DEPARTMENT OF FOOD AND RESOURCE ECONOMICS
UNIVERSITY OF COPENHAGEN

IFRO Working Paper

Assessing the value of surface water and groundwater quality improvements when time lags and outcome uncertainty exist

Results from a choice experiment survey across four different countries

Tobias Holmsgaard Larsen
Thomas Lundhede
Søren Bøye Olsen

2020 / 12

IFRO Working Paper 2020 / 12

Assessing the value of surface water and groundwater quality improvements when time lags and outcome uncertainty exist: Results from a choice experiment survey across four different countries

Authors: Tobias Holmsgaard Larsen, Thomas Lundhede, Søren Bøye Olsen

JEL classification: C83, D60, Q51, Q53

Published November 2020

This report constitutes project deliverable 3.2 in the project "Legacies of Agricultural Pollutants (LEAP): Integrated Assessment of Biophysical and Socioeconomic Controls on Water Quality in Agroecosystems" (https://uwaterloo.ca/legacies-of-agricultural-pollutant/). The authors acknowledge funding from the WaterJPI - WaterWorks2015 ERA-NET CoFund 2016 Joint Call for Transnational Collaborative Research Projects (Innovation Fund Denmark project number 6144-00024B).

See the full series IFRO Working Paper here:
www.ifro.ku.dk/english/publications/ifro_series/working_papers/

Department of Food and Resource Economics (IFRO)
University of Copenhagen
Rolighedsvej 25
DK 1958 Frederiksberg DENMARK
www.ifro.ku.dk/english/

Assessing the value of surface water and groundwater quality improvements when time lags and outcome uncertainty exist

-

Results from a choice experiment survey across four different countries

Tobias Holmsgaard Larsen, Thomas Lundhede, Søren Bøye Olsen

University of Copenhagen

Abstract

This report summarizes the main results from a choice experiment survey addressing peoples' willingness to pay (WTP) for improvements in surface water quality as well as groundwater quality. A particular novel focus is on estimating the extent to which WTP is impacted by the time lags and outcome uncertainties that commonly occur in practice when implementing new policies to improve water quality. The survey is conducted across four different case areas in four different countries, involving responses from more than 3000 respondents. Results generally confirm previous findings that people on average have quite high WTP for improvements in water quality, both in relation to surface water and groundwater. In addition, the results show that the WTPs reduce significantly with increasing time lags and outcome uncertainty in relation to the actual water quality improvements.

JEL classification: C83, D60, Q51, Q53

Keywords: Economic Valuation, Choice Experiment, Water Quality, Outcome Uncertainty, Time Lags

Acknowledgments

This report constitutes project deliverable 3.2 in the project "Legacies of Agricultural Pollutants (LEAP): Integrated Assessment of Biophysical and Socioeconomic Controls on Water Quality in Agroecosystems" (https://uwaterloo.ca/legacies-of-agricultural-pollutant/). The authors are grateful to all the project participants in project LEAP, in particular Maria Cunha, Jerker Jarsjö and Roy Brouwer, for their invaluable help in developing and validating the questionnaires used for data collection in the Portuguese, Swedish and Canadian case areas. Furthermore, Anna Kristina Edenbrandt and Raphael Filippelli are thanked for their help in translating questionnaires. The authors acknowledge funding from the WaterJPI - WaterWorks2015 ERA-NET CoFund 2016 Joint Call for Transnational Collaborative Research Projects (Innovation Fund Denmark project number 6144-00024B).

1. Introduction

Water provides a crucial basis for most biological life, thus providing services of considerable value to people. Yet, human activities have greatly accelerated the nitrogen (N) and phosphorus (P) cycles, with excess N and P leaching into surface water and groundwater, causing eutrophication, aquatic toxicity and drinking water contamination in many areas around the World (Giordano, 2009; Carpenter et al., 2011). Many countries consider this as an important environmental concern, which, among other things, have resulted in adoption of the Water Framework Directive in the European Union in 2000, and the Great Lakes Water Quality Agreement between Canada and USA in 2012.

Several measures are available which can improve the quality of surface water and groundwater (Jacobsen, 2020), and thus benefit nature and society. However, the implementation of such measures is typically costly. The policy and decision makers thus need to consider and compare the societal benefits with the associated costs when deciding which measures to apply. This is of particular interest in relation to the European Water Framework Directive, which states that "disproportional costs" may exempt countries of reaching good ecological status of certain water bodies (Jensen et al., 2013; European Commission, 2009). In order to address this, policy makers ideally need to know the monetary values of the all the ecosystem services provided by the water and potentially affected by changes in water quality.

Due to the non-rival and non-excludable nature of many water-related services, their value to society is not readily available in terms of market prices. Instead, economic valuation methods can provide monetary estimates of these non-market values. Stated preference (SP) methods have been widely applied in the literature to value environmental services, and hence enhanced the basis of policymaking (Johnston et al., 2017b). In this paper, we employ the SP method choice experiment (CE) to estimate the socioeconomic value of surface water and groundwater improvements. The CE method is one of the most popular SP valuation methods, partly due to its ability to estimate marginal values of changes to characteristics of a good or policy (Hanley et al., 2001).

Despite implementation of a range of best management practices, efforts to reduce excess nutrients entering water bodies have often been disappointing, and water quality targets have not been met (Lintern et al., 2020). This can partly be explained by time lags between implementation of measures and resulting water quality changes (Meals et al., 2010; Vero et al., 2018). Such time lags are often caused by legacy nutrient stores that have accumulated in the landscape (Van Meter et al., 2016). Part of the explanation may also be that scientists have imperfect knowledge regarding the exact water quality outcomes that will result from a given policy measure. Both time lags and uncertainty[1] about outcome are likely to affect the benefit people derive from

[1] Here, we use "uncertainty" to refer to situations where the odds of a specific outcome is known, although this has traditionally been termed "risk" (Knight, 1921). As such, we follow the wording of the literature on outcome uncertainty, which also often deal with cases that actually regard outcome risk. In our empirical survey we however use the term "risk".

environmental goods and services and hence socioeconomic welfare. This has previously been treated in the CE literature on time discounting (e.g. Viscusi et al., 2008) and outcome uncertainty (e.g. Glenk and Colombo, 2011). However, none of the existing studies on water quality valuation account simultaneously for both the time lags and the outcome uncertainty that is inherent in water planning and management in practice. If both of these aspects are not taken into account, cost-benefit analyses are at risk of guiding decision makers towards suboptimal policy decisions.

This study contributes to the existing literature of valuation of water quality improvements by incorporating time lags and outcome uncertainty in CEs on surface water and groundwater quality improvements, respectively. Using online questionnaires, the CE survey is conducted in four different case areas; one in each of the countries Canada, Denmark, Portugal and Sweden. This provides a solid basis not only for assessing the value of water quality improvements, but also for exploration of how time lags and outcome uncertainty may affect these values in different countries and settings. Given the limited number of CEs related to groundwater quality (e.g. Hasler et al., 2007; Tentes and Damigos, 2015), the multi-country aspect of this study also offers a unique opportunity to expand the knowledge on the value of different groundwater quality levels in the western world.

Our analysis reveals that both increased time lags and increased outcome uncertainty negatively affect people's willingness-to-pay (WTP) for water quality improvements. It is thus crucial to take these aspects into account when assessing the values to society associated with improvements of water quality. We also find that respondents generally exhibit positive preferences for both surface water and groundwater quality improvements, and hence are willing to pay for such changes. It is found that the size of this WTP differs considerably between the case areas, hence underlining the importance of taking case-specific hydromorphological and population characteristics as well as potential preference heterogeneity into account in environmental valuation studies.

The remainder of this paper is organized as follows. Section 2 reviews the main previous literature on the use of CEs in water quality valuation, and on the inclusion of time aspects and outcome uncertainty within the environmental valuation literature. In section 3, we present our empirical survey in detail. Section 4 contains a theoretical description of the CE analysis method, which we apply to the collected data in section 5. We discuss the results in section 6 and conclude in section 7.

2. Previous literature

One of the earliest applications of CEs in valuation of environmental goods, was used to value the conditions of two rivers, including water quality, in Alberta, USA (Adamowicz et al., 1994). Hanley et al. (2006) were the first to value water quality in terms of ecological status, as specifically embodied within the Water Framework Directive. This was done in CEs concerning two different rivers in the UK. Since then, several CE studies have been conducted to assess the value of surface water quality changes (e.g. Johnston et al., 2017a; Bateman et al., 2011). Similarly, CE studies have been carried out to assess that value of changes in groundwater quality (e.g. Hasler et al., 2007; Tentes and Damigos, 2015). However, groundwater valuation studies have mainly been conducted using the contingent valuation method (see Brouwer and Neverre (2020), for a review and meta-analysis).

The literature on the theory and elicitation of time preferences is vast, particularly in the field of experimental economics (e.g. Harrison et al., 2002; Andersen et al., 2008). Within the valuation literature, time preferences have sometimes been elicited in separate sections of valuation questionnaires (e.g. De Marchi et al., 2016), which makes it hard to link the elicited preferences directly to the valuation exercise. Some studies on environmental valuation include attributes regarding time delays on benefits that are not directly related to the environmental good (e.g. Logar and Brouwer, 2017). Others include time-related attributes in environmental domains not concerned with water quality (e.g. Meerding et al., 2010). Only few studies include a time-related attribute in a CE regarding water quality. In a CE study regarding improvements in water quality Viscusi et al. (2008) include an attribute on the time of achieving water quality improvements. The results reveal that time preferences are consistent with hyperbolic discounting. A CE study on the value of a cleanup in the Minnesota River Basin by Meyer (2013) is less conclusive regarding the exact discounting specification. Yet, it is found that a 5-year delay in basin cleanup leads to a loss of almost half the benefits derived from the cleanup.

The literature on outcome uncertainty is well established within environmental economics. Most classical papers (e.g. Arrow and Fischer, 1974; Henry, 1974) focus on uncertainty in terms of the potential costs and irreversibility of environmental deterioration. However, here we refer to outcome uncertainty in relation to benefits from environmental improvements. Torres et al. (2017) point to the fact that the source of outcome uncertainty may matter in terms of people's preferences. In the present study, the focus is on scientific uncertainty of the models and theories that are used to predict future outcomes of different policies. Several studies have investigated such outcome uncertainty within the CE literature related to valuation of environmental services. These have largely focused on climate change, e.g. in terms of mitigation (Glenk and Colombo, 2011), the effect on the future abundance of birds and bird species (Lundhede et al., 2015; Faccioli et al., 2019), future temperature rises (Akter et al., 2012), and flood risk (Botzen and van den Bergh, 2012). Outcome uncertainty has also been studied in relation to water quality. Roberts et al. (2008) do so with a particular focus on algae blooms and water levels, while Wielgus et al. (2009) focus on the recreational value for anglers and scuba divers. In terms of the source of the outcome uncertainty, the study by Larue et al. (2017) is likely closest related to the present study. Their focus is on valuation of improved water quality as a product of agricultural BMP's in the Chaudière region south of Quebec City, Canada. More specifically a CE is used to investigate rural residents' preferences for decreased phosphorus and coliform levels in the Chaudière and Etchemin watersheds. The CE includes two attributes addressing the probable reductions of phosphorus and coliform bacteria respectively. Each attribute level represents three equally likely uniformly distributed reduction levels. The attribute levels differ in terms of means and spreads of these distributions, hence introducing outcome uncertainty. Outcome uncertainty is also included in several studies on groundwater valuation (Brouwer and Neverre, 2020), yet, to our knowledge, none use a CE to investigate this aspect.

3. Emprirical survey

3.1 Case areas

The present study is part of the LEAP project (https://uwaterloo.ca/legacies-of-agricultural-pollutant/), which investigates the biophysical and socioeconomic aspects of agricultural pollutant legacies within water systems. The project is carried out by partners in Canada, Denmark, Portugal and Sweden. The focus is on four different case areas, located in the partner countries (Table 1). The case areas differ in several ways, but all face the common challenge of maintaining both agricultural productivity and clean water for human and ecosystem health. The particular focus of the LEAP project is on improving the predictive and socioeconomic understanding of the long-term release of legacy nutrient stores that stems from excess nutrient loads (N and P) in agroecosystems. This is done by studying the time lags and uncertainty that are related to reduced N and P pollution from agriculture.

