Economic benefits of the Baltic Sea Action Plan – Revisiting the methodology of the welfare change assessment

Abstract: Although the Baltic Sea is a flagship case for studying major environmental perturbations and the efficacy of various management responses, only few studies to date have examined the monetary value of improvements to the Baltic Sea environment. In this paper, we revisit the estimates of Ahtiainen et al. (2014) – a large-scale contingent valuation study that focused on the social value of the Baltic Sea eutrophication reduction associated with the implementation of the Baltic Sea Action Plan (BSAP). With surveys administered to nearly 10,000 respondents in all Baltic Sea countries, it remains the most comprehensive valuation study of eutrophication to date. However, advances in the methodology of stated preference data analysis allow us to improve these estimates with respect to two important aspects. First, using the data from one country in which surveys were administered through both Computer Assisted Personal Interviews (CAPI) and Computer Assisted Web Interviews (CAWI), we control for the survey administration method and provide estimates that reflect respondents' Willingness To Pay as if in all countries it was elicited with the generally more reliable and recommended CAPI method. Second, we investigate multiple alternative model specifications and identify the one, which is the most flexible and fits the data best, resulting in more reliable and robust estimates of Willingness To Pay. Overall, our paper updates the estimated economic benefits of the implementation of the BSAP, providing a more robust basis for future analyses, various policy considerations and generally – an input for science-based management.

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

JEL codes: D12, Q40, Q48, Q51, Q55

1. Introduction

The Baltic Sea is a flagship case for studying major environmental perturbations and the efficacy of various management responses. Its main environmental problems include eutrophication along with its consequences – deterioration of the water transparency, increased toxic algal blooms, the expansion of oxygen-minimum zones and changes in fish stock (HELCOM, 2014). From the natural science perspective, the Baltic Sea is one of the most intensely studied areas in the world. The high-density data, many long-term data series and coordinated macro-regional research agenda provide crucial inputs for science-based management (Reusch et al., forthcoming). However, the science-based management is not possible without robust inputs from social sciences, particularly without valid estimates of economic benefits and costs of various policy actions.

The monetary value of the improvements to the Baltic Sea environment have been the focus of only a few studies to date. Most notably, Ahtiainen et al. (2014) applied stated preference methods (Contingent Valuation, CV) and estimated the value of alleviating eutrophication in the Baltic Sea at EUR 3.6 billion annually. The economic value of the reductions in eutrophication has earlier been measured in the Stockholm archipelago of Sweden (Söderqvist and Scharin, 2000) and in Lithuania, Poland and Sweden (Markowska and Żylicz, 1999). Tuhkanen et al. (2016) estimate the value of benefits for improvements of water quality in the Estonian waters, while Pakalniete et al. (2017) provide willingness to pay estimates for the improved coastal water quality for recreation in Latvia. The benefits of the improved water quality for recreation were also studied using the Travel Cost Method (TCM). Czajkowski et al. (2015) study the littoral countries recreation patterns and estimate the total economic benefits provided by the Baltic Sea-based recreation at EUR 14.8 billion per year; they also predict they would be nearly EUR 2 billion higher, if the environmental status of the sea improved. These results are in line with more recent spatially-explicit estimates of Czajkowski et al. (forthcoming-b). Other TCM estimates were provided by Vesterinen et al. (2010) for Finland, and Sandström (1996) and Soutukorva (2005) for Sweden.[1]

In this paper, we revisit the estimates of Ahtiainen et al. (2014) – a large-scale contingent valuation study that focused on the social value of the Baltic Sea eutrophication reduction associated with the implementation of the Baltic Sea Action Plan (BSAP; HELCOM, 2013). With surveys administered to nearly 10,000 respondents in all Baltic Sea countries, it remains the most comprehensive and influential valuation study of eutrophication to date. However, advances in the methodology of stated preference data analysis allow us to improve these estimates with respect to two important aspects. First, using the data from one country in which surveys were administered through both Computer Assisted Personal Interviews (CAPI) and Computer Assisted Web Interviews (CAWI), we control for the survey administration method (Menegaki, Olsen and Tsagarakis, 2016) and provide estimates that reflect respondents' Willingness To Pay (WTP) as if in all countries it was elicited with the generally more reliable and recommended CAPI method (Johnston et al., 2017). Second, we investigate multiple alternative model specifications and identify the one, which is the most flexible and fits the data best, resulting in more reliable and robust estimates of WTP. In addition, we use the weighted maximum likelihood method framework to control for the differences in socio-demographic characteristics of each country sample and target population. Finally, despite the open sea focus of the valuation study, there

[1] The literature dealing with estimating the costs of estimating costs of nutrient loading reductions to the Baltic Sea includes Gren, Jannke and Elofsson (1997), Ollikainen and Honkatukia (2001), Schou et al. (2006), COWI (2007), Gren (2008), Wulff et al. (2014), Ahlvik et al. (2014) and Czajkowski et al. (forthcoming-a).

is an indication that some respondents declared WTP expecting improvements in both open sea and coastal waters. We provide WTP estimates adjusted for this embedding effect, noting that they are only valid for the quantified and described improvements to the open sea ecosystems and coastal waters quality improvements (particularly important for recreation) should be a subject of a separate study. Overall, our paper updates the estimated economic benefits of the implementation of the BSAP, providing a more robust basis for future analyses, various policy considerations and generally – an input for science-based management.

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

2. Survey design and administration

The data used for this study comes from a survey conducted in 2011 that aimed at estimating people's WTP for reducing eutrophication in the open-sea areas of the Baltic Sea by 2050, as envisaged by the implementation of the Baltic Sea Action Plan. In total, 10,564 respondents all the nine countries around the Baltic Sea were surveyed (Ahtiainen et al., 2014).

The study used a stated preference approach, namely the CV method, in which respondents' are familiarized with a policy that is considered for implementation, and asked if they would support such a policy if it was associated with a specified cost to them (Hanley and Czajkowski, 2017). By observing the share of respondents who would vote "yes" for policies associated with different cost levels, it is possible to estimate the distribution of people's WTP, and, for example, calculate mean WTP that can be used to compare with the actual costs of the policy. If people's WTP exceeds the cost of policy implementation the policy is considered net beneficial to the society.

