Organisation for Economic Co-operation and Development

ENV/CBC/MONO(2022)42

Unclassified

English - Or. English

2 December 2022

ENVIRONMENT DIRECTORATE
CHEMICALS AND BIOTECHNOLOGY COMMITTEE

Valuing the Impacts of Chemicals on Environmental Endpoints: A Scoping Study

Series on Risk Management
No. 75

JT03509287

OECD Environment, Health and Safety Publications
Series on Risk Management
No. 75

Valuing the Impacts of Chemicals on Environmental Endpoints: A Scoping Study

IOMC

INTER-ORGANIZATION PROGRAMME FOR THE SOUND MANAGEMENT OF CHEMICALS

A cooperative agreement among **FAO, ILO, UNDP, UNEP, UNIDO, UNITAR, WHO, World Bank and OECD**

Environment Directorate

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT

Paris 2022

About the OECD

The Organisation for Economic Co-operation and Development (OECD) is an intergovernmental organisation in which representatives of 38 industrialised countries in North and South America, Europe and the Asia and Pacific region, as well as the European Commission, meet to co-ordinate and harmonise policies, discuss issues of mutual concern, and work together to respond to international problems. Most of the OECD's work is carried out by more than 200 specialised committees and working groups composed of member country delegates. Observers from several countries with special status at the OECD, and from interested international organisations, attend many of the OECD's workshops and other meetings. Committees and working groups are served by the OECD Secretariat, located in Paris, France, which is organised into directorates and divisions.

The Environment, Health and Safety Division publishes free-of-charge documents in twelve different series: **Testing and Assessment; Good Laboratory Practice and Compliance Monitoring; Pesticides; Biocides; Risk Management; Harmonisation of Regulatory Oversight in Biotechnology; Safety of Novel Foods and Feeds; Chemical Accidents; Pollutant Release and Transfer Registers; Emission Scenario Documents; Safety of Manufactured Nanomaterials;** and **Adverse Outcome Pathways**. More information about the Environment, Health and Safety Programme and EHS publications is available on the OECD's World Wide Web site (www.oecd.org/chemicalsafety/).

This publication was developed in the IOMC context. The contents do not necessarily reflect the views or stated policies of individual IOMC Participating Organisations.

The Inter-Organisation Programme for the Sound Management of Chemicals (IOMC) was established in 1995 following recommendations made by the 1992 UN Conference on Environment and Development to strengthen co-operation and increase international co-ordination in the field of chemical safety. The Participating Organisations are FAO, ILO, UNDP, UNEP, UNIDO, UNITAR, WHO, World Bank and OECD. The purpose of the IOMC is to promote co-ordination of the policies and activities pursued by the Participating Organisations, jointly or separately, to achieve the sound management of chemicals in relation to human health and the environment.

This publication is available electronically, at no charge.

Also published in the Series on Risk Management link

**For this and many other Environment,
Health and Safety publications, consult the OECD's
World Wide Web site www.oecd.org/chemicalsafety/**

or contact:

**OECD Environment Directorate,
Environment, Health and Safety Division
2, rue André-Pascal
75775 Paris cedex 16
France**

Fax: (33-1) 44 30 61 80

E-mail: ehscont@oecd.org

Foreword

During the implementation of the valuation work on health impacts of chemicals, countries have continued to indicate an interest in working collaboratively on the valuation of environmental endpoints, recognising that methodological work is needed prior to embarking on any collective survey work. This scoping study aims to identify how to potentially advance towards surveys of willingness-to-pay to avoid negative chemicals-related environmental impacts to inform chemicals regulation.

This draft document was developed by Dennis Guignet and Adan L. Martinez-Cruz and benefited from the review and input of the Surveys of willingness-to-pay to avoid negative chemicals-related health effects (SWACHE) Project Expert Group, the Working Party on Risk Management and OECD secretariat. The document will inform discussions on next steps of the work at the OECD.

The Working Party on Risk Management endorsed the scoping study for publication at their meeting in September 2022 and the document is published under the responsibility of the Chemicals and Biotechnology Committee.

Executive Summary

The objective of this scoping study is to discuss and assess potential paths to advance towards a stated preference (SP) protocol with the aim of estimating the benefits of improvements in environmental endpoints due to chemical management and regulatory decisions. This scoping study is meant as a starting point for discussions on developing and testing a survey instrument. Particular attention is given to the possible generalisability of a survey, and the capacity for future benefit transfer applications. An emphasis is placed on isolating direct values for environmental endpoints and trying to minimise consideration of human health motivations (which are being valued separately under the Organisation for Economic Co-operation and Development's (OECD) Surveys of willingness-to-pay to avoid negative chemicals-related health effects (SWACHE) Project). Several key challenges are discussed, including the identification of widely applicable environmental endpoints that can be linked to human welfare, ecotoxicological models, and policy-levers; the role of scientific uncertainties in estimating ex ante benefits; and communicating baseline conditions in the face of such uncertainties. Alternative valuation questions are posed and discussed, along with practical steps to consider in advancing towards the development and testing of a SP survey instrument.

Table of Contents

Tables

Figures

Chapter 1. Introduction

Benefit-cost analysis (BCA) is used to compare the benefits and costs of a policy under consideration. The calculation of net benefits (i.e., benefits minus costs) across competing public policies allows policy makers to rank alternative options based on which one yields the highest net benefits to society. The usefulness of BCA as a tool for informing policy decisions hinges on whether the most critical benefits and costs are appropriately accounted for; something that cannot be taken for granted in a number of policy-relevant instances. Calculating benefits from improvements in environmental quality is challenging as benefits from these improvements often stem from non-market values – i.e., values assigned to goods and services that are not traded in a market. In other words, environmental changes imply improvements (or damages) that benefit (or cost) society in ways that are outside of traditional markets, and hence a market price to signal the monetized social value of an improvement (or damage) is not observed. Consequently, BCA of policies dealing with changes in environmental quality heavily depend on estimates of non-market values.

Economists implement two approaches to try and estimate non-market values –revealed preferences (RP) and stated preferences (SP) methods. On one hand, RP methods encompass strategies that take advantage of data reflecting decisions people make in existing markets that are indirectly related to the environmental amenity of interest. An instance is as follows: while people who appreciate green spaces in urban contexts cannot directly buy closeness to a green space in a market, they can buy or rent an accommodation close to a green space –and the price of such an accommodation implicitly accounts for closeness to green spaces. This is the fundamental idea behind the hedonic price method, where economists indirectly infer values from the price of related goods (i.e., an accommodation in this case). On the other hand, SP methods encompass strategies to simulate hypothetical markets for goods where there is no indirectly related market or data. For instance, there is no market where people can "buy" preservation of a certain animal species they care about. Environmental economists may develop a survey instrument to elicit monetary values of how much people are willing to pay to preserve this species based on their stated amounts or choices in a carefully crafted and tested scenario. Both RP and SP methods have their advantages and drawbacks. One comparative advantage of SP methods is that they allow to derive non-market values for the entire population and for a wider spectrum of endpoints.

It is against this backdrop that the Organisation for Economic Co-operation and Development (OECD) is interested in a scoping study on potential paths to advance towards a stated preference (SP) study to estimate the benefits of improvements in environmental endpoints due to chemical management and regulatory decisions. This scoping study is meant to be a starting point for discussions on developing and testing a survey instrument. Particular attention is given to the possible generalisability of a survey, and the capacity for future benefit transfer applications. This scoping study assesses the existing literature and proposes potential paths towards developing a SP protocol to estimate the public's willingness to pay (WTP) to avoid damages from chemicals to environmental endpoints. Importantly, the focus is on identifying SP strategies to infer benefits that can reasonably be presented as reflecting only values assigned to environmental changes directly, and that are not motivated by human health concerns. Human health endpoints are being valued through OECD's SWACHE Project (Surveys on Willingness-to-pay to Avoid Negative Chemicals-Related Health Impacts). The objective of the SWACHE Project is to use SP survey methods to estimate internationally comparable values of the health benefits from reduced exposure to toxic chemicals. The SP surveys were developed to elicit WTP values for reduced risks of asthma, fertility loss, IQ loss, chronic kidney disease, very low birth weight, hypertension, thyroid dysfunction, miscarriage, skin sensitisation and non-fatal cancer (OECD, n.d.). Survey implementation in several countries started in 2021, and is expected to be completed in 2023.

The scope of this study – identification and recommendation of SP approaches for valuation of quality changes in environmental endpoints– touches on a number of topics across disciplines. For instance, this report recurrently brings toxicity of chemicals into the conversation. How toxicity of a chemical can be determined is a matter for toxicologists; and how toxicity can or should be communicated becomes an issue requiring a conversation among, at the very least, toxicologists and non-market valuation economists. This example is particularly relevant in the context of this scoping study because the SP methods suggested here require that non-technical audiences report their willingness to contribute monetarily towards environmental improvements from reductions in chemical use, toxicity, etc. The credibility and reliability of stated WTP values heavily depends on respondents grasping technical concepts such as toxicity and implications from regulating it.

Thus, as this scoping study is meant to reach a multidisciplinary audience across natural and social sciences, clarifying what this report does and does not cover is a first step towards an interdisciplinary conversation to come. In particular, this report falls short of defining chemical characteristics with scientific precision. For instance, for purposes of this scope study, it suffices to refer to persistence, toxicity, and bioaccumulation as three different characteristics of chemicals. The reader will also notice that this report does not provide detailed descriptions of context-specific examples that can potentially be valued by the SP protocols suggested. This lack of detailed examples is a direct consequence of the aim of this study, which is to serve as a starting point for discussions among natural scientists, social scientists, and policy practitioners in order to further advance towards developing a SP survey instrument to value quality of specific environmental endpoints – coming up with a list of specific cases is a task that involves experts across several disciplines.

The focus of this report is to describe prototypes of SP protocols aiming to gather preferences of the general population including mostly non-technical audiences for improvements in quality of environmental endpoints. When advancing in the design of such SP protocols, non-market valuation economists will need to coordinate a structured interaction with natural and social scientists so that these protocols reflect two features at once: i) scientific precision in the description of features that characterise chemicals and their impact on the environment, and ii) clarity in description to general, non-technical audiences. These points are revisited in chapter 7 which discusses the role that is foreseen for experts and policy practitioners in shaping the suggested SP protocols.

This scoping study is structured as follows. Additional background is first provided in chapter 2. Key environmental endpoints to consider are then described in chapter 3, followed by the role of scientific uncertainty in chapter 4. Other more common challenges and considerations in designing a SP survey are outlined in chapter 5. Alternative SP designs are proposed in chapter 6. Chapter 7 describes the role of ecotoxicologists, related discipline experts, and policy practitioners in the development of a SP survey instrument and subsequent benefit transfer applications. A companion technical annex to chapters 6 and 7 is provided to illustrate how the proposed SP valuation questions and outputs from expert elicitation exercises can be used for benefits transfer. A structured set of next steps are then proposed in chapter 8, followed by concluding remarks in chapter 9.

Chapter 2. Background

Stated preference (SP) methods have been used extensively to value changes in environmental quality, and there are several applications specifically valuing impacts resulting from the release and clean-up of toxic chemicals (e.g., Alberini et al., 2007, 2012; Tonin et al., 2012; Alberini and Scasny, 2014; Scasny and Zverinova, 2014; IEc, 2016). The elicited willingness to pay (WTP) values for reducing such exposures, however, have been motivated primarily by changes in human health impacts. Human health motivations are one of the contributors to the total benefits of preventing and cleaning up chemical pollution. In fact, human health benefits often compose a large portion of the benefits of environmental regulations (e.g., Petrolia et al. 2021). Much less is known about the benefits due to improvements in environmental endpoints and services that are contributing to human well-being directly, and not through improvements in human health. This limited of knowledge regarding environmental endpoints is illustrated by OECD's SACAME Project (Socio-economic Analysis of Chemicals by Allowing a better quantification and monetization of Morbidity and Environmental Impacts). Several case studies were conducted, examining perfluorooctanoic acid (PFOA) (Gabbert 2018), mercury (Dubourg 2018), phthalates (Holland 2018), 1-Methyl-2-pyrrolidone (Hunt and Dale 2018a) and formaldehyde (Hunt and Dale 2018b). The case studies focused almost exclusively on human health, primarily due to the weaker quantified link between toxic chemicals and impacts on environmental amenities and ecosystem services (Navrud 2018).

Sometimes referred to as final ecosystem goods and services, environmental endpoints are environmental features that directly enter a household's utility or household production function (Boyd and Krupnick 2013, Johnston et al. 2013, Ringold et al. 2013). Thus, for purposes of this report the term environmental endpoints (or endpoints for short), refers to environmental or natural amenities and ecosystem services that humans value directly, such as water clarity, population size of an iconic species, or drinking water purification and protection from storm surges. Such environmental endpoints are not traded directly in a market, and so non-market valuation approaches like SP methods are needed. A well-developed SP study can capture values people hold for improvements in environmental endpoints due to improved aesthetics and recreation, as well as due to non-use motivations, including existence and bequest values. For example, one may experience increased welfare or satisfaction just by knowing an environmental endpoint is in good condition, and or because that endpoint is being preserved for future generations to use and experience.

There are two broader paths OECD and member countries could consider in order to estimate the value of improvements in environmental endpoints. One could conduct numerous smaller-scale and location-specific SP studies, where surveys are catered to a particular context, including the location-specific environmental endpoints that are of most importance. In contrast to human health endpoints like those being valued by the SWACHE project, there is significantly more heterogeneity across locations and contexts in terms of what environmental endpoints are present, are potentially impacted by chemicals, and that are valued by respondents. Although a survey instrument valuing a specific environmental endpoint (or endpoints) across countries may in general not be feasible due to such heterogeneity, the results of such smaller-scale case studies may be useful and generalisable in aggregate. If a large number of context-specific SP case studies are conducted, the results could later be quantitatively synthesised in a meta-analysis and subsequently generalised and transferred to policies using a function transfer approach. The U.S. Environmental Protection Agency and others often use this approach in BCAs informing policy (e.g., Corona et al., 2020). Alternatively, countries could transfer results from an individual study or set of studies that are applicable to their context and priorities. Case-specific primary valuation studies are advantageous because they can cater to location and context specific details. This includes a proper understanding and communication of baseline conditions.

The drawback of a case-study approach, however, is that findings are often not generalisable. Thus, many studies would need to be conducted in order to develop a set of primary study results of sufficient size and coverage (both geographically and in terms of relevant endpoints, scale, etc.). Developing and implementing a high-quality SP survey is costly, time-consuming, and may not be feasible in all settings. Even just measuring and communicating baseline conditions for a single case study survey would likely require significant efforts from natural scientists.

An alternative approach is to develop and implement a widely applicable SP survey, where minimal adjustments are needed when implementing the survey in one location or another. The goal of such a broader survey would not be to measure the value of all relevant endpoints in all countries, but rather to value the most relevant, broadly characterised endpoints across most countries. The advantage of a single streamlined and generalisable SP survey is that it can provide relevant information for BCAs in a cost-effective manner. The trade-off with such generalisability, however, is that important (possibly critical) location-specific characteristics may not be adequately considered. The key question is – what is acceptable when considering this accuracy versus cost trade-off across the two approaches?

If a more immediate result is needed to inform policies regulating and managing chemicals, then consideration should be given to first developing a single, widely applicable survey instrument that can be implemented across many contexts and countries, as is proposed here. This will provide a coarse, yet cost-effective transfer function to estimate benefits and inform policy decisions. Context specific case studies can then be prioritised and carried out as resources allow, with the long-term goal of developing a literature sufficient to support robust benefit transfer. Importantly, in order to pursue such a benefit transfer approach, case studies should be as specific as possible about the impacted medium (e.g., terrestrial or aquatic) and relevant environmental endpoints.

The aim of this scoping study is to examine the feasibility of, and provide a starting point for, the development of a generalisable and widely applicable SP survey to estimate WTP for the most welfare-relevant environmental endpoints affected by chemicals. That said, many of the same challenges discussed throughout this scoping study are also relevant towards more detailed case studies.

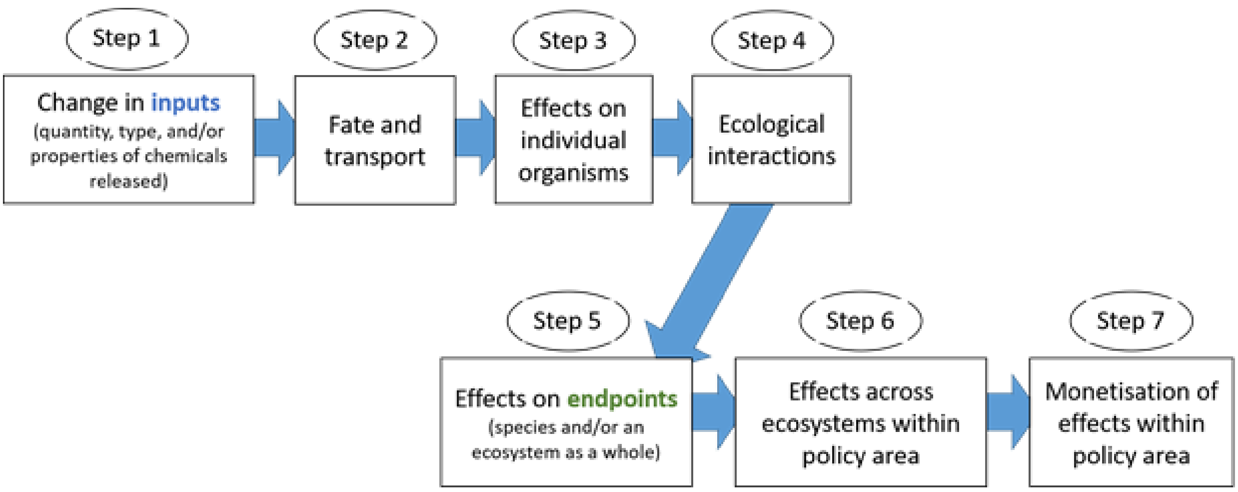
The actual development of a final survey instrument will require numerous and iterative rounds of focus group testing and cognitive interviews with participants from the general public; as well as consultations with ecotoxicologists, ecologists, and environmental risk assessors; and practitioners and policy analysts –an aspect that is discussed in detail in chapter 7. A thoroughly developed SP study can provide a useful input for BCAs of chemical policies and management decisions. Given the high costs associated with the development and implementation of a rigorous SP study, in proposing potential paths forward future benefit transfer applications are kept in mind. Particular attention is given to transferability across chemicals and geographic locations.

The need for iterative, transdisciplinary collaboration can be seen in Figure 1, which displays a stylised description of the steps needed to estimate the benefits from improvements in environmental endpoints due to changes in chemical management and regulatory actions. First, one must identify and somehow quantify the change in chemical emissions, toxicity, etc. from a regulatory intervention (step 1). To do so requires inputs from practitioners and policy experts specialised in hazardous chemicals management and regulations. One must then (ideally) use fate and transport modelling to estimate how this change in chemical inputs migrates through the physical environment and ecological systems (step 2), and ultimately impacts environmental endpoints that people directly care about (steps 3 through 5). Understanding these critical steps requires an interdisciplinary team of natural scientists, including ecotoxicologists, ecologists and environmental risk assessors. These earlier steps, along with focus group input from the general public, help economists identify the most relevant endpoints to value and how to quantify the effects of these endpoints in a scientifically rigorous manner that is, at the same time, understandable to the general public (step 5). The effects on the relevant endpoints must then be aggregated, perhaps across numerous ecosystems, to the broader area that is impacted by the policy (step 6). Again, economists can act as a bridge between natural scientists and the general public to make sure such aggregation occurs in a scientifically rigorous but also understandable manner. This is needed for the final step of monetising (i.e., assigning a monetary value to) the quantified change in environmental endpoints due to the policy intervention (step 7).

It is quite possible that not all links in Figure 1 can be made in a manner that is considered sufficiently thorough across all disciplines. Nonetheless, a key objective of the expert elicitation protocols discussed in chapter 7 is to inform the necessary assumptions to make such links in the most accurate way possible. Economists must often make assumptions when conducting BCAs, and it is best practice to make such assumptions transparent, and when necessary, conduct sensitivity analyses around key assumptions (US EPA, 2014). In short, BCAs can still be completed in a rigorous and defensible fashion, even if the current state of scientific knowledge cannot definitively establish all the necessary links in Figure 1.

Application of SP study results to BCAs of actual policy and management decisions pertains mainly to the final step of monetisation (step 7). However, in order to ensure the usefulness of an SP study for subsequent policy analysis, all steps must be kept in mind and regularly revisited when designing the SP protocol.

Figure 1. Steps to estimate environmental benefits from chemical management and polices



Chapter 3. What to Value?

There are several SP studies valuing impacts on environmental endpoints (e.g. Kosenius 2010; Lundhede et al., 2015, Moore et al., 2018; Wakamatsu et al., 2018; Lew, 2019), and credible benefit transfer from these studies to management and regulatory decisions of a particular chemical may be possible in some cases. There are several unique challenges, however, in the context of chemicals that point to the need for new, original study efforts. First, existing studies often value stocks of pollutants in the environment, and not flows of chemicals, which would be the more appropriate measure in BCAs of chemical regulations (Navrud, 2019). Second, chemicals sometimes present more persistent and toxic effects, but at the same time these effects are often not well-understood by ecotoxicologists and environmental risk assessors, at least not at the level the general public often comprehends and directly values. Understanding whether, and by how much, people value reductions in chemicals and the subsequent environmental improvements in the face of such scientific uncertainty is critical in this setting. Third, heterogeneity and uncertainty across chemicals and impacted ecosystems makes generalisability of existing study results for benefit transfer difficult in most cases. Maintaining generalisability of a SP survey instrument and the results are needed in order to inform BCAs of chemical management and regulatory decisions across OECD member countries in a cost-effective manner. Assessing the feasibility of maintaining such generalisability is a primary focus of this scoping study.

3.1. Valuing environmental endpoints versus intermediate goods or inputs

Economists refer to environmental endpoints as ecosystem services that directly enter people's utility or household production functions (Boyd and Krupnick 2013; Johnston et al. 2013). Such endpoints can include desirable aesthetics (e.g., clear water, no foul odours), recreational services (e.g., well-populated and healthy populations of game fish), and even non-use services from just knowing a particular species or an entire ecosystem is healthy and well-functioning.[1] In contrast, ecosystem inputs and intermediate goods are factors that contribute to the provision or ecological production of the resulting levels of the environmental endpoints. People do not directly value ecosystem inputs and intermediate goods. Improvements in ecosystem inputs and intermediate goods are only valued through how they affect the provision of the welfare-relevant environmental endpoints that people directly value.

Consider a stylised example where we have two ecosystem inputs – the toxicity (*tox*) and bioaccumulation (*bio*) associated with some quantity of a released chemical. Such emissions could occur at any point through the life cycle of the products using that chemical (manufacturing, consumer use, and later disposal). Putting human exposure and possible health concerns aside, people may not care about toxicity or bioaccumulation directly, but they may care about the population of an iconic bird species (*bird*) that is impacted by this chemical and its properties. The ecological production function in this simple example would be $bird=f(tox, bio)$, and a household's utility function would be specified as $U(bird)=U(f(tox, bio))$. Household utility is impacted by changes in the toxicity and bioaccumulation associated with the release of this chemical, but only through how those inputs affect the provision of the ecological endpoint – the iconic bird population in this

[1] Environmental endpoints can also include provisional services, such as erosion control, water purification, improvements in commercial fish stocks, etc., but the focus is not on such endpoints here because they can often be valued using market prices or replacement cost approaches.

instance. The magnitude by which tox and bio affect bird populations is determined by the underlying parameters of the ecological production function *f(·)*, which are based on the baseline state of the ecosystem, including stock pollution levels and other environmental factors and stressors.

