, minutes. Mean/median	87/14	97/14	131/14	0.637/ 0.009	0.139/ 0.500	0.374/ 0.260
Response time[d], minutes. Mean/median	16/14	17/15	22/14	0.000/ 0.003	0.000/ 0.516	0.000/ 0.155
Platform share of respondents	68 %	21 %	11 %			
Observations	3757	1186	592			

a PCs, laptops and netbooks.
b Small, medium and large tablets.
c Smartphones with touch screen.
d Without lowest/highest 5 %

Table 1 shows that smartphone respondents are on average younger than both tablet and PC respondents. The share of females is highest for smartphone, then tablet and lastly PC. The share of highly educated respondents is lower for tablet than both PC and smartphone, and household income is on average higher among smartphone respondents (despite their younger age) than both PC and tablet respondents. Hence, the sample shows some degree of self-selection into the devices. With respect to response time, the mean is significantly different across platforms at the 1 per cent level upon exclusion of the lowest and highest 5 per cent. The mean response time increases from PC through tablet to smartphone. Furthermore, median response time is also significantly lower for PC than for tablet.[10] The standard deviation of the trimmed response time variable is about four times larger for smartphone compared to PC, but approximately equal between PC and tablet respondents. One possible explanation is more multitasking among the smartphone respondents, and therefore interruptions that could affect the response time. On the other hand, smartphone respondents that are able to complete the survey without interruption seem to complete the survey faster.

4.2 Device effects without control for self-selection

[10] Using a non-parametric median test. The standard deviation in the trimmed response time variable (excluding highest/lowest 5%) is approximately equal for PC and tablet respondents, but four times larger for smartphone respondents, which is the likely cause of the vast difference in p-values compared to the small differences in median response time across platforms, as seen in Table 1.

Skeie, Magnus Aa.; Lindhjem, Henrik; Skjeflo, Sofie Waage; Navrud, Ståle.
Smartphone and tablet effects in contingent valuation web surveys – No reason to worry?. Ecological Economics 2019 ;Volum 165.
DOI10.1016/j.ecolecon.2019.106390

We first check the differences across platforms in mean WTP, share of protest zero responses[11] and share of don´t know responses without controlling for self-selection into devices[12]. The results of regression of log WTP on platform dummy variables for the four damage scenarios are given in Table 2, together with shares of protest zero-, don´t know- and inconsistent responses for all three platforms.

[11] A protest zero response is defined by the answer to a WTP follow-up question asking for the response motive, i.e. other reasons than “no utility” or “cannot afford”.

[12] Respondents were given the opportunity to revise their answers after answering all WTP questions based on a hypothetical bias script, and we consistently use the revised values here.

Table 2 Platform effects on mean WTP (NOK)[a], share of zero responses (0)[b] and share of don´t know responses (DK) for four environmental damage levels, and share of inconsistent responses, without controlling for self-selection. PC responses are the baseline

	(1) Small damage			(2) Medium damage			(3) Large damage			(4) Very large damage		
	Reg. coeff.	0 T/P	DK	Reg. coeff.	0 T/P	DK	Reg. coeff.	0 T/P	DK	Reg. coeff.	0 T/P	DK
PC	-	0.17 10/90	0.08	-	0.14 10/90	0.08	-	0.12 10/90	0.08	-	0.11 9/91	0.08
Smart-phone	0.233** (2.24)	0.13 15/85	0.03	0.160 (1.53)	0.12 17/83	0.03	0.210** (2.03)	0.10 14/86	0.03	0.166 (1.54)	0.10 16/84	0.03
Tablet	0.101 (1.18)	0.16 10/90	0.07	-0.011 (-0.13)	0.15 8/92	0.06	-0.028 (-0.32)	0.13 8/92	0.06	-0.091 (-1.00)	0.13 8/92	0.06
Const. term	4.769*** (110.24)	-	-	5.229*** (124.83)	-	-	5.639*** (131.89)	-	-	5.962*** (135.14)	-	-
Obs.	5147	5535	5535	5157	5535	5535	5156	5535	5535	5144	5535	5535

	PC	Tablet	Smartphone
Share of inconsistent responses (as defined across damage levels)[c]	0.07	0.11	0.11
Observations	3757	1186	592

a Don’t know responses removed, all zeros retained. WTP is log transformed and based on payment card interval midpoints. Regression coefficients are estimated with robust standard errors. *t* statistics in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

b The distribution of true- (T) and protest zero responses (P) are given in percent below the respective shares.

c Any respondent for which stated WTP decreases over any part of the four (increasing) environmental damage levels is counted as inconsistent. The given shares are the fractions of the platform responses deemed inconsistent.

