Are preferences stated in web vs. personal interviews different? A comparison of willingness to pay results for a large multi-country study of the Baltic Sea eutrophication reduction

Ewa Zawojska, Mikołaj Czajkowski

University of Warsaw, Faculty of Economic Sciences, Poland

Abstract: We investigate the prevailing view in the stated preference literature that the data collection mode does not significantly affect the value estimates. Based on data from Computer-Assisted Web Interviews and Computer-Assisted Personal Interviews aimed at assessing the social benefits for Poland from meeting the nutrient load reduction targets defined in the HELCOM's Baltic Sea Action Plan (2007), we find that the value estimates obtained from the two modes differ significantly. This evidences the existence of a "pure" mode effect as we control for socio-demographic differences between the web-interviewed and personally-interviewed samples by weighting the observations. The relative difference in the derived values between the two modes is used to update the estimates of the economic values of reducing nutrient loadings to the Baltic Sea provided by Ahtiainen et al. (2014) for every Baltic Sea country. In addition to controlling for the mode effect (as different, web and personal, modes were used in different countries), we examine 18 alternative model specifications to find the distribution that captures best the payment-card willingness-to-pay responses. Overall, our study illustrates the extent of the impact that the choice of a data collection mode can have on valuation results.

Key words: data collection, survey administration, mode effect, Computer-Assisted Web Interviews (CAWI), Computer-Assisted Personal Interviews (CAPI)

Acknowledgements: This study was supported by the National Science Centre in Poland under the grant Preludium (no. 2014/15/N/HS4/01328). Ewa Zawojska obtained funding within a doctoral stipend granted by the National Science Centre in Poland (no. 2016/20/T/HS4/00013).

1. Introduction

Stated preference surveys are administered by various modes, which include mail, phone, web and personal interviews. The prevailing view in the literature is that as long as the samples surveyed via different modes are the same with respect to relevant characteristics, the choice of a data collection mode does not significantly affect the survey results (Johnston et al., forthcoming). However, a closer look into the research that investigated this issue reveals that findings on the mode effect are not that univocal. When the evidence is limited to studies that compared web and personal interviews, the number of studies reporting a significant mode effect is nearly the same as the number of studies reporting this effect to be insignificant (see Table 1 in the next section). Given a substantial rise in the use of web surveys for valuation of public goods in recent years, the investigation of the effect of this mode on respondent's behaviour appears of particular importance. Our study contributes to the literature by testing differences in value estimates of a public good derived from Computer-Assisted Web Interviews (CAWI) and Computer-Assisted Personal Interviews (CAPI).

Personal (face-to-face) interviews have been long acknowledged as a best practice in stated preference research (Arrow et al., 1993; Mitchell and Carson, 1989). The NOAA Panel reasoned that the in-person mode helped respondents understand complex information, for example, through pictures and other visual material provided, and, thus, it fostered collecting data of high quality (that is, data accurately reflecting respondents' preferences). Growth of the Internet use has allowed to administer surveys in a cheaper and faster way, at the same time maintaining the advantage of presenting visual material. With still expanding access to the Internet, web surveys are gaining more and more popularity. The number of web valuation surveys conducted annually more than tripled in the years 2013-2015 in comparison with the years 2001-2007 (Menegaki et al., 2016). The essential question is, therefore, whether, and if so, to what extent, the choice of a data collection mode impinges on survey outcomes.

Differences in value estimates derived from web and personal modes arise from differences in that: who respond and how respond (Fricker and Schonlau, 2002; Stephenson and Crête, 2011). Regarding the first aspect, web surveys may be accessed by different people (particularly, in populations in which Internet access is limited) and may attract different people (self-selection bias). Consequently, a sample of respondents in a web survey is likely to differ from a sample of respondents in a personal survey. This influences the extent to which web-elicited preferences reflect preferences of the population of interest, and, hence, this can undermine the web sample representativeness. Regarding the second aspect, a mode itself may alter respondents' incentives how to answer a survey. This is sometimes referred to as a "pure" mode effect (Jäckle et al., 2010) when the same respondent answers differently surveys which are worded identically but administered by different modes. The "pure" mode effect can be attributed to normative / sociological factors or to cognitive / psychological factors (Dillman, 2000). The normative / sociological factors involve the influence of cultural norms on respondent's behaviour, and this influence may differ across modes. In particular, the presence of the interviewer is likely to affect respondents' perceptions of (and adherence to) cultural norms. In this regard, the most widely recognised source of the mode effect is social desirability: respondents answer in a way they think they ought to answer. The cognitive / psychological factors cover information processing by respondents. A common source of the mode effect in this regard is satisficing behaviour, which means that respondents make shortcuts and choose a satisfactory answer instead of an optimal answer.

In this paper, we use data from a large multi-country study that inquired the social value of the Baltic Sea eutrophication reduction and involved all Baltic Sea countries. In Poland, the survey was administered by two modes: CAWI and CAPI. Based on this data, we verify whether the value estimates differ between the survey modes. We weigh the observations for respondents who participated in each mode to control for differences in socio-demographic characteristics between the mode samples and to make them reflect the characteristics of the general population of Poland. In order to derive mean willingness-to-pay (WTP) estimates for the considered environmental improvement, we use respondents' answers to a payment-card valuation question and identify the distribution that fits best to the data. We find that CAWI respondents are willing to pay on average significantly more for reduction of the Baltic Sea eutrophication than CAPI respondents.