Table 1 Case areas included in the survey

Case area	Agricultural land in watershed	Size of watershed (km^2)	Size of case area water bodies
The Grand River and its main tributaries (Ontario, Canada)	70 %[1]	6,800[1]	300 km (length, Grand River)[1]
Limfjorden (Denmark)	70 %[2]	7,600[2]	1,500 km^2 (area)[2]
The Mondego River and its main tributaries (Portugal)	32 %[3]	6,645[4]	258 km (length, Mondego River)[4]
Lake Mälaren and Hjälmaren (Sweden)	19 %[5]	22,645[6]	1,550 km^2 (area)[6]

Note: Approximate magnitudes based on [1] Grand River Conservation Authority (2014); [2] Miljøministeriet (2011); [3] Teixeira et al. (2014); [4] Agência Portuguesa do Ambiente (2016); [5] SCB (2019); [6] https://viss.lansstyrelsen.se/

3.2 Questionnaire development

The data used in our analysis were collected using a questionnaire-based CE survey. A common template design was used as the basis for development of two questionnaires, one focusing on improvements in surface water quality, and another focusing on improvements in groundwater quality. Four different case area specific versions were made of each of these questionnaires. The case specific versions only varied in terms of the waterbodies that was in focus in the CE, and questionnaires were translated into the relevant language for each case area.

The questionnaires went through a thorough process of development including both qualitative and quantitative pretesting. Well established methods like focus group interviews, cognitive interviews and pilot tests (Johnston et al., 2017b) as well as methods relatively novel to the CE literature like eye tracking (Dudinskaya et al., 2020) and Q-sorting (Armatas et al., 2014; Jensen, 2019) were used over the course of more than a year (Table 2). The testing process ensured that relevant topics were thoroughly covered, and that the respondents understood the content and format. The development process was led by researchers from University of Copenhagen, Denmark. Hence, for practical reasons, some of the initial pretests were only carried out in Denmark. This was deemed suitable as these pretests focused on rather general issues related to survey setup, cognitive load and text-length, rather than case specific aspects. The final focus group interviews and pilot tests were, however, carried out in all case areas, ensuring that any cross-country variation in opinions regarding the close-to-final questionnaire was taken into account. All focus group interviews were carried out by project partners in similar settings, following a detailed interview protocol across the countries. The use of a common questionnaire design meant that some compromises had to made during the process of development; rather than making different questionnaires in which all details were perfectly shaped for each case area, we constructed a more general questionnaire, which was as relevant as possible for all case areas. Due to the extensive development and testing procedure we find that a common basic questionnaire design will not jeopardize the validity of the results but, in turn, it ensure a high degree of cross-case area comparability of data and results as well as provide broad and multifaceted insights about preferences for water quality across waterbodies with different characteristics and locations. The questionnaires used are made up of three distinct parts. The first part consists of behavioral questions concerning the respondents' use of water bodies for recreational purposes (in the surface water questionnaire) and their consumption of bottled water (in the groundwater questionnaire). These are intended to "warm-up" the respondents and make them think about the subject and their preferences. In other words, putting them into the mental frame of surface water and groundwater quality.

Table 2 Questionnaire pretesting process

Time period	Pretest method	Participants from	No. of participants
June-November 2018	Focus group interview 1, including 1st stage of Q-methodology (Development of concourse)	All case areas	5-8 per group/case area
June 2019	Eye-tracking analysis	Denmark	13
June 2019	Expert focus group interview with project partners	All case areas	15
November 2019	2nd stage of Q-methodology (Q-sorting)	All case areas	32 (6-11 per case area)
December 2019	Eye-tracking assisted cued retrospective think-aloud personal interviews	Denmark	18
December 2019	Focus group interview 2	Denmark	8
February 2020	Focus group interview 3	All case areas	6-8 per group/case area
May 2020	Pilot test	All case areas	502 (96-206 per case area)

The second part of the questionnaires is the CE section, beginning with a scenario description. In the surface water quality questionnaire, each case area is divided into two or three separate sub-areas (Figure 1). Most of these vary in terms of current water quality and hydromorphological conditions. Some part of the case areas were already classified as being in good quality, and these were not included in the CE (this was explained to the respondents). The division into sub-areas ensures greater applicability of the results in subsequent analyzes, and a more realistic scenario description. Scenario realism is particularly important in order to avoid respondents making their own assumptions about water quality conditions (Kataria et al., 2012). Based on the pretest process (Table 2), expert consultations, literature reviews, and the overall project focus, a range of attributes were identified to be included in the final CE setup (Table 3).

A water quality ladder is used to define and communicate different surface water quality levels (Figure 2). Each level is represented by a color and described in terms of easily understandable ecosystem services regarding both biological quality elements (abundance of fish and fauna, and suitability for swimming and angling) and physio-chemical elements (water clarity). This is closely related to the setup and definitions suggested by Hime et al. (2009) for use in water quality related SP surveys. The groundwater quality levels are described using a similar water quality ladder with colors to represent each level. A related setup has previously been utilized in a groundwater valuation study by Brouwer et al. (2018). We define the groundwater quality levels in relation to chemical conditions, both in terms of use-values (drinking water) and non-use existence values (pollution of the groundwater).

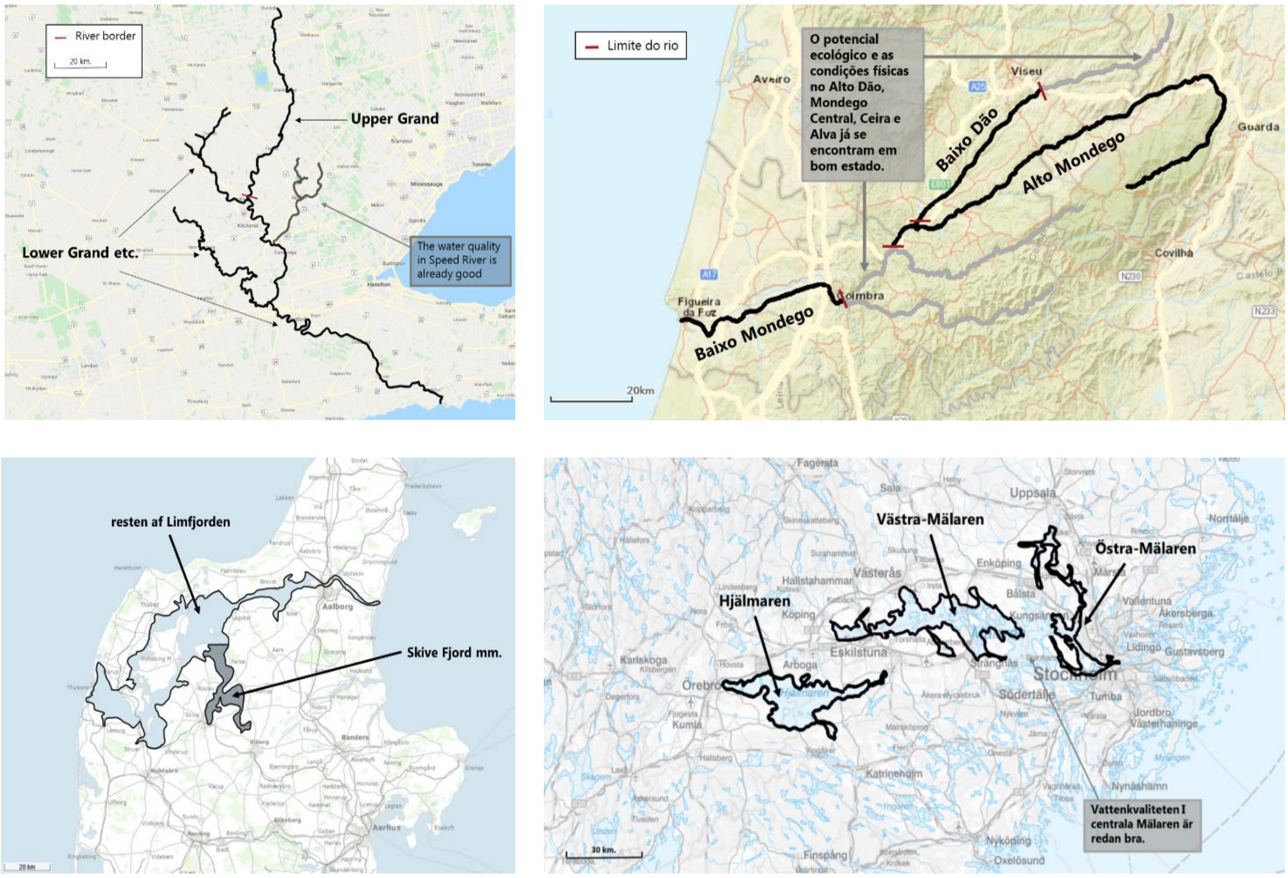

Figure 1 Maps used in the surface water questionnaire to present the sub-areas
Note: Top-left: Canadian case area; Top-right: Portuguese case area; Bottom-left: Danish case area; Bottom-right: Swedish case area.

Table 3 Attribute levels and descriptions presented in the CE

Attribute	Levels	Attribute description shown to respondents
The expected water quality	Poor; Moderate; Good	<u>*Canadian version:*</u> **Three different water quality levels are distinguished**: *Good, Moderate and Poor*. The differences between these levels are described below. The water quality will not affect your drinking water at home or the treatment of this. Each water quality level has a different color. The colors are used on the small map to show how **the water quality is expected to be on average** in Upper Grand and Lower Grand etc. **Without new policy, water quality is expected to be *Poor* in the Lower Grand etc., and *Moderate* in the Upper Grand in 4 years**. However, the expected water quality may improve by implementing new policy measures. <u>*Portuguese version (translated from Portuguese):*</u> In the following, we will distinguish between **three different levels of ecological potential and physical conditions**: *Good, Moderate and Poor*. The differences between these levels are described below. The ecological potential and physical conditions will not affect your drinking water or the treatment of this. The levels are associated with specific colors, which are used on the small map to show how the conditions is expected to be on average in Lower Dão, Upper Mondego and Lower Mondego. **If no new policy is adopted, the ecological potential and physical conditions are expected to be *Poor* in most parts of the Lower Mondego while they are expected to be *Moderate* in Lower Dão and Upper Mondego in 4 years**. However, the expected conditions may be improved by implementing new policy measures.
The risk of water quality not improving	No risk; 10 % risk; 40 % risk	Some measures do not always work as expected. **Some proposals will therefore face a risk that the water quality will not improve,** even though the adopted measures usually works. This risk is based on practical and scientific expert judgement.
The time it takes for water quality to be achieved	<u>*Surface water version:*</u> 4 years; 8 years; 20 years <u>*Groundwater version:*</u> 8 years; 20 years; 50 years	It takes time before the impacts of new measures will take full effect. Scientists can predict the number of years it takes before a new policy leads to a specific water quality. Once this number of years has passed, water quality will stay at the achieved level.
How much your household's annual tax payment increases as a consequence of the proposal	<u>*Canadian version:*</u> \$15; \$30; \$50; \$105; \$210; \$420 <u>*Danish version:*</u> 100 DKK; 200 DKK; 350 DKK; 700 DKK; 1400 DKK; 2800 DKK	Implementation of the proposals for a new policy comes at a cost, which will be covered by a **household tax increase**. Hence, **each policy proposal presented in the following is associated with an increased annual tax payment for your household**. The increase can vary from \$15 to \$420 per year per household. The increase in your household's aggregate annual **tax payment will be implemented in 2020**, even if, as described, there may be uncertainty regarding the expected

	Portuguese version: 7.5 €; 15 €; 25 €; 50 €; 105 €; 210 € *Swedish version:* 115 SEK; 230 SEK; 400 SEK; 800 SEK; 1600 SEK; 3200 SEK	improvements, and it will take several years before they are achieved. The extra tax payment will be the same amount in all future years. If the current policy is continued with no changes, your household's tax payment will not increase. Note that the possible increase in your annual tax payment will be used exclusively for improving water quality in the Grand River. The policy proposals will thus not affect the water quality in other parts of Ontario.

