Discrete and Portfolio Choice Experiments

Methodological Considerations and Health Policy Applications

Sander Boxebeld

Discrete and Portfolio Choice Experiments

Methodological Considerations and Health Policy Applications

Sander Boxebeld

Printing: Ridderprint, ridderprint.nl
Layout and design: Erwin Timmerman, persoonlijkproefschrift.nl

Discrete and Portfolio Choice Experiments: Methodological Considerations and Health Policy Applications

Discrete en portfoliokeuze-experimenten: methodologische overwegingen en toepassingen in gezondheidsbeleid

Thesis

to obtain the degree of Doctor from
Erasmus University Rotterdam
by command of the
rector magnificus

Prof.dr.ir. A.J. Schuit

and in accordance with the decision of the Doctorate Board.
The public defence shall be held on

Friday 14 November 2025 at 13.00 hrs

by

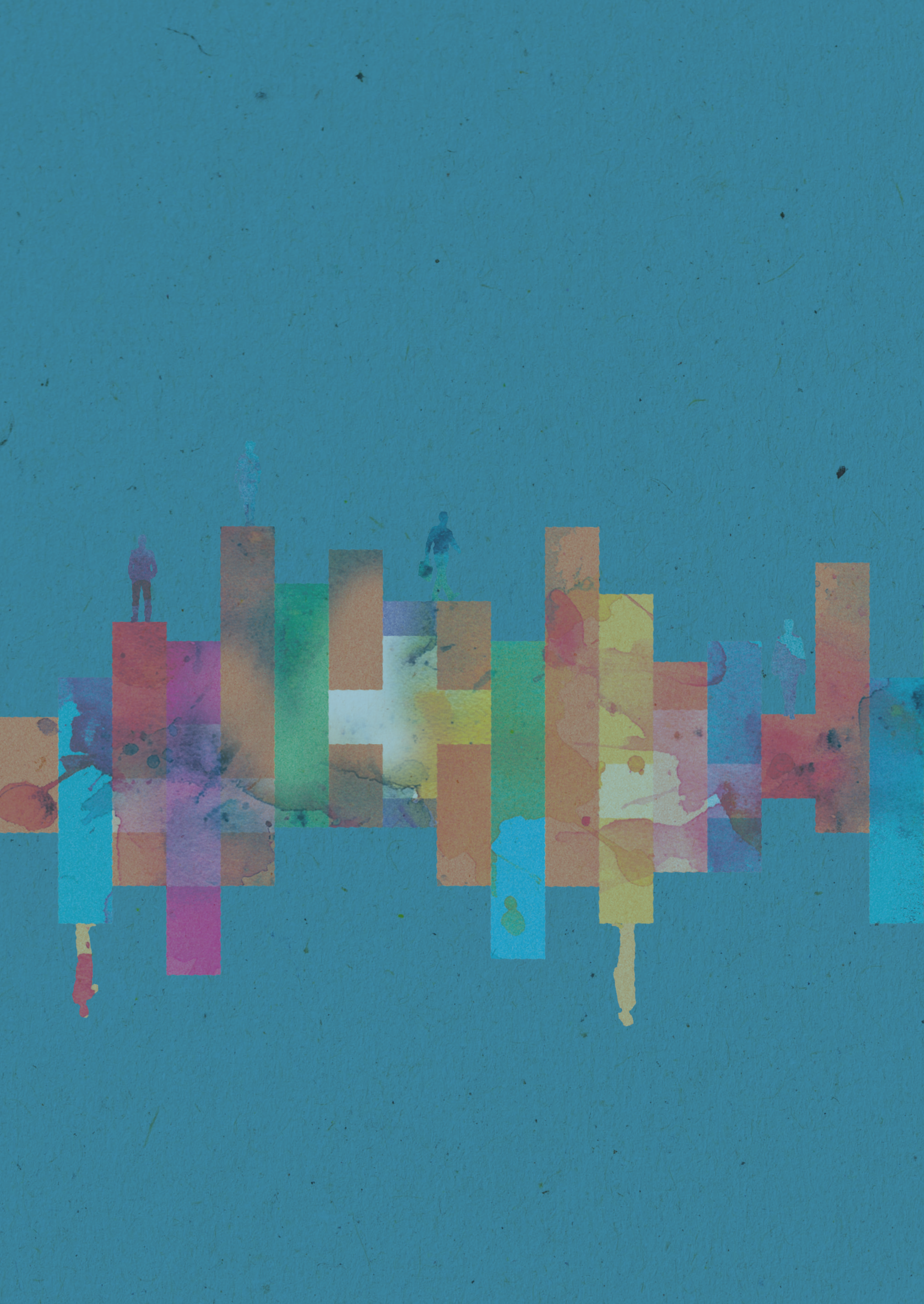
Sander Boxebeld
born in Nijmegen, the Netherlands.