Based on our result evidencing significantly different value estimates from CAWI and CAPI, we use the relative difference between them as a correction factor for calculating the social value of marine eutrophication reduction for every Baltic Sea country. The survey was administered through different modes (web and personal) in different countries. We compare our results with the values reported by Ahtiainen et al. (2014), who evaluated this environmental improvement for each of the Baltic Sea coastal countries, but without controlling for the mode effect. Given our derived correction factor, we recalculate the social values provided by Ahtiainen et al. Our results illustrate the extent of the difference in welfare measures if the data collection mode is accounted for.

The remainder of the paper is structured as follows. Section 2 presents the existing empirical evidence on the mode effect from comparisons of web and personal stated preference surveys. Section 3 provides details about the survey design and its administration. Survey 4 outlines our modelling approach. Section 5 discusses the results, and Section 6 concludes.

2. Previous valuation studies that compared web and personal stated preference surveys

Stated preference literature typically holds that the effect of the data collection mode on the survey results is negligible. Lindhjem and Navrud (2011a) reviewed 17 stated preference studies which had compared web and other-mode surveys in the context of environmental goods and environmentrelated health risks, and concluded that "[t]he SP [stated preference] studies … do generally not find substantial difference" (p. 309). Menegaki et al. (2016) identified 41 economic valuation studies conducted from 2001 through 2015 that had examined differences in value estimates from web and other-mode surveys, and found that "the majority of [these studies] … fail to confirm the existence of mode effects" (p. 40). Finally, the contemporary guidance for stated preference studies (Johnston et al., forthcoming) says that "[r]ecent research suggests that data collection mode does not substantially influence SP [stated preference] study outcomes …". Though, the authors add that the results are mixed and specific to a research context.

A thorough review of existing empirical research leads to a somewhat different conclusion, suggesting that a data collection mode may affect the outcomes considerably. Given the focus of our paper on the comparison of web- and in-person-based value estimates, we refer below to stated preference studies that examined differences in results derived from web and personal surveys. We find 13 such studies. Table 1 summarises our literature review.[1]

As presented in Table 1, the evidence regarding the difference in outcomes derived from web and personal survey modes is mixed. Out of the 13 studies, seven reported a significant mode effect. Findings with respect to the direction of the difference in value estimates from the two modes are not consistent, although the vast majority observed that web-based data generated lower value estimates

[1] Here we do not list studies which involved web and personal data collection modes but: which did not inquire differences between the two modes (for example, Ahtiainen et al., 2014; Hamzaoui-Essoussi and Zahaf, 2012; Reichl et al., 2013); which employed not equivalent value elicitation formats in different modes (Goethals et al., 2012; Ready et al., 2006; Sandorf et al., 2016); which involved no valuation task neither a choice task (Goldenbeld and De Craen, 2013); which evaluated different goods in different modes (Maier et al., 2015).

than in-person-based data. Some studies relate this result to social desirability bias prevalent in personal interviews (Lee et al., 2016; Mjedle et al., 2016).

Table 1. Stated preference studies that compared outcomes from web and personal surveys

Author(s)	Topic	Value elicitation format	Difference in values between modes
Balderas Torres et al. (2013)	Carbon offsetting by local forests	Multiple choice sequence (DCE)	Yes (Web < Personal)
Bell et al. (2011)	Water quality in rivers, lakes and streams	Binary choice sequence (DCE)	Yes (Web < Personal)
Canavari et al. (2005)	Pesticide ban; Organic apples	Yes-No question and Openended question (CV); Open-ended question (CV)	No Yes (Web > Personal)
Cardamone et al. (2014)	Risk of road accident	Ranking (DCE)	No
Covey et al. (2010)	Prevention of railway fatalities	Ranking (DCE)	No
Lee et al. (2016)	Nature preservation	Yes-No question (CV)	Yes (Web < Personal)
Lindhjem and Navrud (2011b)	Biodiversity protection	Payment card (CV)	No
Marta-Pedroso et al. (2007)	Landscape preservation	Open-ended questions (CV)	Yes (Web < Personal)
Mjedle et al. (2016)	Nature preservation	Multiple choice sequence (DCE)	Yes (Web < Personal)
Mulhern et al. (2013)	Health state	Binary choice sequence (DCE)	No
Nielsen (2011)	Gain in life expectancy in the context of air pollution	Open-ended questions (CV)	No
Ščasný and Alberini (2012)	Reduction of mortality risk attributable to a climate change	Multiple choice sequence (DCE)	No
van der Heide et al. (2008)	Alleviation of negative effects of habitat fragmentation	Double-bounded dichotomous choice question (CV)	For one scenario: Yes (Web < Personal) and for another scenario: No

Notes: The abbreviations CV and DCE are used to refer to the nomenclature common in the stated preference literature: CV stands for contingent valuation and DCE stands for a discrete choice experiment. Notation "Web < Personal" implies that the value estimate from a web survey was statistically significantly lower than from a personal survey. "Web > Personal" implies the opposite.

The observation that many studies find a significant mode effect diverges from the commonly held view that the choice of a data collection mode does not affect valuation results. Given the inconsistence in the existing evidence, we provide an additional verification, based on a field study, of differences in value estimates derived from web and personal surveys.