Note: Unless otherwise stated, the descriptions shown here apply to the surface water questionnaire in the Canadian case area, but similar descriptions are used in the other case areas and in the groundwater questionnaire. All versions are available in the study documentation: https://sid.erda.dk/wsgi-bin/ls.py?share_id=bbL1PNQSV5. Short descriptions of each attribute level are included for the water quality and risk attribute (figure 2 and 3 present those for the former), in order to increase credibility and policy consequentiality (Zawojska et al., 2019).
The tax levels were aligned through PPP-adjusted exchange rates.

Water quality	Description of water quality
Good	Good for swimming and angling Many fish, bird and plant species Water visibility is always good, the bottom of the Grand River is visible in most places
Moderate	Acceptable for both swimming and angling The number of fish, bird and plant species is limited Water visibility is limited, water is sometimes turbid and not clear
Poor	In some areas and periods, swimming and angling is not recommended Hardly any or only a few fish, bird and plant species Water visibility is poor, water is always opaque

Figure 2 Water quality ladder included in the Canadian surface water questionnaire

One attribute represents uncertainty regarding the policy outcome, while another represents the time lag between policy implementation and water quality improvements. The levels of the former represent certainty, slight uncertainty and much uncertainty, respectively. The levels of the latter vary over a short, medium and long time range and are chosen to be somewhat it line with actual nutrient time lags (e.g. Meals et al., 2010). The pretesting process revealed that the outcome uncertainty attribute placed a substantial cognitive burden on respondents, which is in line with previous findings (e.g. Logar and Brouwer, 2017). Different versions of the attribute were tested during this process, and a simple text-based design proved most appropriate. Part of the explanation for this seemed to be that the combination of maps and colors used in the water quality attribute already constituted considerable visual information. In order to enable estimation of monetary values, a price attribute is included, which is defined as an annual increase in the tax payment for the respondent’s household. It is not stated how the tax increase would be charged (e.g. through income taxes or municipality taxes), in order to ensure credibility across the different tax systems of the four case areas. The pretesting process showed that tax as a payment vehicle was generally perceived consequential to respondents across the case areas, which is critical in ensuring valid results (Vossler et al., 2012). The price levels were based on previous literature as well as information gathered through the pretesting process.

The attribute definitions and levels in the Portuguese surface water questionnaire differ somewhat from those used in the other case areas. The surface water quality in the natural/unmodified parts of the Mondego River basin is overall good in terms of biological quality elements and physio-chemical elements. However, big parts of the basin is heavily modified in order to regulate the water and ensure flood protection. As a result of a lack of monitoring and maintenance, some of the modified river parts are characterized by issues with both hydromorphological elements (increasing the risk of flooding) and occasionally the biological quality elements (leading to some eutrophication). Hence, in the Portuguese surface water questionnaire, water quality is therefore defined in terms of ecological potential and physical conditions (see figure 3). This puts obvious limits to the comparability of the estimated Portuguese water quality preferences with the estimated water quality preferences in the other case areas.

In the questionnaire version regarding groundwater quality, the setup is overall the same as in the surface water questionnaire. The groundwater quality improvements are described as pertaining to the groundwater in the municipalities where the case areas are located, as well as neighboring municipalities i.e. areas somewhat comparable to the case area watersheds. The groundwater quality was generally considered to differ less within each case area, and the case areas are hence not divided into sub-areas in the groundwater questionnaires. The levels for the time lag attribute differ from those used in the surface water questionnaire to reflect the biophysical fact that time lags between reduced pollutant leaching and resulting improvements in water quality are typically longer for groundwater than for surface water (Meals et al., 2010). Hence, the chosen time lag levels are 8 years, 20 years and 50 years (as opposed to 4 years, 8 years and 20 years in the surface

Level	Description of ecological potential and physical conditions
Good	Dams, dikes, riverbanks, and sedimentation levels are frequently monitored and maintained Minimal risk of algae growth in dams Minimal risk of flooding
Moderate	Dams, dikes, riverbanks, and sedimentation levels are occasionally monitored and maintained Some risk of algae growth in dams Some risk of flooding
Poor	Dams, dikes, riverbanks, and sedimentation levels are rarely monitored and maintained High risk of algae growth in dams High risk of flooding

Figure 3 Ladder with levels of ecological potential and physical conditions included in the Portuguese surface water questionnaire (English translation)

water version). Other than that, the same attributes and levels, as reported in Table 3, are used in the two different versions of the questionnaire.

The choice sets in the groundwater version are made up of the four attributes listed in Table 3. However, as separate average water quality expectations are stated for each of the sub-areas, the surface water versions have five or six attributes depending on the division into sub-areas. Each choice set includes a business-as-usual (BAU) alternative, presented as a certain forecast of the water quality in 4 years' time for the surface water version, and in 8 years' time for the groundwater version. The water quality levels in the BAU scenarios are based on data and reports on the present average quality level, current trends and most likely projections as expected by national water authorities in the partner countries[2]. By presenting these quality levels as future BAU conditions rather than current status quo conditions, we ensure greater scenario-realism in areas where there is wide variation in water quality within each sub-area. During pretesting, it became evident that in these areas, our focus on average water quality levels could otherwise be considered unrealistic. It is assumed (and described to respondents) that there is no risk that new policies will lead to water quality levels lower than those presented in the BAU alternative. Examples of choice sets used in the Canadian surface water and groundwater questionnaires are provided in Figure 4 and 5 (examples from the other case areas are accessible in the study documentation: https://sid.erda.dk/wsgi-bin/ls.py?share_id=bbL1PNQSV5).

[2] Canadian case area: https://www.grandriver.ca/en/our-watershed/resources/Documents/Water-Quality/GRCA_Board_WaterQualityConditions_February-24-2017.pdf
Danish case area: http://miljoegis.mim.dk/cbkort?&profile=vandrammedirektiv2-bek-2019
Portuguese case area: https://sniamb.apambiente.pt/content/planos-de-gestão-de-região-hidrográfica?language=pt-pt
Swedish case area: https://ext-geoportal.lansstyrelsen.se/arcgis/apps/MapSeries/index.html?appid=0d5184a960834906af2e0fc72d8cd99d

Choice situation 1

	Current policy	Proposal 1	Proposal 2
Expected water quality	 *Upper Grand:* **Moderate** *Lower Grand etc.:* **Poor**	 *Upper Grand:* **Good** *Lower Grand etc.:* **Poor**	 *Upper Grand:* **Moderate** *Lower Grand etc.:* **Good**
Risk of water quality not improving	No water quality improvement	**40 % risk** of not improving water quality	**No risk** (Water quality will improve as expected)
Water quality is achieved in	**4 years**	**4 years**	**8 years**
Tax increase for your household	**$0 per year**	**$15 per year**	**$420 per year**

I prefer *(If you find the proposals too expensive relative to the resulting improvements, you should choose the current policy)*

Current policy | Proposal 1 | Proposal 2

Figure 4 Example of a choice set used in the Canadian surface water questionnaire

Choice situation 1

	Current policy	Proposal 1	Proposal 2
Expected water quality	**Poor**	**Moderate**	**Good**
Risk of water quality not improving	No water quality improvement	**40 % risk** of not improving water quality	**No risk** (Water quality will improve as expected)
Water quality is achieved in	**8 years**	**50 years**	**8 years**
Tax increase for your household	**$0 per year**	**$15 per year**	**$105 per year**

I prefer *(If you find the proposals too expensive relative to the resulting improvements, you should choose the current policy)*

Current policy | Proposal 1 | Proposal 2

Figure 5 Example of a choice set used in the Canadian groundwater questionnaire

The attribute levels were assigned to alternatives and paired into choice sets of three alternatives: a zero-priced BAU alternative and two experimentally designed improvement alternatives with an associated tax increase. As a full factorial design comprised up to 594 alternatives, a fractional factorial design consisting of 12 choice sets was identified (Louviere et al., 2000). An efficient design optimized for D-efficiency and assuming a multinomial logit model was developed using the Ngene software version 1.2.1 (ChoiceMetrics, 2018). For the pilot tests, design priors were informed by findings from previous related surveys as well as theoretical expectations and experience gained from the pretesting phase so far. The design was subsequently updated with better informed priors in terms of parameter estimates obtained in multinomial logit models run on the choice data obtained in the pilot tests (about 100 respondents for each case area, see Table 2). Before finalizing the experimental design, as suggested by Huber and Zwerina (1996), a manual swapping procedure was conducted for a few choice sets where an alternative was clearly dominating or where an alternative appeared unrealistic. It was ensured that attribute level balance was maintained during this procedure. For the surface water questionnaire, the experimental design differed across the country-specific versions due to differences in the number of sub-areas and in the BAU conditions. The overall BAU groundwater quality level was assessed to be similar (poor) across the case areas, and hence the same experimental design was used across these. The 12 choice sets were randomized in order to average out any associated ordering effects (Johnston et al., 2017b). In order to minimize the potential presence of hypothetical bias, a cheap talk reminder was inserted prior to the sequence of choice sets, and an opt-out reminder was presented with each choice set (Cummings and Taylor, 1999; Ladenburg and Olsen, 2014; Alemu and Olsen, 2018).

The third and last part of the questionnaire consists of attitudinal and demographic questions. The attitudinal questions primarily consist of two sections of questions aimed at elicitation of time and risk preferences in a monetary domain. The wording and setup of these questions are heavily inspired by those used in the Global Preference Survey (Falk et al., 2016).

3.3 Data collection

The survey data were collected online in May and June 2020 by the professional survey company Userneeds. The survey was answered by a sample of 3344 respondents[3] living in the municipalities/districts where the case areas are located, as well as neighboring municipalities/districts (Table 4). All respondents were enrolled in online panels administered either by Userneeds or by partnering survey companies (Table 5). Userneeds used a standard invitation for the survey, and if the respondents did not react to the invitation, they received a reminder after a few days. The reminder was setup as a copy of the invitation with only few changes, and hence appeared as a re-invitation rather than a warning regarding missing participation. The respondents were invited, so that the total sample (including both respondents

[3] Responses from another 1673 respondents were also collected in the Danish case area using questionnaire versions that had been slightly modified to enable a range of methodological developments. Results from these methodological research investigations are not reported here.

to the pilot survey and respondents to the main survey) were nationally representative in terms of gender and age (non-interlocked). In Canada and Portugal sampling was done progressively, first sending out invitations to respondents in the primary area, and only sending out to the secondary (and later tertiary) areas if the target number of respondents had not been reached few days after re-invitations were sent out. More information on the sampling procedure can be found in the study documentation: https://sid.erda.dk/wsgi-bin/ls.py?share_id=bbL1PNQSV5.