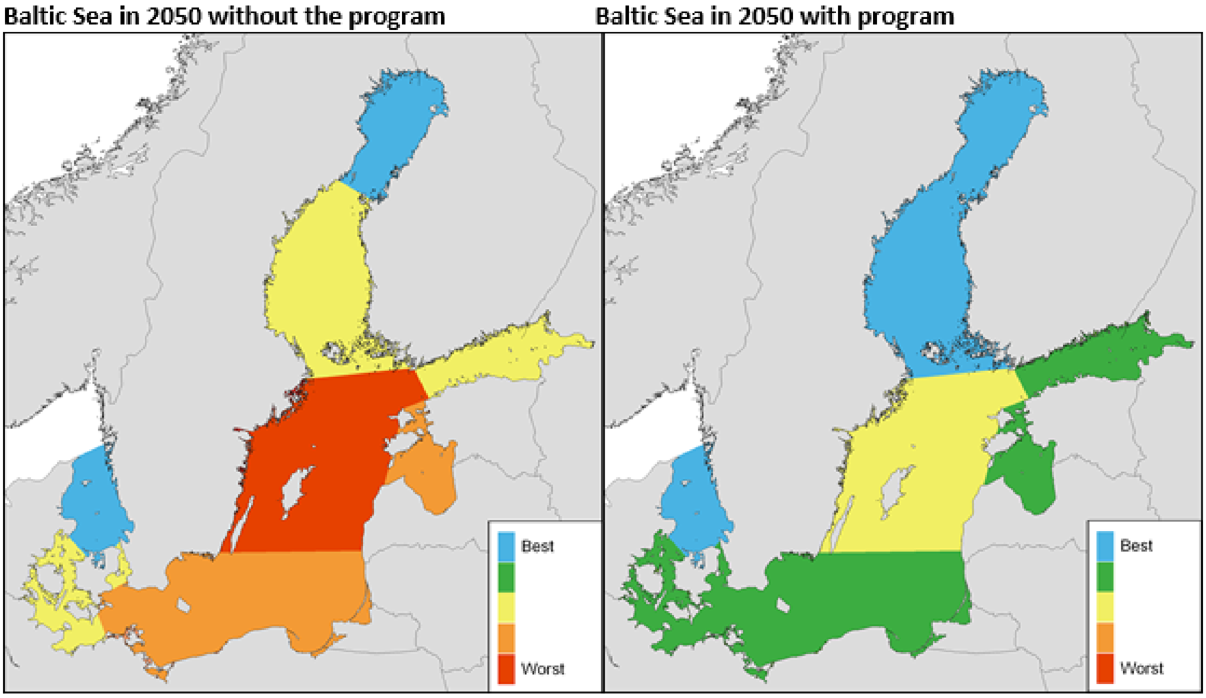
In this case, the policy considered for implementation was the associated with meeting the nutrient load reduction targets defined in the HELCOM's Baltic Sea Action Plan (HELCOM, 2013). The resulting environmental quality improvement was assessed against the eutrophication level predicted for 2050 in the case no new investments in nutrient abatement measures were made. The effects of eutrophication in the improvement scenario and in the no-improvement baseline were assessed using state-of-the-art marine models for the Baltic Sea (Kiirikki et al., 2001; Kiirikki et al., 2006; Maar et al., 2011; Ahlvik et al., 2014) and a professional evaluation by marine ecologists.

The concept of eutrophication was introduced in the survey by linking it to five ecosystem effects: water clarity, blue-green algal blooms, a condition of underwater meadows, a composition of fish species and oxygen content in deep-sea bottoms.[2] Each effect was described on a five-step colored water quality scale, in which colors depicted different levels of the effect intensity and were labelled from "best

[2] Particular emphasis was placed on designing the questionnaire equally relevant and accurate in each coastal country, both in informing about the eutrophication effects and in describing a scenario of the environmental improvement. Pre-testing included five expert reviews, three focus groups, sixteen cognitive interviews in different countries and pilot surveys in all nine littoral countries. This helped develop an identical survey instrument for every country, which was translated into national languages.

possible water quality" to "worst possible water quality".[3] After teaching respondents about the meaning and construction of the scale, the improvement scenarios were shown in a form of color-coded maps that illustrated eutrophication levels in all sub-basins of the Baltic Sea in 2050. The visual representation of the scenario (reproduced in Figure 1) was supported with a verbal description.

Figure 1. A map illustrating changes in the open-sea water eutrophication as a result of implementing the nutrient loadings reduction program; the colors follow a detailed scale defined in the survey describing effects for water clarity, blue-green algal blooms, a condition of underwater meadows, a composition of fish species, and oxygen content in deep-sea bottoms.



After being introduced to the policy for eutrophication reduction the respondents were asked whether in principle they would be willing to pay anything at all for eutrophication reduction in the Baltic Sea. This type of question is typically referred to as an "in-the-market" question, because it shows whether a respondent is interested at all in having the good provided, and thus should be a part of continuous distribution of WTP in the population or represents a share of people with zero WTP, who constitute a jump density of the WTP distribution at zero.

Respondents were asked to indicate the maximum amount they would be willing to pay for the improvement using a payment card – a list of various WTP amounts. The exact wording of the WTP 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 described in the survey was a tax which each individual and each firm in the Baltic Sea countries would need to pay annually upon implementation of the environmental improvement program.[4] The description of the payment

[3] The scale with precise descriptions, as presented to respondents, is available in the Online Supplement A to this paper (for now provided at the end of the manuscript) and in an appendix to Ahtiainen et al. (2014).
[4] 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 WTP for the program if they

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 that the tax vehicle was perceived both credible and acceptable by the interviewed populations.

The payment cards were designed analogically in every country, based on responses from the pilot studies. Each card included 18 positive bids, a zero bid and a "don't know" option.[5] 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 was truncated, as this has been evidenced to possibly affect the welfare estimates (Rowe, Schulze and Breffle, 1996; Roach, Boyle and Welsh, 2002).

The survey informed the respondents that by answering the questionnaire, they could affect the environmental policy projects related to controlling eutrophication levels 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. (2012) and Ahtiainen et al. (2014).

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 modes were employed in Poland. The choice of a mode in every country was guided mainly by a consideration of a share of people in a given country with access to the internet and a consideration of costs of the survey administration.

3. Web vs. personal interviews for stated preference surveys

Stated preference surveys are administered by various modes, which include mail, phone, web and personal (face-to-face) interviews. The prevailing view in the literature is that as long as the samples surveyed via different modes are equivalent with respect to relevant characteristics, a choice of a data collection mode does not affect the survey results significantly. Lindhjem and Navrud (2011) reviewed 17 stated preference studies which had compared web and other-mode surveys in the context of environmental goods and environment-related health risks. They concluded that in general, the studies had not evidenced important differences in value estimates derived from data collected via different modes, and that data from web surveys had not been observed to be of lower quality or validity than data from surveys administered with other modes. Menegaki, Olsen and Tsagarakis (2016) identified 41 economic valuation studies conducted from 2001 to 2015 that had examined differences in value estimates from web and other-mode surveys, and found that the majority of them had not confirmed the existence of mode effects. Finally, the contemporary guidance for stated preference studies (Johnston et al., 2017) says that "[r]ecent research suggests that data collection mode does not substantially influence SP [stated preference] study outcomes …", however, the authors add that the results are mixed and specific to a research context. Indeed, a thorough look into studies that inquired this issue reveals that findings on the mode effect are not univocal. A summary provided in Table 1 shows that when the evidence is limited to research that compared web and personal interviews, the

believed that they would need to pay twice for the improvement – through the firms they worked in and individually.

[5] The only exception is the questionnaire used in Russia, which included 14 positive bids.

number of studies reporting a significant mode effect is nearly the same as the number of studies reporting this effect to be insignificant.[6]

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

Author(s)	Topic	Value elicitation format	Difference in value estimates between modes
Balderas Torres et al. (2015)	Carbon offsetting by local forests	Multiple choice sequence (DCE)	Yes (Web < Personal)
Bell, Huber and Viscusi (2011)	Water quality in rivers, lakes and streams	Binary choice sequence (DCE)	Yes (Web < Personal)
Canavari, Nocella and Scarpa (2005)	Pesticide ban; Organic apples	Yes-no question and open-ended question (CV); Open-ended question (CV)	No Yes (Web > Personal)
Cardamone, Eboli and Mazzulla (2014)	Risk of road accidents	Ranking task (DCE)	No
Covey et al. (2010)	Prevention of railway fatalities	Ranking task (DCE)	No
Lee, Kim and Mjelde (2016)	Nature preservation	Yes-no question (CV)	Yes (Web < Personal)
Lindhjem and Navrud (2011)	Biodiversity protection	Payment card question (CV)	No
Marta-Pedroso, Freitas and Domingos (2007)	Landscape preservation	Open-ended questions (CV)	Yes (Web < Personal)
Mjelde, Kim and Lee (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)	Yes (Web < Personal) and No

Notes: Abbreviations CV and DCE are used to refer to the common nomenclature in the stated preference literature: CV stands for contingent valuation and DCE stands for a discrete choice experiment. Notation

[6] Table 1 does not list studies, which involved web and personal data collection modes but: (1) did not inquire differences between the two modes (e.g., Hamzaoui-Essoussi and Zahaf, 2012; Reichl, Schmidthaler and Schneider, 2013; Ahtiainen et al., 2014); (2) employed inequivalent value elicitation formats in different modes (e.g., Ready et al., 2006; Goethals, Leclercq-Vandelannoitte and Tütüncü, 2012; Sandorf, Aanesen and Navrud, 2016); (3) did not involve valuation question (e.g., Goldenbeld and de Craen, 2013); or (4) evaluated different goods in different modes (e.g., Maier, Wilken and Dost, 2015).