3. Survey design and administration

Our study of the mode effect in web and personal stated preference surveys is based on data from valuation research of eutrophication reduction in the Baltic Sea (Ahtiainen et al., 2014). The survey was conducted from October through December 2011 in all nine countries having access to the sea. Particular emphasis was placed on developing the questionnaire equally relevant and accurate in each coastal country, both in informing about the eutrophication effects and in specifying the valuation scenario. Pre-testing included five expert reviews, three focus groups, sixteen cognitive interviews in different countries and pilot surveys in all nine countries. This helped design an identical survey instrument for every country, which was translated into national languages.

Data collection involved different modes in different countries: web surveys were administered in Denmark, Estonia, Finland, Germany and Sweden; personal interviews were conducted in Latvia, Lithuania and Russia; both web and personal interviewing modes were utilised in Poland. We, therefore, analyse at first the data for Poland to identify the mode effect, and subsequently we use the result to recalculate the values of the Baltic Sea eutrophication reduction in each coastal country accounting for the impact of the survey mode on respondents' choices.

The survey aimed at evaluating a better eutrophication status of the Baltic Sea in 2050. Respondents' preferences towards this improvement were elicited through assessment of two valuation scenarios which considered meeting, respecitively, 100% and 50% of the nutrient load reduction targets defined in the HELCOM's Baltic Sea Action Plan (HELCOM, 2007). The improvement scenarios were assessed against a baseline level of eutrophication predicted for 2050 that assumed no new investments in nutrient abatement and no changes in the wastewater treatment and in practices of agricultural and other economic sectors. The scenarios and the baseline were developed following state-of-the-art Baltic Sea marine models (Ahlvik et al., 2014; Kiirikki et al., 2001, 2006; Maar et al., 2011) and marine ecologists' professional evaluation.

The concept of eutrophication was introduced in the survey by linking it to five ecosystem effects: water clarity, blue-green algal blooms, condition of underwater meadows, composition of fish species and oxygen content in deep sea bottoms. Each effect was described on a five-step coloured water quality scale, in which every colour depicted a different level of the effect intensity (this, in turn, translated into different levels of water quality ranging from "best" to "worst"). After presenting the scale to the respondents, the improvement scenarios were shown in a form of colour-coded maps that illustrated eutrophication levels in all sub-basins of the Baltic Sea in 2050. The visual scenario representation was supported by a verbal description. Both the large (100%) and small (50%) improvement scenarios were included in each questionnaire, and the order in which they were displayed varied to account for potential scope and order effects.

After being introduced to the considered improvement program, the respondents answered whether they would be willing to pay anything in principle for eutrophication reduction in the Baltic Sea (this type of question is typically referred to as an in-the-market question or as a spike question). Next, the respondents were shown the maps illustrating the scenarios and were asked to indicate their WTP for the improvement on the provided payment card. The exact wording of the willingness-to-pay question was: "What is the most you would be willing to pay every year to reduce eutrophication in the Baltic Sea as shown in the maps?"

The payment mechanism was a tax which each individual and each firm in the Baltic Sea countries would need to pay annually upon the introduction of the environmental improvement program.[2] The description of the payment mechanism highlighted that the tax would be used for reducing the Baltic Sea eutrophication. A previous survey (Ahtiainen et al., 2013) revealed that citizens of the Baltic Sea countries preferred payments done by everyone to other means of funding actions. Pre-testing showed the tax vehicle to be perceived both credible and acceptable by the interviewed populations.

[2] Pre-testing showed that mentioning firms was important for the respondents for the reason of fairness. This formulation, however, could have incentivised the respondents to understate their willingness to pay if they believed that they would need to pay twice for the improvement – through the firms they worked in and individually.

The payment cards were designed analogically in every country, using the responses from the pilot studies. Each card included 18 positive bids, a zero bid and a “don’t know” option.[3] Specific bid values differed between the countries. The bid ranges were chosen so that neither the lower nor the upper end of the bid distribution would be truncated, as this has been evidenced to possibly affect the welfare estimates (Roach et al., 2002; Rowe et al., 1996).

The survey informed the respondents that by answering the questionnaire, they would affect the environmental policy programs related to eutrophication being conducted in the Baltic Sea area. Precisely, the survey said that respondents’ “answers will help governments around the Baltic Sea to develop appropriate water quality improvement programs”. Further details of the survey design and implementation are available in Ahtiainen et al. (2014).

4. Econometric approach

In the survey used in this study, respondents indicated their WTP for the Baltic Sea eutrophication reduction on a payment card: respondents were asked to choose the maximum bid they would be willing to pay from a provided set. Hence, their choice reveals that their WTP is equal to or higher than the selected bid and lower than the next (not selected) higher bid. Consequently, the payment-card responses can be viewed as interval-type data, since respondent’s actual WTP lies between the chosen bid and the next higher bid in the payment card (Cameron and Huppert, 1989).

We use the interval WTP responses to fit the parametric distribution of the respondents’ WTP. We do not arbitrary define a specific distribution of WTP, as there is no a priori or theory-driven information about the WTP distribution in the population. Instead, we try 18 commonly used parametric distributions and select the specification which matches the data best in terms of the finite sample corrected Akaike Information Criterion (AICc).[4] To fit the parameters of a selected distribution to the WTP data, we calculate the value of the cumulative distribution function of the selected distribution at the upper bound of each respondent’s WTP interval and subtract the corresponding value evaluated at the lower bound of their WTP interval.[5] This calculation gives the probability that the respondent’s WTP is above the lower bound (the selected bid) and below the upper bound (the next higher bid). The value of this probability is this respondent’s contribution to the likelihood function. By summing up each respondent’s contributions and maximising the resulting function with respect to the distribution parameters, we use the maximum likelihood method to fit a parametric distribution to the interval data. Based on the estimated parameters, we apply the bootstrapping technique proposed by Krinsky and Robb (1986) to simulate the parameters of the WTP distribution.