This example can further be illustrated by the benefit estimation steps described in Figure 1. Step 1 entails measurement of the ecological inputs, in this case the quantity and properties of some chemical that is emitted into the environment. The ecological production function captures steps 2 through 4, which assesses the physical, chemical, and biological processes that impact various intermediate goods (e.g., plant and animal species, water quality, etc.). Ideally, the projections from the natural sciences can then be linked to make quantified projections of changes in endpoints that people directly value (steps 5 and 6), in this case the population of an iconic bird species. Fully connecting all steps in Figure 1 will require the use of any existing environmental models, and expert elicitation of ecotoxicologists and environmental risk assessors, to fill in the gaps, and finally a SP valuation component (step 7). Such linkages have been made in the context of eutrophication from nutrients pollution (e.g., Van Houtven et al., 2014), but to have not been fully implemented in the context of toxic chemicals and environmental endpoints.

Generally, it is desirable to value ecosystem endpoints because survey respondents are asked to value changes in a commodity they directly care about, which minimises the need for complex narratives explaining ecological links (Johnston et al., 2013). SP surveys asking respondents to value ecological inputs directly often require additional text explaining the ecological production function linking ecological inputs to the relevant outputs. Without such information, the survey results could be biased because (i) respondents may not be fully aware of the ecological endpoints that are impacted, and (ii) they may not accurately understand the underlying biophysical relationships (Johnston et al., 2013).

For two reasons, however, consideration could be given to valuing environmental impacts based on changes in ecosystem inputs or intermediate goods. First, ecosystem endpoints in the current context are potentially numerous, and likely location- and chemical-specific. An SP survey focusing on endpoints would need to be highly-catered to a particular study area, chemical, etc., which would deter generalisation and the potential to use the survey results for benefit transfer. For example, von Stackelberg and Hammitt (2005) examine WTP to reduce reproductive risks to wildlife due to exposure to polychlorinated biphenyls. One approach they use to value ecological risks focuses on the reproductive impacts to a single “high-profile” species, in their case Bald Eagles. Their findings are not necessarily generalisable to toxic chemicals yielding other non-reproductive ecosystem impacts, nor to locations where Bald Eagles are not impacted, or even present.

A second drawback to valuing ecological endpoints directly is the level of uncertainty and complexity in the environmental damage function involving toxic chemicals. The general lack of empirical evidence (e.g., Chiu 2017; Navrud 2017, 2018) makes a SP survey valuing endpoints potentially less useful for policy analysis, at least in the immediate term.

In contrast, a SP survey asking respondents to value ecological inputs presents several advantages. For example, ecological inputs can more closely be linked to policy, either directly through the characteristics of the policy and regulated chemical (or group of chemicals), or indirectly via the results (or future results) from environmental risk assessment models and/or expert elicitation. A focus on ecological inputs may also yield results that are more generalisable to other chemicals, locations, baseline conditions, etc., where endpoints differ. Such generalisability would facilitate future benefit transfer applications. Several studies in the toxic pollutants or similar contexts have asked respondents to value reductions in emissions (Hagan et al., 1999; King et al., 2021) or the frequency in which emissions levels exceed environmental standards (Logar and Brouwer, 2017; Logar et al., 2014). Both could be considered as ecological inputs, because these measures are at the very beginning of the ecological production function (i.e., step 1 in Figure 1).

The glaring disadvantage of valuing ecological inputs directly is that instead of concerns regarding uncertainty in the quantified scientific links to endpoints, the results would be confounded by respondents' (heterogeneous and possibly unobserved) perceptions of how changes in the ecological inputs impact endpoints they care about. This could be minimised by better informing respondents–either qualitatively, quantitatively, or both– of the ecological production function, baseline environmental state of the ecosystem, and links between changes in the intermediate good being valued and the ecological endpoints respondents directly care about (Johnston et al. 2013). Indeed, clear and concise description of such information, if it is available given the current science, requires significant testing during focus groups and cognitive interviews with the general public.[2]

Given the numerous challenges associated with valuing ecological inputs or intermediate goods, it is recommended to develop in priority a survey that focuses as close as possible to valuing ecological endpoints.

3.2. Identifying relevant environmental attributes

This section considers the environmental attributes that could be potentially valued in a valuation question or choice scenario. The term environmental attributes is used to not necessarily only reflect preference for environmental endpoints, but also to potentially include ecological inputs and intermediate goods, as well as possible dual commodities (i.e., goods that are simultaneously ecological endpoints and inputs), bundled or compound commodities, and other proxy measures.

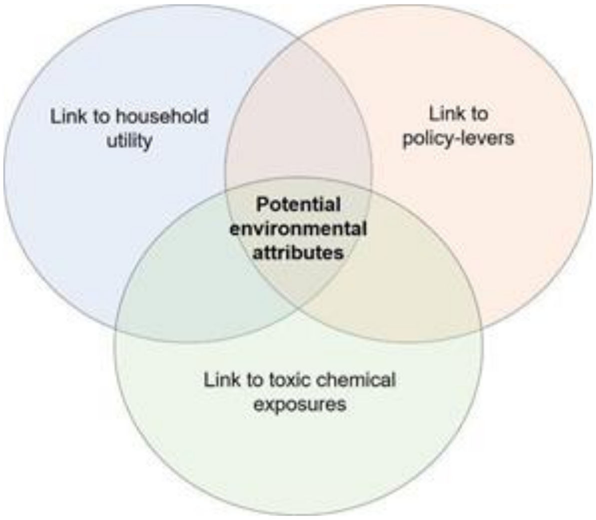
Figure 2 shows the three main criteria for the potential inclusion of an environmental attribute in a SP survey to estimate WTP to avoid environmental damages from chemicals. First and foremost, an included attribute must be directly or indirectly linked to household utility. In other words, the environmental attribute must be something that households value and care about directly, or indirectly through the ecological production function. A second requirement is that ecotoxicologists have established a (ideally quantitative) link between changes in chemical exposures and changes in the environmental attribute's provision (see Figure 1). This may be through a direct link to outputs from environmental risk assessments, or through a series of quantitative links (e.g., Forbes et al., 2017). Third, in order to be relevant for benefit transfer to BCAs of chemical management and policy decisions, the attribute should somehow be able to be linked to policy-levers. For example, suppose the use of a toxic chemical is now banned in some products, but not all. Ideally, that would quantitatively be linked to changes in the environmental attributes included in the survey.

The second and third criteria are related, but not necessarily the same. For example, it may be known households directly value an iconic fish species as an ecological endpoint, and it may also be known how a change in the concentrations of a chemical impacts that species. Therefore, this fish species attribute satisfies criteria one and two. It may not be known, however, how implementation of a new abatement technology or chemical management

[2] Debriefing questions about perceived endpoints and the magnitude of changes could also be asked, perhaps to facilitate ex post adjustments or estimation of a perceived ecological production function. This would be in addition to the standard suite of debriefing questions asking about perceived consequentiality, protest and other biassing behaviours, perceived objectivity of the survey, etc.

process will impact the quantity or concentrations of that chemical, and so the third criterion would not be satisfied. This example illustrates that, given the state of the science, identification of environmental attributes satisfying all three criteria may be quite challenging.

Figure 2. Criteria for Candidate Environmental Attributes



Linking toxic chemical exposures to changes in the environmental attribute is perhaps the most difficult requirement due to uncertainty in such links. While ecotoxicological studies examine impacts from chemicals on individual organisms or at sub-organism levels, these quantified relationships cannot necessarily be directly extrapolated when considering impacts at the population level for the same species, nor to different species or groups of species, let alone to the level of an entire ecosystem. The uncertainty involved in this extrapolation –and sometimes in the individual level studies— implies a significant gap between the outputs from current ecological risk-assessments and environmental attributes valued in SP studies (Chiu 2017; Forbes et al. 2017; Navrud 2017, 2018; OECD 2016).

Although the set of candidate environmental attributes that currently satisfy all three criteria in Figure 2 is likely small, and possibly null, decisions regarding policies and the management of toxic chemicals must be made –even decisions of no action are still policy decisions with social welfare implications.

Therefore, a key objective of this scoping study is to provide options on practical steps forward in developing the most useful stated preference survey instrument possible to help inform decision-makers. In doing so, the current uncertainties and likely future trajectories of the ecotoxicology and environmental risk assessment literature must be considered in order to identify environmental attributes that can potentially satisfy all three criteria in the future. In other words, economists should identify and conduct studies to value the most relevant environmental attributes now, while considering where the results from future ecotoxicological studies may likely later fill in the gaps (Donohue and Kipusi 2016; Navrud 2018). In the meantime, WTP estimates for improvements in environmental attributes that people care about but where a quantitative link to a chemical is not yet established are still informative – granted that a technically sound SP instrument is designed.[3] For instance, the

[3] A clear distinction should be made between what information is needed for estimating a WTP function, versus what is needed for benefit transfer and estimating the benefits of policy and management decisions. The former is one of several inputs needed for the latter. A key advantage

results can be used for illustrative “what-if” scenarios and break-even analyses, in the absence of a full-fledged BCA (US EPA, 2014). Expert elicitation can also be used to help fill in current gaps (Van Houtven et al., 2014), as discussed in chapter 7.

In this paper, potential environmental attributes are broadly categorised for inclusion in an SP survey into three categories –the environmental medium impacted, environmental quality, and the spatial extent of the environmental impacts.

3.2.1. Environmental medium

Environmental mediums affected by the release of a toxic chemical (i.e., air, water, or soil) could be varied as part of a split-sample design, or even as an attribute in a discrete choice experiment (as was done by IEc, 2016). Such considerations would, however, increase cognitive burden on respondents and necessitate a more extensive experimental design.

It may be possible to develop a generalised survey, where environmental attributes encompass impacts across mediums. As a starting point for survey development and early consultation with experts, policy practitioners, and focus groups with the general public, it is suggested to frame the initial survey scenarios in the context of both terrestrial and aquatic impacts. If needed, focus can be later narrowed depending on feedback from the experts, and assessments of what respondents care about most during focus group testing.[4]

3.2.2. Environmental quality

In reviewing the literature to identify attributes to define environmental quality, this paper mainly focused on the few stated preference studies that examined the environmental impacts of toxic chemicals or similar hazardous substances.

Logar et al. (2014) and Logar and Brouwer (2017) conduct a discrete choice experiment (DCE) SP study to estimate households’ WTP to reduce micropollutants in Swiss waterways. They include an ordinal measure of “potential environmental risk” which includes three categories based on the number or fraction of micropollutants that exceed the corresponding environmental quality standard. More specifically, “low” potential environmental risk means that 0 out of 15 micropollutants exceed the environmental quality standard, “medium” implies 1 to 3 exceed the standard, and “high” is when greater than 4 exceed the standard.

Hagan et al. (1999) conduct a contingent valuation (CV) study examining US households’ WTP for a reduction in mercury emissions in Minnesota, U.S. Environmental quality is measured directly as the percent reduction in mercury deposition, and they then qualitatively link reduced mercury to changes in environmental quality.

King et al. (2021) examine British households’ WTP for reductions in the release of microplastics into the environment. They framed a few different valuation scenarios. One scenario was a DCE where respondents trade off quality and the unit price of a cosmetic

---

of the SP survey methodology is that hypothetical situations are posed to elicit respondents’ WTP. The posed situations need not be directly based on ecotoxicological and environmental risk studies; they must only be perceived as credible by respondents from the general public. A WTP function is then estimated with the collected survey data. The parameterised WTP function can then be later used to estimate the benefits of a policy (i.e., benefit transfer). It is for this later benefit transfer exercise that the quantified changes must be based on ecotoxicological and environmental risk studies and experts.

[4] Once feedback is gathered, another possible path is to group countries and develop surveys versions based on the most policy-relevant environmental mediums.

product with reductions in microplastics used in the product and subsequently emitted pollutants.[5] Similar to Hagan et al., their environmental quality attribute was described as the percent reduction in emissions. A second valuation scenario was a dichotomous choice policy referendum, where it was simply stated that microplastic emissions would be eliminated through wastewater plant upgrades.

Overall, the described SP studies have favoured an input-based approach when addressing WTP for improvements in environmental endpoints caused by reductions in chemicals or similar materials. That is, respondents have essentially been asked to report WTP for reductions in emissions. How these reductions translate into changes in environmental quality is described qualitatively (and sometimes quantitatively), to varying degrees. Logar et al. (2014) and Logar and Brouwer (2017) implicitly incorporated factors like toxicity and sensitivity of the ecosystem in their studies because they measured contamination relative to the environmental quality standards, which are presumably based on such considerations.

In a broader SP context, not specific to hazardous chemicals, several studies value specific environmental endpoints directly (e.g. Kosenius 2010; Lundhede et al., 2015, Moore et al., 2018; Wakamatsu et al., 2018; Lew, 2019). In the context of hazardous chemicals, impacts to ecological endpoints and broader measures of water quality and ecosystem health are very site- and chemical-specific. The set of specific, welfare-relevant endpoints to value in a SP study are likely to be quite different across sites, especially at a national or multinational scale. Additionally, models to predict impacts on specific ecological endpoints in response to toxic chemicals may not yet be developed. The key point is that a SP survey valuing endpoints may not be as useful for benefit transfer to policy and management decisions regarding numerous, sometimes novel, toxic substances, and the use of such chemicals across national and multinational scales.

As summarised by Donohue and Kipusi (2016), IEc (2016) conducted a DCE study for Health Canada to estimate household WTP for improvements in human health and the environment from reduced emissions of toxic chemicals. Their choice scenarios explicitly accounted for environmental toxicity, bioaccumulation, and chemical persistence. Their framework is advantageous in that the actual chemicals are left generic, but chemical properties are included as attributes. This representation facilitates estimation of a benefits transfer function where benefit estimates could be catered to a particular policy context based on properties of that specific chemical and ecosystem (i.e., toxicity, bioaccumulation, and persistence). The drawback of IEc's survey, however, is that these attributes should be considered ecological inputs, and are thus subject to the criticisms described in section 3.1.

Nonetheless, building on IEc's (2016) survey is one potentially viable path forward. Explicitly accounting for variation in a chemical's persistence, bioaccumulation, and toxicity is in line with recommendations from a 2013 workshop sponsored by the Royal Society of Chemistry, UK Environment Agency, and UK Chemicals Stakeholder Forum (RPA, 2013). Efforts should be made, however, to frame these attributes in a more endpoint-based setting, ideally in a way that is generalisable across geographic locations. Linking such chemical properties to related endpoints in a credible way will require collaboration with ecotoxicologists, environmental risk assessors, and policy practitioners, as well as iterative focus group testing with participants from the general public.

[5] The focus of this section is on describing how previous studies have addressed environmental quality attributes. In this respect, King et al. (2021) is a useful reference. It is noted, however, that their use of a change in the unit price as the payment vehicle may inhibit subsequent welfare analysis.

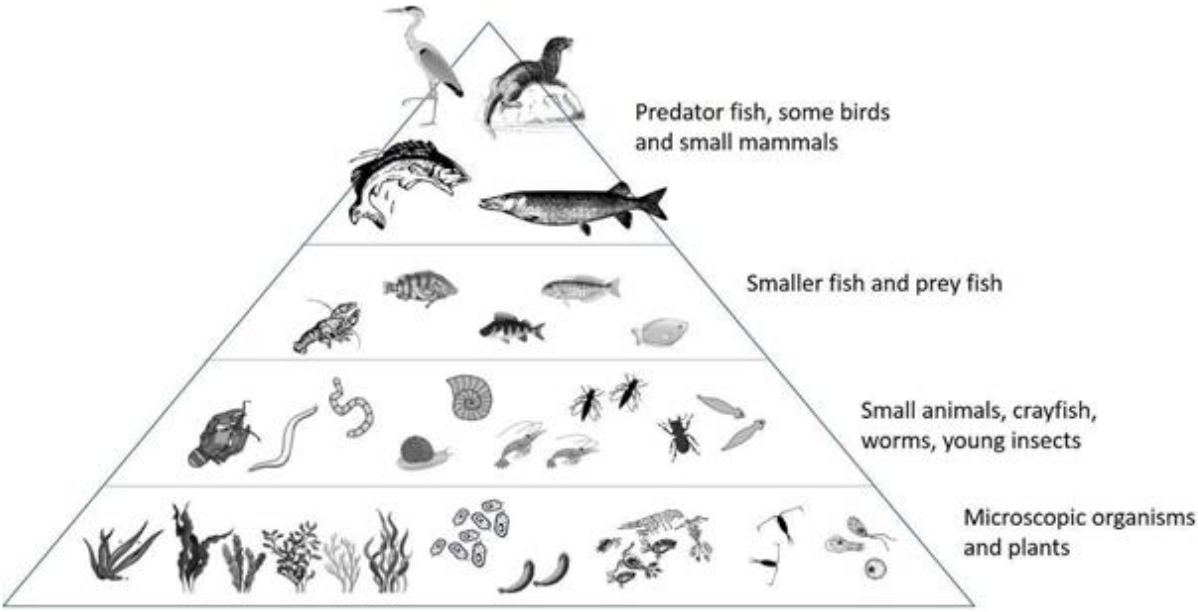
What is impacted?

Consider each of the three chemical property attributes included in IEc's (2016) survey for Health Canada – bioaccumulation, toxicity, and persistence. Bioaccumulation is measured in the IEc (2016) survey as a binary variable (i.e., bioaccumulates or does not bioaccumulate). Bioaccumulative capability is a property of a chemical or category of chemicals, and serves as an ecological input towards the "production" of environmental endpoints that respondents directly care about. What people may directly care about is what plants and animals are ultimately impacted. One issue is that bioaccumulation as an ecological input does not account for and communicate impacts due to ecological interactions. Even if a chemical does not bioaccumulate, if it degrades populations of important microorganisms, then that can lead to a chain reaction in the ecological system and still impact higher order species that households do directly value. Whether due to recreation, other resource use motivations, or non-use values, non-market valuation literature generally suggests that households have a higher WTP to maintain or preserve the well-being of higher-order species, especially iconic and keystone species (e.g., Lew, 2019; Lundhede et al., 2015; Moore et al., 2018; Morse-Jones, 2012; Wakamatsu et al., 2018).

As described in Table 1 (section 3.3), this bioaccumulation attribute can be extended to be framed more as an ecological endpoint that describes what types of microorganism, plant, and animal species are impacted. One could describe individual species, but again this would be very chemical and location specific. To facilitate generalisability and future benefit transfer, framing an attribute instead in terms of what order of species are impacted may be useful. Such effects could be communicated using similar graphics as with bioaccumulation and the food chain pyramid, as shown in Figure 3. What order species are directly and indirectly impacted could then be visually depicted in a valuation scenario (an example is presented in chapter 6).

Focus group testing is needed to see if respondents only differentiate impacts to the lowest versus the highest trophic-level organisms, or if they differentiate between impacts to trophic levels in the middle as well. If the former, then a binary measure similar to the IEc (2016) survey is sufficient, but in the latter case an ordinal or categorical variable denoting how far up the food chain pyramid impacts go would be a useful direction to test. Connecting this to chemicals regulated by policy and projections from ecotoxicologists and environmental risk assessors will require expert elicitation procedures, if quantitative forecasting models are not available. As an alternative, in future benefit transfer exercises where no such projections are available, the bioaccumulative properties of a regulated chemical or group of chemicals can inform a conservative scenario that does not account for indirect effects via ecological interactions.

Figure 3. Stylised diagram to depict directly and indirectly impacted species



How bad are the impacts?

Toxicity is described in the IEc (2016) survey as whether a chemical is harmful to microorganisms, plants, and other animals. Similar to bioaccumulation, such impacts can be direct, based on the toxicity of a chemical, quantity released, an ecosystem's capacity for natural attenuation, etc. Impacts of a chemical can also depend on indirect effects due to interactions and chain reactions in the ecosystem. Identifying an endpoint-based attribute that describes the magnitude of a change in environmental quality, and that is ideally generalisable across locations and ecosystems is perhaps the most challenging aspect in designing a survey instrument in the current context.

Current and past SP studies have used or proposed several different composite, endpoint-based measures. Vossler et al. (2022) recently proposed the use of biological condition gradients (BCGs). BCGs describe overall ecosystem integrity on a 1 (pristine) to 6 (severely degraded) scale. Moving to higher levels of degradation is associated with a decrease in sensitive species, and possibly an increase in more tolerant or invasive species. The BCG is analogous to a field-based dose-response curve, measuring how biological condition is affected by anthropogenic stressors (US EPA, 2016). The US EPA (2021) recently proposed a SP study that will use O/E ratios to measure overall ecosystem quality. The O/E ratio is the ratio of different observed macroinvertebrate species over the expected number under the least disturbed conditions at reference sites in the corresponding ecoregion. Van Houtven et al. (2014) implement a CV study where they take an endpoint-based approach and ask respondents to report WTP for improvements in lakes, which are measured by a eutrophication index constructed by expert judgements. Johnston et al. (2013) use a 0-100 index of biotic integrity (IBI) to reflect overall aquatic ecosystem health. Similar to a BCG, IBIs are meant to characterise the overall condition or naturalness of an ecosystem relative to an undisturbed reference site, and are typically a composite of numerous measures (e.g., species composition, trophic role, reproductive strategy and the abundance/condition of individual organisms). In focus groups, Johnston et al. found participants valued the IBI directly as a holistic measure of ecosystem well-being. A similar argument can potentially extend to the other indices and composite measures like the BCG, eutrophication index, and O/E ratios.[6]

All of the above examples are in the context of nutrient pollution and are not specific to toxic chemicals. Outputs from ecotoxicological studies of toxic chemicals are different, and likely more heterogeneous across different chemicals and locations. Nonetheless, the above examples are useful in illustrating the types of potentially generalisable endpoints that have been used. In addition, to tackle the challenge of linking reductions in emissions and representations of improvements in ecosystem quality, Van Houtven et al. (2014) use an expert elicitation protocol that may serve as a useful template in the current context (see chapter 7).

In terms of potential composite measures in the context of chemicals, two potential measures to explore were identified – risk quotients and species sensitivity distributions. Exploration of whether either of these measures are potentially appropriate requires extensive consultation with ecotoxicologists, environmental risk assessors and policy practitioners.

Chiu (2017) discusses a main output from ecotoxicological studies as the predictive environmental concentration (PEC). The PEC by itself, however, is not that useful because the harmfulness of a chemical, holding volume or concentration constant, varies across chemicals and environments. Chiu (2017) and Navrud (2019) also describe how ecotoxicologists often focus on predicted no effect concentrations (PNEC) when examining chemicals. In other words, at what concentration of a specific chemical, in a particular ecosystem, do we begin to see adverse impacts on a specific organism. The PNEC is often based on when adverse effects are observed on an organism in terms of growth, survival, or reproduction. Together, the PEC and PNEC could provide a useful measure of potential harm to the ecosystem. Some case studies define a risk quotient as the ratio of PEC/PNEC. In this paper, it is interpreted that the risk quotient normalises the concentration levels across different chemicals and environments such that a value less than one signals little adverse impact on the ecosystem, and a value greater than one implies there are potential adverse impacts.

Advantages of the risk quotient are that it directly links to outputs from ecotoxicological studies, and the link to specific chemicals and ambient water concentrations connects it with policy-levers. A key disadvantage is that it provides only a relative measure of harm to the ecosystem, and how the magnitude of such a measure should be communicated to and interpreted by respondents is unclear. Another challenge is in translating the risk quotient into a welfare-relevant attribute (Chiu, 2017). In any subsequent SP survey, based on step 6 in Figure 1, the measure would likely be conveyed as an ecosystem-level risk quotient, and perhaps even a summary measure (e.g., average) across ecosystems in a sometimes large jurisdiction (e.g., a municipality, country, or even a multinational region). In contrast, ecotoxicologists tend to measure environmental impacts from toxics at the individual animal- or organ-level (Chiu 2017; Navrud 2019). Another disadvantage of the

[6] Other SP studies examining water quality more generally have used other composite endpoint measures, such as a water quality ladder linked to designated uses (boatable, swimmable, and drinkable) and related indices (e.g., Bateman et al., 2011; Brouwer et al., 2010; Carson and Mitchell, 1993; Choi and Ready, 2021; Hampson et al., 2017; Schaafsma et al., 2012; US EPA, 2021; Walsh et al., 2022). Walsh and Wheeler (2013) find that the use of different functional forms for such indices can yield large differences in BCA results. Additionally, impacts on recreational uses like whether a waterbody is "safe" for boating, swimming, and or drinking are more health-based, and could be considered outside of the current task of trying to isolate WTP estimates for environmental endpoints that are unrelated to human health.

risk quotient is that the PNEC value is conservative in order to better ensure environmental protection in chemical risk assessments, whereas for purposes of analysing ecological impacts from the life cycle of a chemical, more robust and less conservative measures should be used (Henderson et al, 2011; Fantke et al., 2018).