DOI10.1016/j.ecolecon.2019.106390

Table 2 shows that WTP is significantly higher for smartphone respondents than for PC respondents (baseline) for the small damage level and the large damage level. There is no significant difference for the medium or very large damage levels, nor for tablets. The shares of zero responses are fairly similar across platforms and should therefore have limited impact on any differences in platform effects on WTP. However, smartphone both has the lowest shares of zeros altogether and the highest shares of true zeros. The share of don´t know responses are low and similar for PC and tablet, but for smartphone only about half of that of the others. Since don't know responses are removed in the WTP regression, lower prevalence of such responses in the smartphone group would influence WTP. Shares of inconsistent responses are equal for tablet and smartphone, but somewhat lower for PC. There is no guidance in the literature as to "acceptable" shares of inconsistency in SP surveys[13], but there is a tendency towards lower internal scope responses on mobile devices. Hence, from a first look at the data, there are some differences between responses by device that influence response quality and mean WTP, but WTP is only different for two out of four valuation scenarios and for phone

[13] Normally studies investigate internal scope by testing differences in means for different valuation scenarios. Paired t-tests of mean WTP between each of the damage levels (small versus medium, medium versus large and large versus very large) show that the responses from all three platform subsamples pass this more traditional test of scope. For all three platform subsamples, we can reject that mean WTP for a lower damage level equals mean WTP for a higher damage level, against the alternative hypothesis of greater WTP for the lower damage level at $p<0.05$. The way we measure scope here is not a pass-fail criterion overall, more an indicator of the degree of inconsistency.

DOI10.1016/j.ecolecon.2019.106390

responses only. The question is the extent of the self-selection effect, which we turn to next.

4.3 Device effects with control for self-selection

4.3.1 Differences in WTP (H1)

To address the potential selection bias, we use propensity score matching to compare WTP of smartphone respondents to similar tablet- and PC respondents (see e.g. Liebe et al., 2015). In the first step Logit model, we include variables that are likely to be correlated with both WTP and platform choice and variables that are potentially related to WTP, to reduce the bias of any observed confounders. This includes household income, age, gender and education, distance from the coastline, previous experience with oil spills, trust in the measures to prevent oil spills, use of the area affected in the damage scenarios, membership in an environmental organisation, as well as dummy variables for the five regional samples. Balance analysis on treatment effects show that covariates are fairly balanced in the matched samples of smartphone- and PC respondents, and tablet- and PC respondents. This means that we are comparing respondents that are similar in observable characteristics, but differ in platform choice, with the aim of isolating the effect of platform choice on WTP. As a sensitivity analysis, we have also estimated a Heckman type two-step selection model that aims to control for unobserved factors that are correlated with both selection and stated WTP (see section 4.5). The results from the PSM approach to estimating the effect of platform choice on stated WTP for the four damage levels are shown in Table 3.[14]

[14] We use the Stata teffects package, with WTP coded as the midpoint, cf. section 3.2.

DOI10.1016/j.ecolecon.2019.106390

Table 3 Average treatment effect of smartphone and tablet response on WTP (NOK) to avoid four damage levels.[a] Propensity score matching using nearest neighbour matching

	Log WTP Small damage	Log WTP Medium damage	Log WTP Large damage	Log WTP Very large damage
Dummy for response by smartphone w/PC baseline	0.316** (2.01)	0.224 (1.37)	0.249 (1.27)	0.150 (0.73)
Dummy for response by tablet w/PC baseline	0.0148 (0.14)	-0.138 (-1.32)	-0.0983 (-0.92)	-0.144 (-1.21)
Observations (smartphone/tablet)	3773/4291	3777/ 4296	3774/4297	3766/4290

a Don't knows removed, all zeros retained. *t* statistics in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

We find that the positive association between smartphone and WTP to avoid the small damage level is robust, but there is no significance for the other damage levels, nor for tablets.

4.3.2 Shares of don´t know responses and protest zero responses (H2 and H3)

We make further use of the PSM approach to address selection bias in observed shares of “don´t know” and protest zero responses. The results for the share of “don’t know” responses are shown in Table 4.

DOI10.1016/j.ecolecon.2019.106390

Table 4 Average treatment effect of smartphone response on share of “don’t know” responses. Propensity score matching using nearest neighbour matching

	WTP = “Don´t know” Small damage level	WTP = “Don´t know” Medium damage level	WTP = “Don´t know” Large damage level	WTP = “Don´t know” Very large damage level	WTP= “Don´t know” all damage levels
Dummy for response by smartphone w/PC baseline	-0.0591*** (-8.68)	-0.0596*** (-8.97)	-0.0337** (-1.96)	-0.0586*** (-8.37)	-0.0502*** (-7.80)
Dummy for response by tablet w/PC baseline	-0.0244** (-2.45)	-0.0322*** (-3.37)	-0.0328*** (-3.50)	-0.0363*** (-4.01)	-0.0309*** (-3.64)
Observations (smartphone/tablet)	4041/4612	4041/4612	4041/4612	4041/4612	4041/4612

t statistics in parentheses, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The results show that the share of “don’t know” responses to the WTP questions, somewhat surprisingly, is significantly lower among tablet and smartphone respondents compared to matched PC respondents, for all valuation scenarios. Hence, the higher share we observed among smart phone respondents in Table 2 is more than reversed when controlling for self-selection; in fact, both smartphones and tablets <u>reduce</u> rather than increase the share of don’t know responses as compared to PC among similar-type respondents.