Because WTP is expected to be non-negative and a substantial share of zero WTP responses is identified, we use the spike model approach (Kriström, 1997). In the spike model, the respondents who declared that they were not willing to pay anything in principle (that is, the respondents who were not “in the market”) and the respondents who have an implied probability of holding negative WTP are cumulated in a spike-discontinuity of the WTP distribution at 0. Given our interest in testing the mode effect, we incorporate in the model a binary-coded variable indicating the mode type as an explanatory variable of the WTP responses.

---

[3] The only exception is the questionnaire in Russian, where 14 positive bids were included because of technical problems.

[4] For comparing fit of the models to the data, we use the AICc instead of the log-likelihood values because the considered distributions differ in the number of parameters.

[5] When the calculated difference in the values of the cumulative distribution function is equal to zero, we take the value of the probability density function evaluated at the lower bound of a respondent’s WTP interval.

Formally, the probability that individual i 's WTP lies between the selected bid $b_{i,s}$ and the next higher bid $b_{i,s+1}$ can be expressed as

$$P\left(b_{i.s} < WTP_i < b_{i,s+1}\right) = CDF\left(b_{i,s+1}, \boldsymbol{\beta}_i\right) - CDF\left(b_{i,s}, \boldsymbol{\beta}_i\right), \tag{1}$$

where CDF denotes a cumulative distribution function of the selected WTP distribution, and $\boldsymbol{\beta}_i$ is a vector of the distribution parameters (for example, mean and standard deviation for the normal distribution). By making $\boldsymbol{\beta}_i$ dependent on individual i 's characteristics, the parameter vector becomes individual-specific. In our study, we include only one individual-specific variable, namely a binary-coded variable equal to 1 for the respondents participating in CAPI and 0 for the respondents participating in CAWI.

Taking into account that the probability specified in (1) expresses individual i 's contribution to the likelihood function, we represent the log-likelihood function for the sample of N individuals as

$$\log L = \sum_{i=1}^{N} \log\left[CDF\left(b_{i,s+1}, \boldsymbol{\beta}_i\right) - CDF\left(b_{i,s}, \boldsymbol{\beta}_i\right)\right] \ . \tag{2}$$

Combining (2) with the information about individual i's market participation (captured by a binary-coded variable yes_i equal to 1 if individual i is "in the market" and 0 otherwise), the log-likelihood function for observing a particular set of choices in the sample is

$$\log L = \sum_{i=1}^{N} \left\{ yes_i \cdot \log\left[CDF\left(b_{i,s+1}, \boldsymbol{\beta}_i\right) - CDF\left(b_{i,s}, \boldsymbol{\beta}_i\right)\right] + \left(1 - yes_i\right) \cdot \log\left[CDF\left(0, \boldsymbol{\beta}_i\right)\right] \right\} \ . \tag{3}$$

(3) represents the spike approach. Maximisation of (3) generates estimates of the parameters underlying the WTP distribution.[6]

5. Results

This section discusses the results of our empirical inquiry. First, we examine differences in sociodemographic characteristics of the two samples of respondents in Poland: those who participated in CAWI and those who participated in CAPI. We also check how these samples diverge from the general Polish population in terms of socio-demographic characteristics. Second, we report the results of estimation of the models as described in Section 4. In order to control for the socio-demographic differences between the CAWI and CAPI samples, and in order to make the two samples representative for the Polish general population, we weigh the observations. Third, finding significant differences in the WTP values between the two modes, for each Baltic Sea country we update the value of the marine eutrophication reduction provided by Ahtiainen et al. (2014) who did not take into account the mode effect.

5.1. Socio-demographic characteristics of the CAWI and CAPI samples in Poland

The respondents surveyed in Poland were adult individuals in the age from 20 to 60. Sociodemographic characteristics of the two samples distinguished by mode, along with the characteristics for the general population of Poland, are given in Table 2. The table presents the shares of respondents for every category / level of the characteristics, and the shares for the general population are used in econometric modelling for weighting the observations to make the mode samples reflect the general population.

[6] The models were estimated using a custom code developed in Matlab, available at https://github.com/czaj/DistFit under CC BY 4.0 license.

Table 2. Socio-demographic characteristics for Poland for the mode samples and for the general population (the numbers represent the shares in percentages)

Characteristic	CAWI sample	CAPI sample	General population
Individual income a)			
Below 2,000 PLN	40.5	65.5	60
2,000 - 2,499 PLN	21.8	20	20
2,500 - 3,499 PLN	20.6	8.7	15
Above 3,500 PLN	17.1	5.8	5
Occupational status a)			
Employed	63.9	67.5	50
Self-employed	7.1	5.5	9.6
Unemployed	8.2	11	5.4
Other (Retired, Student, Home-employed)	20.8	16	35
Highest educational level attained b)			
Compulsory	1.3	6.2	23.5
Vocational	7.8	35.8	22.5
High school	41.6	42.3	34.5
University	49.3	15.7	19.5
Household size a)			
1	5.8	7	6.9
2	21.3	20.4	21
3	29.6	29.2	20.5
4	28.2	28.5	23.8
5	11.3	9.4	13.8
6 and more	3.8	5.5	14
Number of household members under 18 c)			
0	50.5	47	58.9
1	24.1	26.6	20.2
2	17.1	21.1	15.3
3 and more	8.3	5.3	5.6
Age b)			
20 - 29	25	25.2	26.9
30 - 39	24	25	25.9
40 - 49	25.5	25	21.4
50 - 60	25.5	24.8	25.8
Gender b)			
Female	49.4	50.1	52
Male	50.6	49.9	48