With these critiques in mind, a second option that is specific to toxic chemicals is the use of species sensitivity distributions (SSDs). von Stackelberg and Hammitt (2005, 2009) focus on polychlorinated biphenyls and explore the use of species sensitivity distributions, which quantify the proportion of species that will be affected with some probability (e.g., there is a 25% probability that 60% of the species will experience some type of adverse effects). This measure is beneficial because it is a widely used output of ecotoxicological studies (Fox et al., 2021; Henderson et al., 2011; Posthuma et al, 2019; Xu et al., 2015), can account for uncertainty in the science, is a concept that is potentially applicable across locations, and describes the magnitude of the impact (at least in terms of quantity or proportion of species impacted). SSDs can also be used to describe impacts at a broader ecosystem level, and there are efforts underway to globally standardise the modelling approaches underlying SSDs (Fantke et al., 2018). The main drawback is that it is difficult for respondents to comprehend a measure with nested probabilities and proportions (von Stackelberg and Hammitt 2005).

In consultation with ecotoxicologists, environmental risk assessors, and policy practitioners, consideration should be given to composite endpoint-based measures like those described above. Care must be taken as to how such measures would be linked to policy and impacts projected by environmental models and/or expert elicitation. This paper suggests the use of species sensitivity distributions as a potential starting point for further discussions. Fantke et al. (2018) describe an additional step of translating the potentially affected fraction (PAF) of species from SSDs to a more field-relevant measure: the potentially disappeared fraction (PDF) of species – i.e., X% of the species will disappear from this ecosystem. Such a measure may serve as a more salient and important endpoint for the general public. Similar metrics based on the number of species preserved or lost, and or the percentage of species relative to current conditions or some other reference level, have been used and tested in the SP literature on biodiversity and ecosystem services (e.g., Johnston et al., 2012, 2013; Morse-Jones et al., 2012; Parsons and Thur, 2008; Breeze et al., 2015; US EPA, 2021). Given the comprehension difficulties reported by von Stackelberg and Hammitt (2005), extensive focus group testing and further development of a survey instrument and communication tools are needed.

One potential approach to simplify communication of the PDF of species may be to use qualitative measures of uncertainty when describing the likelihood of a disappearance of X% of species (e.g., “almost certain”, “likely”, “unlikely”). These qualitative statements can then be linked to quantitative, respondent-specific subjective probabilities, which can be elicited following a longstanding psychology and economics literature. Wallsten et al. (1986) carried out pioneer studies documenting that qualitative statements such as “likely” are interpreted differently by individuals, and a number of strategies have been developed since to gain insights into subjective perceived uncertainty (see Manski (2004) for references in the economic literature). The estimated subjective probabilities could then be explicitly used in modelling respondents’ choices. That said, communication and respondent understanding of quantitative probabilities nested with proportions or percentages (of species) impacted may be possible. For example, Atherton et al. (2020) conduct a choice experiment on persistent chemicals used as flame retardants, and present respondents with both the percent of waters cleaned up, and the probability of whether the contaminant is “safe”. Both are presented as percentages, along with some qualitative and graphical descriptions. The main point is that there are options to consider, and SP survey graphics, electronic modes, and communication techniques have improved substantially since von Stackelberg and Hammitt's (2005) study.

The environmental quality metric that ultimately should be selected, and any corresponding measure of uncertainty, should arise from technically sound consultations with ecotoxicologists and be carefully tested on a non-technical, general audience.

How long do the impacts last?

The last chemical attribute considered in IEc's (2016) survey for Health Canada is persistence, which they describe as how long a chemical remains in the environment, until it disappears through degradation via natural processes. Persistence is still largely focused on properties of the chemical itself, and therefore describes an ecological input (steps 1 and 2 of Figure 1). A respondent may not just care about how long a chemical remains in the environment, but how long it impacts the endpoints they care about. For example, a chemical that tends to persist for just a few months could still have long lasting impacts on the environment due to subsequent ecological interactions. This paper suggests to reframe this persistence attribute as the overall duration of harmful effects (see Table 1). In the absence of formal environmental models or expert elicitation procedures to project the duration of environmental impacts, persistence levels can be used for categorisations for different groups of chemicals to provide a conservative measure for benefit transfer that does not account for ecological interactions.

The appropriate duration levels for a subsequent experimental design should be chosen to reflect chemicals that may dissipate rather quickly, as well as those that could take a generation, two generations, or even never, to naturally attenuate (i.e., "forever" chemicals). The important thing for survey development is to make sure that the experimental design covers the relevant attribute space that experts could foresee in future policy applications. That way the survey results will still have utility for benefit transfer as expert judgements are updated, models are developed/refined, etc.

3.2.3. Spatial extent of environmental impacts

In terms of facilitating benefit transfer across policy contexts, allowing for variation in scale (e.g., the geographic scope of impacted ecosystems) is a worthwhile avenue to explore. The spatial extent of the ecosystem impacts is based on where a toxic chemical is emitted, the quantity that is emitted, the area or jurisdiction corresponding to the posited policy or management change, as well as the mobility of the posited chemical, and subsequent fate and transport modelling results (step 2 in Figure 1). Navrud (2017) emphasises that:

> "New primary valuation studies should be designed with value transfer in mind, and cover several countries, in order to extrapolate and generalise the values to evaluate international chemical regulations. These new primary studies should ideally also cover all relevant scales of the impacts, in order to develop generalised adjustment factors for differences in scale of the impacts between the study site and policy site" (page 4).

The geographic scale of the environmental impacts could be an attribute that is explicitly varied in a valuation question or a split-sample design. If respondents are asked questions pertaining to environmental improvements at a local, regional, national, and or international level, for example, then this would yield a benefit transfer function that could be explicitly adjusted for scale. Logar et al. (2014) and Logar and Brouwer (2017) implement a DCE study to examine WTP for reductions in micropollutants in waterbodies in Switzerland. They vary the geographic scale of the posited environmental improvements as affecting only waters in the canton (or state) where a respondent lives, or for the entire country. As economic theory would suggest, they find a higher WTP for policies that lead to environmental improvements over a larger spatial scale, all else constant.[7]

Accounting for geographic scale using an ordinal variable (e.g., subnational, national, continental) is advantageous in that the actual survey instrument may be more directly applicable to different study areas, and it may be more salient to the general public. The main disadvantage in conducting subsequent benefit transfer to policy or management decisions is that you would be implicitly assuming that un-controlled for characteristics across the study and policy sites are similar, including baseline conditions and the quantity of impacted aquatic and/or terrestrial ecosystems.

A more continuous, quantitative measure would be better for purposes of benefit transfer because the resulting transfer function could cater analyses to more location specific information, such as the quantity of impacted waterbodies at varying geographic scales. Johnston et al. (2013), for example, includes the surface area (acres) of waters improved in their SP study of fish restoration projects in part of the northeast U.S. A proposed SP study by the U.S. Environmental Protection Agency (US EPA, 2021) lays out a similar approach, where the improved waters are varied as an attribute ranging from 3% to 100% of all lakes, rivers, and streams in the U.S. In the context of this scoping study, one could use Geographic Information Systems (GIS) to derive continuous measures of the quantity of waters and acres of impacted terrestrial ecosystems. Some drawbacks of the approach are that survey instruments would have to be tailored to each specific location where they would be implemented, the measures may not be easily understandable to the general public, and what measures of quantity are most relevant to people is unclear, and perhaps context specific (e.g., number of lakes, surface area of lakes, kilometres of river).

All things considered, it is recommended to use a simple ordinal variable to describe spatial scope, such as: subnational (e.g., province, district, state), national, and continental (or sub-continental). There are three main advantages in implementing this strategy. First, it eases communication of scale –people do generally think in terms of province, state, nation, continent, etc. Second, it allows for a survey instrument that can be implemented across different countries with minimal adjustment. Third, it enables benefit transfer to likely policy-relevant boundaries. Policy practitioners and ecotoxicologists must be consulted, however, to ensure that the range of geographic scales presented encompass the range that is relevant to the environmental impacts and policy decisions. For example, if policy decisions are often at a national-level, national changes should be valued, or perhaps even broader impacts if fate and transport models suggest cross-jurisdictional impacts. On the other hand, if the environmental impacts are often very local in nature, then a more local scale is also necessary.

This emphasises the importance of including sufficient variation a priori in the experimental design. Doing so will allow for estimation of a benefit transfer function that encompasses the relevant attribute space for this spatial extent dimension. When later conducting benefit transfer, predictions from fate and transport models, expert elicitation, and based on the mobility of a chemical, can be used to determine the appropriate value to plug in for the spatial extent variable in the parameterized transfer function.

---

[7] A related and equally important variable as geographic scale is distance to the impacted resources. Attempting to disentangle how preferences vary with spatial scale and distance to the resource is a policy-relevant and active area of budding research (e.g., US EPA, 2021). It is not recommended pursuing such an approach here because these approaches are still new and not yet tested, and would substantially increase the complexity of an experimental design where there are higher priority objectives.

One could consider supplementing the SP data ex post with GIS and remote sensing information on the actual quantity of waters or land area affected, and explore the use of that objective quantity measure in subsequent choice model estimation. The assumption in doing so would be that the objective quantity measure is a reasonable proxy for the quantities perceived by respondents. At the very least, such data would serve as a useful sensitivity analysis to examine whether heterogeneity in the quantity of waters within a spatial unit is important, holding the geographic scale constant.

An additional benefit of including an explicit spatial scale variable in general is that it allows for an easy test for scope sensitivity, which serves as a useful check for internal validity. Although tests for scope may be more informative and are generally more desirable in the quality dimension, doing so in the "quantity" dimension will still allow for a test for scope sensitivity. All else constant, economic theory suggests that a household holds a higher WTP for improvements to a greater number of waterbodies, acres of an ecosystem, etc.

3.3. Proposed starting point for a generalisable set of environmental endpoints

Table 1 displays a proposed starting point for potential environmental endpoints to consider for a survey instrument that is widely applicable in terms of both implementation and later benefit transfer.[8] This set of environmental attributes and the corresponding descriptions could serve as initial draft text for a potential survey. Collaboration with practitioners, ecotoxicologists, environmental risk assessors, and related discipline experts is needed to ensure the final set of attributes can (now or in the future) satisfy all criteria described in Figure 2, and thus fulfil all necessary steps described for policy-relevant benefits estimation and benefit transfer outlined in Figure 1. Such collaboration will need to occur in sequence with iterative focus groups with members of the general public. In any subsequent survey instrument, the text in Table 1 would be accompanied by further details regarding the ecosystem type and habitat, the species affected, the geographical scale, and the specific location and duration of harmful effects, etc.

It is emphasised that Table 1 is merely a proposed starting point for later discussions. The proposed attributes could change, be revised, and or be condensed depending on feedback from focus groups of the general public, and advice from policy practitioners and discipline experts. For example, ecotoxicologists may suggest some attributes are redundant, policy practitioners may have opinions on higher or lower priorities for benefit estimation and focus group participants from the general public may view some attributes as more welfare-relevant than others or simply find some attributes too complicated to comprehend.

[8] There are many other environmental attributes that could be considered, but that were not pursued in great depth in this scoping study. For example, policy-relevant temporal lags of up to a few decades between exposure and environmental impacts may sometimes be present, and are likely important for benefits estimation (Navrud, 2018). Adding an additional temporal dimension, however, will further increase the cognitive burden on respondents, and given the uncertainty in environmental impacts from most chemicals (Chiu 2017; Navrud 2017, 2018), any associated time lags are also likely to be very uncertain.

Table 1. Potential environmental endpoints to consider

Attribute	Definition
What is impacted?	What types of organisms, plants, and animals are impacted. A chemical's impact depends on several factors: • Bioaccumulation - Some chemicals accumulate in the tissues of plants and animals, starting with smaller organisms. These organisms are then eaten by larger organisms, and those organisms are then eaten by even larger animals, and so on. A chemical has a high level of bioaccumulation if it is passed up the food chain and remains at high concentrations in larger animals (e.g., fish and birds). • Ecological interactions - Even if not directly impacted by a chemical, plants and animals in an ecosystem can be impacted indirectly if a chemical negatively affects a key microorganism, plant, or animal, that other species in the ecosystem depend on. Such impacts are sometimes measured in terms of the order or category of organisms that are impacted. Lower order organisms are near the bottom of the food chain pyramid, and higher order organisms are near the top. What order or category of microorganisms, plants, and animals that are impacted depends on bioaccumulation and ecological interactions.
How bad are impacts?	How bad the impacts are on the affected microorganisms, plants, and animals depends on several factors: • Toxicity - A chemical is more toxic as its adverse impacts on the environment increases; this includes reduced rates of survival and the ability for organisms to reproduce. • Quantity of chemical in the environment - The greater the amount of a chemical that is in the environment, the worse its impacts will be. • Ecological interactions - Even if not directly impacted by a chemical, plants and animals in an ecosystem can face a greater indirect threat if a chemical has a greater impact on key microorganisms, plants, or animals, that other species in the ecosystem depend on. One way scientists measure ecological impacts is by the proportion of species in that ecosystem that will potentially disappear as a result of the release of a toxic chemical. For example, 3 out of 10 (or 30%) of the species in an ecosystem will disappear.
How long do impacts last?	How long the harmful effects on the environment would last if environmental exposures to a chemical stop in the future. • A chemical can remain in the environment until it disappears through degradation via natural processes. A chemical is persistent if it takes a relatively long time to disappear. • Even after a chemical disappears, the negative effects on the ecosystem can remain for several years due to ecological interactions across different microorganisms, plants, and animals. Scientists usually measure how long ecological impacts last in terms of years.
Where do these impacts occur?	Where or in what areas will the impacts on the environment be experienced. What areas are impacted by a chemical depend on: • Where the chemical is being released into the environment, and how much of the chemical is being released. • How mobile is the chemical. In other words, does the chemical tend to stay in one place or does it easily travel with the wind, in water, or with animals as they migrate. Scientists measure the location and area that will be impacted by a chemical in terms of geographic location, such as a local, national, or international area.

Chapter 4. Risk and Uncertainty

Risk and uncertainty are two concepts that must come into the conversation in this scoping study. Risk refers to a situation that can be characterised in probabilistic terms –i.e. a situation under which a probability of occurrence can be assigned to each and all potential outcomes– and, in this sense, a decision maker can form expectations. In contrast, uncertainty is a situation under which such information is not available to the decision-maker and, therefore, one can only form vague expectations –see Park and Shapira (2017) for further details. Indeed, there are situations for which, while a distribution of outcomes is known, the probabilities of those outcomes are not. This situation is referred to as risk ambiguity (Backus et al., 2015).[9]

To illustrate the difference between risk and uncertainty, let us refer to Table 1 above. In this table, the attributes that a SP protocol may take into consideration are discussed. In particular, when it comes to how bad the impacts from a chemical are or can be, scientists may be able to express the probability that a given proportion of species in an ecosystem will disappear as a result of the release of a toxic chemical. Although motivated by uncertainties in the science, this probabilistic statement reflects risk. However, for a number of chemicals, there is a high level of uncertainty in the current science in terms of a quantitative link between levels of toxicity and environmental endpoints (Chiu 2017; Navrud 2017, 2018), not to mention uncertainty in terms of a causal relationship. This uncertainty could be so high for many chemicals (or groups of chemicals) that making statements regarding explicit probabilities is not possible. For a number of chemicals, risk ambiguity might be the term that best characterises the degree of uncertainty with respect to their environmental toxicity and subsequent impacts. For instance, scientific evidence may allow experts to report a range of potential values of environmental toxicity but evidence may remain limited about measures of central tendency and/or variance.

Uncertainty is at the core of efforts to estimate benefits from the regulation and management of chemicals because environmental effects from the chemicals themselves are often uncertain given the current state of the science. In this context, the distinction between risk and uncertainty comes into the conversation because communication of risk can be addressed in SP protocols, but uncertainty is much more challenging. For instance, there is a large, long-standing literature on the value of a statistical life (VSL), which estimates the value that people attach to reductions in mortality risks –see Viscusi and Aldy (2003) for a comprehensive review of this literature; and Kniesner and Viscusi (2019) for a recent description of the approach. This literature has developed a number of strategies to communicate risk probabilities in ways that general audiences find intuitive and relatable, including textual explanations, square grids or dots, risk ladders, and or other symbols or pictograms (Logar and Brouwer, 2017).

However, communication of uncertainty to general audiences becomes more complex because it can involve nested probability distributions, and uncertainty inherently implies a lack of information. Consequently, SP studies motivated by the presence of uncertainty in the effects from chemicals have mostly relied on attributes that are described in terms of risk probabilities (e.g. Atherton et al., 2020 Logar et al., 2014). SP surveys pose hypothetical scenarios that are meant to simulate a market decision; as a result, non-market valuation economists are generally comfortable positing a hypothetical scenario in which

[9] Shaw (2015) provides an intuitive and comprehensive review of the literature on risk, uncertainty, and risk ambiguity, in the context of environmental issues.

uncertainties have been resolved to a degree, and can be communicated as risk probabilities.

To complicate matters, previous studies have pointed out that respondents' perception of risk are better predictors of choices over risky outcomes than science-based or experts' assessments of risk (Viscusi et al., 1991). In this situation, risk ambiguity provides a framework to model perceived risk by, for instance, treating risk as a random variable with an unknown but estimable probability distribution whose variance reflects uncertainty perceived by individuals (Nguyen et al., 2010). In the context of development of SP protocols informed with attributes reflecting risks, the implication is that even if there was a state of scientific knowledge that allows for risk statements, SP protocols dealing with changes in environmental endpoints should consider risk ambiguity.

This section discusses scientific uncertainty and its implications for the aim of this scoping study. Once scientific uncertainty is defined and how it is present at all steps in estimating the benefits of chemical policies (see Figure 1), the relevance of considering a precautionary principle when designing the SP protocols is discussed. This section finishes with a description of how uncertainty has been addressed in previous SP efforts dealing with toxic chemicals. Note that this section closely relates to chapter 7, which discusses the role of experts in conceptualising, expressing, and communicating uncertainty when designing the SP protocols.

4.1. Uncertainty in the effects of toxic chemicals

The specific source of uncertainty that is relevant in the context of valuing the benefits of chemical management and regulatory decisions is objective uncertainty surrounding the measurement of scientific results. Uncertainty of measurement refers to "the dispersion of the values that could reasonably be attributed to the measurand" (JCGM, 2008, p. 2). The measurand refers to the actual quantity that a researcher is attempting to measure. The true value of a measurand is unknown –otherwise, no research is needed. The true value of a measurand is always unknowable to a certain degree – i.e., there is always measurement uncertainty. "A sound measurement provides a best estimate, but that estimate always leaves uncertainty regarding the value of the measurand" (Rigdon et al., 2020, p. 329).

High uncertainty implies that the measurement is consistent with a wide range of plausible values for the measurand, and researchers may expect to observe a wide range of values across different measurements and across different studies. It is important to clarify that high uncertainty "does not mean that the measurand has multiple values, in the sense of a random coefficient model, but only that the researcher's limited knowledge leaves a range of values as plausible" (Rigdon et al., 2020, p. 329). The implications of this statement can be illustrated with the motivation behind random parameter logit specifications that non-market valuation practitioners regularly estimate on data gathered with DCEs. These models' motivation is the practitioners' suspicion that preferences for attributes in DCE vary across respondents. There is no uncertainty about the presence of heterogeneity in preferences –previous empirical evidence, theoretical tools, and anecdotal daily life events imply that scientific knowledge is mature enough to conclude that preferences vary across individuals. Thus, the distribution representing the variation of preferences does not reflect scientific uncertainty; instead, it reflects the possibility that utility parameters take on different values across individuals.

In the current context of estimating benefits of chemical policies, scientific uncertainty applies to steps 1 through 6 as displayed in Figure 1. This includes uncertainty due to our inability to adequately characterise biological and ecological processes, migration of certain chemicals, etc. Uncertainty is present within and across every step in Figure 1. In general, the more sources of uncertainty at each step, the greater the uncertainty when trying to communicate impacts to the general public and elicit how they value those impacts for the monetisation of benefits (step 7).

Starting with step 1 in Figure 1, even if the properties of a chemical are well-understood in the lab, how that chemical interacts when released into the environment and its ultimate fate and transport (Step 2) introduces uncertainty. Such uncertainty is compounded when considering the available evidence of impacts (if any) at the individual organism level (step 3). Outputs from ecotoxicological studies tend to examine growth, survival, and reproduction impacts on individual organisms from a single species; such studies sometimes even examine effects at the sub-organism level. While there are a number of chemicals for which the current evidence may not face significant uncertainty in steps 1 through 3, there is still significant uncertainty in the impacts on individual organisms for most chemicals (Chiu 2017; Navrud 2017, 2018).

Consequently, uncertainty is even greater when moving to subsequent steps in a benefits analysis because doing so entails scaling up the already uncertain evidence from step 3 and preceding steps, and then adding additional sources of uncertainty. The scientific uncertainties are compounded during the attempt to incorporate ecological interactions and aggregate the effects up to impacts on multiple organisms at the species- and ecosystem-level (steps 4 and 5). A non-exhaustive list of challenges include accounting for: i) nonlinear dynamics that can change over time and space; ii) functional redundancy –for example, a species is lost from a system without any obvious impacts on other ecosystem processes—; iii) threshold behaviours and impacts to keystone species that cause disproportionately large (and possibly irreversible) changes to the system; iv) feedback effects between individual-level versus population-level responses; and v) other context specific variables, such as rainfall and temperature (Forbes et al. 2017). Such uncertainties are then further exacerbated when aggregating across ecosystems in the relevant policy area (step 6).[10]

In general, when it comes to uncertainty, Forbes et al. (2017) emphasise that there are "wide gaps between current ecological risk-assessment endpoints and potential effects on services provided by ecosystems" (p. 845). In practice, the entire process faces high uncertainty of measurement as there is, to begin with, a high level of uncertainty in the current science in terms of a quantitative link between levels of various chemicals and environmental endpoints — which ultimately implies a lack of dose-response functions in steps 1 through 3 of Figure 1 (see Navrud, 2018).

There is promising research underway to fill those gaps. When it comes to linking outputs from environmental risk assessments to ecological endpoints (or final ecosystem goods and services), Forbes et al. (2017) report ongoing development of a framework that uses advances in mechanistic modelling to link three categories of models –i) sub-organism processes to organism responses; ii) organism responses to population responses; and iii) single-species responses to multi-species or ecosystem responses. They illustrate the framework using two case studies. The first case study examines the impacts of an endocrine disruptor that is commonly used in oral contraceptives on a mountain stream. The potential endpoints to be quantified are measures of increased game fish abundance. The second case study focuses on the effects of an insecticide used in agriculture on a

[10] Here, the reference is to combined uncertainty in measurement, which arises from the accumulation of uncertainties originated in different stages within each step. Examples of sources included in the combined uncertainty can be found in Hund et al. (2001).

reservoir ecosystem. The final environmental endpoints being quantified are measures of increased game fish abundance, water clarity, and the frequency of algal blooms.

The Life Cycle Initiative hosted by the United Nations Environment Programme is another significant effort underway to fill the gaps in assessing the adverse impacts of chemicals on ecosystems. Fantke et al. (2018) detail the efforts and provide recommendations to further the science in a consistent way. Fantke et al.'s current framework encompasses steps 1 through 5 in Figure 1, where the final measure is the potentially disappeared fraction (PDF) of species. As discussed in section 3.2.2, the PDF of species is proposed as one potentially generalisable endpoint to consider and test in focus groups when developing a survey instrument.

A related research agenda also filling gaps provoked by uncertainty aims to provide systematic, replicable strategies to estimate ecosystem level effects despite the lack of evidence at such levels. Hemming et al. (2018) reports on how structured elicitation protocols can be adopted when expert judgements are used to inform science. They illustrate their point referring to the case in which expert judgement informs conservation and natural resource management –with applications to threatened species management, environmental impact assessment and structured decision-making.