"Web < Personal" implies that the value estimate from a web survey was statistically significantly lower than its equivalent from a personal survey. "Web > Personal" means the opposite.

Personal interviews have long been acknowledged as a best practice in stated preference research (Mitchell and Carson, 1989; Arrow et al., 1993). The NOAA panel (Arrow et al., 1993), which set recommendations for stated preference studies, reasoned that the in-person mode helped respondents understand complex information, for example, through providing pictures and other visual material, and, hence, the mode fostered collecting data of high quality (that is, data that accurately reflects respondents' preferences). More recent guidelines for stated preference research (Johnston et al., 2017) also emphasize advantages of using personal interviews, but they point to high cost of employing this mode. Growth of the use of internet has allowed researchers to administer surveys in a cheaper and faster way, at the same time retaining the possibility of presenting visual material. With the 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, Olsen and Tsagarakis, 2016). Therefore, an essential question is whether, and if so, to what extent, a choice of a data collection mode impinges on survey outcomes.

Discrepancies in value estimates derived from web and personal modes may arise from differences in the populations that are being reached due to internet access penetration or self-selection bias (Fricker and Schonlau, 2002; Stephenson and Crête, 2011). In other words, a sample of respondents to a web survey is likely to differ from a sample of respondents to a personal survey. These factors can undermine the representativeness of a web sample, and hence, they may influence the extent to which web-elicited preferences reflect preferences of the population of interest. In addition, a mode itself can alter respondents' stated answers to a survey. This is sometimes referred to as a "pure" mode effect (Jäckle, Roberts and Lynn, 2010). The "pure" mode effect can be attributed to normative/sociological factors or to cognitive/psychological factors (Dillman, Smyth and Christian, 2014). The former involve the influence of social norms on respondents' behavior, and this influence may differ between modes. In particular, the presence of an interviewer in personal surveys is likely to affect respondents' perceptions of (and adherence to) social norms. In this regard, a widely recognized source of the mode effect is social desirability, which means that respondents answer a survey in a way they think they ought to answer because of some social considerations.[7] The cognitive/psychological factors pertain to information processing by respondents. For example, a mode effect in this regard can emerge as a result of satisficing behavior (Manski, 2017), which means that respondents make shortcuts and choose a satisfactory answer instead of their best answer.

Overall, as illustrated by the summary presented in Table 1, empirical evidence on the existence of a difference in value estimates derived from web and personal surveys is mixed. Out of the 13 listed studies, 7 reported a significant mode effect. Findings on the sign of this difference are not consistent either, although a majority of the studies observed that web-based data generated lower value estimates than in-person-based data. The observation that many studies found a significant mode effect diverges from the commonly held view that a choice of a data collection mode does not impinge on valuation results. In the face of the inconsistent evidence, in what follows we provide an additional field study verification or whether web and personal surveys lead to equivalent value estimates.

[7] Indeed, some studies observed a stronger social desirability bias in personal interviews than in web surveys (Lee, Kim and Mjelde, 2016; Mjelde, Kim and Lee, 2016).

In our case, the survey was administered through different modes (web and personal) in different countries. In Poland, the valuation survey was administered by two modes: CAWI and CAPI. Based on this data, we verify whether the value estimates differ between the two survey modes. We control for possible differences in socio-demographic characteristics between the mode samples (and to make the samples reflect the general population with respect to these characteristics) by using weighted maximum likelihood method for estimation. In what follows, we find that on average CAWI respondents are willing to pay significantly more than CAPI respondents. We use the estimated difference to control for the survey mode effect on the WTP estimates in other countries, thus improving the estimates reported by Ahtiainen et al. (2014).

4. Econometric approach to modelling payment card data

A payment card is a collection of bids. Respondents are asked to select the highest of the provided bids they would be willing to pay, thus indicating that their true maximum WTP lies in the range between the selected bid (the amount they were willing to pay, lower bound of WTP estimate) and the next bid (the amount they were not willing to pay, upper bound of their true WTP).[8] This information can be used to fit a parametric distribution describing people's WTP. Assuming the WTP distribution is of particular form (e.g., normal) with unknown parameters describing its mean and standard deviation, the probability of observing a particular choice (interval) is equal to the cumulative distribution function (CDF) of this distribution evaluated at the upper bound (i.e., the probability that WTP is lower than the upper bound) less the CDF of this distribution evaluated at the lower bound (i.e., the probability that WTP is lower than the lower bound). The remainder of this subtraction is the probability that true WTP lies between the lower and the upper bound. The parameters of the selected parametric distribution can be found by maximizing the sum of these probability for the observed choices of all respondents.

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

$$P\left(b_{i.s} \leq WTP_i < b_{i,s+1}\right) = CDF\left(b_{i,s+1}, \boldsymbol{\beta}_i\right) - CDF\left(b_{i,s}, \boldsymbol{\beta}_i\right)^{9} \tag{1}$$

where CDF denotes a cumulative distribution function of the considered WTP distribution and $\boldsymbol{\beta}_i$ is a vector of the distribution parameters (for example, for a normal distribution, $\boldsymbol{\beta}_i$ consists of a mean and a standard deviation). By making $\boldsymbol{\beta}_i$ dependent on individual i's characteristics, the parameter vector becomes individual-specific, allowing for identifying variables that influence parameters of the distribution (e.g., survey mode).

The parameters of the distribution can be estimated using the maximum likelihood method. The probability specified in (1) expresses individual i's contribution to the likelihood function, while the log-likelihood function for a sample of N individuals can be formulated as:

---

[8] Payment card responses can therefore be viewed as interval-type data (Cameron and Huppert, 1989).

[9] Actually, $CDF(x)$ is the probability that the random variable will take a value less than or equal to *x*. The difference does not matter, as the absolute likelihood for a continuous random variable to take on any particular value is zero. We operationalize the estimations by using the value of the PDF evaluated at the lower bound instead of the difference of CDFs, whenever the calculated difference in the values of the CDF is numerically equal to zero.

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

The above are conditional on selecting a parametric distribution, whose CDFs are calculated. However, a researcher does not usually know what parametric distribution is best for approximating the distribution of WTP in the population. Instead, it is recommended that many parametric distributions are tried to select the one that fits data best. Because the distributions can vary with respect to the number of parameters, one can use the Akaike information criterion (AIC) or the Bayesian information criterion (BIC), rather than simply the value of the log-likelihood function as a means for comparisons.