Notes: Sources of the statistics for the general population: a) Eurostat, European Union Statistics on Income and Living Conditions (EU-SILC) 2011; b) Central Statistical Office (2016). *Demographic Yearbook of Poland*. Warsaw, Poland, data for 2010; c) Eurostat, Labour Force Survey, data for 2011. According to the National Bank of Poland, the average exchange rate in 2011 was 1 EUR = 4.1196 PLN.

The comparison of the statistics for the two samples reveals that major differences appear with respect to income and attained education: the CAWI respondents have higher incomes and higher education than the CAPI respondents. In regard to other characteristics, the differences between the two modes are not that striking, although we note among the CAWI respondents more retired, home-employed and students and larger variance in the number of household members under 18 years old. Formal verification of the differences with chi-squared tests confirms our observations. The chi-squared tests' results are displayed in Table 3.

Table 3. Results of chi-squared tests of equality of distributions across the CAWI and CAPI samples

Characteristic	Test statistics	P-value	Significant difference between the CAWI and CAPI samples
Individual income	126.89	0.000	Yes
Occupational status	13.57	0.004	Yes
Highest educational level attained	386.68	0.000	Yes
Household size	6.15	0.292	No
Number of household members under 18	12.59	0.006	Yes
Age	0.35	0.950	No
Gender	0.08	0.774	No

5.2. Test of the mode effect

Given no a priory information or theory-grounded expectation about the WTP distribution in the population, we try 18 widely used parametric distributions to see which one fits our data best. The summary of the comparison of various model specifications is presented in the Appendix. Out of the considered distributions, the Birnbaum-Saunders distribution appears to provide the best fit to the data in terms of the lowest value of the finite sample corrected Akaike Information Criterion (AICc). We, thus, use this distribution for the estimation of the annual WTP for the Baltic Sea eutrophication reduction based on CAWI and CAPI responses.[7] All calculations are conducted on the weighted samples so that the socio-demographic characteristics of each CAPI and CAWI subsample reflect the characteristics of the general population of Poland.

The estimation results of the payment-card answers indicating the respondents'' values of the Baltic Sea eutrophication reduction are shown in Table 4. The models in the table assume the Birnbaum-Saunders distribution of WTP. The Birnbaum-Saunders distribution is characterised by two parameters: shape and scale parameters, estimates of which are given in the table. We present results of three specifications: (A) is based on the full sample that pools the CAWI and CAPI responses; (B) uses the full sample and includes a binary-coded explanatory variable indicating the CAWI mode; (C) provides the model estimates separately calculated for CAWI and CAPI.

Specification (B) which includes a binary-coded variable controlling for the mode outperforms specification (A) in terms of the log-likelihood value and the AICc: the log-likelihood ratio test suggests rejecting the null hypothesis that the restrictions (placed in specification (A)) do not significantly deteriorate the model fit (the log-likelihood ratio test statistics = 191.394, p-value = 0.000). Thereby, specification (B) is preferred, which emphasises the significance of the mode effect.

The explanatory variable controlling for the mode in specification (B) affects statistically significantly both distribution parameters and the spike parameter. This evidences substantial differences in the value estimates derived from the CAWI and CAPI responses. We, thus, conclude that a significant mode effect is present.

Descriptive statistics for the fitted distribution in specification (C) illustrate the extent of the mode effect. The annual mean WTP derived from the CAWI responses is at the level of 40.15 PLN, and the annual mean WTP obtained from the CAPI responses is equal to 18.08 PLN: on average, the CAWI respondents stated values more than twice higher than the CAPI respondents. Specification (C) also reveals that the Spike probability is by twice lower for the CAWI respondents.

[7] As mentioned in Section 3, the respondents evaluated two improvement scenarios: meeting 100% and 50% of the nutrient load reduction targets specified in the HELCOM's Baltic Sea Action Plan. Because our findings are consistent across the two scenarios, we focus here on the 50% improvement scenario.

Table 4. Estimation results of the annual WTP of Polish citizens for the Baltic Sea eutrophication reduction

	(A)	(B)		(C)	
	CAWI and CAPI pooled	**CAWI and CAPI pooled, with a binary-coded variable for CAWI**		**CAWI only**	**CAPI only**
	Mean parameter estimates	Mean parameter estimates	Parameter estimates for CAWI	Mean parameter estimates	Mean parameter estimates
Shape parameter	33.006*** (0.560)	26.695*** (1.152)	10.972*** (1.316)	37.675*** (0.639)	26.678*** (1.143)
Scale parameter	1.118*** (0.015)	1.162*** (0.034)	-0.106*** (0.037)	1.055*** (0.015)	1.161*** (0.033)
Spike	-0.100*** (0.013)	0.269*** (0.027)	-0.742*** (0.030)	-0.471*** (0.014)	0.268*** (0.027)
Model characteristics					
Log-likelihood	-4,049.478	-3,953.781		-2,324.955	-1,618.963
AICc/n	4.3787	4.279		5.023	3.511
n	1,851	1,851		927	924
Fitted distribution descriptive statistics					
Mean	29.407 (26.705)	27.201 (26.033)		40.152 (27.367)	18.080 (21.924)
Standard deviation	51.167 (29.408)	48.645 (28.471)		57.902 (28.843)	38.944 (24.881)
Median	8.201 (16.481)	6.008 (16.178)		19.876 (17.602)	0.000 (13.137)
0.025 quantile	0.000 (2.489)	0.000 (2.496)		0.000 (2.896)	0.000 (1.891)
0.975 quantile	177.367 (109.554)	168.546 (106.219)		203.189 (108.083)	132.610 (92.185)
Spike probability	0.460 (0.005)	0.457 (0.006)		0.319 (0.005)	0.606 (0.011)