4.2. The precautionary principle

Due to the high uncertainty around the toxicity of a number of chemicals, several agencies and actors – e.g. UK Royal Society of Chemistry (RPA, 2013) — have put forward the suggestion that a precautionary principle should be considered. The precautionary principle aims to support environmental decision making, and involves four components: "taking preventive action in the face of uncertainty; shifting the burden of proof to the proponents of an activity; exploring a wide range of alternatives to possibly harmful actions; and increasing public participation in decision making" (Kriebel et al., 2001, p. 871).

With the aim of supporting implementation of the precautionary principle, Kuntz-Duriseti (2004) explored how BCA can provide insights about the economic value of precaution. He did so through three approaches: i) a precautionary premium, analogous to an insurance premium; ii) a precautionary response that alters current actions to hedge against possible future negative welfare shocks; and iii) a modification of risk assessment that takes into consideration the implications of low-probability, high-consequence outcomes. Importantly, Kuntz-Duriseti (2004) shows that under each approach, uncertainty in measurement imposes a penalty on top of a risk-aversion penalty.

The implication of Kuntz-Duriseti's theoretical result is that, in practice, people may be willing to pay to avoid uncertainty. Thus, the general public's ex ante values for environmental improvements may be higher if the magnitude of such improvements is uncertain. It is recommended to explore this implication when developing SP protocols in the context of chemicals and environmental endpoints.

Previous SP studies have explored the implications of the precautionary principle when it comes to estimating WTP. Antonio et al. (2022) implemented a CV protocol to explore WTP of residents in Gorizia, Italy, for an accelerated replacement of pipes that contain asbestos. This replacement qualifies as a precautionary measure because asbestos has yet to be scientifically linked to major threats to the quality of drinking water. Their findings confirm the theoretical expectation that people, on average, are willing to pay a premium to avoid uncertainty. However, the authors also document a polarisation in views that makes it difficult to conclude that accelerated replacement of pipes would be supported by a majority of the residents in Gorizia. Motivated by the prevalent uncertainty surrounding genetically modified food, Valente and Chaves (2018) explored the role of informing consumers about potential negative effects. Via a split-sample approach on students of the University of Porto, Portugal, the authors document that respondents presented information about uncertain negative effects report lower WTP for genetically modified food, which they interpret as reflecting stated benefits from implementing a precautionary principle on the adoption of modified food. Torres et al. (2017) presented a DCE to visitors of the S'Albufera wetlands, in Spain. They estimate WTP for a policy aiming to reduce the expected impacts on the wetlands from climate change. Using a split sample approach, the authors explore whether WTP differs depending on if the effects from the policy are presented with certainty or as uncertain. They document that stated WTP under an uncertain scenario is either greater than or equal to the WTP under a scenario where the expected loss is presented as certain. This result is again interpreted as reflecting a premium for precaution.

4.3. Approaches to communicate uncertainty

SP studies have mostly communicated uncertainty as a feature that more closely resembles risk. The outcome to be realised is unknown ex ante due to uncertainties and gaps in the science, but SP studies often assume (at least implicitly) that simulated scenarios reflect a situation in which there is a known probability assigned to each potential outcome. This is a practice consistent with the simulation of hypothetical markets in SP studies. Its realism can be explored and refined in focus groups among non-technical audiences in order to develop a credible scenario that can yield valid welfare estimates.

Whether uncertainty can realistically and validly be expressed in terms of explicit probabilities is a matter for ecotoxicologists and related discipline experts to decide. Otherwise, the resulting WTP estimates might not be useful for benefits analysis and transfer when analysing actual policies, following the steps in Figure 1. In order to utilise such information in policy analysis one needs to be able to quantify scientific uncertainty in a similar probabilistic fashion.

There are generally two broad approaches to incorporating uncertainty into SP protocols. Either (i) present the changes in environmental attributes to respondents as certain, and then model uncertainty after the fact; or (ii) explicitly incorporate uncertainty into the valuation scenario, either qualitatively or by treating the uncertainty as yielding risky outcomes with known probabilities that can be quantitatively conveyed to respondents.

Introducing uncertainty into a survey scenario could unintentionally reduce respondents' confidence in the survey instrument, which in turn may undermine perceived credibility and consequentiality (Banzhaf et al., 2006). In that sense, the first approach of presenting the survey scenario as certain and modelling uncertainty after the fact when projecting benefits is a desirable strategy. Doing so, however, ignores respondents' preferences towards uncertain outcomes (i.e., ignores their willingness to adopt a precautionary principle), which is particularly relevant in the context of chemical management and policy decisions.

The few SP studies that have been conducted on the environmental impacts of chemical exposures generally treat scientific uncertainty in a fashion similar to how the broader SP literature treats risk. Studies posit two or more potential outcomes (or states) as if they will be realised with some known probability. When it comes to quantitatively communicating risks, SP studies on chemical exposures and environmental impacts generally take one of two approaches, by communicating risks as: i) relative frequencies –with or without the help of a risk ladder—; or ii) the probability of failure or success.

Logar and Brouwer (2017) examine WTP to reduce the frequency of waterbodies in Switzerland exceeding the corresponding pollution standard. They communicate this frequency attribute both qualitatively (low, medium, and high), as well as quantitatively (i.e., X out of 15 pollutant standards are exceeded). They implement a split sample design where half the respondents are also presented with a risk ladder to frame those frequencies in the context of the probability of dying from several commonly known causes. For instance, the risk ladder equates "higher risk" to an average risk of death of "1 in 5 people"; and "lower risk" to "1 in 100,000 people".

Atherton et al. (2020) communicate uncertainty in scientific knowledge through a combination of a qualitative scale and probability statements, where probabilities are expressed as percentages, rather than frequencies. Their focus is on estimating WTP for cleaning up flame retardant chemicals in waterbodies in the UK. Atherton et al.'s DCE includes an attribute that describes the scientific uncertainty of whether flame retardants are safe. This attribute is described as, for example: "probably safe (75% chance)" and "probably not safe (25%)".

Other studies outside of just those examining environmental impacts of chemical exposures have taken similar approaches. Glenk and Colombo (2013) used the "risk of failure" to communicate the probability that a soil carbon program might actually fail to deliver net emission reductions. In their application, the attribute "risk of failure to reduce emissions" takes four levels: zero (no risk), 10%, 30%, and 60%. Other studies have used the "probability of success" to communicate uncertainty. Wielgus et al. (2009) implement a CV protocol to elicit WTP for marine conservation policies in the Gulf of California, Mexico. They use a split-sample approach that incorporates risk by varying the probability that improvements in the quality of a marine ecosystem are reached –probabilities take the value of 60% or 90%. Using a split sample approach, Roberts et al. (2008) conduct a DCE on policy measures to improve water quality of a lake in Oklahoma, U.S. They communicate risk as the probability that each level of each attribute is reached. More specifically, the attribute "water level" –takes the values of "normal", 2 ft low, 5 ft low, 8 ft low, and 10 ft low— was intersected with a companion probability attribute –taking the values of 100%, 90%, 50%, 10%, and 0%. Thus, a risk level was expressed as, for instance, a 50% probability that the normal water level is reached if the proposed policies are implemented. Notice that risk (or lack of risk) is part of the experimental design –i.e. 100% probability is included in the attribute space for the companion probability attribute.

Alternatively, when thinking of uncertainty in the strongest sense –where the current state of knowledge is too uncertain to even assign probabilities to potential outcomes– then such uncertainty can, at best, only be communicated in a qualitative scale. Lundhede et al. (2015) used a qualitative scale to communicate uncertainty with which a conservation policy would deliver the conservation of birds endemic to Denmark. They design a DCE that includes an attribute described as the likelihood that the policy will be effective. This attribute is presented with three levels: "very certain", "rather certain" and "rather uncertain".

In consultation with experts and policy practitioners, if probabilities can potentially be assigned in subsequent benefit transfer exercises and policy analysis, then it is recommended to follow the literature and consider the incorporation of probabilities or other quantitative measures representing uncertainty into a valuation question. Indeed, the inclusion of such measures will have to undergo significant focus group testing and rely on insights from the SP and risk communication literature.

Alternatively, if experts and policy practitioners believe quantified probabilities or other measures of uncertainty will not be viable in the foreseeable future, even with extensive sensitivity analysis in subsequent BCAs, then qualitatively communicating such measures may be the best direction to consider. That said, an in-between option would be to communicate uncertainties qualitatively, and then also elicit subjective, respondent-specific quantitative probabilities that individuals attach to such qualitative statements. Procedures exist to elicit respondent-specific, subjective probabilities (e.g. Andersen et al., 2014; Harrison et al., 2022; Scarpa et al., 2021), and in doing so it may be possible to later incorporate quantified probabilities into the econometric models in order to characterise scientific uncertainty.

Chapter 5. Other Challenges and Considerations

This section outlines some of the key challenges that researchers face when developing a SP survey in general, and discusses these considerations specifically in the context of environmental impacts from toxic chemical regulations and management decisions. Common foundational questions to consider include generalisability of the valuation scenario, and thus broader applicability for survey implementation and future benefit transfer; choice of the provision mechanism and payment vehicle; specifying the baseline or status quo conditions, and minimising considerations of unintended endpoints (i.e., minimising the potential for omitted variable bias). In making decisions regarding these common considerations, best practices for survey development should be followed. As described by Johnston et al. (2017), such best practices include: providing a clear explanation of the baseline (or status quo); posing consequential valuation questions; conducting qualitative pretesting during survey development via focus group and/or cognitive interviews; clearly documenting rationale for various survey feature decisions, while keeping in mind the needs of decision makers; and ensuring valuation scenarios are not overly burdensome on respondent cognition.

5.1. Framing the valuation scenario

The central piece of any SP survey is a valuation question(s) where respondents stated responses directly or indirectly reveal how much they value a change in the natural environment. To do so, the valuation question(s) must push respondents to make a trade-off between a hypothetical change in environmental quality and costs to their households. Three key features in constructing a credible SP valuation scenario are the type of valuation question, the provision mechanism, and the payment vehicle. The two main types of SP valuation protocols are contingent valuation and discrete choice experiments (Johnston et al., 2017). Contingent valuation (CV) asks respondents to value a fixed bundle of environmental improvements, whereas the experimental design in a discrete choice experiment (DCE) allows researchers to estimate the various dimensions (or attributes) of a bundle of improvements separately. The provision mechanism is the means by which environmental quality is stated to improve in the hypothetical valuation scenario. The payment vehicle is the context by which respondents are stated to experience a hypothetical increase in costs in order to obtain the improvements posited in the valuation scenario. Here, key considerations of these three features for an SP survey are discussed in the context of improvements in environmental quality due to reduced chemical exposures.

It is important to emphasise that the provision mechanism and payment vehicle specified in a survey do not need to match real-world policies and trade-offs that people face (even if such policies are likely to be the focus of future benefit transfer exercises). These hypothetical features of the valuation scenario are merely a means to try and elicit accurate responses of what respondents' actual WTP might be. In that sense, researchers must simply be sure to follow best practices (Johnston et al., 2017) and design a survey that respondents find credible and consequential. If participants believe their responses will influence policy decisions, then this gives them an incentive to provide truthful statements of their WTP, even though they know the valuation scenario is hypothetical. The fact that the provision mechanism and payment vehicle do not need to match real-world policies is an advantage of the SP protocols suggested in this scoping study. The suggested protocols are meant to describe situations as general as possible so that the gathered information can be widely applicable for benefit transfer to future policy decisions.

5.1.1. Type of valuation protocol

In order for SP protocols and WTP estimates to be useful in informing future management and regulatory decisions, survey results must be as applicable as possible across a range of countries, chemicals, and policy contexts. A critical decision in promoting such flexibility is the type of valuation question to present in a SP survey.

The contingent valuation (CV) approach has several advantages, such as reduced cognitive burden on respondents, reduced econometric complexity, and has been argued to be more incentive compatible (Colombo et al., 2022; Czajkowski et al. 2017b; Vossler et al., 2012; Vossler and Watson, 2013). On the other hand, the CV approach bundles the often high-dimensional, complex features of an environmental commodity, describing them only qualitatively or, if put forth in quantitative terms, forcing environmental attributes to change in conjunction. Environmental improvements are surely quite heterogeneous across different management and policy decisions, pertaining to different chemicals, and affecting different ecosystems in different ways, and over different time horizons.

The increased flexibility of a discrete choice experiment (DCE) in valuing individual, unbundled attributes allows for estimation of a flexible benefit-transfer function that is more applicable to future policies, especially in heterogeneous contexts like environmental impacts from toxic chemicals (IEc, 2016). At the same time, the cognitive burden associated with a complex DCE scenario may be too overwhelming, and could lead respondents to adopt simplifying heuristics that would invalidate any inference of preferences from their stated responses. Therefore, it is a question for focus group testing and survey development as to which type of valuation question format to adopt when considering this simplicity versus applicability trade-off.

5.1.2. Provision mechanism

Provision mechanisms are often framed in terms of yielding a private or public good. A private good provision mechanism that is applicable to the toxic chemical context could entail the purchase of alternative consumer products, where the chemical compounds or characteristics of the chemical compounds vary across alternatives. Such a private good provision mechanism may be appropriate for human health benefits related to chemicals (e.g. Alberini and Chiabai, 2007; Alberini and Scasny, 2014; Scasny and Zverinova, 2014; Morris and Hammitt, 2001; IEc, 2016), which are often framed as a private good, but are not the most valid approach in the context of improvements in environmental endpoints, which are, by their nature, often public goods. Therefore, a scenario describing a hypothetical public policy seems like the most viable provision mechanism. Using a private good provision mechanism to value improvements in a public good could lead to strategic responses that are biased by issues like freeriding and tragedy of the commons, and in general reduce incentive compatibility.

Within the literature, there are two potentially viable approaches in framing a public policy provision mechanism. The first strategy is to posit a scenario involving a specific government intervention, such as a specific infrastructure upgrade that leads to environmental improvements. For instance, Logar et al. (2014) estimate the benefits of a national policy in Switzerland aiming to reduce micropollutants in freshwater bodies (e.g., rivers). This reduction is described as being provided through the upgrading of municipal sewage treatment plants across Switzerland. King et al. (2021) present a CV scenario that describes a reduction in microplastics released in aquatic ecosystems, and this reduction is also reached through an upgrade to wastewater treatment plants.

There is a trade-off in framing a more specific provision mechanism. A detailed and specific provision mechanism can increase credibility and perceived consequentiality of the valuation scenario, but can also be distracting, lead respondents to consider confounding factors (i.e., value features of the provision mechanism rather than the specified changes in the environment), and perhaps most importantly for the current context, reduce the generalisability of the results and potentially inhibit benefit transfer to future policies. Focus group testing and consultation with policy practitioners is needed, but in the current context the most appropriate starting point for a provision mechanism, , is to test a general –purposely abstract— bundle of public policy interventions. This type of provision mechanism is described as several vaguely specified measures that are implemented to deliver an improvement in environmental quality. King et al. (2021) explore preferences for a reduction in releases of microplastics originated in the cosmetics industry, and frame alternative public initiatives where substitute chemicals that are less harmful (but also more expensive) are used in the cosmetic products. IEc (2016) poses a similar public policy, but keeps the bundle of consumer products abstract, and just states that it may include things like personal care products, cleaners, plastic products, batteries, light bulbs, fertilisers and pesticides, automotive products, and construction materials.

Van Houtven et al. (2014) estimate the benefits from reduced eutrophication of lakes in the Southeastern U.S. The stated program is kept vague, and framed as potentially implementing one or several measures (e.g., upgraded sewage plants, increased inspections of septic tanks, improved storm water runoff systems, and improved agricultural practices). Moore et al. (2018) take a similar approach in estimating WTP for improvements in the Chesapeake Bay, a large iconic estuary in the eastern U.S. Although they had a specific policy in mind when developing the survey, they framed the provision mechanism as an abstract policy bundle. Doing so minimises possible protest responses associated with a particular policy, and promotes generalisability for later benefit transfer to future policies.

Purposefully general descriptions of provision mechanisms are also used in studies estimating the benefits from protecting ecosystems from climate change. A general provision mechanism is a desirable feature in the context of climate change measures because practitioners want to avoid people reporting values that are confounded by preferences for how improvements would be achieved –the focus should be, instead, on the specified improvements themselves. For instance, Sandorf et al. (2016) estimate the benefits from the protection of cold-water coral in Norway. Their provision mechanism is described as a regulation that would protect a given area, without providing detail on what the protection measures would entail. Robinson et al. (2022) explore benefits from the mitigation of climate change effects in coral reefs surrounding the Hawaiian Islands. Their provision mechanism consisted of non-specified mitigation measures.

Given the objective of assessing paths forward in developing an SP study to estimate WTP to avoid environmental damages from the use of chemicals through their entire life cycle, and the ultimate goal of benefit-transfer, an abstract public policy bundle is the recommended starting point for a provision mechanism. Focus group testing is then needed to assess whether an abstract bundle of public interventions is credible and understandable to the general public, and if it is the most appropriate provision mechanism for minimising the consideration of confounding factors, as well as protest and strategic responses.

5.1.3. Payment vehicle

The selected payment vehicle should be realistic and credible, and perceived as binding to all respondents to the greatest extent possible (Johnston et al., 2017). The payment vehicle need not necessarily match how similar programs are actually funded in reality, even for programs for which the WTP estimates may later be applied. The payment vehicle is only intended to pose a credible situation in which respondents must make a trade-off between improvements in environmental endpoints and costs to their household.

Conditional on an abstract public policy as the provision mechanism, there are three obvious choices from the literature in terms of an appropriate payment vehicle: (i) an increase in specific fees or bills, (ii) a general cost of living increase, or (iii) an increase in taxes.

The first payment vehicle refers to an increase in a specific service bill – e.g., water bill — which is consistent with a specific description of a provision mechanism – e.g. upgrading water treatment systems, as is the case in the studies by Logar et al. (2014), King et al. (2021), and Atherton et al. (2020), for example.

A second strategy is referring to an increase in the cost of living as the payment vehicle – understood as a broad category that includes an increase in prices of goods and services that respondents pay for on a regular basis, utility bills, taxes, etc. (e.g., Hagan et al. 1999; Magat et al. 2000; Moore et al. 2018). This general characterisation is consistent with a general description of the provision mechanism. Viscusi et al. (2008) estimate the value of water quality in the U.S. by presenting respondents with a trade-off between an increase in annual costs of living and the percentage of lake acres and river miles with "good" water quality. They offer no explicit explanation to how water quality would be achieved in their survey.

Similarly, Van Houtven et al. (2014) paired a broadly characterised increase in annual cost of living with an abstract program that would improve water quality. Their phrasing is illustrative:

"The changes required by the program would have a cost for all home state households. Some of the basic things people spend money on would become more expensive. For example, for homeowners, water bills or costs for maintaining septic systems would go up. For renters, rent or utility bills would go up. Imagine that for households like yours, starting next year, the program would permanently increase your cost of living by $X per year, or $X/12 per month." (p. 46)

A purposefully abstract payment vehicle is often found to work well for the same reasons that an abstract provision mechanism is often used, it can help minimise consideration of confounding factors, and possibly reduce protest and strategic responses. At the same time, a vaguely framed payment vehicle can reduce credibility and perceived consequentiality.

Studies using a generally framed provision mechanism could also use a specific payment vehicle, such as a tax. Banzhaf et al. (2006), for instance, value ecological improvements due to reductions in acid rain, and ask respondents to vote on a policy referendum that would yield the stated environmental improvement by increasing taxes by $X for 10 years. Choi and Ready (2021) pose an increase in annual taxes to respondents, but do not specify the type of tax. Abate et al. (2020) describe a purposefully abstract initiative to reduce impacts of marine plastic pollution, and describe a payment vehicle consisting of mandatory annual taxes. Their phrasing of the payment vehicle is illustrative:

> "Considering the anticipated results of the initiative outlined before, would you vote for this initiative if it would cost your household and annual tax of NOK $X for the next ten years?" (p. 4)

Given the goal of generalisation and future benefit-transfer across policies and study areas, a general cost of living increase may be the most appropriate payment vehicle. It can also deal with concerns about protest responses –as a general rule, for instance, payment in the form of taxation produces protest zeros in CV protocols (Rankin and Robinson, 2018). Determination of the appropriate payment vehicle, however, must ultimately be determined through focus group testing and survey development; especially given that respondent acceptance of a particular payment vehicle (e.g., a tax versus a general cost of living increase) may vary across countries.

An additional consideration is the duration of the stated increase in cost to respondents. On one hand, surveys can present respondents with a one-time increase in taxes, a hypothetical bill, or cost of living. For example, von Stackelberg and Hammitt (2005) pose a one-time increase in income taxes. Moore et al. (2018) provide an example on the other extreme, and present respondents with an indefinite increase in their annual cost of living. In the middle are studies like Abate et al. (2020) and Banzhaf et al. (2006), who present an increase in annual costs for some fixed time period; in these cases, ten years. An indefinite cost increase may be perceived as not credible or be open to strategic thoughts to circumvent costs (e.g., a respondent anticipates moving in a few years, or plans on undertaking some other cost mitigating behaviour). At the same time, experience has shown that focus group participants have sometimes questioned the credibility of the provision mechanism when the posited costs are stated to only be experienced for a few years. The rationale sometimes expressed by respondents is that the posited environmental improvements will not be fully realised or maintained if the costs funding those provisions are not continued over many years. In the end, focus group testing with representative participants from the general public is needed to choose the payment vehicle time period that is realistic, credible, and perceived as most binding to respondents.

5.1.4. Specifying a baseline

Specifying the baseline or the status quo conditions is an equally important part of framing a credible and accurate valuation scenario. The baseline defines the situation or state of the world if a respondent decides to "do nothing." Under the baseline scenario, the cost stated to be incurred by the respondent is zero, and the specified environmental conditions are often characterised as being similar to the current conditions, although some SP studies have specified future baselines that are different from the current state (Banzhaf et al., 2006; Soto Montes de Oca and Bateman, 2006; Lew et al., 2010; Maguire et al., 2018).

In the context of environmental endpoints and the impacts from chemical management and policy decisions, baseline conditions are very location specific, depending on the nature of the local ecosystem, emissions from nearby pollution sources, and stock levels of the relevant pollutants, including both chemicals of interest and other pollutants. The lack of data and substantial scientific uncertainty would, from a practical standpoint, make it extremely difficult to gather baseline level data for the relevant environmental attributes. Even if such data could be gathered, such baseline conditions are location specific, and so specifying the same baseline in a survey instrument would deter its implementation in different study areas, and hinder later benefit transfer to policy sites.

Specifying an accurate and credible baseline is actually more difficult than positing the proposed environmental changes, provision mechanisms, and payment vehicle. The latter are all hypothetical features of a simulated scenario, whereas the baseline is often based on the current conditions (or current trends) and must be as accurate as possible for two reasons. First, it increases the credibility of the survey and the valuation scenario. If a survey specifies a baseline condition that is not in line with respondents' prior perceptions, then respondents will question the credibility of the entire survey. Second, specifying an accurate baseline allows respondents to link their own experiences to the specified baseline attribute levels, thus providing a reference point to help them understand the attribute levels and stated changes in those levels.

Two broader approaches in which one could address the issue of specifying a baseline are considered. First, researchers could attempt to gather the necessary data and expert input to accurately characterise baseline conditions at each specific location where the survey will be implemented. Information on such baseline conditions would also be later needed for any policy areas where one intends to apply the results using benefit transfer. Therefore, acquiring current and possibly projected status quo quantified levels for all attributes (e.g., endpoints) included in the valuation scenario is a key criterion for identifying locations where such a SP survey could be implemented, as well as where the results could be used for future benefit transfer applications. If overcoming such a hurdle seems feasible, and or necessary, then this is the preferred approach. It is speculated, however, that the data constraints and scientific uncertainty make this option infeasible, at least in the near term.