Table 4 Sampling areas

Country	Sampling areas
Denmark - Municipalities	Herning, Holstebro, Ikast-Brande, Lemvig, Randers, Skive, Struer, Viborg, Brønderslev, Frederikshavn, Hjørring, Jammerbugt, Mariagerfjord, Morsø, Rebild, Thisted, Vesthimmerlands, Aalborg
Canada - CMA Districts	*Primary:* Brantford, Guelph, Kitchener-Cambridge-Waterloo *Secondary:* Hamilton *Tertiary:* Toronto
Portugal - Districts	*Primary*: Coimbra, Guarda, Viseu *Secondary*: Aveiro *Tertiary*: Bragança, Castelo Branco, Leiria, Portalegre, Porto, Santarém, Vila Real
Sweden - Municipalities	Stockholm, Botkyrka, Ekerö, Salem, Upplands-Bro, Strängnäs, Våsterrås, Halstahammar, Kungsör, Köping, Håbo, Järfälla, Huddinge, Nykvarn, Södertalje, Örebro, Katrineholm, Arboga, Eskilstuna, Vingåker

Table 5 Online panels used to sample respondents

Country	Panel provider
Denmark	Userneeds, Cint, YouGov, Eovendo
Canada	Cint, Dynata
Portugal	Cint
Sweden	Userneeds, Cint, SampleXpress

The elicited data consist of eight different datasets, two for each of the four case areas; one with data from the surface water questionnaire, and one with data from the groundwater questionnaire. When analyzing data from a SP survey, it is common practice to account for protest bidders (Johnston et al., 2017b) also known as serial non-participants. These are respondents who always choose the opt-out alternative, for example because they oppose to the survey and hence indicate a zero value for a good that they actually value (von Haefen et al., 2005; Halstead et al., 1992). We included a follow-up question asking respondents who choose the BAU alternative in all choice sets, to state their reason for doing so. Based on responses to this question, we identified respondents who we suspected were protest bidders and removed these from the datasets. Protest bidders made up 2-6 % of all respondents. We also identified respondents who were responding very fast, as we would suspect that these were not able to indicate their true preferences. Identifying a response time that is too fast is not straightforward, and we thus took a very precautious approach to this. Campbell et al. (2017) find that some respondents spend as little as 2.5 seconds per choice set, including the time required to load a webpage (the next choice set), which likely takes 1-2 seconds. With this lower bound as our reference, we decided to remove respondents spending less than 2.5 seconds in average on the three first choice sets, as previous studies have found that the first couple of choice sets may be particularly important for respondents to learn the CE setup (Day et al., 2012; Carlsson et al., 2012; Hess et al., 2012). Once the respondents have understood the setup, it is possible that fast choices reflect genuine preferences, if it e.g. is the case that only one attribute is decisive for their choice. Although the condition applied here may seem somewhat arbitrary, we argue that this conservative approach most likely did not result in removal of respondents whose choices reflected actual preferences. In total, the protest bidders and speeders made up 4-8 % of all respondents. This share of respondents, identified as not complying with the utility assumptions underlying our model (see section 4), is much smaller than what is found in similar water quality studies in Europe (see e.g. Kataria et al. (2012) or Pinto et al. (2016)).

Sociodemographic statistics for the final samples used for analyses and for the sample area populations are presented in Appendix 1. The sample gender distributions are representative of all populations but those in the Danish case area, where an overweight of females is found. None of the samples are perfectly representative of the populations when considering age, education and income distributions. Specifically, across all case areas there is a tendency that the sampled respondents are on average younger, have longer educations, and have higher incomes than the population. Especially the latter could have consequences for the generalizability of WTP estimates since one of the core tenets in economic theory, namely decreasing marginal utility of income, predicts a positive correlation between income and WTP. In other words, elicited WTP estimates from all case areas might be overestimating the actual WTPs in the populations.

4. Theoretical framework for the CE analysis

The CE data is analyzed based on Lancaster's characteristics theory (Lancaster, 1966) and Random Utility Theory (McFadden, 1974). The utility that respondent *n* derives from choosing alternative *i* in choice situation *k*, can be specified as:

$$U_{nik} = -\alpha p_{nik} + \beta x_{nik} + \epsilon_{nik} \tag{1}$$

where p denotes the cost attribute, *x* the vector of non-cost attributes, and α and β are the corresponding coefficients to be estimated. The error term ϵ, is added as the analyst never directly observes utility. In (1), utility is specified in preference space. However, here we are interested in the respondents' WTP rather than preference coefficients. WTP is calculated as the ratio of the non-cost attribute coefficient to the cost coefficient. We denote this ratio as $w=\beta/\alpha$. Hence, we rewrite (1) to instead specify utility in WTP space, meaning that the ratios of the cost and the non-cost attributes are estimated directly (Train and Weeks, 2005):

$$U_{nik} = -\alpha p_{nik} + (\alpha w)x_{nik} + \epsilon_{nik} \quad (2)$$

Assuming that the error term, ϵ, is independently and identically distributed and follows a Gumbel distribution, the probability of respondent *n*'s sequence of choices, *y,* can be represented by the conditional logit model:

$$\Pr(y_n|p_n, x_n) = \prod_{k=1}^{K} \frac{\exp(-\alpha p_{nik}+(\alpha w)x_{nik})}{\sum_{j=1}^{J} \exp(-\alpha p_{njk}+(\alpha w)x_{njk})} \quad (3)$$

where it is observed that alternative *i* is preferred of all alternatives, *J,* in choice task *k*. However, the conditional logit model is restrictive as it assumes the same preferences (and thus WTP) for all respondents. This restriction, as well as the property of the independence of irrelevant alternatives, have made more flexible models such as the random parameters logit (RPL) model popular. The RPL model avoids these restrictions by allowing random preference variation, i.e. random variation in α and *w*. This is done by denoting the joint density of $[\alpha_n, w_{n1}, w_{n2}, \dots, w_{nT}]$ by $f(\theta_n, \Omega)$, where θ_n represents the vectors of individual-specific random parameters and Ω is a vector of parameters of their distribution, such as the mean and the variance. The choice probability from (3) now becomes the integral over all possible values of α_n and w_n:

$$\Pr(y_n|p_n, x_n, \Omega) = \int \prod_{k=1}^{K} \frac{\exp(-\alpha_n p_{nik}+(\alpha_n w_n)x_{nik})}{\sum_{j=1}^{J} \exp(-\alpha_n p_{njk}+(\alpha_n w_n)x_{njk})} f(\theta_n, \Omega)d(\theta_n) \quad (4)$$

The probability in (4) cannot be calculated exactly, as the integral does not have a closed form. The coefficients are instead approximated through simulated maximum likelihood estimation by taking draws from the density. Further technical and theoretical details of the RPL model and the simulated maximum likelihood procedure are available in Train (2009).

It is a well-established phenomenon that people evaluate well-known alternatives (e.g. status quo or BAU alternatives) differently from other alternatives when presented with a decision-making task (e.g. Samuelson and Zeckhauser, 1988; Meyerhoff and Liebe, 2009). In order to take this into account, we include an alternative specific constant in our model, which captures all other determinants of a choice of the BAU alternative that are not captured by the attributes. We, furthermore, include an error component, additional to the usual Gumbel-distributed error term, to capture any remaining BAU effects in the stochastic part of utility. The error component, which is implemented as an individual-specific zero-mean normally distributed random parameter, is assigned exclusively to the two experimentally designed policy alternatives. By specifying a common error component across these two alternatives, correlation patterns in utility can be accommodated (Scarpa et al., 2005). This results in the following utility structure:

$$U_{nik} = \begin{cases} V(p_{nik}, x_{nik}, \tilde{\alpha}_n, \widetilde{w}_n, \mu_n) + \epsilon_{nik}, & j = 1, 2; \\ V(ASC, p_{nik}, x_{nik}, \tilde{\alpha}_n, \widetilde{w}_n) + \epsilon_{nik}, & j = BAU \end{cases} \quad (5)$$

where the indirect utility, V, is a function of the price variable, p, the vector of non-cost attributes, *x,* the individual-specific random cost parameter, $\tilde{\alpha}$, and the vector of individual-specific random non-cost parameters, $\widetilde{w}$. For the two experimentally designed policy alternatives, a common individual-specific error component, μ, is also included, in addition to the unobserved error term, ϵ. In the BAU alternative an ASC is included. We continue to apply the same subscript definitions as in Eq. (1)-(4). All the random non-cost parameters are specified as normally distributed, while the random cost parameter is assumed to be log-normally distributed.

5. Results

In this section, we present the results from the specified RPL model on the eight datasets (see section 3.3). In order to improve the convergence of the likelihood function of the RPL models, we used starting values obtained from conditional logit models based on the same data. Due to the relatively large numeric differences in the estimated parameters, parameters were scaled prior to estimation, making sure that these were of a similar size. The stability of the results have been tested by running models repeatedly with different numbers of draws (up to 10 000), different draw types, different starting values, and different scaling. Also, different model specifications have been tested. The majority of these tests neither changed the key results and conclusions markedly nor significantly improved model fit, suggesting that the presented results can be considered relatively stable. As the cost parameter is assumed log-normally distributed, the cost estimate represents the mean and standard deviation (SD) of the logarithm of this parameter. From these estimates we have derived the mean and SD of the actual parameter (as exemplified by e.g. Train, 2009), and these are the estimates presented in the tables in this section. As all the models are estimated in WTP space, the coefficients are directly interpretable as the marginal annual WTP per household for changes in the attribute levels, as compared to the BAU levels. This model specification furthermore addresses concerns regarding potential differences in error variance across the datasets (Swait and Louviere, 1993) as the associated scale parameters cancel out in WTP space, and the WTP estimates can thus be compared directly (Train and Weeks, 2005).

Goodness-of-fit statistics suggest that all presented models fit data well. There is, furthermore, a clear tendency of sensitivity to scope of all parameters in all models, i.e. that higher water quality/risk/time lag levels lead to numerically larger estimates. In all the models, the ASC estimate is significantly negative, indicating that respondents regardless of attribute levels tend to choose the experimentally designed water quality improving policy alternatives more often than the BAU alternative. In the groundwater models, it is, however, worth noting that the WTP for a change from poor water quality to moderate water quality by design is embedded in the ASC estimate, entailing that all new policies at a minimum results in a change from poor to moderate water quality. This leads to relatively large negative ASC estimates in these models – which can also be interpreted as relatively strong preferences to avoid the BAU situation. It is also worth noting that the error component is highly significant in all the models, indicating that the unexplained variance is generally larger for the two experimentally designed policy alternatives, than for the BAU alternative.

We divide the following section into two subsections; one focusing on models for the surface water data, and one focusing on models for the groundwater data. Note that the results are not directly comparable across the four countries due to geographical and hydromorphological as well as sociodemographic differences between the four case areas. As noted above, the definition of water quality improvements is also different in the Portuguese surface water questionnaire, as compared to the surface water questionnaire for the three other countries. In this section, we will therefore describe the results for each country isolated. Common patterns and differences will be further discussed in section 6.

5.1 Surface water models

5.1.1 Canadian case area

The results from the Canadian surface water data are presented in Table 6.

Table 6 Results from RPL model in WTP space for the Canadian surface water data

Canada	Mean WTP (Euro/household/year)		Standard Deviation (SD)	
	Estimate	**Std. Error**	**Estimate**	**Std. Error**
Upper Grand, Good WQ	39***	14	19***	7.3
Lower Grand etc., Moderate WQ	159***	23	32**	13
Lower Grand etc., Good WQ	160***	25	149***	13
10 % risk of no improvement	17**	9.4	1.8	4.5
40 % risk of no improvement	-80***	9.7	56***	13
Reached in 8 years	-21***	8.1	2.7	13
Reached in 20 years	-38***	8.1	8.3	9.9
Cost	-0.0098***	0.0012	0.0113***	0.0027
ASC (business as usual)	-356***	37		
Error component (alt1, alt2)			476***	53
Model characteristics				
No. of respondents (obs.)	434 (5208)			
LL(0)	-5722			
Final LL	-4160			
Adjusted pseudo R^2	0.270			

Note: '***' indicates significance at the 0.01 level; '**' at the 0.05 level; '*' at the 0.1 level.
Simulations done using 1000 draws based on the scrambled Sobol sequence. Standard errors constructed using the robust sandwich estimator.
Estimates converted to Euro using the 2019 exchange rates reported by the OECD.
BAU water quality levels: Upper Grand - Moderate; Lower Grand etc. - Poor
BAU risk level: No risk
BAU time lag: Reached in 4 years

All mean WTP estimates are significant and indicate that outcome uncertainty and time lags have a negative effect on WTP for water quality improvements. They furthermore indicate that the WTP is higher for an improvement to moderate or good water quality in the “Lower Grand etc.” (covering the lower part of the Grand River, the Nith River and the Conestogo River) than for an improvement to a good quality in the Upper Grand. It is worth noting that the sign of the estimate for “10 % risk of no improvement” is positive, as we *a priori* expected this to be negative. Interestingly, the SD estimate for this parameter is insignificant, and hence we do not find heterogeneity in preferences within the sample for this attribute level. One reason for the unexpected sign may be that it is cognitively hard for the respondents to fully comprehend and account for the risk attribute, something also indicated during the questionnaire pretesting phase. It is also unexpected that the difference between the WTP estimates for moderate and good water quality in "Lower Grand etc." is not statistically significant. One explanation for this may be that the respondents would like water quality improvements in this area, and yet care little about the extent of these improvements. We further analyzed these results with latent class models (not shown here), which suggested that these water quality estimates possibly reflect that a minority of the respondents display what seems to be irrational preferences[4], and they spent significantly less time answering the choice sets than did other respondents. Reasons for such behavior may be that the respondents lacked interest in the topic, experienced response fatigue, or were worried that policies addressing water quality improvements would limit their land use options (although it was clearly stated in the scenario description that they should ignore this aspect when completing the choice tasks).