Notes: Standard errors are given in brackets. *** indicates 10% significance level. n denotes the number of the observations.

5.3. Values of the eutrophication reduction in all Baltic Sea countries corrected for the mode effect

Given our finding from the previous section implying a significant mode effect, we recalculate the values of the Baltic Sea eutrophication reduction reported by Ahtiainen et al. (2014) for each country having access to the Baltic Sea. Ahtiainen et al. did not take into account the differences in the values arising from using various preference elicitation modes such as web surveys and personal interviews.

The summary of the results is provided in Table 5. The table contains the average WTP for every Baltic Sea country: as reported by Ahtiainen et al. (2014) and as calculated by us with calibrating for the mode effect. The WTP values are expressed in euro.[8] We also include the Spike probabilities in the table. The results are derived based on the same modelling approach as used above for Poland, with the

[8] The questionnaires in Denmark, Latvia, Lithuania, Poland, Russia and Sweden used national currencies. For the purpose of the comparison, we convert national currencies into euro using the PPP corrected exchange rates for 2011.

exception of no sample weighting because in the countries other than Poland a single mode was employed (either a web survey or personal interviews). Using the relative difference in the mean WTP values between CAWI and CAPI for Poland, we calibrate the results for the other countries to calculate the value estimates assuming that the other data collection mode would have been implemented than the one actually used (for example, for the countries where the data was collected through a web survey, we calibrate the obtained value to learn what the value would be if the data was collected through personal interviews).

Table 5. Annual mean WTP (in euro) for the marine eutrophication reduction for every Baltic Sea country corrected for the mode effect

	Ahtiainen et al. (2014)			**Our results**		
	Annual mean WTP based on CAWI	**Annual mean WTP based on CAPI**	**Spike probability**	**Annual mean WTP based on CAWI**	**Annual mean WTP based on CAPI**	**Spike probability**
Poland	12.2	12.2	0.47	15.3	7.0	CAWI: 0.25, CAPI: 0.60
Denmark	31.7	---	0.48	36.3	*16.6*	0.40
Estonia	24.0	---	0.48	15.8	*7.2*	0.40
Finland	41.8	---	0.37	40.7	*18.6*	0.36
Germany	25.0	---	0.46	21.7	*9.9*	0.43
Latvia	---	5.5	0.52	*9.8*	4.5	0.50
Lithuania	---	8.8	0.50	*14.6*	6.7	0.51
Russia	---	8.5	0.69	*17.9*	8.2	0.69
Sweden	75.7	---	0.33	70.5	*32.3*	0.20

Notes: We refer to the results of Ahtiainen et al. (2014) obtained from a Spike model which most closely resembles our modelling approach. The numbers in italics are calibrated WTP values assuming that the other data collection mode would have been used than the one actually implemented.

The annual mean WTP values do not differ substantially between our results and those provided in Ahtiainen et al. (2014). The observed (small) differences in the value estimates come from differences in the assumed distribution of WTP (Ahtiainen et al. assumed the log-normal distribution, while we use the distribution that fits best to the data) and from differences in the explanatory variables used (Ahtiainen et al. incorporated in the model several explanatory variables). The corresponding Spike probabilities have also similar values across our models and those of Ahtiainen et al.

The crucial finding from this analysis is the extent to which the WTP estimates are affected by the data collection mode. Table 5 illustrates this issue by providing the value estimates from the data for the actually conducted mode and the calibrated values for the other mode. We observe large discrepancies between the values which underlines how considerably the mode impinges on the valuation results. Importantly, Table 5 displays differences in the average WTP values, while for policy assessments the aggregate value for the entire population is typically used. Aggregation of the average WTP values for the whole population will result in even larger differences in the value estimates derived from the two modes. Consequently, the choice of the mode may substantially impinge on the evaluation of benefits from the policy, and subsequently may affect the decision following the cost-benefit analysis whether the policy should be introduced or not.

6. Discussion

Stated preference surveys play an important role in the cost-benefit analysis of public policies, as well as in litigation over environmental damages. Given their widespread use for policy and legal purposes, it is crucial that survey-based value estimates provide valid welfare measures. For long, guidelines in stated preference research suggested using personal interviews to obtain relevant public good values (NOAA Panel's recommendations; Arrow et al., 1993). Recently, however, the use of web interviews has increased considerably. Hence, the important question is whether, and if so, to what extent, the choice of a data collection mode impinges on obtained value estimates. We inquire this question in a field study that evaluates economic benefits from meeting the targets of nutrient load reduction defined in the HELCOM's Baltic Sea Action Plan (HELCOM, 2007).