The second option is to posit a valuation scenario where the baseline attribute levels are not explicitly specified in quantified, absolute terms. For example, in the illustrative valuation scenarios proposed later in chapter 6, a baseline where a chemical is allowed to continue to be used by industry and consumers is proposed. The quantified baseline attribute levels are specified as a change relative to the current conditions. Although specifying absolute levels may be preferable, the characterisation allows the survey to be implemented across study areas with minimal adjustments. The policy scenario is defined as a situation where the use of a chemical is reduced, and thus where the specified baseline environmental damages are avoided. A potential drawback from this approach is that respondents may not perceive baseline conditions in the same way, which ultimately will influence their preference for the status quo. While less than ideal, an alternative to control for such differences in perceptions is through the modelling of preferences for the status quo alternative as a function of individual-specific factors that would include variables arising from debriefing questions capturing indicators associated with perception of status quo –for instance, perceptions of baseline conditions of freshwater bodies may be associated with number of annual visits to lakes and rivers.[11]

Given the need for benefit transfer, this simplification is seen as a practical necessity towards developing a generalisable and widely applicable SP survey. Specifying this simplified choice scenario and baseline facilitates applicability of a common survey instrument across study areas, as well as enables future benefit transfer with minimal need to collect and model current environmental conditions.

In deciding what approach to take in specifying the baseline conditions, and in better pursuing that approach, significant consultation and feedback from practitioners, policy experts, and ecotoxicologists and environmental risk assessors is needed, as well as focus group and cognitive testing with members of the general public. Best practices in defining the baseline should be considered to the extent possible (Johnston et al., 2017; Whittington and Adamowicz, 2011).

5.2. Controlling for health risks: strategies to minimise omitted variable bias

The objective of this scoping study is to examine the feasibility of a generalisable survey instrument to elicit people's WTP for environmental improvements due to the regulation

[11] Econometrically, modelling preferences for status quo involves a two-stages process in which a random utility model is first empirically estimated via a Random Parameters Logit and then individual-specific preference parameters are modelled as a function of individual-specific variables via Ordinary Least Squares. This strategy was first proposed by Campbell (2007) who focused on rural landscape improvements in Ireland. Subsequent applications of this approach have analysed preferences for recreational use of forests in Lorraine, France (Abildtrup et al., 2013); biodiversity enhancement in New Zealand's planted forests (Yao et al., 2014); forest management and protection program in Poland (Czajkowski et al., 2017a); power outages in Mekelle, Ethiopia (Zemo et al., 2019); demand for crop insurance in India (Ghosh et al., 2021); and coastal and marine conservation in Nha Trang Bay, Vietnam (Börger et al, 2021).

and management of chemicals, with a specific focus on values directly for environmental endpoints and not related to human health. In the context of hazardous chemicals, however, health risks are important and likely at the forefront of respondents' minds when they make decisions. For example, Atherton (2019) found that people most concerned about potential human health impacts had the highest WTP for remediation of persistent chemical pollution in surface waters. Given the goal of isolating direct values for improvements in environmental endpoints, health risks are an omitted variable that must be controlled for in some fashion. If a key factor is omitted from a survey scenario, then respondents tend to fill in missing information with their own default assumptions, and the perceived values respondents fill in for an omitted variable could differ greatly from that of the researchers (Carson 1998; Banzhaf et al., 2006; Johnston et al. 2013). By filling in the gaps themselves, respondents may use the specified environmental improvements as a proxy for improvements in other omitted variables they care about, in this case human health. This highlights the classic omitted variable issue. The aim of the scoping study is to isolate WTP for some environmental endpoint X, but when making decisions respondents correlate X with an unobserved variable that only enters through the error term in the econometric model, in this case human health risks.

Two common approaches to address omitted variable bias in SP studies are to (i) not mention the issue in the valuation scenario; or (ii) to explicitly state that respondents should not consider the omitted variable or frame a scenario to limit such considerations. Under these two approaches debriefing questions are often included to test for the presence of and or control for any residual consideration of the omitted variables. A third approach is to explicitly include the otherwise omitted variable in the valuation scenarios so it can then be statistically conditioned out of the WTP estimates of interest. Each approach has its own strengths and weaknesses.

Explicitly instructing respondents not to consider a particular factor or framing a scenario where it would naturally not be a consideration may be appropriate in some contexts, but such approaches are not valid in all settings. For example, when valuing ecological improvements in lakes due to reduced acid rain, Banzhaf et al. (2006) found that focus group participants often considered unfounded health effects, and had "expansive priors" regarding impacts to forests and birds. They address the latter by simply stating that any impacts to forests and birds are minimal. To break any over-perceived links to health, they told survey respondents that the level of the acidity in lakes was similar to that of orange juice, and so there are no health effects. Framing a scenario where human health is not impacted in the context of toxic chemicals management would be difficult. Such a scenario would have to allow for non-use and use values in order to capture the full benefits of environmental improvements, but also eliminate considerations of health risks that are associated with use of natural amenities in a credible way. In the context of toxic chemicals, such a scenario does not seem plausible if one wants to include use values.

Moore et al. (2018) conducted a DCE to examine how people value improvements in ecological endpoints in a large iconic estuary in the eastern U.S. In focus group testing they found that respondents considered omitted factors like impacts on seafood markets and improvements to waterbodies outside of the watershed. As a result, they explicitly state that the hypothetical policy bundle would not affect seafood markets or waterbodies outside of the watershed. The potential drawback of doing so, however, is that by calling out a particular factor you may draw attention to it and possibly exacerbate the issue. In fact, based on responses to debriefing questions, Moore et al. found that about 36% and 50% of their respondents, respectively, still considered these omitted variables. In their econometric models, Moore et al. included interaction terms to allow for heterogeneity across respondents who did and did not consider these omitted variables (i.e., seafood markets and lakes outside the watershed). In doing so, they were able to estimate WTP premiums associated with considerations of omitted variables and separate that from the WTP estimates of interest. A similar approach could be taken in the context of health risks. A potential drawback with this approach is that respondents may not have considered health risks when responding to a valuation question, but when prompted in the later debriefing questions they may exhibit “yea-saying,” and state they did. Keeping the survey as succinct and simple as possible may facilitate more accurate recall and reduce the potential for “yea-saying,” but ultimately focus group testing would be needed to assess the potential for such behaviour.

The third approach is to explicitly include the confounding variable in the valuation questions. In the current context, this means including a human health attribute in the valuation question. Even though WTP to reduce human health risks is not of primary interest, by explicitly including this variable it can later be conditioned out when econometrically estimating WTP for environmental improvements. The drawback of this approach is that human health is a complex, multidimensional good itself, and the addition of even a simple health attribute can complicate an already complex valuation scenario and further increase the cognitive burden on respondents. Focus group testing is needed to see if a simple health metric can be included to isolate such considerations from the welfare estimates of interest. Any WTP estimates for health risks from such a survey would not be credible for subsequent policy analysis. A health attribute would merely be included to purge respondents’ health considerations from the WTP estimates for environmental endpoints. Health benefits are still important for policy analysis, but such benefits are being estimated separately under OECD’s SWACHE Project, and are outside of the efforts being discussed in this scoping study.

In the context of hazardous chemicals, human health is always going to be at the forefront of people’s minds and likely composes a large portion of their WTP for reductions in chemical releases. IEc’s (2016) SP study for Health Canada included two coarse human health attributes – a binary variable denoting whether a chemical is carcinogenic, and a categorical variable of other human health effects (i.e., no effects, respiratory or cardiovascular effects, reproductive effects, and developmental effects). Given the interest in solely including a human health variable in order to minimise confounding effects, a single, coarse human health attribute is proposed. It is questionable whether respondents would find a binary health variable as credible, so as a starting point for focus group testing, an ordinal 0 to 5 human health variable, where 0 = no human health effects, 1 = minimal acute health effects (e.g., minor headache, nausea), and so on until 5=severe non-fatal health effects (e.g., cancer, cardiovascular disease, infertility, etc.) is proposed.

If focus group testing reveals that the addition of a human health variable increases cognitive burden to the point where it detracts from estimating WTP for the environmental improvements of interest, then a simple binary variable can be considered, or some type of split-sample design where human health effects are not an explicit attribute that varies. If all else fails, a reasonable backup plan is to adopt the approach of relying on debriefing questions and controlling for health considerations ex post in the econometric model, as done by Moore et al. (2018).

Chapter 6. Alternative Approaches

Considering all of the challenges discussed in this scoping study, valuation questions where the provision mechanism is an abstract policy bundle is proposed. The stated environmental improvements can include aquatic and terrestrial environmental endpoints, to start, but depending on later feedback with experts and from focus group participants from the general public, the focus may be narrowed. A general cost of living increase as a starting point for the payment vehicle is proposed, but focus group testing may suggest a specific tax or fee is more credible among the general public in some countries. Moreover, the hypothetical scenarios described in the SP valuation questions need not be exactly the same as actual regulations and management decisions. The SP scenarios are merely meant to present respondents with credible trade-offs from which WTP values can be inferred. It is the improvements in environmental endpoints and WTP that are of most importance for benefit transfer, and not necessarily how similar the features of the constructed survey scenario are to later policy applications.

Most importantly, given the reality that there is a vast amount of scientific uncertainty regarding the environmental impacts of different chemicals on environmental endpoints, and the likelihood that people have a positive WTP for reducing impacts even in the face of this uncertainty, such considerations must be at the forefront of any survey design. A key question is whether the science is certain enough to communicate and elicit values for impacts on environmental endpoints. That decision will depend on extensive consultation and collaboration with ecotoxicologists, environmental risk assessors, and policy practitioners. With those future discussions in mind, two alternative valuation questions are proposed. These questions need not be mutually exclusive, and the most appropriate path forward may lie somewhere in the middle.

On one extreme, a valuation question that explicitly includes improvements in environmental endpoints is proposed (Proposal 1). A question similar to that suggested would deliver estimates of benefits from improving the quality of environmental endpoints via policies that regulate chemicals for which scientific knowledge is mature enough (or will be in the foreseeable future) to communicate probabilistic statements about the expected outcomes.

On the other end, in the case that experts report that uncertainty is so large that it deters us from expressing probability statements now or in the foreseeable future, a valuation protocol that brings uncertainty to the forefront of the exploration of WTP is proposed (Proposal 2). Given the scientific uncertainty, respondents would be asked whether they are willing to bear an increase in costs to avoid the potential for future, unknown environmental damage. This approach would provide benefit estimates that rest squarely on the precautionary principle –i.e. policies taking preventive action in the use of chemicals in the face of uncertainty about the impacts.

6.1. Proposal 1: A Generalised Endpoint Approach

This first proposal is a discrete choice experiment (DCE) that covers many of the generalisable environmental endpoints discussed. A DCE is an appealing option for a valuation question in the context of chemical management because of the possibility to value environmental improvements based on individual characteristics of that chemical and how it interacts in the environment. Various studies and workshops involving both researchers and practitioners have expressed the desire to examine preferences towards chemical management decisions based on characteristics related to a chemical's persistence, bioaccumulation, and toxicity (e.g., RPA 2013; Donohue and Kipusi 2016; IEc, 2016). Doing so would, in theory, allow for estimation of a flexible transfer function across a variety of contexts based on specific chemicals and study areas, while also potentially accounting for what can be projected with relative certainty or not in subsequent benefit transfer applications.

A starting point for the proposed DCE is illustrated in Figure 4. In an actual survey, text would precede this question to introduce and describe each attribute in detail. The survey developed for Health Canada by IEc (2016) and Table 1 in section 3.3 of this scoping study provide a good starting point for language that communicates these attributes in a manner that is understandable to the general public.

Given the complexities discussed throughout this scoping study, including the communication of numerous technical attributes, incorporation of scientific uncertainty, and controlling for human health and geographic scale, the choice question should be as simple as otherwise possible. For this reason, it is recommended that respondents only be presented two alternatives in each choice question – a status quo option and a policy option. As seen in Figure 4, the status quo and policy options are not explicitly represented as two separate columns. This is done partly due to the difficulties in explicitly defining a baseline given the lack of data and desire for a generalisable survey across geographic locations (see section 5.1.4 for details).

The provision mechanism is specified as a policy to reduce the use of a hazardous chemical in industrial processes and in consumer products. The latter is intended to minimise protest responses by respondents who believe industry is solely responsible for preventing adverse environmental impacts.

The baseline in the proposed DCE is a state of the world where a chemical will continue to be used. The resulting adverse environmental impacts that would be experienced (with some probability) under this baseline are expressed in terms relative to the current conditions. This approach is proposed as a starting point, with the hopes of facilitating implementation of the survey across multiple countries, and also to increase the generalisability of the results for future benefit transfer applications (see section 5.1.4 for further discussion).

The proposed policy scenario is a state of the world without those adverse impacts. In the proposed scenario, the "Policy" is posited to avoid the impacts described in the "No Policy" column in Figure 4. The reference level in the valuation scenario is specified to be a future state of the world with greater pollution levels compared to current conditions. Respondents are asked about the trade-offs for a gain in environmental quality compared to that reference level, and so WTP (as opposed to willingness to accept) is the appropriate measure for welfare changes in this instance (Knetsch et al., 2012; Nguyen et al., 2021).

An underlying assumption in the proposed DCE is that in avoiding the adverse impacts, future conditions remain the same as current environmental conditions. This may not be realistic in many contexts, but this simplifying assumption gives respondents a useful benchmark, while still presenting trade-offs from which values for incremental changes can be inferred. The motivation for this simplifying "constant" policy scenario is analogous to the simplifying assumption of a constant baseline that is conventionally assumed in SP studies (Maguire et al., 2018).

There are seven attributes in the choice question proposed in this paper, –a number in the upper end of previous DCE studies (Martinez-Cruz, 2015; Soekhai et al., 2019), but that has been documented to yield statistical efficiency of findings similar to designs with a lower number of attributes (Caussade et al., 2005; Meyerhoff et al., 2015). Again, however, the proposed valuation scenario in Figure 4 is meant merely as a starting point. Consultation with experts, policy practitioners, and participants from the general public are needed to help prioritise attributes and ideally narrow the focus. Focus group testing is also needed to assess the potential for cognitive fatigue and attribute non-attendance (Lew and Whitehead, 2020; Meyerhoff et al., 2015; Olsen et al., 2017).

The first attribute "What is impacted?" describes what types of organisms, plants, and wildlife potentially would be impacted, based on the corresponding trophic levels. Such impacts could be direct through bioaccumulation, or indirect through ecological interactions, as described in the proposed text in Figure 4 and discussed in section 3.2.2. In a subsequent econometric model of respondents' indirect utility function, this attribute could be modelled as an ordinal variable denoting each trophic level that is impacted. For example, an attribute level of two would imply that the species in the lowest two trophic-levels are affected. If nonlinear preferences are of potential importance in this dimension, then a series of non-exclusive binary variables to account for all impacted trophic levels could be included. Based on feedback from focus groups, it may also be possible to aggregate some trophic levels and simplify the experimental design. For example, preferences among the general public may only differentiate between whether higher-order species are impacted or not, suggesting a simple binary attribute may be reasonable. The estimated equation and experimental design would need to be devised with such considerations in mind.

The levels for the "What is impacted" attribute would be hypothetical in the posited choice scenarios, and randomly assigned based on the experimental design. One must just ensure that the designated attribute space covers the range of plausible values, as informed by consultation with ecotoxicologists, environmental risk assessors, etc. For subsequent benefit transfer exercises and policy analysis based on the results, expert elicitation methods would ideally be relied on (see chapter 7). At the very least, knowledge of a chemical's bioaccumulative capabilities can inform a conservative scenario that ignores indirect interaction effects within an ecosystem (e.g. a chemical may not directly affect the highest order species, but such species could be indirectly impacted due to a depletion in food or suitable habitat).

The second attribute "How bad are the impacts?" is the most challenging, especially given the desire to have a measure that is generalisable across locations to encourage consistent implementation of a common survey instrument, as well as applicability for future benefit transfer. As described in section 3.2.2, there are numerous composite endpoint measures of ecosystem health to consider, but the most promising starting point for the discussion is the use of species sensitivity distributions (SSD), and more specifically, perhaps the potentially disappeared fraction (PDF) of species (Fantke et al., 2018). Similar measures have been used in the SP literature (Johnston et al., 2012, 2013; Morse-Jones et al., 2012; Parsons and Thur, 2008; Breeze et al., 2015; US EPA, 2021), and can be linked to the projections from ecotoxicologists (Fox et al., 2021; Henderson et al., 2011; Posthuma et al, 2019; Xu et al., 2015). This attribute is currently presented as a proportion, expressed as X out of 10 species would disappear from the ecosystem, on average. This does not necessarily mean that the species are extinct, but rather that they are no longer present in the ecosystems of the impacted area. As with all attributes and metrics, extensive focus group testing is needed. The current measure of communication was chosen mainly to distinguish this attribute from the companion uncertainty probability attribute discussed next.

An added advantage of a SSD or PDF-based endpoint measure is that it could be accompanied by a corresponding attribute reflecting the uncertainty behind whether the stated outcome would be realised, as done by von Stackelberg and Hammitt (2005) in the context of SSDs, and by Atherton et al. (2020) in a related context. Scientific uncertainty is presented as an explicit attribute in Figure 4 – "How likely are these impacts?". In this example, uncertainty is communicated in probability or percentage terms –as has been done by other SP studies previously dealing with uncertainty in the context of chemical regulations (e.g., Logar and Brouwer, 2017; Atherton et al., 2020). Effectively, by presenting probability statements, this example communicates the risk of adverse baseline impacts.

An alternative approach to communicating uncertainty, and one that appears to have not been tested in the SP literature, is to convey a range of potential outcomes (e.g., a 95% confidence interval) based on an assumed distribution (perhaps informed by expert elicitation procedures). Such a range could be communicated to respondents as, for example: "the range in which scientists are confident that X will fall." The range of values in the posited scenarios would be generated from assumed variance values that are randomly assigned as part of the experimental design, which again would be informed by feedback from experts. The subsequent econometric model would then include the variance or some other measure of the spread of that distribution as an explanatory variable. The convenient feature of this strategy is that it is consistent with the definition of risk ambiguity discussed in chapter 4, it does not present such ambiguity as an "all or nothing" type of outcome (as is the case with the probability measure in Figure 4), and it may prove more salient and easier to communicate to respondents.

The decision of whether a quantitative uncertainty attribute can be included in the DCE in the first place should be made in consultation with ecotoxicologists and related discipline experts –and correspondingly, the scale and range of values that describes uncertainty must also be informed by discipline experts. To ensure applicability of the results for future benefit transfer, the attribute space for the uncertainty variable needs to reflect the range of plausible values, and, should include a zero or near zero uncertainty value to reflect a scenario with (near) certainty. Such a "certainty" scenario serves as a useful benchmark and ensures that benefit transfer is possible in future cases where natural scientists are fairly confident in the projected scenarios.

Indeed, the design of an attribute that communicates uncertainty requires not only consultation with experts but with the general public as well. Accurately communicating uncertainty and risk in SP studies is a well-studied challenge. For instance, in the context of SSDs, von Stackelberg and Hammitt (2005) found it difficult for respondents to comprehend nested probabilities and proportions (von Stackelberg and Hammitt 2005). Extensive focus group testing is needed to develop this attribute. If proven to be too cognitively burdensome, then this is a reason to consider qualitative statements regarding uncertainty. Such statements are generally easier for respondents to comprehend. Additionally, qualitative statements can be linked to quantified respondent-specific subjective probabilities based on a battery of questions that allow for estimation of subjective probabilities – a number of available strategies enable researchers to estimate these probabilities or beliefs (e.g., Andersen et al., 2014; Harrison et al., 2022; Lundhede et al., 2015). If ecotoxicologists and related experts are confident in projecting quantified probabilities to reflect uncertainty, perhaps based on the distribution of responses to expert elicitation exercises (see chapter 7), then the respondent-specific, subjective probabilities could be included in the econometric models, and subsequently used for benefit transfer.

The fourth attribute is "How long do the impacts last?" As described in section 3.2.2, how long environmental impacts are experienced depends on the persistence of the chemical of interest, as well as subsequent ecosystem dynamics. The attribute values in the hypothetical scenarios would be randomly assigned based on the experimental design. One must just ensure that the relevant attribute space includes the range of plausible values for subsequent benefit transfer, and so this should likely cover just a few months or a year, all the way to the upper end of "forever," or at least some fairly high value (e.g., 500 years). Responses to an SP survey scenario will reflect the total value respondents' hold for an improvement in the environmental endpoints, including non-use values like existence and bequest values. An adequately long time horizon is needed to ensure such values are fully captured. When estimating the corresponding indirect utility function based on respondents' choices, this duration attribute could be modelled continuously. Values estimated in this way would capture the stream of benefits that respondents derive, including utility gained from improved quality for future generations (i.e., bequest values). And of course, as with all proposed attributes, extensive focus group testing is needed.

Valuing human health effects are outside of the current task, and are therefore not of primary interest. Nonetheless, human health is an important factor that must be controlled for. In the simplest case, the human health attribute could be modelled as a continuous variable, but if nonlinearities are important confounding factors to control for, then perhaps modelling it with a series of indicators denoting each level would be needed. Either way, this attribute is not of direct interest, and it is not needed to estimate WTP estimates for human health improvements based on this variable. Human health benefits due to improved chemical management and regulations is the focus of OECD's SWACHE Project (see chapter 1 for details). Here the goal is to make sure human health motivations do not confound more direct reasons for why people value environmental endpoints. In fact, if included in a survey instrument, one could consider not varying the human health attribute at all although including such variation may allow for a nice validity check and increase perceived credibility by respondents.

Finally, the geographic scale of the potential impacts avoided by the policy could be accounted for using a split-sample design, or as an attribute that varies across choice questions, as shown in Figure 4. Such decisions depend on the findings from focus group testing. This variable could be accounted for in the econometric model using a series of indicators denoting the corresponding scale.

When developing the experimental design and determining the necessary sample size, number of choice questions, etc., one should consider the possible inclusion of interaction effects between several of the attributes. For example, a respondent may value a policy that covers a larger area (i.e., a direct effect), but they may also have a higher marginal WTP for other attributes (e.g., reduced losses in species) as the policy area size increases (i.e., an interaction effect). Similar interaction effects should be considered with respect to other variables, and in particular, with respect to any uncertainty attribute that may be included. Illustrative economic models that formally incorporate such interaction effects are provided in the Annex.

Figure 4. Discrete Choice Experiment (DCE) Question with Generalisable Endpoints

Suppose the following policy is being considered to limit the use of a chemical in manufacturing and consumer products (e.g., cleaning, personal care, and cooking products) over the next 10 years. This chemical is sometimes released into the environment by industries, as well as when products are used and later put in the trash by households like yours. By limiting the use of this chemical and exposures in the environment, this policy would increase the costs of various products you buy, and thus increase your annual cost of living for the next 10 years.

If the policy is NOT enacted, then:

- The chemical would continue to be used by you and others at its current rate, and would impact the environment as described in the table below.
- You would experience no increase in costs to your household.

If the policy is enacted, then:

- The chemical would be used less and releases into the natural environment would be reduced. The potential impacts described in the table would be avoided, and the environment in terms of chemical contamination would remain like how it is today.
- Your annual cost of living would go up for the next 10 years.

Would you prefer the chemical continue to be used as it is currently and experience no increase in costs? Or would you prefer the policy be enacted and your household experiences **a $XXX increase in annual costs ($XX per month) for the next 10 years**.
Please check one option below.

	No Policy
What is impacted? What types of species (plants, animals, and other organisms) would potentially be impacted. (Levels 1 through 4)	Levels 1 and 2  (microorganisms, plants, small animals, worms, and insects)
How bad are impacts? How many species on average would potentially disappear from this ecosystem.	3 out of every 10 species in the ecosystem will disappear.
How likely are these impacts? Probability that above number of species would disappear.	45% likelihood that these impacts will occur.
How long do impacts last? How long it would take the ecosystem to recover back to current conditions if exposure is reduced in the future.	10 years
Human health impacts. How harmful is the chemical to humans. (0 = no health effects to 5 = severe non-fatal health effects)	1 minimal acute health effects (e.g., minor headaches, nausea)
Where do these impacts occur? The area where the impacts would occur.	National 

○ I prefer NO POLICY, and will pay $0.