Additionally, it is usually found that there is a large share of respondents whose WTP is equal to zero. This can be represented by a jump discontinuity in a probability density function (PDF) of any parametric distribution and is typically called a spike (Kriström, 1997) or a zero-inflated model (Greene, 2011). In the zero-inflated model, respondents' WTP is modelled as a mixture of a Bernoulli distribution (a point mass at zero) and a given parametric distribution, allowing for over-proportional share of zero responses (Gurmu and Trivedi, 1996).

Finally, for datasets in which it is known with certainty that some respondents' WTP is equal to zero (for example, because they declared that they are not "in the market") their lower and upper bounds can both be equal to zero.[10] As a result, combining (2) with the information about individual i's market participation captured by a binary-coded variable yes_i equal to one if individual i is "in the market" and zero otherwise, the log-likelihood function for observing a particular set of choices in the sample is given by:

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

Maximization of (3) with respect to $\boldsymbol{\beta}_i$ generates estimates of the parameters of the assumed WTP distribution.[11]

5. Results

Using the same data as Ahtiainen et al. (2014), we now turn to estimating the country-specific WTP for the improvements in the water quality in open sea areas associated with implementing the Baltic Sea Action Plan. We improve the existing estimates by controlling for the differences in survey administration mode between countries and by applying a more flexible approach to modelling the distribution of WTP in the population – we try 16 parametric distributions[12] augmented with the zero-inflation component to find the one that fits our data best.

[10] Similarly, for respondents whose true WTP is known precisely (for example, because they truthfully stated their WTP in an open-ended question) their lower and upper bound can be equal to the known value.

[11] The models presented here were estimated using a custom code developed in Matlab, available at [removed for the review] under CC BY 4.0 license. The code and data for estimating the models presented in this paper are available from [removed for the review].

[12] Online Supplement C (for now provided at the end of the manuscript) lists all distributions which we consider. We have tried several other distributions, including Nakagami, Poisson and Weibull distribution, however, in the case of this dataset (which incorporated a relatively wide range of bids) they resulted in numerical problems (such as parameters reaching bounds of the parameter space) and hence were excluded for further analysis.

5.1. The survey mode effect

Using the data from Poland, where both administration modes were used, we investigate whether there are significant differences in WTP between samples. To make the comparison fair, we start with testing if the samples differ with respect to observable socio-demographic characteristics. We found statistically significant differences with respect to respondents' income and education: the CAWI respondents had higher incomes and higher education than the CAPI respondents. The differences in other characteristics were less sound, although we observed that there were relatively more CAWI respondents, who were retired, home-employed, students, and with more household members under 18 years old.[13] Having found significant differences between survey modes, in what follows we use weighted maximum likelihood estimation framework to control for observable differences. We do this not only for the observations from Poland, where two survey modes were used, but also for all the other countries, to correct for the differences in socio-demographic characteristics between the sample and the general population of the country.

We compared the performance of 16 parametric distributions, in terms of the fit to payment card data collected in Poland. Each model was estimated using pooled data from CAWI and CAPI samples, and included a binary-coded variable that controlled for the effect of the CAWI mode, which was interacted with every distribution parameter in a model. We found that the Birnbaum-Saunders distribution augmented with the zero-inflation component fits best, closely followed by the inverse Gaussian distribution.[14]

The results of the fitted Birnbaum-Saunders distribution are shown in Table 2. The Birnbaum-Saunders distribution is characterized by two parameters: shape and scale. We consider two specifications, (A) and (B), which differ in that specification (B) includes a binary-coded explanatory variable associated with the CAWI mode, while specification (A) does not. The binary variable representing the CAWI mode turns out to be a significant explanatory variable of the shape and scale parameters, as well as the zero-inflation constant. In other words, not accounting for survey mode leads to significant decrease in the model fit (LR-test statistic 191.394; p-value < 0.0001).

These results translate to differences in the mean WTP for the two samples. Using the estimated parameters of the fitted distribution, we simulate mean WTP along with the 95% confidence interval and the probability of WTP equal to zero.[15] The results are provided in the lower panel of Table 2.

[13] A detailed comparison of the socio-demographic characteristics of the two samples and the general population along with the results of the statistical equality tests are provided in the Online Supplement B (for now provided at the end of the manuscript).

[14] The estimation results for Poland and other countries are summarized in the Online Supplement C (for now provided at the end of the manuscript).

[15] Using the estimated parameters, we apply the bootstrapping technique proposed by Krinsky and Robb (1986) to simulate the descriptive statistics of the WTP distribution.

Table 2. Estimation results of the annual WTP of Polish citizens for the Baltic Sea eutrophication reduction (the Birnbaum-Saunders distribution of WTP)

	(A)	(B)	
	Parameter estimates	Parameter estimates	Parameter estimates for CAWI (interaction)
Shape parameter	12.992*** (0.396)	10.508*** (0.551)	4.319*** (0.773)
Scale parameter	1.118*** (0.026)	1.162*** (0.044)	-0.106*** (0.054)
Zero-inflation constant	-0.100*** (0.029)	0.269*** (0.042)	-0.742*** (0.060)
Model characteristics			
AIC/*n*	4.379	4.279	
BIC/*n*	4.388	4.296	
Log-likelihood	-4,049.478	-3,953.781	
n (observations)	1,851	1,851	
k (parameters)	3	6	
Simulation results for the fitted distribution			
	Pooled (CAWI and CAPI)	CAPI only	CAWI only
Annual mean WTP per person (in EUR)	11.916 (0.710)	6.438 (0.542)	16.100 (0.937)
95% confidence interval for the mean WTP	10.656-13.429	5.502-7.612	14.275-17.918
Share of WTP = 0	0.460 (0.012)	0.605 (0.016)	0.320 (0.016)

Notes: Standard errors are given in brackets. *** indicates a 1% significance level. n denotes the number of observations. k denotes the number of a model's parameters.

We find that on average CAWI respondents are willing to pay significantly more than CAPI respondents. In what follows, we use the estimated difference to control for the survey mode effect on the WTP estimates in other countries, thus improving the estimates reported by Ahtiainen et al. (2014).

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

In the face of our findings evidencing a significant mode effect, we recalculate the values of the Baltic Sea eutrophication reduction provided by Ahtiainen et al. (2014). They estimated the value of this environmental improvement for every country with access to the Baltic Sea, however, they did not take into account the influence of a data collection mode on the WTP amounts stated by the respondents. With the exception of Poland, where both modes were used, the data was collected through a single mode within a country: either through CAWI or through CAPI. We show how the value estimates of the marine eutrophication reduction could vary for each Baltic Sea country if the mode of data collection was different than the one actually employed.

The summary of the results is presented in Table 3. The table contains the estimates of the annual mean WTP in EUR per person for the marine eutrophication reduction in every Baltic Sea country.[16] We report

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

the values for both modes: CAWI and CAPI. The values of the considered environmental improvement for the mode actually used are obtained from the Krinsky-Robb simulation for a fitted distribution on 1,000 random draws, and the fitting of a parametric distribution to data was based on the econometric approach as outlined in Section 4. The values for the other mode, that is, for the mode which was not actually used, are derived from calibration of the value estimate obtained for a given country by the relative difference in the mean WTP between CAWI and CAPI for Poland. For example, for the countries where the data was collected through CAWI, we calibrate the mean WTP value derived from the CAWI responses to learn what the value would be if the data had been collected through CAPI. The calibrated values are presented in italics in Table 3.