The stated preference survey aimed at the assessment of the benefits from reducing nutrient loadings to the Baltic Sea was conducted in every country that has access to the Baltic Sea. In different countries, different data collection modes were used: web or personal. Poland is the only country in which both types of the data collection modes were employed. Thus, based on the data for Poland, we verify whether the mode affects the value estimates, and find that the web respondents are willing to pay on average significantly more for this environmental improvement than the respondents interviewed in-person. Given this result, we recalculate the values of this improvement for every Baltic Sea country reported by Ahtiainen et al. (2014) who did not control for the mode effect. This illustrates the substantial impact that the choice of a data collection mode may have on valuation results.

Our research emphasises the need for caution when choosing a data collection mode. Although the predominant view in the stated preference literature suggests that the mode does not significantly affect the valuation results (Johnston et al., forthcoming), we show that using different modes can lead to considerably different results. This finding is particularly important in the light of the use of surveybased value assessments for policy purposes. Our results imply that employing different data collection modes, the policy efficacy can be differently evaluated.

References

Ahlvik, L., Ekholm, P., Hyytiäinen, K., and Pitkänen, H. (2014). An economic–ecological model to evaluate impacts of nutrient abatement in the Baltic Sea. *Environmental Modelling and Software*, 55:164-175.

Ahtiainen, H., Artell, J., Czajkowski, M., Hasler, B., Hasselström, L., Huhtala, A., Meyerhoff, J., Smart, J. C. R., Söderqvist, T., Alemu, M. H., Angeli, D., Dahlbo, K., Fleming-Lehtinen, V., Hyytiäinen, K., Karlõševa, A., Khaleeva, Y., Maar, M., Martinsen, L., Nõmmann, T., Pakalniete, K., Oskolokaite, I., and Semeniene, D. (2014). Benefits of meeting nutrient reduction targets for the Baltic Sea - A contingent valuation study in the nine coastal states. *Journal of Environmental Economics and Policy*, 3(3):1-28.

Ahtiainen, H., Artell, J., Czajkowski, M., Hasler, B., Hasselström, L., Hyytiäinen, K., Meyerhoff, J., Smart, J., Söderqvist, T., Zimmer, K., Khaleeva, J., Rastrigina, O., and Tuhkanen, H. (2013). Public preferences regarding use and condition of the Baltic Sea - An international comparison informing marine policy. *Marine Policy*, 42:20-30.

Arrow, K., Solow, R., Portney, P. R., Leamer, E. E., Radner, R., and Schuman, H. (1993). Report of the NOAA panel on contingent valuation. *Federal Register*, 58:4601-4614.

Balderas Torres, A., MacMillan, D. C., Skutsch, M., and Lovett, J. C. (2013). The valuation of forest carbon services by Mexican citizens: the case of Guadalajara city and La Primavera biosphere reserve. *Regional environmental change*, 13(3):661-680.

Bell, J., Huber, J., and Viscusi, W. K. (2011). Survey mode effects on valuation of environmental goods. *International Journal of Environmental Research and Public Health*, 8(4):1222-1243.

Cameron, T. A., and Huppert, D. D. (1989). OLS versus ML estimation of non-market resource values with payment card interval data. *Journal of Environmental Economics and Management*, 17(3):230-246.

Canavari, M., Nocella, G., and Scarpa, R. (2005). Stated willingness-to-pay for organic fruit and pesticide ban: an evaluation using both web-based and face-to-face interviewing. *Journal of Food Products Marketing*, 11(3):107-134.

Cardamone, A. S., Eboli, L., Mazzulla, G. (2014). Drivers' road accident risk perception. A comparison between face-to-face interview and web-based survey. *Advances in Transportation Studies*, 33:59-72

Covey, J., Robinson, A., Jones-Lee, M., and Loomes, G. (2010). Responsibility, scale and the valuation of rail safety. *Journal of Risk and Uncertainty*, 40(1):85-108.

Dillman, D. (2000). *Mail and Internet Surveys: The Tailored Design Method*, John Wiley and Sons, Inc.

Fricker, R. D., and Schonlau, M. (2002). Advantages and Disadvantages of Internet Research Surveys: Evidence from the Literature. *Field Methods* 14:347-367.

Goethals, F., Leclercq-Vandelannoitte, A., and Tütüncü, Y. (2012). French consumers' perceptions of the unattended delivery model for e-grocery retailing. *Journal of Retailing and Consumer Services*, 19(1):133-139.

Goldenbeld, C., and De Craen, S. (2013). The comparison of road safety survey answers between webpanel and face-to-face; Dutch results of SARTRE-4 survey. *Journal of Safety Research*, 46:1320.

Hamzaoui-Essoussi, L., and Zahaf, M. (2012). Canadian Organic Food Consumers' Profile and Their Willingness to Pay Premium Prices. *Journal of International Food and Agribusiness Marketing*, 24:1-21.

HELCOM (2007). Helcom Baltic Sea Action Plan. Adopted on 15 November 2007 in Krakow, Poland by the HELCOM Extraordinary Ministerial Meeting. Helsinki Commission, Helsinki.

Jäckle, A., Roberts, C., and Lynn, P. (2010). Assessing the effect of data collection mode on measurement. *International Statistical Review*, 78(1):3-20.

Johnston, R., Boyle, K., Adamowicz, W., Bennett, J., Brouwer, R., Cameron, T. A., Hanemann, W. M., Hanley, N., Ryan, M., Scarpa, R., Tourangeau, R., Vossler, C. (forthcoming). Contemporary

Guidance for Stated Preference Studies. *Journal of the Association of Environmental and Resource Economists*.