The results for the share of protest zero responses are shown in Table 5. We find that the share of protest zeros is somewhat <u>lower</u> for smartphone (small damage level), but higher for tablet (medium and large damage levels).

Table 5 Average treatment effect of smartphone response on share of protest zero responses.[a] Propensity score matching using nearest neighbour matching

	Zero WTP, small damage level	Zero WTP, medium damage level	Zero WTP, large damage level	Zero WTP, very large damage level	Zero WTP, all damage levels
Dummy for response by smartphone w/PC baseline	-0.0456** (-2.07)	-0.0109 (-0.52)	-0.0123 (-0.57)	-0.00859 (-0.35)	-0.0193 (-0.89)
Dummy for response by tablet w/PC baseline	0.00759 (0.47)	0.0316** (2.01)	0.0199 (1.37)	0.0238* (1.76)	0.0141 (0.99)
Observations (smartphone/tablet)	3707/4218	3724/4242	3730/4249	3727/4292	3425/3898

a True zero-responses are excluded. *t* statistics in parentheses, * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

4.3.3 Response randomness (H4)

We have defined response randomness as the variance of the unexplained variation in WTP after controlling for relevant observable characteristics. In our view this is analogous to using scale in random utility models for analysis of response randomness in CE (see e.g. Liebe et al., 2015). Firstly, we estimate the following regression model: $\log WTP_i = x_i{}' \beta + u_i$, where $x_i{}'$ is a vector of explanatory variables with associated estimated parameters β and $u_i \sim N(0, \sigma)$. Included explanatory variables are shown in Appendix A. We do not control for the platform used, as we want the platform choice to be left in the residual, representing the random component of WTP. Secondly, regression residuals u_i are predicted and kept for subsequent analysis. Plots of their distribution show that $u_i \sim N(0, \sigma)$. This means we have isolated a (practically) random component of stated WTP.

To compare response randomness across platforms we use Levene's test for homogeneity of variances to compare the variances of the predicted residuals from our regression model of WTP for each damage level. The results are shown in Table 6.

Table 6 Levene's test[a] for homogeneity of variance in predicted residuals[b]

	Small damage level	Medium damage level	Large damage level	Very large damage level
Smartphone vs. PC	- (p=0.000) ***	- (p=0.035) **	- (p=0.002) **	- (p=0.006) **
Tablet vs. PC	- (p=0.672)	- (p=0.766)	+ (p=0.676)	- (p=0.599)

a The table reports p-values from Levene's test, testing H_0: Variances are equal, against H_1: (At least one of the) variances are unequal (to the others.) The minus and plus signs report lower and higher variance, respectively, for smartphone/tablet users than for PC users, as found in the test.

b Don't knows removed. All zeros retained in order to capture platform variance irrespective of WTP response motive. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The results show, somewhat surprisingly, that response randomness is significantly lower for smartphone respondents than for PC respondents. We do not find any significant difference between tablet respondents and PC respondents.

4.3.4 Response inconsistency (H5)

Our final quality indicator is the share of response inconsistency across platforms, using internal scope as an indicator. We test differences with a definition of inconsistency that only demands non-decreasing WTP over increasing ES damage. We again use the PSM approach to estimate the effect of platform choice on the probability of inconsistent response. The results are shown in Table 7.

Table 7 Average treatment effect of smartphone and tablet response on share of inconsistent responses[a]. Propensity score matching using nearest neighbour matching

	Share of inconsistent responses
Dummy for response by smartphone w/PC baseline	0.120**
	(2.67)
Dummy for response by tablet w/PC baseline	0.0506***
	(3.93)
Observations (smartphone/tablet)	4041/4612

a Don´t knows removed, all zeros retained. *t* statistics in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Smartphone respondents have a significantly higher share of inconsistent responses than PC respondents, after matching on propensity scores. The same finding holds for tablet users. Hence, the tendency observed in Table 2 is robust when controlling for self-selection.

4.4 Summary of test results

Table 8 sums up our tests of platform effects on WTP level and response quality.

Table 8 Test results for mean WTP and indicators of response quality.