Table 3 provides simulation results for three specifications: one assuming the inverse Gaussian distribution of WTP, one assuming the Birnbaum-Saunders distribution of WTP and another assuming for each country such a distribution of WTP that gives the lowest value of AIC. The choice of these specifications is based on the following. The inverse Gaussian distribution fits the data best in terms of average values of AIC and BIC for all countries. The Birnbaum-Saunders distribution fits the data best in terms of the sum of log-likelihood values for all countries. The distributions best matching the WTP data for each country in terms of AIC are: Birnbaum-Saunders for Denmark; inverse Gaussian for Estonia, Germany, Lithuania and Russia; log-normal for Finland and Sweden; and generalised Pareto for Latvia. A comparison of different parametric distributions fitted to the WTP data for every Baltic Sea country is presented in the Online Appendix C. All estimations are conducted on weighted samples: the samples for each Baltic Sea country are weighted with respect to the shares of females, unemployed people and people with higher education, so that these shares in the study samples represent the actual shares in the respective country populations. The socio-demographic characteristics of every country population are taken from Ahtiainen et al. (2014).[17] Important to mention, Ahtiainen et al. used unweighted samples for calculating the value estimates.

The consideration of several model specifications allows us to verify robustness of our results. The value estimates shown in Table 3 are very similar across the three specifications taken into account. Overlapping 95% confidence intervals for the mean WTP estimates indicate that there are no statistically significant differences. This evidences that our findings are not driven by a choice of a WTP distribution.

Table 3 also lists the value estimates obtained by Ahtiainen et al. (2014). They considered a few modelling approaches, and we refer here to their results from a spike model because of two reasons. First, although Ahtiainen et al. assumed in the spike model a different distribution of WTP (namely, log-normal) than we do, this specification most closely resembles our modelling approach and, therefore, it constitutes the most relevant point for comparison. Second, Ahtiainen et al. argued that the spike model gave the appropriate treatment of respondents with zero WTP and, hence, they used the values from the spike model for estimating the aggregate benefits from the marine eutrophication reduction.

2011, as provided by OECD.Stat (retrieved June 12, 2017 from https://stats.oecd.org/Index.aspx?DataSetCode=SNA_TABLE4).

[17] The socio-demographic statistics reported by Ahtiainen et al. (2014) are retrieved from various sources for different countries, namely they are based on: Statistics Denmark 2011, Statistics Estonia 2011, Statistics Finland 2010, Statistisches Bundesamt 2010 (Germany), Population Census 2011 (Latvia), Statistics Lithuania 2011, Rosstat 2010 (Russia) and Statistics Sweden 2010.

Table 3. Annual mean WTP (in EUR) per person for the marine eutrophication reduction in every Baltic Sea country corrected for the mode effect

	Results of Ahtiainen et al. (2014) from a spike model		Our results assuming the inverse Gaussian distribution of WTP		Our results assuming the Birnbaum-Saunders distribution of WTP		Our results assuming for each country the distribution of WTP best fitting to data in terms of AIC	
	CAWI	CAPI	CAWI	CAPI	CAWI	CAPI	CAWI	CAPI
Poland	12.2 (0.14) 11.9-12.4	12.2 (0.14) 11.9-12.4	16.4 (1.08) 14.5-18.8	6.4 (0.64) 5.3-7.8	16.1 (0.94) 14.3-17.9	6.4 (0.54) 5.5-7.6	16.1 (0.94) 14.3-17.9	7.2 (0.59) 6.1-8.4
Surveyed through CAWI								
Denmark	31.7 (1.88) 28.1-35.4	---	35.5 (2.85) 30.5-41.5	*13.9* *(1.69)* *11.2-17.2*	36.4 (2.50) 31.8-41.5	*14.5* *(1.44)* 12.2-17.6	36.4 (2.50) 31.8-41.5	*16.3* *(1.57)* 13.6-19.5
Estonia	24.0 (2.29) 19.5-28.5	---	29.5 (3.43) 23.6-37.2	*11.5* *(2.03)* *8.6-15.4*	28.1 (2.55) 23.5-33.2	*11.2* *(1.46)* 9.1-14.1	29.5 (3.43) 23.6-37.2	*13.2* *(2.15)* 10.1-17.5
Finland	41.8 (0.76) 40.33-43.3	---	40.8 (2.78) 35.8-46.6	*15.9* *(1.65)* *13.1-19.3*	41.8 (2.50) 37.2-46.8	*16.62* *(1.44)* 14.3-19.9	42.6 (2.87) 37.6-48.7	*19.1* *(1.80)* 16.0-22.9
Germany	25.0 (0.79) 23.4-26.5	---	26.2 (1.80) 23.0-30.0	*10.2* *(1.07)* *8.4-12.4*	26.7 (1.64) 23.8-30.1	*10.6* *(0.94)* 9.2-12.8	26.2 (1.80) 23.0-30.0	*11.7* *(1.13)* 9.8-14.1
Sweden	75.7 (8.12) 59.8-91.6	---	82.6 (5.80) 72.0-94.0	*32.2* *(3.44)* *26.3-39.0*	84.3 (5.08) 75.1-94.6	*33.5* *(2.92)* 28.9-40.2	80.7 (5.62) 70.7-93.0	*36.1* *(3.53)* 30.2-43.6
Surveyed through CAPI								
Latvia	---	5.5 (0.06) 5.3-5.6	*13.8* *(1.13)* *11.8-16.6*	5.4 (0.67) 4.3-6.9	*13.1* *(0.87)* 11.2-14.8	5.2 (0.50) 4.3-6.3	*12.8* *(1.27)* 10.3-16.2	5.7 (0.80) 4.4-7.6
Lithuania	---	8.8 (0.26) 8.3-9.3	*24.3* *(1.60)* *21.3-27.7*	9.5 (0.95) 7.8-11.5	*24.4* *(1.38)* 21.3-26.9	9.7 (0.79) 8.2-11.4	*21.2* *(1.51)* 18.3-24.5	9.5 (0.95) 7.8-11.5
Russia	---	8.5 (0.19) 8.1-8.9	*21.8* *(2.46)* *17.0-28.4*	8.5 (1.46) 6.2-11.8	*20.9* *(1.58)* 17.4-24.3	8.3 (0.91) 6.7-10.3	*19.0* *(2.33)* 14.5-25.1	8.5 (1.46) 6.2-11.8

Notes: The numbers in each cell are, respectively, a mean WTP estimate, a standard error of this estimate (in a bracket) and a 95% confidence interval for the mean WTP estimate. The numbers in italics are calibrated WTP values assuming that the other data collection mode would have been used than the one actually implemented.

For the actually employed mode, the annual mean WTP values from our calculations are statistically indistinguishable from those reported by Ahtiainen et al. (2014): in each case, 95% confidence intervals for the mean WTP overlap, implying lack of statistically significant differences. The reason why our results are not identical to those of Ahtiainen et al. lies in at least three differences in modelling: different assumed distributions of WTP (Ahtiainen et al. assumed the log-normal distribution, while we use the distribution that fits best to the data), different explanatory variables used (Ahtiainen et al. included several explanatory variables in the model, while we do not) and differences in observations' weighting (Ahtiainen et al. did not weigh the observations, while we do).