○ I prefer the POLICY to avoid these impacts,
and would *pay an $XX every year for the next 10 years*.

The key advantage of a DCE like that proposed in Figure 4 is that the estimated transfer function will allow subsequent estimates for policy applications to be catered to specific chemicals and policy areas. However, two potential disadvantages could diminish the utility of a DCE question like that in Figure 4 in informing BCAs of chemical policies. First, a question like that in Figure 4 may be overly complex and impose a high level of cognitive burden on respondents. If so, respondents will resort to general rules of thumb when making decisions, which will invalidate their responses and the resulting WTP estimates. Extensive testing through focus groups and cognitive interviews with participants from the general public of the proposed study areas would be needed to ensure comprehension, perceived consequentiality, and minimal protests and other biasing behaviours. Second, perhaps the current state of the science is too uncertain to accurately communicate potential impacts, let alone project future benefit transfer scenarios based on actual policies. Extensive consultation with ecotoxicologists, environmental risk assessors, and policy practitioners is needed to decide if the uncertainty is just too great (and will remain so in the near and medium-term). If that is the case, then a more accurate, qualitative depiction of the current science and environmental conditions is more appropriate. Following the precautionary principle, respondents may still hold valid values for preventing uncertain adverse outcomes in the future. Elicited WTP values from a well-developed, thoroughly tested survey instrument could still serve as a useful ex ante welfare measure for BCA. In the next section, a much simpler valuation question is proposed, focused primarily on scientific uncertainty.

6.2. Proposal 2: Valuation Given Significant Scientific Uncertainties

In the contexts of interest here, it is quite possible that consultation with experts and focus group testing may reveal that a generalised endpoint approach like the one in Figure 4 is not feasible due to insurmountable uncertainties in the science, at least in the foreseeable future, and or high cognitive burden on respondents from the general public. In such cases, a valuation question that describes the scenario in mainly qualitative terms may be a necessary simplification. Proposals 1 and 2 are not necessarily mutually exclusive, and some combination of the two approaches may ultimately turn out to be the best anticipated path forward.

The simplest possible case is presented in Figure 5 below, where respondents are presented with a single dichotomous choice question. The starting point for this contingent valuation (CV) question is that in the face of uncertainty, preferences for precautionary actions are likely the biggest driver of benefits. Therefore, uncertainty is the only feature that is varied across the dichotomous choice question each respondent would receive. However, how split-sample designs based on various dimensions could be incorporated to provide validation tests and allow for greater flexibility for later benefit transfer are discussed.

Responses to a valuation question like that in Figure 5 will reflect the total value respondents hold, including use and non-use values, for a qualitatively and (quite possibly) vaguely defined avoided decrease in environmental quality. Such an ill-defined commodity is generally not deemed satisfactory among non-market valuation economists and practitioners, but it may be an honest depiction of the current science in many cases. If the best science truly cannot even provide informed guesses about what the potential avoided damages might be, then that is presented as accurately as possible to respondents and attempt to elicit their WTP given those uncertainties. This approach is in stark contrast to the preceding valuation approach in section 6.1, where an attempt is made to define the changes in environmental quality more concretely.

As currently framed, the uncertainty variable is presented qualitatively. One could envision an experimental design where this qualitative text varies on some scale to reflect different levels of uncertainty. One could also consider a more quantitative measure, as in the proposed DCE question in Figure 4, but if determined to be too cognitively burdensome based on focus group testing, then a qualitative uncertainty measure may be the most appropriate path forward. This would not necessarily lead one to sacrifice scientific rigour, however. Prior to the valuation scenario questions can be posed to assess respondents' subjective interpretation of qualitative probability statements. In other words, "likely affect" can be converted into a respondent-specific quantitative, perceived probability (Wallsten et al., 1986; Manski, 2004; Spiegelhalter et al., 2011). This quantitative measure could then be explicitly included in the econometric model. Such an approach builds on previous efforts on collecting subjective probabilities (e.g. Andersen et al., 2014; Harrison et al., 2022; Lundhede et al., 2015; Scarpa et al., 2021).

Figure 5. Example Dichotomous Choice Question Focused on Uncertainty

The government is considering a regulation on a specific chemical that is currently used by a wide range of industries. The chemical is safe for humans. In fact, this chemical can be found in cleaning, personal care, and cooking products commonly found in your home.

However, when these products are discarded and the chemical is released into the environment, it has been determined to likely affect microorganisms, plants, and wildlife in your country. Such harm potentially includes reduced rates of survival and the ability for affected species to reproduce.

The words "likely affect" are emphasised because the current scientific evidence is not enough to determine with certainty that the chemical does actually harm these ecosystems.

The proposed regulation will make manufacturers reduce the use of this chemical for 10 years, when it is expected that the scientific evidence will be strong enough to determine how harmful the chemical is to the environment. If the regulation is enacted, this means that this chemical would not potentially harm the microorganisms, plants, and other wildlife.

There are alternative chemicals that industries can use in the products that you buy and use, but these alternative chemicals cost more and so it would result in some of the products you buy being more expensive. The increase in costs to your household would be **$XXX per year (or $XX per month) over the next 10 years.**

Would you vote in favour of the proposed regulation? *Please check one.*

- ○ Yes
- ○ No

Although the simplicity of the proposed CV question leads one to sacrifice dimensions that could otherwise be used to further cater benefit estimates to a particular context when conducting benefit transfer, some flexibilities through cross-respondent variation in a split-sample design can potentially be incorporated. There are several dimensions to consider, but in order of priority it is proposed to first consider a split sample design where the qualitative (or quantitative) uncertainty measure varies. If qualitative values are used, for example, variation in the text could range from: “possible, but not likely affect” to “very likely affect.” A second dimension to consider varying in a split-sample design would be geographic scale. This would allow for tests of scope sensitivity, and as discussed by Navrud (2017), would allow for adjustment factors to account for differences in the study versus policy sites when conducting benefit transfer. The geographic scale of the study area is underlined in Figure 5 above, and this could be varied across subsamples. A third consideration for a split-sample design is to include the “no human health effects” text for half the sample, and not include this text for the other half. This would allow for a validation test to determine if respondents were really focusing only on environmental benefits, which is the objective of the proposed research.

Chapter 7. Role of Experts and Practitioners

In the context of chemical management and policy decisions, the goal of this scoping study is to discuss the feasibility of SP protocols to enable estimation of non-market benefits (or costs) from changes in welfare-relevant environmental endpoints, and to suggest potential paths towards that goal. The two SP protocols suggested in chapter 6 are prototypes that follow empirical and theoretical conventions in Economics –particularly, in the non-market valuation field. These prototypes, however, are meant as informed starting points for the development and testing of a SP survey to implement in the field. In order to be useful towards informing policy, both practitioners and experts across fields must help shape the posited hypothetical "markets" behind the valuation scenarios.

Experts and practitioners will play important roles in four realms. The first three areas pertain to survey development, and the fourth surrounds future benefit transfer applications (which occur after a SP survey has been implemented and the data analysed). First, there are scientific terms that require precise, non-technical descriptions to general audiences – this challenge can only be overcome after several iterations in which non-market valuation economists and experts test communication strategies to accurately depict the concepts to general audiences. Second, experts will be essential in deciding whether devising and expressing quantitative risks (i.e. probabilistic statements about a number of outcomes) to represent scientific uncertainty is feasible, or whether one can only accurately express uncertainty in qualitative terms (i.e. the notion that the evidence is not yet sufficient to quantify and communicate probabilities to reflect uncertainty).

Third, ecotoxicologists, environmental risk assessors, and related discipline experts, as well as policy practitioners, are needed to inform the SP study experimental design. More specifically, the number and range of values for all attributes (or endpoints) must be informed by the experts' informed guesses of plausible ranges that could be experienced in the future. The simulated SP scenarios that respondents receive are hypothetical and randomly assigned, and so in that sense the actual values do not matter. However, the corresponding attribute space (or range) from which those random values are selected is important, and should cover all potentially relevant future policy and baseline outcomes. Such coverage increases the credibility of future benefit transfer exercises by ensuring that the estimates are based on within sample predictions. Extrapolating WTP estimates outside of the range of observed environmental endpoint values would still be possible, but is less credible. In addition, for discrete variables like the impacted area, the actual values matter. Natural science experts and policy practitioners are needed to ensure that the impacted areas posited in the hypothetical scenarios match the policy-relevant areas to which future benefit estimates will be extrapolated. For example, if future policies are mainly at the national level, it should be ensured that "nationwide" is one included value for the "Where do these impacts occur?" attribute.

The fourth area for expert involvement pertains to subsequent benefit transfer exercises to inform decision makers. After the SP survey is implemented and a benefits transfer function has been estimated, the help of ecotoxicologists and related discipline experts who are familiar with the relevant chemical(s), policy areas, and potential impacts on the environment and corresponding uncertainties around those impacts will be needed. Elicitation of potential values and changes in endpoints from these experts is required in order to derive policy-specific attribute values to be plugged in when predicting benefit estimates for BCA. In the absence of scientific models to link all steps outlined in figure 1, expert elicitation protocols are needed to fill in any gaps.

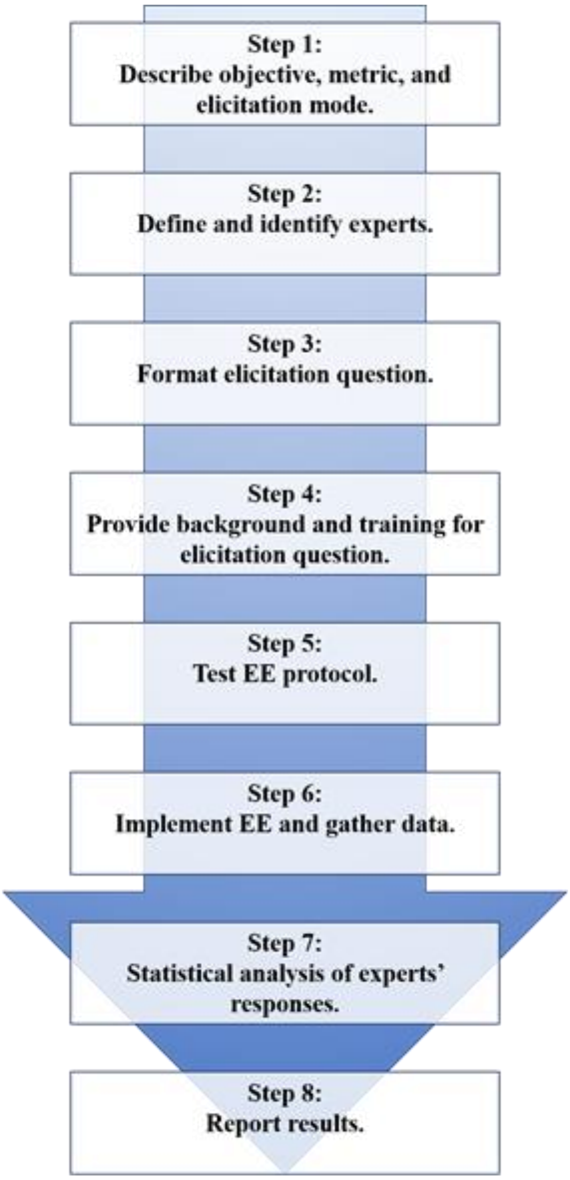
Opinions and suggestions from experts, practitioners and policy makers can be gathered in a systematic way through a number of strategies that include formal and informal conversations, focus groups, semi-structured interviews, and stakeholder analysis, among others. While these strategies are useful means of communication among professionals of different disciplines, the specific challenge in such an effort is that opinions and suggestions from experts, practitioners and policy makers must be expressed not only in a qualitative manner but eventually (and ideally) in quantitative terms. For instance, if the SP scenario communicates a quantitative measure of uncertainty about the impacts of a chemical on an environmental endpoint (e.g., PDF of species), then experts would be required to report their informed opinion regarding central tendency and dispersion of a distribution reflecting the possible values of changes in that endpoint that could be realised.

In this context, expert elicitation is suggested as a useful tool to systematically represent opinions of experts in a quantitative manner. Expert elicitation (EE) is a structured process that collects scientific and technical judgements from experts (Morgan, 2014; Bosetti et al., 2016). EE is deemed useful to gather empirical data when other strategies are expensive, limited or unreliable (Bolger and Rowe, 2015; James et al., 2010). EE has also been used to inform key sources of uncertainty in policy analyses where the scientific data and models are inadequate or unavailable (e.g., Morgan, 2014; IEc, 2006).

EE conventionally consists of eight iterative steps, as shown in Figure 6. In the first step, EE describes the objective and mode of the elicitation. A goal's definition includes an unambiguous description of the metric that experts are requested to use when reporting their answers. The elicitation mode ranges from face-to-face interviews to protocols that can be self-administered via a web-based platform –and the selection of elicitation mode comes with trade-offs that must be weighed (see Baker et al., 2014; Verdolini et al., 2015). In the second step, the type of expertise needed is defined, and individuals with such expertise are identified. In the third step, the format of the elicitation question is determined. In the fourth step, experts are provided with background material and trained in the rationale behind the elicitation question. In the fifth step, the EE protocol is piloted –making sure that the metric of interest is clear to the experts and that scenarios are precisely described. The sixth step consists in gathering the data via the elicitation protocol. In the seventh step, statistical analysis of the gathered data is carried out. The final step consists of reporting experts' opinions (see Bosetti et al. (2016) for details about these steps). Usually EE consists of several iterations of steps 1 through 5, so that when the EE is fully implemented and the data gathered, there is no need to go back and re-design the protocol.

Exploration of synergies between SP protocols and expert elicitation tools is a relatively recent area of research, but with substantive work on a number of fronts. For instance, Alberini et al. (2006) use a DCE to gather experts' opinions on country-level adaptive capacity to climate change. Leon et al. (2003) elicit opinions of experts to inform benefit transfer exercises. Strand et al. (2017) have taken the EE approach one step further by asking experts on environmental valuation to provide their opinion on the WTP of populations in their countries for Amazon Forest protection; and Ahtiainen and Martinez-Cruz (2017), taking the approach even further, asked experts in non-market valuation to carry out a benefit transfer themselves. Martinez-Cruz et al. (2017) adapted a double-bounded dichotomous CV question to elicit experts' opinions on expected impacts from climate change on potato yields in the Bolivian Altiplano. Sainz-Santamaria and Martinez-Cruz (2019) implemented this adapted CV question to elicit experts' opinion on impacts from irrigation policies on recharge of an aquifer in Aguascalientes, Mexico.

Figure 6. Expert elicitation (EE) steps



EE is seen as a way to first inform the survey development in terms of what endpoints to include, how to communicate those endpoints, and what the relevant values should be in the SP study experimental design. Second, given the significant scientific uncertainties regarding the impact of different chemicals on environmental endpoints, EE can be used to fill in the benefits analysis gaps (as illustrated in Figure 1) when later performing benefits transfer for policy analysis. Ultimately, EE would inform estimation of policy-specific projected baseline and policy scenario values of the relevant endpoints, which practitioners could then plug in to the parameterised valuation functions estimated from a SP study. Illustrative economic models and technical details on plugging in endpoint values for benefit transfer are provided in the Annex.

A CV study by Van Houtven et al. (2014) provides a useful example of combining EE and SP protocols, and demonstrates how EE can help tackle the uncertainties illustrated in Figure 1. They asked respondents to report their WTP for improvements in lakes due to a hypothetical State-level regulation on the use of nitrogen and phosphorus. To tackle the challenge of linking reductions in emissions to lake quality endpoints, Van Houtven et al. use a eutrophication index constructed by expert judgements. In particular, the changes in lake quality presented to respondents were validated by water quality experts through an EE protocol that aimed to reduce the uncertainties involved in extrapolating how reductions in the use of nitrogen and phosphorus affect lakes.

Implementation of a similar EE protocol for benefit transfer in the context of chemical management and regulatory decisions seems like the most viable path forward, at least until the science catches up. If experts and practitioners have concerns with making premature jumps, the counter is that policy decisions must still be made, and no action in and of itself is a policy decision. Practitioners can introduce additional sensitivity analyses into their BCAs to transparently convey analytical uncertainties, and ultimately to inform decision makers to the greatest and most accurate extent possible.

Chapter 8. Proposed Path Forward

In this section the possible next steps in methodological development and advancing OECD's ability to value environmental endpoints related to chemical regulations and management are described. The next steps include consultation with ecotoxicologists, environmental risk assessors, and practitioners at OECD and regulating agencies; focus groups and cognitive interviews with the general public to iteratively test and refine draft survey instruments in a small number of countries; a pilot study to implement the survey in one or a few countries; and an actual or illustrative benefit-transfer exercise to serve as a proof-of-concept.

This paper elaborates on each suggested step below, and categorises these steps into three sequential stages. Ideally, all three stages would be carried out, but to account for the potential desire for incremental investment and possible adjustments, each stage could be implemented sequentially. Table 2 outlines proposed deliverables and possible metrics or goals to measure the success of each step.

8.1. Stage 1: Survey development

Step 1.1: First, the research team would want to further elaborate on both survey instrument proposals in chapter 6 and more formally sketch out the econometric models that are proposed in the Annex of this scoping study. An expert elicitation plan could then be developed, and semi-structured interviews and focus groups with ecotoxicologists, environmental risk assessors, and practitioners would be held. Consultations with the former are needed to make sure the science is accurately characterised and that the most critical dimensions are being touched on –here is where the EE protocols would be first utilised. Importantly, experts will be essential in determining whether scientific uncertainty is too large that it precludes us from treating uncertainty in a similar fashion as risk, and making probabilistic statements regarding the likelihood of an outcome being realised. Interaction with policy makers and experts on non-market valuation is needed to ensure that the proposed econometric models are in fact useful for later benefit transfer. The proposed survey instruments would then be revised accordingly.

Consultation with practitioners is also needed for choosing potential countries for survey development. One would ideally want to ensure that the survey is developed for and tested in a few countries to reflect general differences in languages, culture, and variation in the distributions of income and education. Ideally, survey development and testing would take place in at least one country in Europe, the Americas, and in particular, Asia. Navrud (2018) notes that there are few studies of relevant impacts in Asia, but at the same time, countries in Asia often have the highest production of toxic chemicals. Variation in the distribution of education and income across country populations is important. If the goal is to have a survey instrument that can be implemented in multiple countries, then one needs to make sure comprehension, credibility, and perceived consequentiality of the survey instrument hold across populations. Testing across the full range of education levels in the potential sample frame is critical. Another more practical consideration is whether the countries have robust and credible internet panels available for purchase. In order to capture a representative convenience sample for subsequent pilot tests.

Step 1.2. Hold a series of focus groups and 1-on-1 cognitive interviews in the chosen pilot countries. Initial focus groups could be used to examine whether the overall structure of the draft survey instruments, and in particular the valuation scenarios, are viable. Most importantly, focus group participants would help identify what policy-relevant endpoints are most relevant to them. The survey instrument and valuation questions would be revised accordingly, and iteratively, based on several rounds of testing. Early on one would be able to decide whether a DCE or CV approach is more viable in terms of respondent comprehension – with the degree of underlying scientific uncertainty playing a key role in this decision.

It is possible that the proposed DCE, or even a simpler variant of it, are just too challenging for respondents, especially given the scientific uncertainty. If so, one may then focus on the CV proposal. Both Abate et al. (2020) and Van Houtven et al. (2014) initially wanted to pursue a DCE, but for the exact aforementioned reasons, reverted to a more understandable CV format.

Later focus groups and cognitive interviews would be used to refine the survey text to ensure comprehension and accurate interpretation by the respondents, minimise respondent fatigue, test functionality of the electronic survey, and establish the relevant range for the cost attribute space.

Step 1.3. Conduct expert elicitation with ecotoxicologists, environmental risk assessors, and related discipline experts, and then refine the survey experimental design based on their feedback. Experts' opinions are necessary at this step to address two issues. First, once the attributes found to be most relevant to the general population are established, one would then want to determine the relevant attribute space (or range) in the experimental design for each variable. To maximise the applicability of subsequent results for benefit transfer, it must be ensured that the attribute value ranges in the experimental design capture all plausible values. Second, additional discussion and relationship-building is needed at this stage in order to develop a plan and confirm that later expert elicitation will be possible for purposes of benefit transfer. For example, if a dose-response function is not available, experts can provide opinions to gain insights (as discussed in chapter 7). Based on the feedback from the expert elicitation, and simulations and statistical power analysis, the draft electronic survey and experimental design would be finalised, and the necessary sample sizes would be determined. A memo formally outlining the final econometric models and procedures for later benefit transfer should be compiled at this stage, in order to ensure all the proper pieces are in place before investing in fielding the survey.

8.2. Stage 2: Survey pilot

For a pilot study as the one proposed here, a convenience sample from a compiled internet panel (like those often maintained by marketing and research firms) is the most appropriate path forward. Although this would clearly be a convenience sample and is open to criticism in terms of possible selection biases compared to other sample frames and modes, it is a suitable and cost-effective option given that the objective of the pilot study would be to test the methodology and valuation scenarios. Additionally, one can still ensure that the sample is representative of the sample frame in terms of key sociodemographic characteristics (e.g., age, gender, region within the country, income, and education).

Step 2.1. For the chosen pilot country (or ideally countries) a small pilot (with just n = 50 to 100 respondents, for example, depending on the complexity of the survey) could first be conducted to ensure comprehension and functionality of the internet-based survey, as well as viability of the responses based on intuition and economic theory. Examining the viability of the responses at this stage will be judged purely on sign, magnitude, and some simple descriptive statistics of key variables. More formal statistical judgements would be saved for the full pilot in step 2.2. In carrying out a survey on an internet panel, it would be necessary to i) coordinate with the internet panel provider to make sure that the visuals are well visible on the scripted version for the various machines used (e.g., laptop or smartphone); ii) verify language translations made by internet panel provider(s); and iii) test the scripted survey to check for typos and mistakes. Some of these tasks may also be included as part of the cognitive interview testing in stage 1.

Step 2.2. The full pilot for the chosen country or countries could then be administered. The pilot surveys could ideally be administered at the same time in a few different countries, but could also be rolled out sequentially based on cost constraints or other considerations. The data would then be empirically analysed, and a series of unit values and benefits functions would be estimated. A report documenting the survey development, implementation, econometric methods and results would be written up.

8.3. Stage 3: A proof-of-concept benefit transfer exercise

An actual or illustrative benefit-transfer exercise could be carried out to serve as a proof-of-concept. The objective is to illustrate the steps and feasibility of the approach. For purposes of this pilot study of a single country (or a few countries), the benefit transfer would entail a within-country application, rather than extrapolating benefits to a policy in a separate (out-of-sample) country. The latter exercise may prove viable at some point, but in order to assess the validity of such an exercise, the survey would need to be implemented across many countries.

Step 3.1. First, a structured expert elicitation procedure could be carried out to project the quantified changes in the relevant attributes (Van Houtven et al., 2014; Hemming et al., 2018). This step is essential because, until science catches up, a well-developed expert elicitation is a relatively cheap, fast, and reliable approach to bridging the gap between the outputs from ecotoxicological studies and policy-relevant endpoints. Therefore, at least for now, a solid and replicable expert elicitation protocol should be developed for future benefit transfer applications.

Step 3.2. Carry out a benefit-transfer exercise based on the quantified changes in the endpoints and estimated benefits transfer functions from steps 3.1 and 2.2, respectively. For purposes of this proof-of-concept illustration, one would transfer the estimates based on the data from one of the pilot countries to a policy scenario in that same country.

This benefit transfer exercise could then potentially be extended and improved by adding a calibration procedure relying on a Bayesian strategy that infers prior distributions of WTP measures via an expert elicitation of non-market valuation economists. This additional step would closely follow previous work by Leon et al. (2003), Strand et al. (2017), and Ahtiainen and Martinez-Cruz (2017). Such a Bayesian calibration strategy would aim to decrease transfer errors. The precision of benefit transfer estimates remains a limitation of the methodology – absent this calibration procedure, average and median transfer errors have been documented in the range of 35% and 39%, and even as high as 400% (Boyle et al., 2010; Kaul et al., 2013; Kosenius and Ollikainen, 2015).