Kiirikki, M., Inkala, A., Kuosa, H., Pitkänen, H., Kuusisto, M., and Sarkkula, J. (2001). Evaluating the effects of nutrient load reductions on the biomass of toxic nitrogenfixing cyanobacteria in the Gulf of Finland, Baltic Sea. *Boreal Environment Research*, 6:1-16.

Kiirikki, M., Lehtoranta, J., Inkala, A., Pitkänen, H., Hietanen, S., Hall, P. O. J., Tengberg, A., Koponen, J., and Sarkkula, J. (2006). A simple sediment process description suitable for 3D-ecosystem modelling - Development and testing in the Gulf of Finland. *Journal of Marine Systems*, 61(12):55-66.

Krinsky, I., and Robb, A. L. (1986). On approximating the statistical properties of elasticities. *Review of Economics and Statistics*, 68(4):715-719.

Kriström, B. (1997). Spike Models in Contingent Valuation. *American Journal of Agricultural Economics*, 79(3):1013-1023.

Lee, C. K., Kim, T. K., and Mjelde, J. W. (2016). Comparison of preservation values between Internet and interview survey modes: The case of Dokdo, South Korea. *Journal of Environmental Planning and Management*, 59(1):22-43.

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*, 5:309-351.

Lindhjem, H., and Navrud, S. (2011b). Are Internet surveys an alternative to face-to-face interviews in contingent valuation? *Ecological economics*, 70(9):1628-1637.

Maar, M., Møller, E. F., Larsen, J., Madsen, K. S., Wan, Z., She, J., Jonasson, L., and Neumann, T. (2011). Ecosystem modelling across a salinity gradient from the North Sea to the Baltic Sea. *Ecological Modelling*, 222(10):1696-1711.

Maier, E., Wilken, R., and Dost, F. (2015). The double benefits of consumer certainty: combining risk and range effects. *Marketing Letters*, 26(4):473-488.

Marta-Pedroso, C., Freitas, H., and Domingos, T. (2007). Testing for the survey mode effect on contingent valuation data quality: A case study of web based versus in-person interviews. *Ecological Economics*, 62(3):388-398.

Menegaki, A. N., Olsen, S. B., and Tsagarakis, K. P. (2016). Towards a common standard - A reporting checklist for web-based stated preference valuation surveys and a critique for mode surveys. *Journal of Choice Modelling*, 18:18-50.

Mitchell, R. C., and Carson, R. T. (1989). *Using surveys to value public goods: the contingent valuation method*. Resources for the Future.

Mjelde, J. W., Kim, T. K., and Lee, C. K. (2016). Comparison of Internet and interview survey modes when estimating willingness to pay using choice experiments. *Applied Economics Letters*, 23(1):74-77.

Mulhern, B., Longworth, L., Brazier, J., Rowen, D., Bansback, N., Devlin, N., and Tsuchiya, A. (2013). Binary choice health state valuation and mode of administration: head-to-head comparison of online and CAPI. *Value in Health*, 16(1):104-113.

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(1):119-129.

Ready, R., Fisher, A., Guignet, D., Stedman, R., and Wang, J. (2006). A pilot test of a new stated preference valuation method: Continuous attribute-based stated choice. *Ecological Economics*, 59(3):247-255.

Reichl, J., Schmidthaler, M., and 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.

Roach, B., Boyle, K. J., and Welsh, M. P. (2002). Testing Bid Design Effects in Multiple Bounded Contingent Valuation. *Land Economics*, 78(1):121-131.

Rowe, R. D., Schulze, W. D., and Breffle, W. S. (1996). A Test for Payment Card Biases. *Journal of Environmental Economics and Management*, 31(2):178-185.

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

Ščasný, M., and 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(12):4760-4781.

Stephenson, L. B., and Crête, J. (2011). Studying Political Behavior: A Comparison of Internet and Telephone Surveys. *International Journal of Public Opinion Research*, 23:24-55.

van der Heide, C. M., van den Bergh, J. C., van Ierland, E. C., and Nunes, P. A. (2008). Economic valuation of habitat defragmentation: A study of the Veluwe, the Netherlands. *Ecological Economics*, 67(2):205-216.

Appendix

A comparison of different parametric distributions fitted to the interval WTP data on the Baltic Sea eutrophication reduction

Distribution	**AICc/n**	**Log-likelihood**
Birnbaum-Saunders	4.2786	-3,953.781
Inverse Gaussian	4.2787	-3,953.889
Log-normal	4.2923	-3,966.498
Log-logistic	4.3182	-3,990.499
Exponential	4.3530	-4,024.661
Gamma	4.3630	-4,031.915
Negative binomial	4.3660	-4,034.712
Generalized Pareto	4.4523	-4,112.582
Generalized extreme value	4.6165	-4,264.568
Logistic	4.7607	-4,400.047
Normal	4.9253	-4,552.304
t location-scale	4.9278	-4,552.673
Johnson SU	4.9568	-4,577.420
Rayleigh	5.0019	-4,625.281
Rician	5.0127	-4,633.195
Extreme value	5.4159	-5,006.387
Uniform	5.6050	-5,181.433
Poisson	23.6494	-21,883.471

Notes: Each model specification is estimated on pooled data for the CAWI and CAPI responses, including a binary-coded variable indicating the CAWI mode. The presented results are ordered according to the finite sample corrected Akaike Information Criterion (AICc) divided by the number of observations (n).