The most important finding from Table 3 is the extent to which the WTP estimates are affected by the data collection mode. Calibration of the mean WTP values based on the relative difference in the value estimates between modes for Poland illustrates how much the value estimates vary across CAWI and CAPI. For none of the Baltic Sea countries, the 95% confidence intervals for the mean WTP estimates overlap for web and personal surveys, which implies that the two modes generate significantly different values. The observed large discrepancies in the value estimates between the modes underline how considerably the mode impinges on the valuation results. Notably, Table 3 displays differences in the mean WTP values, while for policy assessments the aggregate value for the entire population is typically used. Aggregation of the mean WTP value for the whole population will result in even larger discrepancies in the value estimates derived from the two modes. Consequently, the choice of a data collection mode may importantly impinge on evaluation of benefits from a considered policy, which, in turn, may affect the authorities' decision whether the policy will be introduced or not. Given our findings that different modes can result in substantially different value estimates, stated preference researchers should be cautious when choosing a data collection mode and try to select such one that will allow them to collect data appropriately reflecting the benefits to the society from a considered policy.

In addition to significant differences in the value estimates between the two modes, we observe differences between the modes in the shares of respondents who declared that they were not willing to pay any additional cost for the marine eutrophication reduction. Based on the summary presented in Table 4, the shares of respondents not willing to pay anything for the considered improvement are smaller in the countries surveyed through CAWI than in the countries surveyed through CAPI. An analogical difference between the modes is observed for Poland, where both, CAWI and CAPI, were used: 23% of the respondents interviewed on the web declared they did not want to pay anything for the eutrophication reduction, while among the respondents interviewed in-person this percentage was 57%. Results of the econometric analysis reflect the reported shares: the probabilities of spike (that is, the probabilities of a spike discontinuity at a zero value of WTP) are substantially smaller for CAWI samples than for CAPI samples. This outcome is robust to the choice of the assumed WTP distribution, as the results from the three considered model specifications show (see Table 4). The estimation results of Ahtiainen et al. (2014) reveal the same relationship between the spike probabilities and the modes, although the differences in the spike probabilities between CAWI and CAPI data are smaller than those indicated by our results.

Table 4. Shares of respondents not willing to pay for the marine eutrophication reduction and spike probabilities

	Shares of respondents not willing to pay anything	Spike probabilities			
		Ahtiainen et al. (2014)	Inverse Gaussian distribution of WTP	Birnbaum-Saunders distribution of WTP	The best distribution of WTP for each country in terms of AIC
Poland	CAWI: 0.23 CAPI: 0.57	0.47 (0.0001)	CAWI: 0.32 (0.0152); CAPI: 0.61 (0.0163)	CAWI: 0.32 (0.0155), CAPI: 0.61 (0.0158)	CAWI: 0.32 (0.0155), CAPI: 0.61 (0.0158)
Surveyed through CAWI					
Denmark	0.37	0.48 (0.0002)	0.39 (0.0161)	0.39 (0.0164)	0.39 (0.0164)
Estonia	0.34	0.48 (0.0005)	0.38 (0.0229)	0.38 (0.0226)	0.38 (0.0229)
Finland	0.34	0.37 (0.0000)	0.35 (0.1221)	0.35 (0.0124)	0.35 (0.0121)
Germany	0.38	0.46 (0.0001)	0.42 (0.0132)	0.42 (0.0132)	0.42 (0.0134)
Sweden	0.19	0.33 (0.0002)	0.20 (0.0131)	0.20 (0.0131)	0.19 (0.0132)
Surveyed through CAPI					
Latvia	0.47	0.52 (0.0002)	0.50 (0.0195)	0.51 (0.0189)	0.47 (0.0287)
Lithuania	0.48	0.50 (0.0004)	0.48 (0.0209)	0.48 (0.0206)	0.48 (0.0209)
Russia	0.64	0.69 (0.0001)	0.67 (0.0126)	0.67 (0.0135)	0.67 (0.0126)

Note: Standard errors are given in brackets.

One could argue that the found differences in the value estimates and in the spike probabilities could be attributed to differences between countries, for example, in terms of culture or income, rather than to differences in the influence of the data collection modes. Although we cannot exclude this possibility, we find it unlikely as we do not observe systematic differences in characteristics (such as cultural or income differences) of the two group of countries distinguished by a mode. Estonia constitutes a good example confirming our claim. Respondents in Estonia were surveyed through CAWI, however, the country is culturally and historically closer to the countries surveyed through CAPI (in particular to Latvia and Lithuania) than to other countries surveyed through CAWI. With respect to economic indicators, Estonia also diverges from other countries surveyed through CAWI, as the country is at a lower level of economic development than other web-interviewed countries in our study sample.

6. Discussion and conclusions

Stated preference surveys play an important role in cost-benefit analyses of public policies, as well as in litigation over environmental damages. Given their widespread use for policy and legal purposes, it is crucial that the survey-based value estimates provide valid welfare measures. For long, the recommendations for stated preference research suggested using personal interviews to obtain relevant values of public goods (Arrow et al., 1993). The most recent guidelines also point to advantages of the in-person data collection mode, but at the same time they mention high cost related to the use of this mode (Johnston et al., 2017). Spreading of the internet use has offered researchers a new mode of data collection, which has two unquestionable advantages – it is cheap and fast. As a result, over the last years, the use of surveys administered via the internet has increased considerably. An unsolved question in the stated preference literature is whether, and if so, to what extent, the choice of a data collection mode impinges on the value estimates. We inquire this issue in a field study that evaluates economic benefits in Baltic Sea countries 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 with 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 modes of data collection were employed. Thus, based on the data for Poland, we verify whether the mode affects the value estimates. Our results show that the web respondents are willing to pay on average significantly more for the considered environmental improvement than the respondents interviewed in-person. In the face of 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. Our estimates illustrate a substantial impact that the choice of a data collection mode may have on valuation results.

Our research emphasizes 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 affect the valuation results significantly (Johnston et al., 2017), our empirical results evidence that using different modes can lead to considerably different outcomes. This finding is particularly important in the light of using the survey-based value assessments for policy purposes. From our analysis, it follows 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 & Software*, 55(0):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. C. R., 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.

Ahtiainen, H., Hasselström, L., Artell, J., Angeli, D., Czajkowski, M., Meyerhoff, J., Alemu, M., Dahlbo, K., Fleming-Lehtinen, V., Hasler, B., Hyytiäinen, K., Karlõseva, A., Khaleeva, Y., Maar, M., Martinsen, L., Nõmmann, T., Oskolokaite, I., Rastrigina, O., Pakalniete, K., Semeniena, D., Smart, J., and Söderqvist, T., 2012. Benefits of meeting the Baltic Sea nutrient reduction targets - Combining ecological modelling and contingent valuation in the nine littoral states. MTT Discussion Papers 1/2012. Available at: http://www.mtt.fi/dp/DP2012_1.pdf.

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., 2015. 'Yes-in-my-backyard': Spatial differences in the valuation of forest services and local co-benefits for carbon markets in México. *Ecological Economics*, 109(0):130-141.

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.

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. *Journal of Food Products Marketing*, 11(3):107-134.

Cardamone, A., Eboli, L., and 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.

COWI, 2007. Economic analysis of BSAP with focus on eutrophication - Interim Report. Report 2(1) to HELCOM, COWI Group Consulting.

Czajkowski, M., Ahtiainen, H., Artell, J., Budziński, W., Hasler, B., Hasselström, L., Meyerhoff, J., Nõmmann, T., Semeniene, D., and Söderqvist, T., 2015. Valuing the commons: An international study on the recreational benefits of the Baltic Sea. *Journal of Environmental Management*, 156:209-217.