Chapter 9. Conclusion

One of the conclusions of the 2013 workshop for the Royal Society of Chemistry, UK Environment Agency, and UK Chemicals Stakeholder Forum was that scientific uncertainty and the general lack of data may imply that there will never be enough information for detailed benefit-cost analysis (BCA), but decision makers must still make decisions. With that motivation in mind, that scientific uncertainty has been embraced and several potential paths have been identified to pursue in order to better inform BCAs of regulatory and management decisions of chemicals. Although not perfect, well-thought out and appropriately caveated benefit estimates are better than no estimates at all. The proposed valuation scenarios and key decision points discussed in this scoping study, as well as the proposed incremental steps, will help make strides towards monetising the environmental benefits from the policy and management of chemicals.

Table 2. Proposed steps and measures of success

Steps		Measures of success.
Stage 1: Survey Development		
1.1	Initial draft paper survey(s) for focus group testing.	Draft survey(s) completed and reasonably agreed upon by team of experts and practitioners.
	Memo outlining potential pilot countries with stakeholder interest, and describing firms with internet-based respondent panels in those countries.	Completed memo with lists of potential study areas and respondent-panel providers.
1.2	Summary focus group report.	Completed report.
	Revised and thoroughly tested survey instrument.	Focus group report and data from debriefing questions will provide information on respondent comprehension, perceived consequentiality, biasing behaviours, etc.
1.3	Report formally describing econometric models, experimental design (i.e., attribute values, statistical power analysis, sample size, etc.) and procedures for subsequent benefit transfer.	Completed memo that is reasonably agreed upon by team of experts and practitioners (possibly peer-reviewed by non-market valuation experts).
Stage 2: Survey Pilot.		
2.1	Fully functional computer-based survey instrument, translated to necessary language(s).	Review by stakeholders and expert team.
	Small initial pilot dataset of responses (n = 50 to 100) to confirm functionality.	Informal analysis of item completeness, and consistency with economic theory.
2.2	Full pilot dataset of survey responses.	Target sample size reached. Quality of data in terms of representativeness, completeness, etc. will be examined in econometric analysis.
	Final report of econometric analysis and results.	Analysis will assess representativeness and completeness of data based on summary statistics. The full benefits function will be estimated, and then assessed based on whether estimated parameter signs and magnitudes match economic theory (e.g., tests for scope sensitivity, diminishing marginal utility of income, etc.).
Stage 3: Proof-of-concept benefit transfer exercise.		
3.1	Memo detailing expert elicitation plan for benefit transfer.	Completion of memo, identify and recruit relevant experts.
3.2	Project baseline and policy endpoint values and corresponding subjective distributions based on expert elicitation.	Completion of expert elicitation protocol and achievement of main aim of eliciting relevant endpoint values and distributions.
	Final report detailing full expert elicitation procedures, formal benefit transfer calculations, and final benefit results.	Completion of memo and satisfactory review by research team and relevant stakeholders.

References

Abate, T. G., T. Börger, M. Aanesen, J. Falk-Andersson, K.J. Wyles, and N. Beaumont (2020), "Valuation of marine plastic pollution in the European Arctic: Applying an integrated choice and latent variable model to contingent valuation," *Ecological Economics*, Vol. 169, 106521.

Abildtrup, J., Garcia, S., Olsen, S. B., and Stenger, A. (2013), "Spatial preference heterogeneity in forest recreation," *Ecological Economics*, Vol. 92, pp. 67-77.

Ahtiainen, H., and A.L. Martinez-Cruz (2017), "Using expert elicitation to reduce benefit transfers: An application to the Baltic Sea," available at https://www.researchgate.net/publication/310464541_Using_expert_elicitation_to_reduce_benefit_transfer_errors_An_application_to_the_Baltic_Sea.

Alberini, A., and A. Chiabai (2007), "Discount Rates in Risk Versus Money and Money versus Money Tradeoffs," *Risk Analysis*, Vol. 27(2), pp. 483-198.

Alberini, A., A. Chiabai, and L. Muehlenbachs (2006), "Using expert judgment to assess adaptive capacity to climate change: evidence from a conjoint choice survey," *Global Environmental Change*, Vol. 16 (2), pp. 123–144.

Alberini, A., S. Tonin, M. Turvani, A. Chiabai (2007), "Paying for permanence: Public preferences for contaminated site cleanup," *Journal of Risk and Uncertainty*, Vol. 34, pp. 155-178.

Alberini, A., M. Scasny, D. Guignet, and S. Tonin (2012), "Cancer values of prevented fatalities (VPFs), one size does not fit all: The benefits of contaminated site cleaups in Italy," *Journal of the Air and Waste Management Association*, Vol. 67(1), pp. 783-798.

Alberini, A. and M. Scasny (2014), "Stated preference study to examine the economic value of benefits of avoiding selected adverse human health outcomes due to exposure to chemicals in the European Union," report prepared for the European Chemicals Agency, available at: https://echa.europa.eu/documents/10162/13630/study_economic_benefits_avoiding_adverse_health_outcomes_3_en.pdf/ccf1db6e-fa50-45e3-a0b4-178a05cc8b96

Andersen, S., J. Fountain, G.W. Harrison, and E.E. Rutström (2014), "Estimating subjective probabilities," *Journal of Risk and Uncertainty*, Vol. 48 (3), pp. 207-229.

Antonio, M., G. Roder, and S. Troiano (2022), "Better safe than sorry? Stated preferences and the precautionary principle for securing drinking water quality in an Italian district," *Utilities Policy*, Vol 76, 101355.

Atherton, J., G. Atkinson, D. Contu, S. Georgiou, and S. Mourato (2020), "Valuing Public Preferences for Complex and Unfamiliar Environmental Bads: The case of chemical substances in surface waters in the UK," Unpublished Manuscript.

Backus, D., A. Ferriere, and S. Zin (2015), "Risk and ambiguity in models of business cycles," *Journal of Monetary Economics*, Vol. 69, pp. 42-63.

Banzhaf, H.S., D. Burtraw, D. Evans, A. Krupnick (2006), "Valuation of Natural Resource Improvements in the Adirondacks", *Land Economics,* Vol. 82(3), pp. 445-464.

Bolger, F., and G. Rowe (2015), "The aggregation of expert judgment: do good things come to those who weight?," *Risk Analysis*, Vol. 35 (1), pp.5–11.

Bosetti, V., L.D. Anadón, E. Baker, L.A Reis, and E. Verdolini (2016), "The future of energy technologies: an overview of expert elicitations," GGKP Working Paper 01, http://www.greengrowthknowledge.org/sites/default/files/downloads/resource/The_Future_of_Energy_Technologies_An_Overview_of_Expert_Elicitations_0.pdf (Accessed June 24, 2022).

Boyd, J. and A. Krupnick (2013), "Using Ecological Production Theory to Define and Select Environmental Commodities for Non-market Valuation", *Agricultural and Resource Economics Review*, Vol. 42(1), pp. 1-32.

Boyle, K.J., N.V. Kuminoff, C.F. Parmeter, and J.C. Pope (2010), "The benefit-transfer challenges," *Annu. Rev. Resour. Econ*., Vol 2(1), pp. 161-182.

Breeze, T.D., A.P. Bailey, S.G. Potts, and K.G. Balcombe (2015), "A stated preference valuation of the non-market benefits of pollination services in the UK", *Ecological Economics*, Vol. 111, 76-85.

Brouwer, R., J. Martin-Ortega, and J. Berbel (2010), "Spatial Preference Heterogeneity: A Choice Experiment," *Land Economics*, Vol. 86 (3), pp. 552-568.

Börger, T., Q. T. K. Ngoc, L. Kuhfuss, T. T. Hien, N. Hanley, and D. Campbell (2021), "Preferences for coastal and marine conservation in Vietnam: Accounting for differences in individual choice set formation," *Ecological Economics,* Vol. 180, 106885.

Campbell, D. (2007), "Willingness to pay for rural landscape improvements: Combining mixed logit and random-effects models," *Journal of Agricultural Economics*, Vol. 58(3), pp. 467-483.

Carson, R. T. (1998), "Valuation of tropical rainforests: philosophical and practical issues in the use of contingent valuation", *Ecological Economics*, Vol. 24, pp. 15-29.

Caussade, S., J. de Dios Ortúzar, L.I. Rizzi, and D.A. Hensher (2005), "Assessing the influence of design dimensions on stated choice experiment estimates," *Transportation research part B: Methodological*, Vol. 39 (7), pp-621-640.

Choi, D.S., and R. Ready (2021), "Measuring the benefits from spatially-explicit surface water quality improvements: The roles of distance, scope, scale, and size," *Resource and Energy Economics*, Vol. 63, 101108.

Chiu, W.A. (2017), "Chemical risk assessment and translation to socioeconomic assessments", *OECD Environment Working Papers*, No. 117, OECD Publishing, Paris, https://dx.doi.org/10.1787/a930054b-en.

Colombo, S., W. Budziński, M. Czajkowski, and K. Glenk (2022), "The relative performance of ex-ante and ex-post measures to mitigate hypothetical and strategic bias in a stated preference study," *Journal of Agricultural Economics*, DOI: 10.1111/1477-9552.12484.

Corona, J., T. Doley, C. Griffiths, M. Massey, C. Moore, S. Muela, B. Rashleigh, W. Wheeler, S. D. Whitlock, and J. Hewitt (2020), "An Integrated Assessment Model for Valuing Water Quality Changes in the United States," *Land Economics*, Vol. 96 (4), pp. 478-492.

Czajkowski, M., W. Budziński, D. Campbell, M. Giergiczny and N. Hanley (2017a), "Spatial heterogeneity of willingness to pay for forest management," *Environmental and Resource Economics*, Vol. 68(3), pp.705-727.

Czajkowski, M., C.A. Vossler, W. Budziński, A, Wiśniewska, and E. Zawojska (2017b), "Addressing empirical challenges related to the incentive compatibility of stated preferences methods," *Journal of Economic Behavior & Organization*, Vol. 142, pp. 47-63.

Donohue, M., and W. Kipusi (2016), "Recent Valuation Research on Environmental and Human Health Impacts Linked to Harmful Chemicals," presentation for the OECD Workshop on Socioeconomic Impact Assessment and Chemical Management, July.

Dubourg, R. (2018), "Economic assessments of the benefits of regulating mercury: A review", *OECD Environment Working Papers*, No. 132, OECD Publishing, Paris, https://dx.doi.org/10.1787/77045f1a-en.

Fox, D.R., R. A. van Dam, R. Fisher, G. E. Batley, A. R. Tillmanns, J. Thorley, C. J. Schwarz, D. J. Spry, and K. McTavish. (2021), "Recent Developments in Species Sensitivity Distribution Modeling," *Environmental Toxicology and Chemistry*, Vol 40 (2), pp. 293-308.

Forbes, V.E., C.J. Salice, B. Birnir, R. Bruins, P. Calow, V. Ducrot, N. Galic (2017), "A framework for predicting impacts on ecosystem services from (sub) organismal responses to chemicals," *Environmental Toxicology and Chemistry*, Vol. 36 (4), pp.845-859.

Gabbert, S. (2018), "Economic assessment and valuations of environmental and health impacts caused by Perfluorooctanoic acid (PFOA) and its salts", *OECD Environment Working Papers*, No. 128, OECD Publishing, Paris, https://dx.doi.org/10.1787/f5250745-en.

Ghosh, R. K., S. Gupta, V. Singh, and P.S. Ward (2021), "Demand for crop insurance in developing countries: New evidence from India," *Journal of Agricultural Economics*, Vol. 72(1), pp. 293-320.

Glenk, K., and S. Colombo (2013), "Modelling outcome-related risk in choice experiments," *Australian Journal of Agricultural and Resource Economics*, Vol. 57 (4), pp. 559-578.

Hagan, D.A., J.W. Vincent, and P.G. Welle (1999), "Economic Benefits of Reducing Mercury Deposition in Minnesota," Report for the Minnesota Pollution Control Agency and the Legislative Commission on Minnesota Resources.

Hampson, D.I., S. Ferrini, D. Rigby, and I.J. Bateman. (2017), "River Water Quality: Who Cares, How Much and Why?," *Water,* Vol. 9 (621), w9080621.

Harrison, G.W., A. Hofmeyr, H. Kincaid, B. Monroe, D. Ross, M. Schneider, and J.T. Swarthout (2022), "Subjective beliefs and economic preferences during the COVID-19 pandemic," *Experimental economics*, 1-29.

Hemming, V., M.A Burgman, A.M. Hanea, M.F. McBride, and B.C. Wintle (2018), "A practical guide to structured expert elicitation using the IDEA protocol," *Methods in Ecology and Evolution*, Vol. 9 (1), pp. 169-180.

Holland, M. (2018), "Socio-economic assessment of phthalates", *OECD Environment Working Papers,* No. 133, OECD Publishing, Paris, https://dx.doi.org/10.1787/a38a0e34-en.

Holmes, Thomas P., and Wiktor L. Adamowicz. (2003). "Attribute-based methods." In A Primer on Non-market Valuation, Edited by Patricia A. Champ, Kevin J. Boyle, and Thomas C. Brown, Dordrecht, The Netherlands: Kluwer Academic Publishers.

Hund, E., D.L. Massart, and J. Smeyers-Verbeke (2001), "Operational definitions of uncertainty," *Trac trends in analytical chemistry,* Vol. 20 (8), pp. 394-406.

Hunt, A. and N. Dale (2018a), "Economic valuation in formaldehyde regulation", *OECD Environment Working Papers,* No. 134, OECD Publishing, Paris, https://dx.doi.org/10.1787/c0f8b8e5-en.

Hunt, A. and N. Dale (2018b), "Economic valuation in 1-Methyl-2-pyrrolidone (NMP) regulation", *OECD Environment Working Papers,* No. 131, OECD Publishing, Paris, https://dx.doi.org/10.1787/ef67fa08-en.

Industrial Economics (IEc, 2006), "Expanded Expert Judgement Assessment of the Concentration-Response Relationship Between PM2.5 Exposure and Mortality," report prepared for the U.S. Environmental Protection Agency. Accessed 23 June 2022, at https://www3.epa.gov/ttnecas1/regdata/Uncertainty/pm_ee_report.pdf.

Industrial Economics (IEc, 2016), "Economic Valuation Study for Canada's Chemicals Management Plan: Final Survey Implementation and Summary Analysis," report prepared for Health Canada.

James, A., S.L. Choy, and K. Mengersen (2010), "Elicitator: an expert elicitation tool for regression in ecology," *Environmental Modelling & Software*, Vol. 25(1), pp. 129-145.

Joint Committee for Guides in Metrology (JCGM, 2008), "Evaluation of measurement data—Guide to the expression of uncertainty in measurement," available at: https://www.bipm.org/en/committees/jc/jcgm/publications

Johnston, R.J., E.T. Schultz, K. Segerson, E.Y. Besedin, and M. Ramachandran (2012), "Enhancing the Content Validity of Stated Preference Valuation: The Structure and Function of Ecological Indicators", *Land Economics*, Vol. 88(1), pp. 102-120.

Johnston, R.J., E.T. Schultz, K. Segerson, E.Y. Besedin, and M. Ramachandran (2013), "Stated Preferences for Intermediate versus Final Ecosystem Services: Disentangling Willingness to Pay for Omitted Outcomes", *Agricultural and Resource Economics Review*, Vol. 42(1), pp. 98-118.

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

Kaul, S., K.J. Boyle, N.V. Kuminoff et al. (2013), "What can we learn from benefit transfer errors? Evidence from 20 years of research on convergent validity," *Journal of Environmental Economics and Management* Vol. 66, pp. 90-104.

King, P., A. Hunt, S. Georgiou, C. Rainsford, and R. Dean (2021), "Economic valuation of benefits from the proposed REACH restriction of intentionally added microplastics," *Bath Economics Research Paper No. 81 /21*, available at: https://researchportal.bath.ac.uk/en/publications/economic-valuation-of-benefits-from-the-proposed-reach-restrictio.

Knetsch, J.L., Y.E. Riyanto, and J. Zong (2012), "Gain and Loss Domains and the Choice of Welfare Measure of Positive and Negative Changes," *Journal of Benefit-Cost Analysis* Vol. 3 (4), pp. 1-18.

Kniesner, T.J., and W.K. Viscusi (2019), "The Value of a Statistical Life." In Oxford Research Encyclopaedia of Economics and Finance.

Kosenius, A.K. (2010), "Heterogeneous preferences for water quality attributes: The Case of eutrophication in the Gulf of Finland, the Baltic Sea," *Ecological Economics*, Vol. 69, pp. 528-538.

Kosenius, A.K., and M. Ollikainen (2015), "Ecosystem Benefits from Coastal Habitats – a Three-Country Choice Experiment in WTP Space," *Marine Policy*, Vol. 58, pp. 15-27.

Leon, J.C., F. Vazquez-Polo, and R. Leon-Gonzalez, R (2003), "Elicitation of expert opinion in benefit transfer of environmental goods," *Environmental and Resource Economics*, Vol. 26, pp. 199–210.

Lew, D.K. (2019), "Place of residence and cost attribute non-attendance in a stated preference choice experiment involving a marine endangered species," *Marine Resource Economics*, Vol. 34, pp. 225-245.

Lew, D.K., D.F. Layton, and R.D. Rowe. (2010), "Valuing Enhancements to Endangered Species Protection Under Alternative Baseline Futures: The Case of the Steller Sea Lion," *Marine Resource Economics*, Vol. 25, pp. 133-154.

Lew, D. K., and J.C. Whitehead (2020), "Attribute non-attendance as an information processing strategy in stated preference choice experiments: Origins, current practices, and future directions," *Marine Resource Economics*, Vol. 35(3), pp. 285-317.

Logar, I., R. Brouwer, M. Maurer, and C. Ort (2014), "Cost-benefit analysis of the Swiss national policy on reducing micropollutants in treated wastewater," *Environmental science & technology*, Vol. 48 (21), pp. 12500-12508.

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

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

Magat, W.A., Huber J., Viscusi, W.K., and Bell, J. (2000), "An Iterative Choice Approach to Valuing Clean Lakes, Rivers, and Streams," *Journal of Risk and Uncertainty*, Vol 21(1), pp. 7-43.

Maguire, K.B., C. Moore, D. Guignet, C. Dockins, and N.B. Simon (2018), "Muddying the Water? An Analysis of Non-Constant Baselines in Stated Preference Surveys," US EPA National Center for Environmental Economics Working Paper, 18-02, Washington, DC, May.

Manski, C.F. (2004), "Measuring expectations," *Econometrica*, Vol. 72(5), pp. 1329-1376.

Martinez-Cruz, A.L. (2015), "Latent class logits and discrete choice experiments: Implications for welfare measure," *Revue d'économie politique*, Vol. 125 (2), pp. 233-251.

Martinez-Cruz, A. L., M. Juárez-Torres, and S. Guerrero (2017), "Assessing impacts from climate change on local social-ecological systems in contexts where information is lacking: An Expert Elicitation in the Bolivian Altiplano," *Ecological Economics*, Vol. 137, pp. 70-82.

Meyerhoff, J., M. Oehlmann, and P. Weller (2015), "The influence of design dimensions on stated choices in an environmental context," *Environmental and Resource Economics*, Vol. 61 (3), pp: 385-407.

Moore, C., D. Guignet, C. Dockins, K.B. Maguire, and N.B. Simon (2018), "Valuing ecological improvements in the Chesapeake Bay and the importance of ancillary benefits," *Journal of benefit-cost analysis*, Vol. 9 (1), pp. 1-26.

Morgan, M.G. (2014), "Use (and abuse) of expert elicitation in support of decision making for public policy," Proceedings of the National Academy of Sciences of the United States of America, Vol. 111 (20), pp. 7176-7184.

Morris, J., and J.K. Hammitt, (2001), "Using life expectancy to communicate benefits of health care programs in contingent valuation studies," *Medical Decision Making*, Vol. 21 (6), pp. 468-478.

Morse-Jones, S., I. J. Bateman, A. Kontoleon, S. Ferrini, N. D. Burgess, and R. Turner. (2012), "Stated preferences for tropical wildlife conservation amongst distant beneficiaries: Charisma, Endemism, scope and substitution effects", *Ecological Economics*, Vol. 78, pp. 9-18.

Navrud, S. (2017), "Possibilities and challenges in transfer and generalization of monetary estimates for environmental and health benefits of regulating chemicals", *OECD Environment Working Papers*, No. 119, OECD Publishing, Paris, https://dx.doi.org/10.1787/95c2b371-en.

Navrud, S. (2018), "Assessing the economic valuation of the benefits of regulating chemicals: Lessons learned from five case studies", *OECD Environment Working Papers*, No. 136, OECD Publishing, Paris, https://dx.doi.org/10.1787/9a061350-en.

Navrud, S. (2019), "Economic Valuation of Environmental Amenities Negatively Affected by Chemical Exposure: Key Insights and Challenges from Environmental Economists and Ecotoxicologists", Unpublished summary report from the joint ECHA/FRAM Workshop: Valuing environmental bads: a dialogue between ecotoxicologists and environmental economists on the valuation of environmental impacts of chemicals. Accessed at https://echa.europa.eu/documents/10162/17241/SETAC_workshop_report2.pdf/19270505-891d-1d84-b650-a847260666f6, 15 December 2021.

Nguyen, T.N., P.M. Jakus, M. Riddel, and W.D. Shaw (2010), "An empirical model of perceived mortality risks for selected US arsenic hot spots," *Risk Analysis*, Vol. 30(10), pp. 1550-1562.

Nguyen, K. T., J.L. Knetsch, and P. Mahasuweerachai (2021), "WTP or WTA: a means of determining the appropriate welfare measure of positive and negative changes when preferences are reference dependent," Environmental and Resource Economics, Vol. 78(4), pp. 615-633.

Olsen, S. B., Meyerhoff, J., Mørkbak, M. R., & Bonnichsen, O. (2017), "The influence of time of day on decision fatigue in online food choice experiments," *British Food Journal*, Vol. 119(3), pp.497-510.

Organisation for Economic Co-operation and Development (OECD) (n.d.), "The costs and benefits of regulating chemicals", https://www.oecd.org/chemicalsafety/costs-benefits-chemicals-regulation.htm, accessed 26 February 2022.

Organisation for Economic Co-operation and Development (OECD) (2016), "Workshop Report: OECD Workshop on Socioeconomic Impact Assessment of Chemical Management", Series on Risk Management, Unclassified Report No. 32, ENV/JM/MONO(2016)68.

Parsons, G.R. and S.M. Thur (2008), "Valuing Changes in the Quality of Coral Reef Ecosystems: A Stated Preference Study of SCUBA Diving in the Bonaire National Marine Park, *Environmental and Resource Economics*, Vol. 40, pp. 593-608.

Petrolia, D.R., D. Guignet, J. Whitehead, C. Kent, C. Caulder, and K. Amon (2021), "Non-market valuation in the Environmental Protection Agency's Regulatory Process", *Applied Economic Perspectives and Policy*, Vol. 43(3), pp. 952-969.

Posthuma, L. et al. (2019), "Species Sensitivity Distributions for Use in Environmental Protection, Assessment, and Management of Aquatic Ecosystems for 12 386 Chemicals," *Environmental Toxicology and Chemistry*, Vol. 38 (4), pp. 905-917.

Rankin, J., and A. Robinson, (2018). Accounting for protest zeros in contingent valuation studies: a review of literature.

Rigdon, E.E., M. Sarstedt, J.M. Becker (2020), "Quantify uncertainty in behavioral research," *Nature Human Behaviour*, Vol. 4 (4), pp. 329-331.

Ringold, P., J. Boyd, D. Landers, and M. Weber (2013), "What Data Should We Collect? A Framework for Identifying Indicators of Ecosystem Contributions to Human Wellbeing", *Frontiers in Ecology and the Environment, Vol.* 11(2), pp. 98–105.

Risk & Policy Analysts (RPA, 2013), "Socioeconomic Valuation of Hazardous Chemicals Control: State of the Art and Future Prospects," report for the Royal Society of Chemistry, the Environment Agency, and the UK Chemicals Stakeholder Forum, October.