		Test approach	Result (p<0.05)
H1	**Level of WTP differs across platforms**	Propensity score matching	Partially confirmed for smartphone; higher WTP for smartphone at small ES damage level.
H2	**Greater share of “don´t know” responses among mobile device respondents than PC respondents**	Propensity score matching	Rejected for both smartphone and tablet; lower share for all ES damage levels
H3	**Greater share of protest zero responses among mobile device respondents than PC respondents**	Propensity score matching	Partially rejected for smartphone (lower share for small damage level, otherwise no significance), some indication of confirmation for tablet for medium and very large ES damage levels.
H4	**Greater response randomness among mobile device respondents than PC respondents**	Levene’s test of homogeneity of variance of residuals	Rejected for both smartphone and tablet; lower response randomness for smartphone for small, large and very large ES damage levels.
H5	**Greater share of inconsistent responses (internal scope insensitivity) among mobile device respondents than PC respondents.**	Propensity score matching	Confirmed for both smartphone and tablet.

4.5 Robustness and further checks

We found in section 4.3.2 that smartphone respondents are less likely to answer protest zero. To isolate any effects of zero responses, we have therefore done the same PSM analysis for positive WTP responses only (see Table 9).

Table 9 Average treatment effect of smartphone and tablet response on WTP (NOK) to avoid four damage levels. Propensity score matching using nearest neighbour matching, positive WTP only[a]

	(1) Log WTP midpoint, small damage	(2) Log WTP midpoint, medium damage	(3) Log WTP midpoint, large damage	(4) Log WTP midpoint, very large damage
Dummy for response by smartphone w/PC baseline	0.200[*] (1.65)	0.226[**] (2.31)	0.144 (1.42)	0.136 (1.04)
Dummy for response by tablet w/PC baseline	0.0362 (0.62)	0.0140 (0.26)	0.0355 (0.60)	-0.000530 (-0.01)
Observations (smartphone/tablet)	3095/3502	3231/3652	3292/3717	3310/3743

a Don´t knows and all zeros removed. *t* statistics in parentheses. [*] $p < 0.10$, [**] $p < 0.05$, [***] $p < 0.01$

DOI10.1016/j.ecolecon.2019.106390

We still find significantly higher WTP to avoid the small damage for smartphone, and in addition, the coefficient on medium damage is also significant. There is still no difference between tablet and PC. We have also rerun the analysis of response randomness using only positive WTP, and we still find significantly lower residual variance for the smartphone responders than the PC responders, but no difference for the tablet responders.[15]

Finally, the PSM approach matches respondents based on their propensity to choose platform, given the information about observed respondent characteristics from the survey. However, there may still be unobserved characteristics of the respondents that are both correlated with platform choice and WTP, creating an endogenous selection bias. As a sensitivity analysis, we have used the Heckman two-step selection model, which aims to control for endogenous selection effects by estimating the platform decision in a first step, and the WTP response in a second step, allowing for the two parts of the model to depend on each other (see for instance Cameron and Trivedi, 2009). In the selection step for smartphone versus PC and tablet versus PC, we include the same explanatory variables as we used in the PSM approach (cf. section 4.3.1). In the model for the outcome of interest the same explanatory variables are included, but also the term known as the non-selection hazard from the selection model, to account for correlation between the error term in the selection equation and the main model. Table 10 shows the results of estimating the model for WTP.

[15] Results available upon request.

DOI10.1016/j.ecolecon.2019.106390

Table 10 Linear regression with endogenous treatment effects of platform choice on WTP (NOK)[a] (two-step model)[b]

	(1) Log WTP midpoint, small damage	(2) Log WTP midpoint, medium damage	(3) Log WTP midpoint, large damage	(4) Log WTP midpoint, very large damage
Dummy for response by smartphone w/PC baseline	-1.082 (-1.55)	-1.265 (-1.86)	-0.908 (-1.31)	-1.119 (-1.55)
Dummy for response by tablet w/PC baseline	-1.600 (-0.69)	-0.605 (-0.26)	-0.625 (-0.26)	-0.325 (-0.13)
Observations (smartphone/tablet)	3773/4291	3777/4296	3774/4297	3766/4290

a Don´t knows removed, all zeros retained.

b Using the Stata command -etregress- with the two-step option. The same explanatory variables are used in the selection step and the main model. *t* statistics in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

We do not find any significant effects of smartphone or tablet response on the level of WTP; the previously negative effect for the small damage level for smartphone is now insignificant (cf. Table 3). Table 11 summarizes the results of equivalent model estimation for the quality indicators.[16]

[16] Estimation results are available upon request. Response randomness (H4) is not eligible for this analysis as we exclude platform choice from the model (see section 4.3.3).