5.1.2 Danish case area

In Table 7, we present results for the Danish surface water data.

Table 7 Results from RPL model in WTP space for the Danish surface water data

Denmark	Mean WTP (Euro/household/year)		Standard Deviation (SD)	
	Estimate	Std. Error	Estimate	Std. Error
Rest of Limfjorden, Moderate WQ	209***	15	18	19
Rest of Limfjorden, Good WQ	338***	20	204***	12
Skive Fjord etc., Moderate WQ	84***	9.2	12*	6.5
Skive Fjord etc., Good WQ	130***	12	64***	9.9
10 % risk of no improvement	-42***	6.1	0.1	3.9
40 % risk of no improvement	-86***	10	86***	12
Reached in 8 years	-15**	6.2	0.6	6.6
Reached in 20 years	-53***	8.0	25	24
Cost	-0.0127***	0.0015	0.0082***	0.0022
ASC (business as usual)	-98***	26		
Error component (alt1, alt2)			253***	27
Model characteristics				
No. of respondents (obs.)	383 (4596)			
LL(0)	-5049			
Final LL	-3365			
Adjusted pseudo R^2	0.330			

Note: '***' indicates significance at the 0.01 level; '**' at the 0.05 level; '*' at the 0.1 level.
Simulations done using 1000 draws based on the scrambled Sobol sequence. Standard errors constructed using the robust sandwich estimator.
Estimates converted to Euro using the 2019 exchange rates reported by the OECD.
BAU water quality levels: Rest of Limfjorden - Poor; Skive Fjord etc. - Poor
BAU risk level: No risk
BAU time lag: Reached in 4 years

[4] The latent class models indicate that a minority of respondents has either insignificant preferences for most attributes (including the cost attribute) or significantly negative preferences for water quality improvements.

All mean WTP estimates are significant and all signs are as expected, with outcome uncertainty and time lags clearly affecting the respondents' WTP for water quality improvements negatively. The results indicate that the value of water quality improvements in the "Rest of Limfjorden" (covering the part of Limfjorden not included in "Skive Fjord etc.") higher than improvements in "Skive Fjord etc." (covering Skive Fjord, Lovns Bredning, Risgård Bredning and Hjarbæk Fjord). An obvious reason for this may be that the "Rest of Limfjorden" by far covers the greatest area of the two. It is worth noting that the SD estimate is insignificant for moderate water quality in the "Rest of Limfjorden" and only significant on a 0.1 level of significance for moderate water quality in "Skive Fjord etc.". Hence, we do not find much variation among the respondents in terms of preferences for smaller quality improvements. The same is true in terms of preferences for "10 risk % of no improvement" as well as time preferences, as indicated by the insignificant SD estimates for these attributes.

5.1.3 Portuguese case area

We present the results from the Portuguese surface water data in Table 8.

Table 8 Results from RPL model in WTP space for the Portuguese surface water data

Portugal	Mean WTP (Euro/household/year)		Standard Deviation (SD)	
	Estimate	Std. Error	Estimate	Std. Error
Lower Dão, Good EP&PC	27***	5.2	29***	5.5
Upper Mondego, Good EP&PC	42***	4.4	10*	5.7
Lower Mondego, Moderate EP&PC	126***	9.7	41***	5.9
Lower Mondego, Good EP&PC	148***	13	106***	8.8
10 % risk of no improvement	-17***	5.1	11	6.7
40 % risk of no improvement	-98***	10	69***	8.9
Reached in 8 years	-30***	5.8	21	16
Reached in 20 years	-24***	5.5	11	9.0
Cost	-0.0177***	0.0021	0.0182***	0.0052
ASC (business as usual)	-156***	22		
Error component (alt1, alt2)			230***	22
Model characteristics				
No. of respondents (obs.)	379 (4548)			
LL(0)	-4996			
Final LL	-3736			
Adjusted pseudo R^2	0.248			

Note: '***' indicates significance at the 0.01 level; '**' at the 0.05 level; '*' at the 0.1 level.
Simulations done using 1000 draws based on the scrambled Sobol sequence. Standard errors constructed using the robust sandwich estimator.
BAU water quality levels: Lower Dão - Moderate; Upper Mondego - Moderate; Lower Mondego - Poor
BAU risk level: No risk
BAU time lag: Reached in 4 years
EP&PC: Ecological potential and physical conditions

As described above, these results concern the ecological potential and physical conditions (EP&PC) of the water (rather than the water quality). All mean WTP estimates are significant and have signs as expected, hence outcome uncertainty and time lags have a negative effect on WTP for EP&PC improvements. The results furthermore indicate that the respondents prefer an improvement to a moderate or good EP&PC level in Lower Mondego to an improvement to a good EP&PC level in the Lower Dão or Upper Mondego River. It is worth noting that the respondents presumably have larger negative preferences for a time lag of 8 years compared to a time lag of 20 years before a given EP&PC level is reached. This difference is, however, not statistically significant, which indicates that respondents dislike time lags beyond 4 years, yet do not care too much about these beyond 8 years. The insignificant SDs for both of the time lag attribute levels suggest that preferences do not vary much among the respondents.

5.1.4 Swedish case area

In Table 9, we present results for the Swedish surface water data.

Table 9 Results from RPL model in WTP space for the Swedish surface water data

Sweden	Mean WTP (Euro/household/year)		Standard Deviation (SD)	
	Estimate	**Std. Error**	**Estimate**	**Std. Error**
Eastern Mälaren, Good WQ	80***	9.1	1.3	6.5
Western Mälaren, Good WQ	123***	7.5	6.0	66
Hjälmaren, Moderate WQ	207***	17	7.2	9.9
Hjälmaren, Good WQ	268***	22	129***	20
10 % risk of no improvement	-42***	7.3	3.0	3.9
40 % risk of no improvement	-114***	15	110***	23
Reached in 8 years	33***	8.7	3.8	16
Reached in 20 years	-48***	9.1	4.3	24
Cost	-0.0085***	0.0011	0.0074***	0.0011
ASC (business as usual)	-201***	66		
Error component (alt1, alt2)			452***	50
Model characteristics				
No. of respondents (obs.)	385 (4620)			
LL(0)	-5076			
Final LL	-3599			
Adjusted pseudo R^2	0.287			

Note: '***' indicates significance at the 0.01 level; '**' at the 0.05 level; '*' at the 0.1 level.
Simulations done using 1000 draws based on the scrambled Sobol sequence. Standard errors constructed using the robust sandwich estimator.
Estimates converted to Euro using the 2019 exchange rates reported by the OECD.
BAU water quality levels: Eastern Mälaren - Moderate; Western Mälaren - Moderate; Hjälmaren - Poor
BAU risk level: No risk
BAU time lag: Reached in 4 years

Once again, all mean WTP estimates are significant and indicate that outcome uncertainty and time lags negatively affect the respondents' WTP for water quality improvements. The results also show that respondents value an improvement to a moderate or good water quality in Hjälmaren higher than an improvement to a good quality in Eastern or Western Mälaren. It is unexpected that the estimate for a time lag of 8 years is positive, as we *a priori* would expect this to be negative. We examined this issue further with latent class models (not shown here), yet this offered us no insights or logical explanations for this result. During the pretesting face we noted a viewpoint that it could be positive to receive water quality improvements further into the future, as the chance of actually reaching these were perceived to be greater. Although a respondent with such a viewpoint has not understood the setup, as we explicitly describe the risk of not reaching the stated water quality, this may have played a role in producing this unexpected estimate. It is interesting that most of the SD estimates are insignificant, even for the water quality improvements. We thus find relatively little evidence of preference heterogeneity among these respondents.

5.2 Groundwater models

5.2.1 Canadian case area

Results for the Canadian groundwater data are presented in Table 10.

Table 10 Results from RPL model in WTP space for the Canadian groundwater data

Canada	Mean WTP (Euro/household/year)		Standard Deviation (SD)	
	Estimate	**Std. Error**	**Estimate**	**Std. Error**
Groundwater, Good WQ	104***	7.7	115***	9.4
10 % risk of no improvement	-32***	5.3	8.6*	4.4
40 % risk of no improvement	-114***	9.9	70***	8.0
Reached in 20 years	-13**	5.1	7.7*	4.5
Reached in 50 years	-50***	10	70***	8.9
Cost	-0.0172***	0.0022	0.0229***	0.0058
ASC (business as usual)	-339***	25		
Error component (alt1, alt2)			317***	33
Model characteristics				
No. of respondents (obs.)	426 (5112)			
LL(0)	-5616			
Final LL	-4057			
Adjusted pseudo R^2	0.275			

Note: '***' indicates significance at the 0.01 level; '**' at the 0.05 level; '*' at the 0.1 level.
Simulations done using 1000 draws based on the scrambled Sobol sequence. Standard errors constructed using the robust sandwich estimator.
Estimates converted to Euro using the 2019 exchange rates reported by the OECD.
BAU groundwater quality level: Poor
Baseline groundwater quality level for model parameters: Moderate
BAU risk level: No risk
BAU time lag: Reached in 8 years

All mean WTP estimates are significant and of the expected sign. The results reveal that groundwater quality improvements have a positive value to the respondents. Yet, it is also evident that the introduction of outcome uncertainty and time lags beyond 8 years overall reduces welfare. The negative ASC suggests that the respondents have a positive WTP for improving water quality from poor to moderate water quality, although we are not able to conclude on the exact size of this. All the SD estimates are significant indicating that respondents' preferences are heterogeneous in terms of all the included attributes.

5.2.2 Danish case area

The results for the Danish groundwater data are presented in Table 11.

Table 11 Results from RPL model in WTP space for the Danish groundwater data

Denmark	Mean WTP (Euro/household/year)		Standard Deviation (SD)	
	Estimate	**Std. Error**	**Estimate**	**Std. Error**
Groundwater, Good WQ	84***	10	148***	8.9
10 % risk of no improvement	-56***	7.3	1.4	2.8
40 % risk of no improvement	-169***	17	163***	18
Reached in 20 years	-55***	6.4	8.1	5.8
Reached in 50 years	-139***	14	84***	21
Cost	-0.0119***	0.0015	0.0127***	0.0030
ASC (business as usual)	-546***	41		
Error component (alt1, alt2)			446***	46
Model characteristics				
No. of respondents (obs.)	383 (4596)			
LL(0)	-5049			
Final LL	-3522			
Adjusted pseudo R^2	0.300			

Note: '***' indicates significance at the 0.01 level; '**' at the 0.05 level; '*' at the 0.1 level.
Simulations done using 1000 draws based on the scrambled Sobol sequence. Standard errors constructed using the robust sandwich estimator.
Estimates converted to Euro using the 2019 exchange rates reported by the OECD.
BAU groundwater quality level: Poor
Baseline groundwater quality level for model parameters: Moderate
BAU risk level: No risk
BAU time lag: Reached in 8 years

Again, all mean WTP estimates are significant and show the expected signs. The results indicate that respondents are willing to pay for ensuring good quality of the groundwater. Once again, it is worth noting that outcome uncertainty and time lags contribute negatively to the overall welfare associated with policies to improve groundwater quality. This is particularly the case when risks are high or time lags are long. The negative ASC indicates that respondents generally value a groundwater quality improvement from the current poor condition to a moderate condition regardless of the other attributes. The majority of the SD estimates are significant and, thus, indicate the presence of preference heterogeneity. However, for "10 % risk of no improvement" as well as for a 20 year time lag the insignificant SD estimates indicate that we do not find heterogeneity in preferences across the sample.

5.2.3 Portuguese case area

In Table 12, we present results for the Portuguese groundwater data. All mean WTP estimates are significant and exhibit the expected sign. This indicates that respondents overall have a positive WTP for groundwater quality improvements, and that this is negatively affected when adding outcome uncertainty and time lags beyond 8 years. The negative ASC again suggests that the respondents have a positive WTP for moderate water quality, although we cannot fully ascertain to what degree the estimate confounds expressions of preference with other behavioral effects. There is significant heterogeneity among the respondents' preferences for most attribute levels, though this is not evident for the 20-year time lag.