Czajkowski, M., Andersen, H. E., Blicher-Mathiasen, G., Elofsson, K., Hagemejer, J., Hasler, B., Humborg, C., Smart, J., Smedberg, E., Stålnacke, P., Thodsen, H., Wąs, A., Wilamowski, M., Wulff, F., Żylicz, T., and Hanley, N., forthcoming-a. Improving the cost-effectiveness of water quality improvements through spatial scale changes to target-setting.

Czajkowski, M., Zandersen, M., Angelidis, I., Aslam, U., Becker, T., Bertram, C., Budziński, W., and Zagórska, K., forthcoming-b. Spatially-explicit model of the Baltic Sea-based recreation demand – identifying recreational hotspots and the influence of environmental conditions.

Dillman, D. A., Smyth, J. D., and Christian, L. M., 2014. Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method. 4 Ed., Wiley.

Fricker, R. D., and Schonlau, M., 2002. Advantages and Disadvantages of Internet Research Surveys: Evidence from the Literature. *Field Methods*, 14(4):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 web-panel and face-to-face; Dutch results of SARTRE-4 survey. *Journal of Safety Research*, 46(Supplement C):13-20.

Greene, W. H., 2011. Econometric Analysis. 7 Ed., Prentice Hall, Upper Saddle River, NJ.

Gren, I.-M., 2008. Costs and benefits from nutrient reductions to the Baltic Sea. Swedish Environmental Protection Agency Report no. 5877, Stockholm, Sweden.

Gren, I.-M., Jannke, P., and Elofsson, K., 1997. Cost-Effective Nutrient Reductions to the Baltic Sea. *Environmental and Resource Economics*, 10(4):341-362.

Gurmu, S., and Trivedi, P. K., 1996. Excess Zeros in Count Models for Recreational Trips. *Journal of Business & Economic Statistics*, 14(4):469-477.

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

Hanley, N., and Czajkowski, M., 2017. Stated Preference valuation methods: an evolving tool for understanding choices and informing policy. University of Warsaw, Department of Economics Working Paper 1(230).

HELCOM, 2013. Taking Further Action to Implement the Baltic Sea Action Plan - Reaching Good Environmental Status for a healthy Baltic Sea. HELCOM Copenhagen Ministerial Declaration, Copenhagen, Denmark.

HELCOM, 2014. Eutrophication status of the Baltic Sea 2007-2011 - A concise thematic assessment. *Baltic Sea Environment Proceedings*, 143.

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. 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.

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(1–2):55-66.

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., 2011. 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.

Manski, C. F., 2017. Optimize, satisfice, or choose without deliberation? A simple minimax-regret assessment. *Theory and Decision*, 83(2):155-173.

Markowska, A., and Żylicz, T., 1999. Costing an International Public Good: The Case of the Baltic Sea. *Ecological Economics*, 30:301-316.

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(Supplement C):18-50.

Mitchell, R. C., and Carson, R. T., 1989. Using Surveys to Value Public Goods: the Contingent Valuation Method. Resources for the Future, Washington, DC.

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.

Ollikainen, M., and Honkatukia, J., 2001. Towards Efficient Pollution Control in the Baltic Sea: An Anatomy of Current Failure with Suggestions for Change. *AMBIO: A Journal of the Human Environment*, 30(4):245-253.

Pakalniete, K., Aigars, J., Czajkowski, M., Strake, S., Zawojska, E., and Hanley, N., 2017. Understanding the distribution of economic benefits from improving coastal and marine ecosystems. *Science of The Total Environment*, 584-585:29-40.

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(Supplement C):256-261.

Reusch, T., Dierking, J., Andersson, H., Bonsdorff, E., Carstensen, J., Casini, M., Czajkowski, M., Hasler, B., Hinsby, K., Hyytiäinen, K., Johannesson, K., Jomaa, S., Jormalainen, V., Kuosa, H., Laikre, L., MacKenzie, B., Margonski, P., Oesterwind, D., Ojaveer, H., Refsgaard, J., Sandström, A., Schwarz, G., Tonderski, K., Winder, M., and Zandersen, M., forthcoming. The Baltic Sea: a time machine for the future coastal ocean.

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(Supplement C):50-61.

Sandström, M., 1996. Recreational Benefits from Improved Water Quality: A Random Utility Model of Swedish Seaside Recreation. Working Paper No. 121, Stockholm School of Economics, The Economic Research Institute.

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.

Schou, J. S., Neye, S. T., Lundhede, T., Martinsen, L., and Hasler, B., 2006. Modelling Cost-efficient reductions of nutrient loads to the Baltic Sea. Model specification, Data and Cost functions. NERI Technical Report no. 592, Copenhagen, Denmark.

Söderqvist, T., and Scharin, H., 2000. The regional willingness to pay for a reduced eutrophication in the Stockholm archipelago. The Beijer Institute, Beijer Discussion Paper Series No. 128.

Soutukorva, Å., 2005. The value of improved water quality. A Random Utility Model of Recreation on the Stockholm Archipelago. Beijer International Institute of Ecological Economics, The Royal Swedish Academy of Sciences.

Stephenson, L. B., and Crête, J., 2011. Studying Political Behavior: A Comparison of Internet and Telephone Surveys. *International Journal of Public Opinion Research*, 23(1):24-55.

Tuhkanen, H., Piirsalu, E., Nõmmann, T., Karlõševa, A., Nõmmann, S., Czajkowski, M., and Hanley, N., 2016. Valuing the benefits of improved marine environmental quality under multiple stressors. *Science of The Total Environment*, 551-552:367-375.

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

Vesterinen, J., Pouta, E., Huhtala, A., and Neuvonen, M., 2010. Impacts of changes in water quality on recreation behavior and benefits in Finland. *Journal of Environmental Management*, 91(4):984-994.

Wulff, F., Humborg, C., Andersen, H. E., Blicher-Mathiesen, G., Czajkowski, M., Elofsson, K., Fonnesbech-Wulff, A., Hasler, B., Hong, B., Jansons, V., Mörth, C.-M., Smart, J. C. R., Smedberg, E., Stålnacke, P., Swaney, D. P., Thodsen, H., Was, A., and Żylicz, T., 2014. Reduction of Baltic Sea Nutrient Inputs and Allocation of Abatement Costs Within the Baltic Sea Catchment. *AMBIO: A Journal of the Human Environment*, 43(1):11-25.

Online supplement A. Description of eutrophication-related water quality levels used in the survey

The effects of eutrophication on water quality in open sea areas

Marine scientists have prepared a colour scale to show how serious eutrophication is in the different parts of open Baltic Sea. Before answering to the following questions, we would like you to familiarise with the colour scale below.

Water quality	Description of the effects of eutrophication					Water quality
	Water clarity	*Blue-green algal blooms*	*Underwater meadows*	*Fish species*	*Deep sea bottoms*	
Best possible water quality	Clear	Seldom	Excellent condition Good for fish spawning and feeding	Cod, herring and perch common	No oxygen deficiency Bottom animals common	**Best possible water quality**
	Mainly clear	Sometimes	Patchy vegetation Good for fish spawning and feeding	Cod, herring and perch common	Oxygen deficiency in large areas Bottom animals common	
	Slightly turbid	In most summers	Cover a small area Less good for fish spawning	Fewer cod, but herring and perch common More roach, carp and bream	Oxygen shortages often in large areas Some bottom animals rare	
	Turbid	Every summer	Cover a small area Bad for fish spawning	Fewer cod, herring and perch More roach, carp and bream	Oxygen shortages often in large areas Some bottom animal groups have disappeared	
Worst possible water quality	Very turbid	On large areas every summer	Almost gone Not suitable for fish spawning	Almost no cod, fewer herring and perch Lots of roach, carp and bream	Oxygen shortages always in large areas No bottom animals in many areas	**Worst possible water quality**

Online supplement B. Comparison of socio-demographic characteristics of the sample collected using CAWI, CAPI, and the general population of Poland.