Skeie, Magnus Aa.; Lindhjem, Henrik; Skjeflo, Sofie Waage; Navrud, Ståle.
Smartphone and tablet effects in contingent valuation web surveys – No reason to worry?. Ecological Economics 2019 ;Volum 165.
DOI10.1016/j.ecolecon.2019.106390

Table 11 Summary of test results for the quality indicators using linear regression with endogenous treatment effects (two-step model)

	Dependent variable	Smartphone vs. PC[a]	Tablet vs. PC[b]
H2	Dummy for don´t know response	Significant, negative effect for all damage levels	No significant effects
H3	Dummy for protest zero response[c]	No significant effects	No significant effects
H5	Dummy for inconsistent response[d]	No significant effects	No significant effects

a Regression coefficient on dummy for smartphone, four ES damage levels.
b Regression coefficient on dummy for tablet, four ES damage levels.
c True zero responses removed.
d Don´t knows removed, all zeros retained.

For the quality indicators, the previously statistically significant impacts reported in the main results are either no longer statistically significant or with weaker significance. However, in the absence of a valid instrumental variable, i.e. a variable that predicts platform choice but is uncorrelated with WTP, the selection model may suffer from high collinearity between the selection correction term and the explanatory variables in the WTP model, yielding inconsistent estimates (Puhani, 2000). Since we do not have a valid instrument for platform choice in our data, we cannot solve the potential endogenous selection problem; only an experimental approach randomly allocating respondents to different platforms would.

5. Discussion and conclusions

We have investigated whether the choice of technical platform or device – smartphone, tablet or laptops and stationary computers (PCs) – from which to answer a CV web survey affects estimates of WTP and data (response) quality. When provided with the choice of their device of preference, younger, female and higher income earners have a tendency to choose smartphones over PCs to answer the survey. Tablet and smart phone users spend more time on the survey. These results are similar to the CE study of Liebe et al. (2015). There are also some differences in other characteristics of respondents. To disentangle self-selection effects from device effects, we use propensity score matching. In estimates of mean WTP, controlling for potential self-selection bias in this way, there is no clear evidence in the data of systematic differences between PC, tablet and smartphone responses. Admittedly, mean WTP for smartphone is found to be significantly different (higher), but only for the first out of four sequential WTP questions each

Skeie, Magnus Aa.; Lindhjem, Henrik; Skjeflo, Sofie Waage; Navrud, Ståle.
Smartphone and tablet effects in contingent valuation web surveys – No reason to worry?. Ecological Economics 2019 ;Volum 165.
DOI10.1016/j.ecolecon.2019.106390 CC-BY-NC-ND

respondent answers, i.e. WTP for avoiding the smallest ES loss[17]. For tablets, which have both technical characteristics (e.g. screen size) and usage more similar to PCs, there is, perhaps more as expected, no difference in mean WTP compared to PC responses. Similarly, Liebe et al. (2015) find in their CE study, also using PSM, some differences in implicit prices but not in a unidirectional way. Hence, both from their findings and ours, it seems likely that choice of device does not have systematic or large effects on estimated welfare measures, even for relatively complex SP surveys with much visual and textual information and WTP questions that require some technical skill to respond to, especially on smartphones. In our CV case some degree of scrolling and zooming on smartphone was required and the payment card slider needed to be moved to reply to the WTP questions using the touch screen.

Comparing survey (response) quality between devices, results are mixed, starting with response randomness, defined as the variance of the unexplained variation in WTP after controlling for relevant observable characteristics. This is similar to using scale in random utility models for analysis of response randomness in CE. For this indicator we find lower response randomness for smartphone responses, indicating somewhat higher response quality. Where Liebe et al. (2015) find a U-shaped relationship between error variance and screen size using (which implies that tablet error term variance is smaller than smartphone error term variance), we find no such relationship. It is difficult to interpret the underlying mechanisms of our results, and it may be that our

[17] This effect disappears when we use a Heckman two-step selection model as an alternative to the PSM approach (cf. Section 4.5).

use of payment card could also tempt low-effort respondents to choose midpoints on the payment scale, as e.g. suggested by Lindhjem and Navrud (2011b), giving less variation in the data, including random variation. Regarding the choice of don’t knows, and to a lesser extent for protest zeros, we also find a similar result, i.e. that the tendencies to choose these responses are lower on smartphones than on PCs. Tablet users also tend to choose don’t knows less often than PC users, but protest zeros more for some WTP questions. Again, it is difficult to interpret the underlying mechanisms, as we have not investigated how exactly respondents use their devices to answer the surveys (e.g. through observing them or using eye tracking etc.). Regarding choice inconsistency, interpreted as insensitivity to internal scope, we find a relatively clear and robust result in favour of PC responses, but again we do not know the reasons why and can only speculate. In any case, the shares of such responses are not high enough to question the validity of the overall data from mobile devices. Overall, results on response quality is not consistently or clearly in the disfavour of smartphones or tablets[18], much in the same way Lindhjem and Navrud (2011a,b) concluded that Internet responses appeared to be of no lower quality or validity compared to other survey modes, and especially compared to the gold standard of personal interviews. Liebe et al. (2015) conclude in much the same way in terms of CE data quality on smartphones and tablets. The result that mobile devices seem not to reduce data quality much, is also supported by other studies from the general survey methodology literature, though there are not yet many such studies (see e.g. Antoun et al. 2017; De Bruijne and Wijnant, 2013; Drewes, 2014).