Table 12 Results from RPL model in WTP space for the Portuguese groundwater data

Portugal	**Mean WTP (Euro/household/year)**		**Standard Deviation (SD)**	
	Estimate	**Std. Error**	**Estimate**	**Std. Error**
Groundwater, Good WQ	51***	5.1	66***	5.3
10 % risk of no improvement	-36***	4.0	19***	4.6
40 % risk of no improvement	-87***	6.9	61***	6.3
Reached in 20 years	-41***	3.3	3.6	4.5
Reached in 50 years	-73***	4.7	39***	6.1
Cost	-0.0270***	0.0029	0.0237***	0.0050
ASC (business as usual)	-230***	20		
Error component (alt1, alt2)			182***	20
Model characteristics				
No. of respondents (obs.)	379 (4548)			
LL(0)	-4996			
Final LL	-3429			
Adjusted pseudo R^2	0.311			

Note: '***' indicates significance at the 0.01 level; '**' at the 0.05 level; '*' at the 0.1 level.
Simulations done using 1000 draws based on the scrambled Sobol sequence. Standard errors constructed using the robust sandwich estimator.
BAU groundwater quality level: Poor
Baseline groundwater quality level for model parameters: Moderate
BAU risk level: No risk
BAU time lag: Reached in 8 years

5.2.4 Swedish case area

Finally, in Table 13 we present the results for the Swedish groundwater data. All mean WTP estimates are significant and with the expected sign. Once again, respondents display a positive WTP for water quality improvements, and this is negatively affected by outcome uncertainty and time lags. The negative impact on WTP is particularly large if there is 40 % risk that a policy alternative will not improve the water quality. The negative ASC is also worth noting, as this suggests that the respondents have a positive WTP for a groundwater quality improvement from poor to moderate. The sample exhibit heterogeneous preferences for most attributes, though not for the 10 % risk level and the 20-year time lag.

Table 13 Results from RPL model in WTP space for the Swedish groundwater data

Sweden	Mean WTP (Euro/household/year)		Standard Deviation (SD)	
	Estimate	Std. Error	Estimate	Std. Error
Groundwater, Good WQ	73***	9.1	103***	15
10 % risk of no improvement	-62***	5.8	4.2	6.9
40 % risk of no improvement	-168***	13	116***	9.7
Reached in 20 years	-21***	6.4	0.7	8.9
Reached in 50 years	-89***	13	97***	8.4
Cost	-0.0159***	0.0021	0.0233***	0.0064
ASC (business as usual)	-622***	62		
Error component (alt1, alt2)			426***	45
Model characteristics				
No. of respondents (obs.)	395 (4740)			
LL(0)	-5207			
Final LL	-3667			
Adjusted pseudo R^2	0.293			

Note: '***' indicates significance at the 0.01 level; '**' at the 0.05 level; '*' at the 0.1 level.
Simulations done using 1000 draws based on the scrambled Sobol sequence. Standard errors constructed using the robust sandwich estimator.
Estimates converted to Euro using the 2019 exchange rates reported by the OECD.
BAU groundwater quality level: Poor
Baseline groundwater quality level for model parameters: Moderate
BAU risk level: No risk
BAU time lag: Reached in 8 years

6. Discussion

As mentioned above, caution should be taken when comparing the results for the case areas presented in section 5, as there are considerable cross-country differences in terms of the hydromorphological characteristics of the water bodies in focus, and also in terms of the sociodemographic characteristics of the respondents. In this section, we will, however, try to point to some general results that are common across the four case areas.

A major focus of this study has been to extend the rather classic water quality valuation setup by explicitly including the uncertainty that inevitably will surround the outcomes of most water related policy alternatives, as well as the presence of time lags between implementation of policy measures and actual improvement of water quality. The results clearly show that these aspects matter a great deal for people’s WTP for new policy alternatives. The total sum that the respondents on average are willing to pay to achieve good surface water quality in all water bodies within a case area decrease with up to 45 % when outcome uncertainty is introduced, and up to 19 % when time lags are introduced[5]. The fact that this effect is evident across all the case areas, despite the differences between these, points to the universality of this result. This has implications for the evaluation of the socioeconomic consequences of various water related policy measures, and could potentially mean that wrong conclusions are drawn if not taken into account. The results also give some indications on how respondents' risk and time preferences fit with theory. It is e.g. somewhat surprising to discover how well many of the risk estimates may align with expected utility theory (see Appendix 2), as this theory has been criticized widely both theoretically and empirically (e.g. Kahneman and Tversky, 1979). Moreover, the time lag estimates show signs of respondent time preferences being in line with hyperbolic discounting, which follows the conclusions drawn from a CE study by Viscusi et al. (2008). Further investigations of our data are needed, before firm conclusions can be drawn regarding these risk and time preference aspects. The above, however, points to the usefulness of our collected data in a broader perspective to investigate the validity of the assumptions of many cost benefit analyzes.

For all countries and water bodies, our results indicate that respondents on average have a positive WTP for water quality improvements, irrespective of whether these improvements concern surface water or groundwater. This is in line with previous empirical studies. In general, it seems that improvements from a poor to a moderate water quality matter more to the respondents than improvements from moderate to good water quality. From an economic theory point of view, this corresponds to the concept of decreasing marginal utility of goods. It is also in line with previous findings of other water quality valuation studies (e.g. Bateman et al., 2011; Pinto et al., 2016). In terms of cross-case area comparisons, an advantage of this study is that we strongly reduce the effect of varying questionnaire designs, as we use a common questionnaire design across the case areas. In spite of this, we find considerable cross-case area differences in the WTP estimates for water quality improvements. As mentioned, there are other reasons for this as respondent and waterbody characteristics vary considerably across the case areas. Yet, it underlines the importance of taking these characteristics into account, both when designing new valuation studies, as well as when conducting benefit transfer studies.

The results from the groundwater models are somewhat more comparable across the case areas than the results from the surface water models, as the former have identical BAU levels in all case areas. Although the case areas differ in size, it is unlikely to matter much for respondents both in terms of their use value of groundwater as a source of their own drinking water and their non-use existence value of groundwater. The results suggest that respondents in Denmark and Sweden have the largest WTP for better groundwater quality, based on estimates for good water quality and the ASC (which embeds the WTP for moderate quality). That groundwater is highly valued in the Limfjorden case area (Denmark) is expected, as this is the only one of the case areas where nearly all drinking water originate from extracted groundwater. However, we also see large WTP estimates in the Mälaren/Hjälmaren case area (Sweden) where only a small fraction of the

[5] Due to design restrictions, we are not able to derive the exact sizes of these decreases for the groundwater cases as the WTP to go from poor to moderate groundwater quality is embedded in the ASC estimate, and as such it is confounded with other behavioral effects that potentially affect the ASC estimate.

drinking water is currently coming from groundwater (SCB, 2017). This may indicate that the respondents in this area appreciate the non-use values of groundwater particularly highly.

All our models result in significantly negative ASC estimates, indicating that the respondents in general had a higher tendency to choose the water quality improving policy alternatives rather than the BAU alternative, no matter what the attribute levels were. This is at odds with the status-quo effect commonly found in behavioral experiments, namely that people disproportionally often choose well-known alternatives, such as status quo or BAU alternatives (e.g. Samuelson and Zeckhauser, 1988; Kahneman et al., 1991). It is, however, not uncommon to find the opposite in environmental valuation studies (e.g. Lehtonen et al., 2003; Scarpa et al., 2005; Marsh et al., 2011), where respondents are often found to want improvements, irrespective of the actual details regarding these. Scarpa et al. (2005) note that it may be consistent with a perception of under-provision of the public good in question. Yet, many drivers may affect the size and nature of the status quo effects, making it hard to conclude the extent to which this reflects respondent preferences, and the extent to which it is induced by the choice context (Boxall et al., 2009; Oehlmann et al., 2017; Barreiro-Hurle et al., 2018).

As mentioned in section 3.3., our respondents are not entirely representative of the populations in the sampling areas. On average, our respondent samples seem to be younger, have longer educations, and receive a higher annual income than the populations, from which we sampled (see Appendix 1 for sociodemographic statistics). Although not uncommon in the SP literature, unrepresentative samples put obvious limitations to the ability to generalize results and conclusions to the population. This is particularly relevant in relation to providing guidance for policy makers (Johnston et al., 2017b). Especially the fact that our samples have higher income than the populations is likely to cause the estimated mean WTPs to be higher than the WTPs in the population. This should be kept in mind if using the results for generalizations to the population level. It may, however, be accommodated for through scaling of the sample mean WTP estimates to the relevant population, e.g. using weights for various demographic and sample features (Johnston et al., 2017b).

7. Conclusion

Our data analysis reveals that the introduction of outcome uncertainty and time lags have a considerable negative effect on the value that people derive from water quality improvements. This has implications for the evaluation of the socioeconomic consequences of various water related policy measures, and it is thus crucial to take these aspects into account when assessing the welfare effects associated with water quality improvements. The effects are evident across four different case areas, in terms of improvements to both surface water and groundwater quality. The large geographical variation and the wide range of different types of water bodies incorporated in our study allows us to point to the universality of these effects. As expected, we find that respondents to our CE survey generally exhibit positive preferences for water quality improvements, and hence are willing to pay for such changes. The results are particularly interesting in terms of the groundwater value, as only few previous studies have employed a CE to estimate this. We, however, also find that the size of all the parameters differs considerably between the case areas, hence underlining the importance of taking case-specific hydromorphological and population characteristics into account in environmental valuation studies.

References

Adamowicz, W., Louviere, J., and Williams, M. (1994). Combining revealed and stated preference methods for valuing environmental amenities. *Journal of Environmental Economics and Management*, *26*(3), 271-292.

Agência Portuguesa do Ambiente (2016). *Plano de gestão de região hidrográfica. Parte 1 – enquadramento e aspetos gerais. Região hidrográfica do Vouga, Mondego e Lis (RH4).* Agência Portuguesa do Ambiente.

Akter, S., Bennett, J., and Ward, M. B. (2012). Climate change scepticism and public support for mitigation: Evidence from an Australian choice experiment. *Global Environmental Change*, *22*(3), 736-745.

Alemu, M. H., and Olsen, S. B. (2018). Can a Repeated Opt-Out Reminder mitigate hypothetical bias in discrete choice experiments? An application to consumer valuation of novel food products. *European Review of Agricultural Economics*, *45*(5), 749-782.

Andersen, S., Harrison, G. W., Lau, M. I., and Rutström, E. E. (2008). Eliciting risk and time preferences. *Econometrica*, *76*(3), 583-618.

Armatas, C. A., Venn, T. J., and Watson, A. E. (2014). Applying Q-methodology to select and define attributes for non-market valuation: A case study from Northwest Wyoming, United States. *Ecological Economics*, *107*, 447-456.

Arrow, K. J., and Fisher, A. C. (1974). Environmental Preservation, Uncertainty, and Irreversibility. *The Quarterly Journal of Economics*, *88*(2), 312-319.

Barreiro-Hurle, J., Espinosa-Goded, M., Martinez-Paz, J. M., and Perni, A. (2018). Choosing not to choose: A meta-analysis of status quo effects in environmental valuations using choice experiments. *Economía Agraria y Recursos Naturales - Agricultural and Resource Economics*, *18*(1), 79-109.

Bateman, I. J., Brouwer, R., Ferrini, S., Schaafsma, M., Barton, D. N., Dubgaard, A., Hasler, B., Hime, S., Liekens, I., Navrud, S., De Nocker, L., Ščeponavičiūtė, R., and Semėnienė, D. (2011). Making benefit transfers work: deriving and testing principles for value transfers for similar and dissimilar sites using a case study of the non-market benefits of water quality improvements across Europe. *Environmental and Resource Economics*, *50*(3), 365-387.

Botzen, W. J., and van den Bergh, J. C. (2012). Monetary valuation of insurance against flood risk under climate change. *International Economic Review*, *53*(3), 1005-1026.