Robinson, P. J., P. van Beukering, L. Brander, R. Brouwer, W. Haider, M. Taylor, and P. Mau (2022), "Understanding the determinants of biodiversity non-use values in the context of climate change: Stated preferences for the Hawaiian coral reefs," *Ecosystem Services*, Vol. 53, 101393.

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

Sainz-Santamaria, J., and A.L. Martinez-Cruz (2019), "How far can investment in efficient irrigation technologies reduce aquifer overdraft? Insights from an expert elicitation in Aguascalientes, Mexico," *Water resources and economics*, Vol. 25, pp. 42-55.

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

Scarpa, R., Bazzani, C., Begalli, D., and Capitello, R. (2021), "Resolvable and Near-epistemic Uncertainty in Stated Preference for Olive Oil: An Empirical Exploration," Journal of Agricultural Economics, Vol. 72(2), pp. 335-369.

Scasny, M. and I. Zverinova (2014), "Stated preference study to examine the economic value of benefits of avoiding selected adverse human health outcomes due to exposure to chemicals in the European Union," report prepared for the European Chemicals Agency, available at: https://www.researchgate.net/publication/274045509_Stated-preference_study_to_examine_the_economic_value_of_benefits_of_avoiding_selected_adverse_human_health_outcomes_due_to_exposure_to_chemicals_in_the_European_Union_FD7_Final_Report_Part_II_Fertili.

Schaafsma, M., R. Brouwer, and J. Rose (2012), "Directional heterogeneity in WTP models for environmental valuation," *Ecological Economics*, Vol. 79, pp. 21-31.

Shaw, W.D. (2015), "Environmental and Natural Resource Economics Decisions Under Risk and Uncertainty: A Survey," *International Review of Environmental and Resource Economics*, Vol. 9, pp. 1-130.

Soekhai, V., E.W. de Bekker-Grob, A.R. Ellis, and C.M. Vass (2019), "Discrete choice experiments in health economics: past, present and future," *Pharmacoeconomics*, Vol. 37(2), pp.201-226.

Strand, J., R.T. Carson, S. Navrud, A. Ortiz-Bobea, and J.R. Vincent (2017), "Using the Delphi method to value protection of the Amazon rainforest," *Ecological Economics*, Vol. 131, pp. 475–484.

Soto Montes de Oca, G and I.J. Bateman (2006), "Scope Sensitivity in Households' Willingness to Pay for Maintained and Improved Water Supplies in a Developing World Urban Area: Investigating the Influence of Baseline Supply Quality and Income Distribution Upon Stated Preferences in Mexico City," *Water Resources Research*, Vol. 42 (7), pp. 1-15.

Spiegelhalter, D., M. Pearson, and I. Short (2011), "Visualizing uncertainty about the future," *Science*, Vol. 333(6048), pp. 1393-1400.

Tonin, S., A. Alberini, and M. Turvani (2012), "The Value of Reducing Cancer Risks at Contaminated Sites: Are More Knowledgeable People Willing to Pay More?", *Risk Analysis*, Vol. 32(7), pp. 1157-1182.

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

US EPA (United States Environmental Protection Agency, 2021), "Supporting Statement for Information Collection Request for Survey To Improve Benefit-Cost Analysis of Water Quality Regulations: Instrument, Pre-test, and Implementation," accessed 7 April 2022 at https://www.regulations.gov/document/EPA-HQ-OA-2019-0292-0002.

US EPA (United States Environmental Protection Agency, 2016), "A Practitioner's Guide to the Biological Condition Gradient: A Framework to Describe Incremental Change in Aquatic Ecosystems," EPA 842-R-16-001, Washington, DC, USA. Accessed 15 June 2022 at https://www.epa.gov/sites/default/files/2016-02/documents/bcg-practioners-guide-report.pdf.

US EPA (United States Environmental Protection Agency, 2014), "Guidelines for Preparing Economic Analyses" accessed 7 April 2022 at https://www.epa.gov/sites/production/files/2017-08/documents/ee-0568-50.pdf.

Valente, M., and C. Chaves (2018), "Perceptions and valuation of GM food: A study on the impact and importance of information provision," *Journal of cleaner production*, Vol. 172, 4110-4118.

Van Houtven, G., C. Mansfield, D.J. Phaneuf, R. von Haefen, B. Milstead, M.A. Kenney, and K.H. Reckhow (2014), "Combining expert elicitation and stated preference methods to value ecosystem services from improved lake water quality," *Ecological Economics*, Vol. 99, pp. 40-52.

Viscusi, W.K., and J.E. Aldy (2003), "The value of a statistical life: a critical review of market estimates throughout the world." *Journal of risk and uncertainty*, Vol. 27(1), pp. 5-76.

Viscusi, W.K., J. Huber, and J. Bell (2008), "The economic value of water quality," *Environmental and Resource Economics*, Vol. 41 (2), pp. 169-187.

Viscusi, W.K., W.A. Magat, and J. Huber (1991), "Communication of ambiguous risk information," *Theory and Decision*, vol. 31(2), pp. 159-173.

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

Vossler, C.A., C.L. Dolph, J.C. Finlay, et al. (2022), "Valuing Improvements in the Ecological Integrity of Local and Regional Waters using the Biological Condition Gradient," Draft manuscript under review.

Vossler, C.A., and S.B. Watson (2013), "Understanding the consequences of consequentiality: Testing the validity of stated preferences in the field," *Journal of Economic Behavior & Organization*, Vol. 86, pp.137-147.

von Stackelberg, K., and J. K. Hammitt, (2005), "Use of Stated Preference Methods to Value the Benefits of Ecological Risk Reductions: A Case Study of Exposure to Polychlorinated Biphenyls," Unpublished manuscript, available at: https://www.researchgate.net/publication/242745466.

Wakamatsu, M., K.J. Shin, C. Wilson, and S. Managi (2018), "Exploring a gap between Australia and Japan in the economic valuation of whale conservation," *Ecological Economics*, Vol. 146, pp. 397-407.

Wallsten, T. S., D.V. Budescu, A. Rapoport, R. Zwick, and B. Forsyth (1986), "Measuring the vague meanings of probability terms," *Journal of Experimental Psychology: General*, Vol. 115(4), pp. 348-365.

Walsh, P.J., and W.J. Wheeler (2013), "Water quality indices and benefit-cost analysis," *Journal of Benefit-Cost Analysis*, Vol. 4(1), pp. 81-105.

Walsh, P.J., D. Guignet, and P. Booth (2022), "National Water Quality Values in New Zealand: Policy-Relevant Stated Preference Estimates," *US EPA NCEE Working Papers*, No. 22-02, Washington, DC, USA, https://www.epa.gov/system/files/documents/2022-05/2022-02_1.pdf.

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

Whittington, D., and W. Adamowicz (2011), "The Use of Hypothetical Baselines in Stated Preference Surveys," *Environment for Development Initiative: Discussion Paper Series*, EfD DP 11-11, https://www.jstor.org/stable/resrep14951?seq=1.

Xu, F., et al. (2015), "Key Issues for the Development and Application of the Species Sensitivity Distribution (SSD) Model for Ecological Risk Assessment," *Ecological Indicators*, Vol. 54, pp. 227-237. Yao, R., R. Scarpa, J. Turner, T. Barnard, J. Rose, J. Palma, and D.

Zemo, K., H. Kassahun, and S. Olsen (2019), "Determinants of willingness-to-pay for attributes of power outage-An empirical discrete choice experiment addressing implications for fuel switching in developing countries," *Energy* Vol. 174, pp. 206-215.

Annex A. Illustrative Model Equations

This annex illustrates the theoretical motivation and possible econometric models for each of the valuation questions proposed in chapter 6 of the main report. First presented is an illustrative model for the generalised endpoint approach (Proposal 1), and then for the valuation approach when faced with nearly insurmountable scientific uncertainty (Proposal 2). The primary objectives here are to (i) formally illustrate how willingness to pay (WTP) can be calculated from the responses to a SP survey, and in turn (ii) clarify how benefit transfer of actual policies could be carried out. In the case of Proposal 1, the equations to formally demonstrate (iii) how interactions between environmental attributes could be accounted for, and (iv) how considerations of human health can be purged from the WTP estimates of primary interest are used. When describing Proposal 2, it is also (v) elaborated on how quantified individual-specific subjective probabilities can be linked to qualitative probability statements, and subsequently used when estimating the econometric models and WTP. The equations presented are meant to demonstrate what the potential econometric models may look like, but there are numerous variations that could be considered, and those types of details should be fleshed out in conjunction with the development of a full SP survey study.

A.1 Illustrative Model for Proposal 1: A Generalised Endpoint Approach

Proposal 1 entails a generalised endpoint approach, using a discrete choice experiment (DCE) framework. Details can be found in section 6.1 of the main report. As a starting point, the proposed DCE contains seven attributes, including cost. Following this proposal, it is posited that a household's utility depends on the types of species in the surrounding natural environment (s), the number of species (b), the probability that these species are in "good" condition (p), the duration of any adverse environmental impacts due to chemical contamination (d), human health status (h), and the geographic extent of any adverse impacts on the environmental endpoints due to chemical contamination (g). Let y denote a household's annual income, and $v(\cdot)$ is their indirect utility function.

Following standard economic theory, a household will choose a policy if it yields a higher level of utility than the baseline (no policy) scenario. More formally, a household will choose the policy scenario (*policy=1*) if:

$$v(s_1, b_1, p_1, d_1, h_1, g_1, y_1) \geq v(s_0, b_0, p_0, d_0, h_0, g_0, y_0) \qquad \text{(A1)}$$

where the subscripts denote the attribute levels without (0) and with (1) a policy. The decision rule can be re-expressed as that a household will choose a policy if the change in utility Δv is positive.

$$\Delta v = v(s_1, b_1, p_1, d_1, h_1, g_1, y_1) - v(s_0, b_0, p_0, d_0, h_0, g_0, y_0) \geq 0 \qquad \text{(A2)}$$

Under the Proposal 1 valuation question, the baseline levels for several of the attributes are expressed as a loss relative to current levels (see Figure 4 in the main report). The specified policy levels under Proposal 1 are stated as being the same as current levels –i.e. as if the policy avoided the losses implied by baseline levels. Put simply, under this valuation question the posited policy would avoid losses in the environmental endpoints. For the attribute referring to duration of adverse impacts, the policy scenario implies that $d_1 = 0$; similarly, the geographic extent of adverse impacts implies that $g_1 = 0$. Let Δ denote the specified loss in an attribute relative to current levels. Thus, one can re-express equation (A2) as:

$$\Delta v = v(s_1, b_1, p_1, d_1 = 0, h_1, g_1 = 0, y_0 - c) \\ - v(s_1 - \Delta s, b_1 - \Delta b, p_1 - \Delta p, d_0, h_1 - \Delta h, g_0, y_0) \geq 0 \qquad \text{(A3)}$$

where s_1, b_1, p_1, h_1 and y_0 equal the current attribute levels and income, and c is the increase in a household's annual cost of living under the policy scenario. The values for Δs, Δb, Δp, Δh, d_0, and g_0 are individual-specific attribute levels that would be randomly assigned based on the underlying experimental design of the SP study, as would c.

For purposes of this illustration, a linear-in-parameters functional form for the change in indirect utility function is assumed, which then implies the following equation:

$$\Delta v = \beta_s s_1 + \beta_b b_1 + \beta_p p_1 + \beta_d g_1 + \beta_h h_1 + \beta_g g_1 + \beta_y (y_0 - c) \\ - \{\beta_s (s_1 - \Delta s) + \beta_b (b_1 - \Delta b) + \beta_p (p_1 - \Delta p) + \beta_d d_0 + \beta_h (h_1 - \Delta h) + \beta_g g_0 \\ + \beta_y y_0\} \geq 0$$

$$\Delta v = \beta_s \Delta s + \beta_b \Delta b + \beta_p \Delta p - \beta_d d_0 + \beta_h \Delta h - \beta_g g_0 - \beta_y c \geq 0 \qquad \text{(A4)}$$

Equation (A4) illustrates that the larger the avoided loss in an environmental attribute (e.g., as Δs increases), the greater the gain in utility that would result from choosing that policy. At the same time, the greater the cost of the policy, the lesser the gain in utility that would result, all else constant. Although most attributes are framed as amenities, it is hypothesised that $\beta_d < 0$ and $\beta_g < 0$, which based on equation (A4), would suggest that the gain in utility from choosing a policy increases with i) the geographic extent of the avoided impacts, and ii) the avoided duration of those adverse impacts.

A more flexible functional form can be assumed that allows for interactions across the environmental attributes. Such interaction terms should be introduced in the initial specification of the indirect utility function, but is not done so here for notational ease. What interactions are most important to ultimately include depends on consultations with experts and policy practitioners, feedback from focus groups, and cost and sample size constraints when implementing a survey. But purely for illustrative purposes here, let us suppose that how a loss in species (Δb) affects indirect utility depends on what species are impacted (Δs), and on the likelihood of those impacts occurring (Δp). One can therefore expand on equation (A4) by adding the interaction terms $\Delta b \times \Delta s$ and $\Delta b \times \Delta p$.

$$\Delta v = \beta_s \Delta s + \beta_b \Delta b + \beta_p \Delta p - \beta_d d_0 + \beta_h \Delta h - \beta_g g_0 - \beta_y c \\ + \beta_{bs} (\Delta b \times \Delta s) + \beta_{bp} (\Delta b \times \Delta p) \geq 0 \qquad \text{(A5)}$$

In equations (A4) and (A5), β_y is the marginal utility of income and is a parameter to be estimated. The other β terms represent the marginal utility of the corresponding environmental attributes. Interacted β terms describe how marginal utilities vary depending on variations captured by the corresponding interaction of attributes. For instance, the coefficient β_{bs} captures how the marginal utility for an avoided loss in species Δb differs depending on the trophic order of the species/organisms that are impacted. If people care more about avoiding adverse impacts to higher order species, one may hypothesise that $\beta_{bs} > 0$. Similarly, the coefficient β_{bp} captures how the marginal utility of an avoided loss in species varies depending on how likely a chemical is to yield the stated impacts if its use continues unchecked. It is hypothesised that $\beta_{bp} > 0$, reflecting that the gain in utility from an avoided environmental loss increases as the occurrence of that loss under the baseline becomes more likely.

One can model the probability of a policy being chosen (i.e., *policy=1*) as the probability that $\Delta v \geq 0$. More formally,

$$Pr(policy = 1) = Pr(\Delta v \geq 0) = F\Big(\beta_s \Delta s + \beta_b \Delta b + \beta_p \Delta p + \theta_d d_0 + \beta_h \Delta h + \theta_g g_0 + \gamma c + \beta_{bs}(\Delta b \times \Delta s) + \beta_{bp}(\Delta b \times \Delta s)\Big) \quad \text{(A6)}$$

where γ is the negative of the marginal utility of income $(\gamma = -\beta_y)$, $\theta_d = -\beta_d$, $\theta_g = -\beta_g$, and $F(\cdot)$ is a cumulative distribution function (CDF) following some assumed distribution. For example, assuming $F(\cdot)$ is a normal CDF allows us to estimate equation (A6) as a probit model.

The data gathered from a stated preference survey would be used to estimate a model similar to equation (A6) and derive estimates for the β, θ and γ coefficients. These estimates are denoted as $\hat{\beta}, \hat{\theta}$ and $\hat{\gamma}$. The ratio of the marginal utility with respect to an attribute over the marginal utility of income yields the marginal willingness to pay (MWTP) for that attribute.

More importantly for purposes of benefit transfer, a household's total WTP can be estimated by calculating the change in indirect utility with respect to a specified change in all policy-relevant environmental attributes, divided by the marginal utility of income (Holmes and Adamowicz, 2003). More specifically:

$$WTP = \frac{\hat{\beta}_s \Delta\tilde{s} + \hat{\beta}_b \Delta\tilde{b} + \hat{\beta}_p \Delta\tilde{p} + \theta_d \tilde{d}_0 + \theta_g \tilde{g}_0 + \hat{\beta}_{bs}(\Delta\tilde{b} \times \Delta\tilde{s}) + \hat{\beta}_{bp}(\Delta\tilde{b} \times \Delta\tilde{s})}{-\hat{\gamma}} \quad \text{(A7)}$$

Notice that any human health impacts (Δh) are excluded from the WTP estimates in equation (A7), essentially assuming $\Delta h = 0$ in a subsequent benefit transfer application. The focus is in estimating the benefits of improvements to the natural environment, independent of any human health considerations. Human health impacts are proposed for inclusion in the DCE simply to help control for such confounding considerations by respondents, but are not included in the WTP estimates of interest. In other words, this procedure allows us to "condition out" any human health considerations, thus minimising the associated omitted variable bias when trying to calculate WTP for improvements in the natural environment. The policy and chemical-specific attribute levels $\Delta\tilde{s}$, $\Delta\tilde{b}$, $\Delta\tilde{p}$, $\tilde{d}_0$, and $\tilde{g}_0$ would be plugged into equation (A7) based on scientific models, expert elicitation of ecotoxicologists and related discipline experts, and details from practitioners that are specific to the policy application.

To recap, the data gathered from a SP survey allows economists to statistically estimate the parameters in a model like that shown in equation (A6). Once parameterized, the resulting estimates can be used for benefit transfer – i.e., to predict average household WTP for real-world policies and management decisions of specific chemicals. This would be done by plugging chemical and policy specific values into equation (A7) for the types of organisms or species that are believed to be impacted if that chemical continues to be used ($\Delta\tilde{s}$), how bad those impacts are believed to be ($\Delta\tilde{b}$) (e.g., the predicted potentially disappeared fraction of species), the probability of those impacts occurring ($\Delta\tilde{p}$), the duration of any negative environmental impacts should that chemical continue to be used ($\tilde{d}_0$), and the geographic extent of the avoided impacts ($\tilde{g}_0$). These chemical and policy-specific values are different from the randomly assigned attribute levels given in the survey, hence the "~" notation, and would be based on scientific models, expert elicitation, and insights from policy practitioners. Plugging these values into a calculation like that shown in equation (A7) would yield estimates of, for example, average annual household WTP, which could then be multiplied by the total number of impacted households to estimate the total environmental benefits resulting from a policy.

Of course, more complicated functional forms to account for additional interaction effects and heterogeneity in preferences and income across households could be incorporated. Additionally, consideration should be given to functional forms that allow MWTP to decline as baseline conditions improve (i.e., a downward sloping inverse demand curve). Doing so will allow for more appropriate application of the results when estimating the benefits of incremental policy decisions, and thus facilitate more valid benefit transfers into the future. The main purpose of this simple illustration was to show how estimates obtained from a SP survey, and in particular a valuation question like that under Proposal 1, can provide a coarse, but broadly applicable means for benefit transfer to actual chemical policy and management decisions.

A.2. Illustrative Model for Proposal 2: Valuation Given Significant Uncertainties

A similar model is underlying the framework under Proposal 2, where scientific uncertainty in the potential environmental impacts of a chemical is deemed to be too large to assign quantitative levels to attributes such as those in Proposal 1. Proposal 2 entails a contingent valuation (CV) question that, in the simplest case, only varies costs and the qualitative level of uncertainty across respondents. Details can be found in section 6.2 of the main report. As before, a respondent will only choose a policy if it yields a higher level of utility compared to the status quo:

$$v(p_1, y_1) \geq v(p_0, y_0) \qquad \text{(A8)}$$

In this setting, indirect utility is posited to depend on the probability of a relatively "good" level of environmental quality (p), and numeraire consumption. The "1" and "0" subscripts refer to the policy and baseline scenarios, respectively. The probability of a good quality environment under the baseline scenario –i.e., when a toxic chemical continues to be used and released into the environment– can be expressed as a reduction in probability (Δp) relative to the policy level probability. In other words, $p_0 = p_1 - \Delta p$. Under the baseline, no additional costs are incurred and numeraire consumption is just equal to a household's income y_0. Under the policy scenario, the probability of good quality is higher (i.e., $p_1 > p_1 - \Delta p$), but a household must incur some increase in their cost of living c. Together, these points imply that one can re-express the decision rule in equation (A8) as:

$$v(p_1, y_0 - c) \geq v(p_1 - \Delta p, y_0) \qquad \text{(A9)}$$

As the CV question posed in Section 6.2 is currently framed, the reduction in the probability of a good quality environment is specified in qualitative terms –i.e. how likely it is that a chemical will adversely impact the environment. In Figure 5, it is stated that a chemical will "likely affect" microorganisms, plants, and wildlife. Although randomly assigned qualitative probability statements may be used in a valuation question, responses to questions preceding the CV scenario can be used to link those qualitative statements to individual-specific quantitative probabilities (see Section 6.2). Therefore, Δp can be thought of as an individual-specific quantitative probability that is plugged into the empirical model of respondents' stated decisions. Assuming a linear-in-parameters specification allows us to re-express equation (A9) as:

$$\Delta v = \beta_p p_1 + \beta_y (y_0 - c) - \{\beta_p (p_1 - \Delta p) + \beta_y y_0\} \geq 0$$

$$\Delta v = \beta_p \Delta p - \beta_y c \geq 0 \qquad \text{(A10)}$$

As β_y reflects the marginal utility of income and economic theory indicates that $\beta_y > 0$, equation (A10) suggests that as the cost of a policy increases, a household's gain in utility from choosing the policy will decrease. Furthermore, it is hypothesised that respondents' utility increases with the probability of a good quality (i.e., $\beta_p > 0$), then as the probability of the avoided adverse impacts increases, an individual's gain in utility from choosing the policy will also increase.

Similar to before, one can model the probability of a policy being chosen (i.e., *policy=1*) as the probability that $\Delta v \geq 0$. More formally,

$$Pr(policy = 1) = Pr(\Delta v \geq 0) = F(\beta_p \Delta p + \gamma c) \qquad \text{(A11)}$$

where γ is the negative of the marginal utility of income $(\gamma = -\beta_y)$, and $F(\cdot)$ is a CDF following some assumed distribution. For example, assuming $F(\cdot)$ is a normal CDF allows us to estimate equation (A11) as a probit model.

The data gathered from a SP survey with a valuation question like that posed in Proposal 2 would be used to estimate a model similar to equation (A11) and, more specifically, estimate the β_p and γ parameters. These parameter estimates are denoted as $\hat{\beta}_p$ and $\hat{\gamma}$. With the estimated parameters, a household's marginal WTP (*MWTP*) for a marginal change in the probability of an avoided adverse impact can be calculated as $MWTP_p = \frac{\hat{\beta}_p}{-\hat{\gamma}}$.

More importantly for purposes of benefit transfer, a household's total WTP to avoid uncertain environmental impacts that would have occurred with some probability can be calculated in this simplest case as:

$$WTP = \frac{\hat{\beta}_p \Delta\tilde{p}}{-\hat{\gamma}} \qquad \text{(A12)}$$

where $\Delta\tilde{p}$ is a policy- and chemical-specific projected probability that the chemical(s) in question would adversely impact the environment. In practice, given the current gaps in scientific evidence for many groups of chemicals, such projected probabilities would be based on formal expert elicitation, as described in chapter 7 of the main report. Estimates derived from an equation like (A12) reflect a household's WTP for adopting a precautionary principle, and avoiding uncertain environmental impacts. Households' WTP would be calibrated based on their perceived probabilities of that underlying uncertainty (Δp), but the change in risk used for the actual estimation of benefits ($\Delta\tilde{p}$) is calculated based on natural scientists' collective determination of such probabilities. $\Delta\tilde{p}$ would be estimated through technically sound expert elicitation protocols. In other words, a benefit transfer exercise would be based on changes in perceived probabilities as determined by scientists, and not based on the general public's perceptions.

Again, more complicated variations of these models could be carried out to incorporate preference heterogeneity, split-sample designs allowing for variation in other relevant dimensions (e.g., geographic extent), etc. The simplest case is presented here just to demonstrate how survey data gathered from a valuation question like that under Proposal 2 can be used to parameterize a WTP function for purposes of benefit transfer to future chemical management and policy decisions.