[18] This conclusion also holds when we use the Heckman two-step selection model.

There are some possible weaknesses with our study. Firstly, in the absence of a randomized controlled experiment in the assignment of platform the respondents should use, there may be unobserved, confounding effects on responses due to self-selection that cannot be controlled for by use of PSM. However, as noted earlier, it is also not easy to conduct a high-quality randomized experiment, as people may refuse to follow instructions or not reply (as has been found in some studies, e.g. Drewes 2014, De Bruijne and Wijnant 2013). Secondly, we have not been able to distinguish between effects related to differences in response context and technical attributes of the devices. It would have been an advantage to have some information, either in the form of paradata from the survey company or direct questions in the survey, to investigate features like multitasking, answers “on the go”, presence of other people etc. Finally, as pointed out by Lindhjem and Navrud (2011a), much is still unknown in the literature about what causes survey mode effects, and more work should be put into understanding this question. In the meantime, even if it is early days for understanding response behaviour and effects of the entry of mobile devices into SP research, we can conclude from this study and that of Liebe et al. (2015) that results do not seem to support early fears of significant loss of quality and the need to discourage so-called unintended mobile respondents (e.g. Peterson, 2012, Peytchev and Hill, 2010)

Appendix A

Table A1 shows the regression results from the regression specified in section 4.3.3 that is used to predict residuals for the test of response randomness.

Table A1 Regression analyses of platform effects on WTP (NOK) to avoid ecosystem service losses from accidental marine oil spills from ships. Linear regression with midpoint estimates of WTP and robust standard errors

	(1) Log WTP midpoint, small damage	(2) Log WTP midpoint, medium damage	(3) Log WTP midpoint, large damage	(4) Log WTP midpoint, very large damage
Log household gross annual income, NOK	0.0397	0.0981^{*}	0.129^{**}	0.186^{***}
	(0.66)	(1.68)	(2.23)	(3.09)
Log age of respondent	-7.295^{***}	-9.967^{***}	-10.82^{***}	-11.22^{***}
	(-3.68)	(-5.45)	(-6.01)	(-6.05)
Log age of respondent squared	0.974^{***}	1.308^{***}	1.401^{***}	1.443^{***}
	(3.68)	(5.34)	(5.80)	(5.79)
Dummy variable for female	0.594^{***}	0.535^{***}	0.469^{***}	0.421^{***}
	(8.29)	(7.65)	(6.62)	(5.75)
Dummy variable for higher educ, MA or PhD	0.432^{***}	0.505^{***}	0.558^{***}	0.579^{***}
	(4.54)	(5.50)	(6.00)	(6.01)
Log distance from ocean, km	-0.0177	-0.0349	-0.0269	-0.0521^{*}
	(-0.61)	(-1.25)	(-0.94)	(-1.78)
Used area affected by very large damage last 12 months	0.252^{***}	0.314^{***}	0.295^{***}	0.325^{***}
	(3.08)	(3.96)	(3.64)	(3.83)
Previous experience with oil spill damage	0.277^{***}	0.243^{***}	0.246^{***}	0.256^{***}
	(3.06)	(2.76)	(2.75)	(2.75)
Membership in recr. and/or environm. org	0.373^{***}	0.391^{***}	0.435^{***}	0.495^{***}
	(4.43)	(4.77)	(5.21)	(5.75)
Very important to prevent oil spills	0.530^{***}	0.416^{***}	0.399^{***}	0.408^{***}
	(7.37)	(5.87)	(5.55)	(5.49)
High trust in measures to prevent oil spills	0.707^{***}	0.733^{***}	0.704^{***}	0.704^{***}
	(9.59)	(10.09)	(9.52)	(9.20)
Southern Norway sample	-0.207^{*}	-0.237^{**}	-0.277^{**}	-0.337^{***}
	(-1.83)	(-2.15)	(-2.45)	(-2.87)
Western Norway sample	0.157	0.131	0.0772	0.00965
	(1.51)	(1.31)	(0.77)	(0.09)
Mid Norway sample	0.0817	0.0454	-0.0243	-0.0454
	(0.74)	(0.43)	(-0.23)	(-0.41)
Northern Norway sample	0.138	0.0149	-0.0447	-0.0617
	(1.16)	(0.13)	(-0.37)	(-0.49)
Constant	16.57^{***}	21.62^{***}	23.59^{***}	24.12^{***}
	(4.69)	(6.66)	(7.34)	(7.27)
Observations	4828	4835	4833	4825

t statistics in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The association between the control variables and stated WTP seems reasonable and is evidence of general validity. The income elasticity is increasing in the severity of the environmental damage level. WTP is decreasing with age, but at a decreasing rate, for all damage levels. Female respondents and highly educated respondents have significantly higher WTP for all damage levels. We find no evidence of distance decay from the coast, which is not implausible due to the likely presence of non-use values. Previous use of the area affected by the very large damage scenario is positively associated with WTP, as is previous experience with oil spill damage, for all damage levels. Membership in environmental and/or recreational organizations is positively associated with WTP for all damage levels. Respondents who state that it is very important to prevent oil spills, and who claim to have high trust in measures proposed to prevent oil spills, also on average state a higher WTP.