Boxall, P., Adamowicz, W. L., and Moon, A. (2009). Complexity in choice experiments: choice of the status quo alternative and implications for welfare measurement. *Australian Journal of Agricultural and Resource Economics*, *53*(4), 503-519.

Brouwer, R., Ordens, C. M., Pinto, R., and de Melo, M. T. C. (2018). Economic valuation of groundwater protection using a groundwater quality ladder based on chemical threshold levels. *Ecological Indicators*, *88*, 292-304.

Brouwer, R., and Neverre, N. (2020). A global meta-analysis of groundwater quality valuation studies. *European Review of Agricultural Economics*, *47*(3), 893-932.

Campbell, D., Mørkbak, M. R., and Olsen, S. B. (2017). Response time in online stated choice experiments: the non-triviality of identifying fast and slow respondents. *Journal of Environmental Economics and Policy*, *6*(1), 17-35.

Carlsson, F., Mørkbak, M. R., and Olsen, S. B. (2012). The first time is the hardest: A test of ordering effects in choice experiments. *Journal of Choice Modelling*, *5*(2), 19-37.

Carpenter, S. R., Stanley, E. H., and Vander Zanden, M. J. (2011). State of the world's freshwater ecosystems: physical, chemical, and biological changes. *Annual review of Environment and Resources*, *36*, 75-99.

ChoiceMetrics (2018). *Ngene 1.2 User manual and reference guide.* Australia.

Cummings, R. G., and Taylor, L. O. (1999). Unbiased value estimates for environmental goods: A cheap talk design for the contingent valuation method. *American Economic Review*, *89*(3), 649-665.

Day, B., Bateman, I. J., Carson, R. T., Dupont, D., Louviere, J. J., Morimoto, S., Scarpa, R., and Wang, P. (2012). Ordering effects and choice set awareness in repeat-response stated preference studies. *Journal of environmental economics and management*, *63*(1), 73-91.

De Marchi, E., Caputo, V., Nayga Jr, R. M., and Banterle, A. (2016). Time preferences and food choices: Evidence from a choice experiment. *Food Policy*, *62*, 99-109.

Dudinskaya, E. C., Naspetti, S., and Zanoli, R. (2020). Using eye-tracking as an aid to design on-screen choice experiments. *Journal of choice modelling*, *36*, 100232.

European Commission (2009). Common Implementation Strategy for the Water Framework Directive (2000/60/EC). Guidance Document No. 20. Guidance document on exemptions to the environmental objectives, Brussels

Faccioli, M., Kuhfuss, L., and Czajkowski, M. (2019). Stated preferences for conservation policies under uncertainty: insights on the effect of individuals' risk attitudes in the environmental domain. *Environmental and Resource Economics*, *73*(2), 627-659.

Falk, A., Becker, A., Dohmen, T. J., Huffman, D., and Sunde, U. (2016). The preference survey module: A validated instrument for measuring risk, time, and social preferences. *Netspar Discussion Paper No. 01/2016-003.*

Giordano, M. (2009). Global groundwater? Issues and solutions. *Annual review of Environment and Resources*, *34*, 153-178.

Glenk, K., and Colombo, S. (2011). How sure can you be? A framework for considering delivery uncertainty in benefit assessments based on stated preference methods. *Journal of Agricultural Economics*, *62*(1), 25-46.

Grand River Conservation Authority (2014). *Grand River watershed - water management plan*. Grand River Conservation Authority, Cambridge, ON.

Halstead, J. M., Luloff, A. E., and Stevens, T. H. (1992). Protest bidders in contingent valuation. *Northeastern Journal of Agricultural and Resource Economics*, *21*(2), 160-169.

Hanley, N., Mourato, S., and Wright, R. E. (2001). Choice modelling approaches: a superior alternative for environmental valuation?. *Journal of Economic Surveys*, *15*(3), 435-462.

Hanley, N., Wright, R. E., and Alvarez-Farizo, B. (2006). Estimating the economic value of improvements in river ecology using choice experiments: an application to the water framework directive. *Journal of Environmental Management, 78*, 183-193.

Harrison, G. W., Lau, M. I., and Williams, M. B. (2002). Estimating individual discount rates in Denmark: A field experiment. *American Economic Review*, *92*(5), 1606-1617.

Hasler, B., Lundhede, T., and Martinsen, L. (2007). Protection versus purification-assessing the benefits of drinking water quality. *Hydrology Research*, *38*(4-5), 373-386.

Henry, C. (1974). Investment decisions under uncertainty: the" irreversibility effect". *The American Economic Review*, *64*(6), 1006-1012.

Hess, S., Hensher, D. A., and Daly, A. (2012). Not bored yet–revisiting respondent fatigue in stated choice experiments. *Transportation research part A: policy and practice*, *46*(3), 626-644.

Hime, S., Bateman, I. J., Posen, P., and Hutchins, M. (2009). A transferable water quality ladder for conveying use and ecological information within public surveys. *CSERGE working paper EDM 09-01*.

Huber, J., and Zwerina, K. (1996). The importance of utility balance in efficient choice designs. *Journal of Marketing Research*, *33*(3), 307-317.

Jacobsen, B. (2020). Working paper on environmental measures, costs and implementation in LEAP case areas - Limfjorden (Denmark), Lake Erie (Canada), Mondego (Portugal) and Mälaren (Sweden)*. LEAP working paper D3.1.*

Jensen, A. K. (2019). A structured approach to attribute selection in economic valuation studies: using q-methodology. *Ecological Economics*, *166*, 106400.

Jensen, C. L., Jacobsen, B. H., Olsen, S. B., Dubgaard, A., and Hasler, B. (2013). A practical CBA-based screening procedure for identification of river basins where the costs of fulfilling the WFD requirements may be disproportionate–applied to the case of Denmark. *Journal of Environmental Economics and Policy*, *2*(2), 164-200.

Johnston, R. J., Besedin, E. Y., and Stapler, R. (2017a). Enhanced geospatial validity for meta-analysis and environmental benefit transfer: an application to water quality improvements. *Environmental and Resource Economics*, *68*(2), 343-375.

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. (2017b). Contemporary guidance for stated preference studies. *Journal of the Association of Environmental and Resource Economists*, *4*(2), 319-405.

Kahneman, D., Knetsch, J. L., and Thaler, R. H. (1991). Anomalies: The endowment effect, loss aversion, and status quo bias. *Journal of Economic perspectives*, *5*(1), 193-206.

Kahneman, D., and Tversky, A. (1979). Prospect theory: an analysis of decision under risk. *Econometrica*, *47*(2), 263-292.

Kataria, M., Bateman, I., Christensen, T., Dubgaard, A., Hasler, B., Hime, S., Ladenburg, J., Levin, G., Martinsen, L. and Nissen, C. (2012). Scenario realism and welfare estimates in choice experiments – A non-market valuation study on the European water framework directive. *Journal of Environmental Management*, *94*(1), 25-33.

Knight, F. H. (1921). *Risk, uncertainty and profit*. Boston: Houghton Mifflin Company.

Ladenburg, J., and Olsen, S. B. (2014). Augmenting short cheap talk scripts with a repeated opt-out reminder in choice experiment surveys. *Resource and Energy Economics*, *37*, 39-63.

Lancaster, K. J. (1966). A new approach to consumer theory. *Journal of Political Economy*, *74*(2), 132-157.

Larue, B., West, G. E., Singbo, A., and Tamini, L. D. (2017). Risk aversion and willingness to pay for water quality: The case of non-farm rural residents. *Journal of Environmental Management*, *197*, 296-304.

Lehtonen, E., Kuuluvainen, J., Pouta, E., Rekola, M., and Li, C. Z. (2003). Non-market benefits of forest conservation in southern Finland. *Environmental Science and Policy*, *6*(3), 195-204.

Lintern, A., McPhillips, L., Winfrey, B., Duncan, J., and Grady, C. (2020). Best Management Practices for Diffuse Nutrient Pollution: Wicked Problems Across Urban and Agricultural Watersheds. *Environmental Science and Technology*, *54*(15), 9159-9174.

Logar, I., and Brouwer, R. (2017). The effect of risk communication on choice behavior, welfare estimates and choice certainty. *Water Resources and Economics*, *18*, 34-50.

Louviere, J. J., Hensher, D. A., and Swait, J. D. (2000). *Stated choice methods: Analysis and applications*. Cambridge University Press.

Lundhede, T., Jacobsen, J. B., Hanley, N., Strange, N., and Thorsen, B. J. (2015). Incorporating outcome uncertainty and prior outcome beliefs in stated preferences. *Land Economics*, *91*(2), 296-316.

Marsh, D., Mkwara, L., and Scarpa, R. (2011). Do respondents' perceptions of the status quo matter in non-market valuation with choice experiments? An application to New Zealand freshwater streams. *Sustainability*, *3*(9), 1593-1615.

McFadden, D., 1974. Conditional Logit Analysis of Qualitative Choice Behavior. In *Frontiers in Econometrics* (pp. 105-142). Academic Press, New York.

Meals, D. W., Dressing, S. A., and Davenport, T. E. (2010). Lag time in water quality response to best management practices: A review. *Journal of environmental quality*, *39*(1), 85-96.

Meerding, W. J., Bonsel, G. J., Brouwer, W. B., Stuifbergen, M. C., and Essink-Bot, M. L. (2010). Social time preferences for health and money elicited with a choice experiment. *Value in health*, *13*(4), 368-374.

Meyer, A. (2013). Intertemporal valuation of river restoration. *Environmental and Resource Economics*, *54*(1), 41-61.

Meyerhoff, J., and Liebe, U. (2009). Status quo effect in choice experiments: Empirical evidence on attitudes and choice task complexity. *Land Economics*, *85*(3), 515-528.

Miljøministeriet (2011). *Vandplan 2010-2015. Limfjorden. Hovedvandopland 1.2. Vanddistrikt: Jylland og Fyn.* Miljøministeriet, Naturstyrelsen.

Oehlmann, M., Meyerhoff, J., Mariel, P., and Weller, P. (2017). Uncovering context-induced status quo effects in choice experiments. *Journal of Environmental Economics and Management*, *81*, 59-73.

Pinto, R., Brouwer, R., Patrício, J., Abreu, P., Marta-Pedroso, C., Baeta, A., Franco, J.N., Domingos, T., and Marques, J. C. (2016). Valuing the non-market benefits of estuarine ecosystem services in a river basin context: Testing sensitivity to scope and scale. *Estuarine, Coastal and Shelf Science*, *169*, 95-105.

Roberts, D. C., Boyer, T. A., and Lusk, J. L. (2008). Preferences for environmental quality under uncertainty. *Ecological Economics*, *66*(4), 584-593.

Samuelson, W., and Zeckhauser, R. (1988). Status quo bias in decision making. *Journal of risk and uncertainty*, *1*(1), 7-59.

Scarpa, R., Ferrini, S., and Willis, K. (2005). Performance of error component models for status-quo effects in choice experiments. In *Applications of simulation methods in environmental and resource economics* (pp. 247-273). Springer, Dordrecht.

SCB (2017). *Vattenanvändingen i Sverige 2015*. Statistics Sweden, Unit of Environmental Accounts and Natural Resources.

SCB (2019). *Markanvändningen i Sverige, sjunde utgåvan*. Statistics Sweden, Regions and Environment Department.

Swait, J., and Louviere, J. (1993). The role of the scale parameter in the estimation and comparison of multinomial logit models. *Journal of marketing research*, *30*(3), 305-314.

Teixeira, Z., Teixeira, H., and Marques, J. C. (2014). Systematic processes of land use/land cover change to identify relevant driving forces: Implications on water quality. *Science of the Total Environment*, *470*, 1320-1335.

Tentes, G., and Damigos, D. (2015). Discrete choice experiment for groundwater valuation: Case of the Asopos River basin, Greece. *Journal of Water Resources Planning and Management*, *141*(7), 04014089.

Torres, C., Faccioli, M., and Font, A. R. (2017). Waiting or acting now? The effect on willingness-to-pay of delivering inherent uncertainty information in choice experiments. *Ecological Economics*, *131*, 231-240.

Train, K., and Weeks, M. (2005). Discrete choice models in preference space and willingness-to-pay space. In *Applications of simulation methods in environmental and resource economics* (pp. 1-16). Springer, Dordrecht.