Table B1. Socio-demographic characteristics of the study samples in Poland and of the Polish general population (the numbers represent the shares in percentage points)

Characteristic	CAWI sample	CAPI sample	General population of Poland
Individual income [a]			
Below 787 EUR	40.5	65.5	60
787 - 983 EUR	21.8	20	20
984 - 1,377 EUR	20.6	8.7	15
Above 1,377 EUR	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, Home-employed, Student)	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
Number of observations:	**927**	**924**	

Notes: Sources of the statistics for the general population of Poland: [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. Data about individual income in the survey was collected in Polish zloty (PLN) and was converted into euro (EUR) using the PPP corrected exchange rate for 2011 as reported by OECD.Stat: 1 EUR = 2.5406 PLN (retrieved June 12, 2017, from https://stats.oecd.org/Index.aspx?DataSetCode=SNA_TABLE4).

Table B2. 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

Online supplement C. Comparison of different parametric distributions fitted to the payment card data for each Baltic Sea country (continued on the next page)

	Denmark			Estonia			Finland			Germany		
Distribution	AIC/n	BIC/n	LL	AIC/n	BIC/n	LL	AIC/n	BIC/n	LL	AIC/n	BIC/n	LL
Normal	5.358	5.374	-2375.808	5.805	5.833	-1291.599	5.479	5.489	-4275.834	5.242	5.253	-3645.391
Logistic	5.168	5.184	-2291.475	5.543	5.571	-1233.110	5.121	5.131	-3996.387	4.979	4.991	-3462.497
Exponential	4.694	4.705	-2082.092	4.877	4.895	-1085.456	4.637	4.644	-3619.265	4.510	4.517	-3136.645
Log-normal	4.654	4.671	-2063.509	4.759	4.786	-1058.165	**4.574**	**4.584**	**-3568.889**	4.431	4.442	-3081.015
Log-logistic	4.686	4.702	-2077.395	4.765	4.793	-1059.678	4.595	4.605	-3585.647	4.456	4.467	-3098.031
Rayleigh	5.469	5.479	-2426.062	6.087	6.105	-1355.360	5.611	5.617	-4379.800	5.347	5.355	-3719.650
Gamma							4.653	4.663	-3630.595	4.531	4.542	-3150.363
Birnbaum-Saunders	**4.636**	**4.652**	**-2055.258**	4.771	4.798	-1060.822	4.577	4.588	-3571.818	4.425	4.436	-3076.596
Generalised Pareto	4.685	4.707	-2076.305	4.790	4.827	-1064.122	4.616	4.630	-3601.083	4.490	4.505	-3121.276
Inverse Gaussian	4.650	4.666	-2061.671	**4.750**	**4.777**	**-1056.160**	4.587	4.597	-3579.405	**4.416**	**4.427**	**-3070.631**
Extreme value	5.925	5.941	-2627.623	6.324	6.351	-1407.209	6.313	6.323	-4927.492	5.921	5.932	-4117.666
Rician	5.471	5.487	-2426.062	6.091	6.119	-1355.360	5.595	5.605	-4366.603	5.349	5.360	-3719.651
Generalised extreme value	4.693	4.715	-2079.823	4.763	4.800	-1058.202	4.607	4.621	-3593.870	4.446	4.461	-3090.265
Negative binomial	4.735	4.751	-2099.301	4.789	4.817	-1065.016	4.631	4.641	-3613.614	4.554	4.566	-3166.850
t location-scale	5.059	5.081	-2242.375	5.116	5.153	-1136.943	4.916	4.929	-3835.099	4.800	4.815	-3336.423
Uniform	5.960	5.976	-2643.167	6.105	6.132	-1358.351	6.619	6.629	-5166.294	6.072	6.083	-4223.026

	Latvia			Lithuania			Poland			Russia			Sweden		
Distribution	AIC/n	BIC/n	LL	AIC/n	BIC/n	LL	AIC/n	BIC/n	LL	AIC/n	BIC/n	LL	AIC/n	BIC/n	LL
Normal	4.615	4.635	-1545.155	5.030	5.052	-1450.554	4.918	4.936	-4,545.270	3.597	3.609	-2426.961	6.318	6.334	-2875.032
Logistic	4.471	4.491	-1496.842	4.877	4.899	-1406.361	4.752	4.770	-4,391.877	3.432	3.444	-2315.370	5.817	5.833	-2646.705
Exponential	3.917	3.930	-1312.119	4.316	4.331	-1245.252	4.319	4.331	-3,993.485	2.990	2.998	-2017.669	5.120	5.131	-2330.223
Log-normal	3.962	3.982	-1326.159	4.254	4.276	-1226.334	4.292	4.310	-3,966.498	2.886	2.898	-1946.689	**4.988**	**5.004**	**-2268.919**
Log-logistic	3.993	4.013	-1336.469	4.280	4.303	-1233.926	4.318	4.336	-3,990.499	2.911	2.923	-1963.507	5.001	5.016	-2274.768
Rayleigh	4.887	4.901	-1637.630	5.201	5.216	-1501.194	4.998	5.010	-4,621.289	3.978	3.986	-2685.302	6.584	6.595	-2997.107
Gamma							4.363	4.381	-4,031.915				5.137	5.153	-2336.795
Birnbaum-Saunders	3.937	3.957	-1317.880	4.233	4.255	-1220.233	**4.279**	**4.296**	**-3,953.781**	2.853	2.865	-1924.427	5.010	5.025	-2278.815
Generalized Pareto	**3.857**	**3.884**	**-1289.959**	4.279	4.310	-1232.750	4.408	4.432	-4,071.633	2.880	2.896	-1941.648	5.063	5.084	-2302.279
Inverse Gaussian	3.967	3.987	-1327.755	**4.210**	**4.232**	**-1213.590**	4.279	4.297	-3,953.889	**2.851**	**2.862**	**-1922.570**	5.002	5.018	-2275.336
Extreme value	5.058	5.078	-1693.946	5.549	5.572	-1600.685	5.415	5.433	-5,005.707	3.932	3.944	-2653.096	7.310	7.326	-3326.779
Rician	4.890	4.910	-1637.630	5.205	5.228	-1501.194	4.999	5.017	-4,620.599	3.980	3.992	-2685.753	6.586	6.602	-2997.111
Generalized extreme value	3.959	3.985	-1324.101				4.303	4.326	-3,973.994				5.016	5.037	-2280.752
Negative binomial				4.395	4.417	-1267.051	4.292	4.310	-3,966.577				5.133	5.149	-2335.225
t location-scale	4.354	4.381	-1456.644	4.575	4.605	-1318.158	4.599	4.623	-4,248.122	3.179	3.195	-2143.569	5.403	5.424	-2456.826
Uniform	5.402	5.422	-1809.248				5.594	5.612	-5,171.120	3.902	3.914	-2633.032	7.528	7.544	-3426.198

Notes: n denotes the number of observations. LL is the value of the log-likelihood function. The results in bold indicate the parametric distribution best matching the data for a given country with respect to a given measure of fit (AIC/n, BIC/n or LL). Missing results represents identification problems (such as parameters reaching bounds of the parameter space).