Acknowledgements

We acknowledge funding from the Norwegian Coastal Administration and the Norwegian Research Council's (project number 255777). We thank Kristin Magnussen for significant contributions to survey design and testing, and participants at the Annual Meeting of the Norwegian Association of Economists 2017 (Oslo) and the Nordic Annual Environmental and Resource Economics (NAERE) Workshop 2017 (Helsinki) for helpful comments and suggestions for the work on this paper. We also thank Magnus Söderberg, Steinar Strøm, Raúl Brey Sánchez and Olvar Bergland for helpful comments in the revision process.

References

Antoun, C., Couper, M. P. and Conrad, F. G. (2017). Effects of Mobile versus PC Web on Survey Response Quality. A Crossover Experiment in a Probability Web Panel. Public Opinion Quarterly 81 (Special Issue) 280-306.

Bateman, I. J., Cole, M., Cooper, P., Georgiou, S., Hadley, D. and Poe, G. L. (2004). On visibility of choice sets and scope sensitivity. Journal of Environmental Economics and Management 47 (1) 71-93

Boyle, K. J., Morrison M., MacDonald, D. H., Duncan, R. and Rose, J. (2016). Investigating Internet and Mail Implementation of Stated-Preference Surveys While Controlling for Differences in Sample Frames. Environmental and Resource Economics 64 (3) 401-419

Callegaro, M., Manfreda, K. L. and Vehovar, V. (2015). Web Survey Methodology. London: Sage Publications, Ltd.

Callegaro, M., Villar, A., Yeager, D. and Krosnick, J. A. (2014). A critical review of studies investigating the quality of data obtained with online panels based on probability and nonprobability samples in Online Panel Research: A Data Quality Perspective, edited by Callegaro, M., Baker, R., Bethlehem, J., Görits, A. S., Krosnick, J. A. and Lavrakas, P. J., 23-53. Hoboken, NJ: John Wiley and Sons, Ltd.

Cameron, A. C., & Trivedi, P. K. (2009). *Microeconometrics using stata* (Vol. 2). College Station, TX: Stata press.

Campbell, R. M., Venn, T. and Anderson, N. M. (2018). Cost and performance tradeoffs between mail and internet survey modes in a nonmarket valuation study. Journal of Environmental Management 20: 316-327

Skeie, Magnus Aa.; Lindhjem, Henrik; Skjeflo, Sofie Waage; Navrud, Ståle.
Smartphone and tablet effects in contingent valuation web surveys – No reason to worry?. Ecological Economics 2019 ;Volum 165.
DOI10.1016/j.ecolecon.2019.106390 CC-BY-NC-ND

Couper, M. P., Antoun, C and Mavletova, A. M. (2017). Mobile Web Surveys. In: A Total Survey Error Perspective in Total Survey Error in Practice: Improving Quality in the Era of Big Data, edited by Biemer, P. P., de Leeuw, E., Eckman, S., Edwards, B., Kreuter, F., Lyberg, L. E., Tucker, N. C. and West, B. T., 133-153. New York, NY: John Wiley and Sons, Ltd.

De Bruijne, M. and Oudejans, M. (2015). Online Surveys and the Burden of Mobile Responding in Survey Measurements: Techniques, Data Quality and Sources of Error, edited by Engel, U., 130-145. Frankfurt: Campus Verlag GmbH.

De Bruijne, M. and Wijnant, A. (2013). Comparing Survey Results Obtained via Mobile Devices and Computers: An Experiment With a Mobile Web Survey on a Heterogeneous Group of Mobile Devices Versus a Computer-Assisted Web Survey. Social Science Computer Review 31 (4) 482-504.

Dillman, D. A., Smyth, J. D. and Christian, L. M. (2014). Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method, 4th Edition. Hoboken, NJ: John Wiley and Sons, Ltd.

Drewes, F. (2014). An empirical test of the impact of smartphones on panel-based online data collection in Online Panel Research: A Data Quality Perspective, edited by Callegaro, M., Baker, R., Bethlehem, J., Görits, A. S., Krosnick, J. A. and Lavrakas, P. J., 367-386. Hoboken, NJ: John Wiley and Sons, Ltd.

Freeman, A.M., Herriges, J., Kling, C. (2014) The measurement of environmental and resource values. RFF Press.