Train, K. E. (2009). *Discrete choice methods with simulation*. Cambridge university press.

Van Meter, K. J., Basu, N. B., Veenstra, J. J., and Burras, C. L. (2016). The nitrogen legacy: emerging evidence of nitrogen accumulation in anthropogenic landscapes. *Environmental Research Letters*, *11*(3), 035014.

Vero, S. E., Basu, N. B., Van Meter, K., Richards, K. G., Mellander, P. E., Healy, M. G., and Fenton, O. (2018). The environmental status and implications of the nitrate time lag in Europe and North America. *Hydrogeology journal*, *26*(1), 7-22.

Viscusi, W. K., Huber, J., and Bell, J. (2008). Estimating discount rates for environmental quality from utility-based choice experiments. *Journal of Risk and Uncertainty*, *37*(2-3), 199-220.

von Haefen, R. H., Massey, D. M., and Adamowicz, W. L. (2005). Serial nonparticipation in repeated discrete choice models. *American Journal of Agricultural Economics*, *87*(4), 1061-1076.

Vossler, C. A., Doyon, M., and Rondeau, D. (2012). Truth in consequentiality: theory and field evidence on discrete choice experiments. *American Economic Journal: Microeconomics*, *4*(4), 145-71.

Wielgus, J., Gerber, L. R., Sala, E., and Bennett, J. (2009). Including risk in stated-preference economic valuations: Experiments on choices for marine recreation. *Journal of Environmental Management*, *90*(11), 3401-3409.

Zawojska, E., Bartczak, A., and Czajkowski, M. (2019). Disentangling the effects of policy and payment consequentiality and risk attitudes on stated preferences. *Journal of Environmental Economics and Management*, *93*, 63-84.

Appendix 1

Comparison of sociodemographic statistics for the samples (including respondents to the surface water and the groundwater questionnaire versions) and the population in each case/sample area.

Canadian case area	Sample		Population			
	#	%	#	%	χ2	p
Age					199.1	<0.0001
18-34 years	343	39.9%	2 075 580	34.0%		
35-49 years	375	43.6%	1 676 064	27.4%		
50-75 years	142	16.5%	2 355 325	38.6%		
Gender					2.98	0.0842
Female	458	53.6%	3 091 040	50.6%		
Male	397	46.4%	3 015 929	49.4%		
Other	5	0.6%				
Education					147.5	<0.0001
Primary education	13	1.5%	1 022 865	16.6%		
Secondary education	239	27.8%	1 624 380	26.4%		
Tertiary education	608	70.7%	3 500 690	56.9%		
Income (after tax)					70.1	<0.0001
Less than $ 10 000	18	2.3%	98 000	3.6%		
$ 10 000 - 19 999	27	3.4%	165 815	6.0%		
$ 20 000 - 29 999	44	5.6%	213 525	7.8%		
$ 30 000 - 39 999	37	4.7%	235 830	8.6%		
$ 40 000 - 49 999	48	6.1%	246 385	9.0%		
$ 50 000 - 59 999	78	10.0%	234 215	8.5%		
$ 60 000 - 69 999	74	9.5%	215 140	7.8%		
$ 70 000 - 79 999	75	9.6%	199 610	7.3%		
$ 80 000 - 89 999	70	8.9%	178 730	6.5%		
$ 90 000 - 99 999	63	8.0%	154 350	5.6%		
$ 100 000 - 124 999	102	13.0%	295 485	10.8%		
$ 125 000 - 149 999	66	8.4%	190 490	6.9%		
$ 150 000 or more	81	10.3%	314 000	11.5%		
Not stated	77	9.0%				

Note: For gender and income, percentages calculated against the sum of respondents not reporting "other" or "not stated", respectively. For "other" and "not stated", percentages calculated against the total of respondents in that region sample.

For income, the population counts refer to the number of households.

Population statistics refer to the population in the sampled Census Metropolitan Areas.

Population statistics extracted from Statistics Canada's "2016 census" and "Population estimates, July 1".

Danish case area	Sample		Population			
	#	%	#	%	χ2	p
Age					7,27	0.0264
18-34 years	230	30.0%	252 300	28.1%		
35-49 years	215	28.1%	225 851	25.2%		
50-75 years	321	41.9%	418 753	46.7%		
Gender					86.2	<0.0001
Female	504	65.8%	439 729	49.0%		
Male	262	34.2%	457 175	51.0%		
Other	0	0.0%				
Education					260.6	<0.0001
Primary education	82	10.7%	235 994	27.8%		
Secondary education	280	36.6%	376 323	44.3%		
Tertiary education	404	52.7%	236 768	27.9%		
Income (after tax)					194.8	<0.0001
Less than 200 000 kr.	89	14.1%	180 873	28.7%		
Kr. 200 000 - 299 999	89	14.1%	138 708	22.0%		
Kr. 300 000 - 399 999	105	16.6%	88 423	14.1%		
Kr. 400 000 - 499 999	99	15.6%	67 511	10.7%		
Kr. 500 000 - 599 999	70	11.1%	59 291	9.4%		
Kr. 600 000 - 699 999	66	10.4%	41 339	6.6%		
Kr. 700 000 - 799 000	34	5.4%	23 835	3.8%		
Kr. 800 000 - 899 000	30	4.7%	12 135	1.9%		
Kr. 900 000 - 999 000	18	2.8%	6 211	1.0%		
1 million kr. or more	33	5.2%	10 997	1.7%		
Not stated	133	17.4%				

Note: For gender and income, percentages calculated against the sum of respondents not reporting "other" or "not stated", respectively. For "other" and "not stated", percentages calculated against the total of respondents in that region sample.

For income, the population counts refer to the number of households.

Population statistics refer to the population in the sampled municipalities.

Population statistics extracted from Statistics Denmark's "FOLK1A", "HFUDD11" and "INDKF132".

Portuguese case area	Sample		Population			
	#	%	#	%	χ2	p
Age					316.0	<0.0001
18-34 years	281	37.1%	1 080 046	25.0%		
35-49 years	370	48.8%	1 249 677	28.9%		
50-75 years	107	14.1%	1 991 240	46.1%		
Gender					0.04	0.8363
Female	399	52.6%	2 258 291	52.3%		
Male	359	47.4%	2 062 672	47.7%		
Other	0	0.0%				
Education					1 186.2	<0.0001
Primary education	21	2.8%	2 718 100	57.8%		
Secondary education	265	35.0%	1 085 500	23.1%		
Tertiary education	472	62.3%	895 600	19.1%		
Income					395.3	<0.0001
Less than 10 000 €	101	16.1%	1 281 033	44.7%		
€ 10 000 - 19 999	219	34.9%	882 976	30.8%		
€ 20 000 - 29 999	141	22.5%	471 914	16.5%		
€ 30 000 or more	166	26.5%	231 285	8.1%		
Not stated	131	20.9%				

Note: For gender and income, percentages calculated against the sum of respondents not reporting "other" or "not stated", respectively. For "other" and "not stated", percentages calculated against the total of respondents in that region sample.

For income, the population counts refer to the number of households.

Population statistics refer to the population in the NUTS regions that cover the sampled areas.
Population statistics extracted from Statistics Portugal's "Annual estimates of resident population", "Labour force survey" and "Income Statistics at local level".
The population income data are based on somewhat different income intervals than those presented here, as it have not been possible to find data for income intervals that exactly fits the data elicited in this survey. The population data are thus based on the income intervals: "Less than 10 000 €", "€ 10 000 - 18 999", "€ 19 000 - 32 499", "€ 32 500 or more".

Swedish case area	Sample		Population			
	#	%	#	%	χ2	p
Age					194.2	<0.0001
18-34 years	321	41.2%	436 265	33.3%		
35-49 years	349	44.7%	381 881	29.1%		
50-75 years	110	14.1%	492 609	37.6%		
Gender					1.90	0.1682
Female	404	52.1%	649 993	49.6%		
Male	372	47.9%	660 762	50.4%		
Other	4	0.5%				
Education					88.1	<0.0001
Primary education	30	3.8%	205 504	16.2%		
Secondary education	344	44.1%	478 346	37.8%		
Tertiary education	406	52.1%	582 782	46.0%		
Income (before tax)					51.8	<0.0001
Less than 100 000 kr.	63	9.4%	153 718	12.7%		
Kr. 100 000 - 199 999	71	10.6%	171 317	14.2%		
Kr. 200 000 - 299 999	83	12.4%	197 879	16.4%		
Kr. 300 000 - 399 999	134	19.9%	241 847	20.0%		
Kr. 400 000 - 499 999	122	18.2%	189 856	15.7%		
Kr. 500 000 - 599 999	80	11.9%	104 429	8.6%		
Kr. 600 000 - 799 999	82	12.2%	89 316	7.4%		
Kr. 800 000 - 999 000	21	3.1%	29 830	2.5%		
1 million kr. or more	16	2.4%	30 232	2.5%		
Not stated	108	13.8%				

For gender and income, percentages calculated against the sum of respondents not reporting "other" or "not stated", respectively. For "other" and "not stated", percentages calculated against the total of respondents in that region sample.

Population statistics refer to the population in the sampled municipalities.
Population statistics extracted from Statistics Sweden's "Population statistics", "Educational attainment of the population" and "Income and tax statistics".

Appendix 2

Expected utility theory deals with the analysis of situations where individuals must make a decision without knowing which outcomes may result from that decision. The theory assumes that in such situations, individuals base decisions on the probabilities and utilities derived from different outcomes. In terms of our study, the individual's decision would thus be based on the following (here we disregard the time lag):

$$E(U) = p_{PQ} * U_{PQ} + p_{MQ} * U_{MQ} + p_{GQ} * U_{GQ} \qquad \text{(A1)}$$

where p denotes the probability of a specific outcome, U denotes the utility derived from this outcome, and the subscripts PQ, MQ and GQ denotes poor quality, moderate quality and good quality, respectively. Our survey design allows estimation of the effect of outcome uncertainty on WTP estimates for water quality improvements. If the respondents' decisions are in line with expected utility theory, we should find the following to be true regarding an improvement to good water quality:

$$-10\% * WTP_{GQ} = RISK10\%_{GQ} \qquad \text{(A2)}$$

$$-40\% * WTP_{GQ} = RISK40\%_{GQ} \qquad \text{(A3)}$$

where RISK10% and RISK40% denotes the estimates for a 10 % and 40 % risk of no water quality improvements, respectively. These risk estimates are likely to depend on the water quality level (hence the subscript added to this parameter), yet as we do not include interaction effects in our models, we have only one estimate representing all water quality levels. In order to investigate if (A2) and (A3) is true, below we calculate the impact of outcome uncertainty on total WTP as implied by our risk estimates. We do this in relation to an improvement to good water quality in all the surface water bodies included in our case areas. Similar calculations for smaller improvements, would lead to higher impacts on total WTP. We do not make the calculations for the results from our groundwater models, as we do not have exact WTP estimates for improvements from the poor quality in the BAU scenarios.

Canadian case area	WTP estimate	Impact on total WTP
Upper Grand, Good WQ	39	
Lower Grand etc., Good WQ	160	
Total case area, Good WQ	*199*	
10 % risk of no improvement	17	9%
40 % risk of no improvement	-80	-40%

Danish case area	WTP estimate	Impact on total WTP
Rest of Limfjorden, Good WQ	338	
Skive Fjord etc., Good WQ	130	
Total case area, Good WQ	*468*	
10 % risk of no improvement	-42	-9%
40 % risk of no improvement	-86	-18%

Portuguese case area	WTP estimate	Impact on total WTP
Lower Dão, Good EP&PC	27	
Upper Mondego, Good EP&PC	42	
Lower Mondego, Good EP&PC	148	
Total case area, Good EP&PC	*217*	
10 % risk of no improvement	-17	-8%
40 % risk of no improvement	-98	-45%

Swedish case area	WTP estimate	Impact on total WTP
Eastern Mälaren, Good WQ	80	
Western Mälaren, Good WQ	123	
Hjälmaren, Moderate WQ	268	
Total case area, Good WQ	*471*	
10 % risk of no improvement	-42	-9%
40 % risk of no improvement	-114	-24%