Jäckle, A., Roberts, C. and Lynn, P. (2010). Assessing the Effect of Data Collection Mode on Measurement. International Statistical Review 78 (1) 3-20.

Jødestøl, K. A., Fredheim, B., Hoell, E. E., Wakili, S., Vinnem, J. E., Myhrvold, A. U., Hasle, J. R. and Ytterborg, G. (2001). Achieving an Industry Standard in the Assessment of Environmental Risk: Oil Spill Risk Management and the Mira Method. International Oil Spill Conference Proceedings (1): 155-165.

Johnston R. J., Boyle, K. J., Adamowicz, W., Bennett, J., Brouwer, R., Cameron, T. A., Hanemann, W. M., Hanley, N., Ryan, M., Scarpa, R., Tourangeau, R. and Vossler, C. A. (2017). Contemporary guidance for stated preference studies. Journal of the Association of Environmental and Resource Economists 4(2): 319-405.

Krosnick, J. A., A. L. Holbrook, M. K. Berent, R.T. Carson, W. M. Hanemann, R. J. Kopp, R. C. Mitchell, S. Presser, P. A. Ruud, and V.K. Smith, W. R. Moody, M. C. Green, M. Conaway (2002). The impact of “no opinion” response options on data quality: Non-attitude reduction or an invitation to satisfice? Public Opinion Quarterly 66:371–403.

Liebe, U., Genk, K., Oehlmann, M. and Meyerhoff, J. (2015). Does the use of mobile devices (tablets and smartphones) affect survey quality and choice behavior in web surveys? The Journal of Choice Modelling 14: 17-31.

Lindhjem, H. and Navrud, S. (2011a). Using Internet in Stated Preference Surveys: A Review and Comparison of Survey Modes. International Review of Environmental and Resource Economics 5 (4) 309-351.

Lindhjem, H. and Navrud, S. (2011b). Are Internet Surveys an Alternative to Face-To-Face Interviews in Contingent Valuation? Ecological Economics 70: 1628-1637.

Lindhjem, H., Magnussen, K., Navrud, S., Skjeflo, S. W. and Brude, O. W. (2016). Verdsetting av miljørelatert velferdstap ved oljeutslipp fra skip: Kalkulasjonspriser for samfunnsøkonomiske analyser. Vistarapport 2016/22. Oslo: Vista Analyse AS.

Mahieu, P-A., Riera, P. and Giergiczny, M. (2012). Determinants of willingness-to-pay for water pollution abatement: A point and interval data payment card application. Journal of Environmental Management 108: 49-53.

Navrud, S., Lindhjem, H. and Magnussen, K. (2017). Valuing Marine Ecosystem Services Loss from Oil Spills for Use in Cost-Benefit Analysis of Preventive Measures in Handbook on the Economics and Management of Sustainable Oceans edited by Paulo A. L. D. Nunes, Lisa Emelia Svensson and Anil Markandya, 124-137. Cheltenham, UK: Edward Elgar Publishing Limited.

Olsen, S.B. (2009). Choosing between internet and mail survey modes for choice experiment surveys considering non-market goods. Environmental and Resource Economics 44(4):591-610.

Parush, A. and Yuviler-Gavish, N. (2004). Web navigation structures in cellular phones: the depth/breadth trade-off issue. International Journal of Human-Computer Studies 60 (5-6) 753-770.

Payne, J. W., Bettman J. R., Schkade D. A. (1999) Measuring Constructed Preferences: Towards a Building Code. Journal of Risk and Uncertainty, 19:1-3: 243-270.

Peterson, G. (2012). Unintended mobile respondents. Paper presented at the annual Council of American Survey Research Organizations Technology Conference, New York, NY.

Peterson, G., Griffin, J., LaFrance, J. and Li, J. (2017). Smartphone participation in web surveys. In: A Total Survey Error Perspective in Total Survey Error in Practice: Improving Quality in the Era of Big Data, edited by Biemer, P. P., de Leeuw, E., Eckman, S., Edwards, B., Kreuter, F., Lyberg, L. E., Tucker, N. C. and West, B. T., 133-153. New York, NY: John Wiley and Sons, Ltd.

Peytchev, A. and Hill, C. A. (2010). Experiments in Mobile Web Survey Design: Similarities to Other Modes and Unique Considerations. Social Science Computer Review 28 (3) 319-335.

Puhani, P. (2000). The Heckman correction for sample selection and its critique. Journal of economic surveys, 14 (1), 53-68.

Sandorf, E. D., Aanesen, M. and Navrud, S. (2016). Valuing unfamiliar and complex goods: A comparison of valuation workshops and internet panel surveys with video. Ecological Economics 129: 50-61.

Whitehead, J. (2016). Plausible responsiveness to scope in contingent valuation. Ecological Economics 128: 17-22