Erasmus University Rotterdam

Doctoral Committee:

Promotors:	Prof.dr. N.J.A. van Exel
	Dr.mr. N.M. Mouter
Other members:	Prof.dr. E.W. de Bekker - Grob
	Prof.dr. J.L.W. van Kippersluis
	Prof.dr. V. Watson

Table of Content



Chapter 1

General introduction

While it may be an uncomfortable truth, one cannot deny that we are often constrained in the fulfilment of our goals and the achievement of our desired outcomes due to scarcity of resources, most importantly, time and money. This is probably relatable on the level of individual behaviour. For example, I am sure that I am not the only one who feels like always ending up compromising on physical exercise, social activities, or sleep, because 24 hours per day is just not enough to do all the things I need and would like to do. The same principle holds for collective resources: governments are constrained in their policy decisions by limits to our collective resources, whether financial, human or natural resources. This scarcity of resources prompts us, individually and as society, to consider the opportunity costs of our resource allocation decisions. After all, resources allocated to a specific purpose cannot be used for another purpose. For example, the time you now spend reading this dissertation could have been spent on something else. Similarly, the monetary budget spent by governments on healthcare cannot be spent on education, infrastructure, or any other purpose.

Valuing resource allocations

To maximize welfare, it is thus important to obtain insights into the relative value of alternative resource allocations. In their valuation of alternative resource allocations, economists typically prefer to rely on revealed preferences. That is, preferences and values for resource allocations are preferably derived from individuals' real-life behaviour (e.g., Mendelsohn, 2019). For example, you reading this dissertation is assumed to reveal your preference for doing so relative to any other activity you could have done in the same period. Likewise, the fact that you pay a certain amount of money for your coffee-to-go is assumed to reveal the value you attach to this coffee relative to keeping the money in your bank account (to spend on anything else). Along the same lines, we can observe many real-life behaviours to derive values for specific resource allocations.

However, in some situations, we cannot observe real-life behaviours. This typically holds for newly developed goods and services that are not yet available on the market, as well as for goods and services that are not traded on markets at all. In the health domain, the domain of focus in this dissertation, many examples of both are available. For example, we cannot observe transactions for drugs or technologies that still have to be developed.

In these cases, researchers typically resort to the use of stated preferences. Stated preferences cannot be derived from behaviour and, therefore, need to be elicited; in a survey, researchers present one or more questions, to which respondents' answers contain information on their preferences. Thus, stated preferences are inferred from individuals' answers to survey questions, while revealed preferences are inferred from their behaviours (Carson & Louviere, 2011).

Preference-elicitation methods

A variety of methods has been developed and applied to elicit preferences and valuations in the health domain (e.g., Ryan et al., 2001; Soekhai et al., 2019b). Originally, the focus within non-market valuation, including in health, has predominantly been on opinion surveys and contingent valuation (CV). In CV, respondents are asked for their valuation of a clearly defined good or service, typically in the form of willingness to pay (WTP) (e.g., Carson & Hanemann, 2005; Carson & Czajkowski, 2014). In recent decades, discrete choice experiments (DCEs) have become more popular in use (e.g., Haghani et al., 2021c; Mahieu et al., 2017; Soekhai et al., 2019a). DCEs present respondents with a sequence of choice tasks, each composed of two or more alternatives.[1] The alternatives consist of a number of attributes, capturing the characteristics and (expected) outcomes of the alternatives on offer, of which the levels are experimentally varied between alternatives and choice sets.

The method of DCE, sometimes also referred to as stated choice experiment or choice-based conjoint analysis, has a strong foundation in choice behaviour theory. The comparative nature of the DCE choice tasks is grounded in Thurstone's 'Law of Comparative Judgment' (Thurstone, 1927), which introduced the idea of obtaining scale values of preference for stimuli based on pairwise comparisons. The multi-attribute nature of DCEs has its foundations in Lancaster's theory that the utility derived from a good is the sum of the utilities derived from the characteristics of that good (Lancaster, 1966). Finally, the modelling of respondents' choices in DCEs dates to McFadden (1974), who combined previous insights into a tractable econometric choice model, embedded in Random Utility Theory (RUT). Under RUT, the utility derived from

1 Given the variety of elicitation formats in both CV and DCE, there is confusion in the literature about the boundaries between CV and DCE (e.g., Carson & Czajkowski, 2014; Carson & Louviere, 2011). Generally, DCE separates itself from CV with its multi-attribute nature, whereas the focus in CV is on the cost attribute and maximally one other specific attribute of interest (not more than two attributes). Also, while in DCE respondents are offered two or more alternatives (potentially plus an opt-out or status quo alternative) to choose from, the focus in CV is on a single alternative.

choosing an alternative can be divided into a deterministic component, consisting of the utilities derived from the attribute levels of that alternative, and a stochastic component (i.e., an error term) (e.g., Manski, 1977; McFadden, 2001). McFadden's model, known as the conditional logit or multinomial logit model, is still the cornerstone for choice modelling applications (e.g., Hensher et al., 2015; Mariel et al., 2021). In Figure 1.1, a stylized example of a DCE choice task is presented. This typical example contains two (unlabelled) alternatives, each described by a combination of three attribute levels. Under RUT and a utility-maximizing decision rule, respondents are expected to choose the alternative resulting in the largest utility to them. By experimentally varying the attribute levels of alternatives between choice tasks (and respondents), the role of the attribute level changes in respondents' choices can be studied and their utility values can be derived. These values are often presented in the form of (marginal) welfare measures (see Textbox 1.1).

Textbox 1.1: (Marginal) welfare measures from choice experiments

The estimated coefficients from a choice model (in preference space) do not have an absolute (meaningful) interpretation in and of themselves, as utility does not have a scale. Therefore, analysts typically derive a marginal rate of substitution (MRS) between attributes. The MRS indicates the rate at which an individual is willing to trade one attribute for another, as to keep utility constant (Dekker, 2014). As an example, it may denote the money an individual is willing to forego for obtaining a given improvement (i.e., a desirable level change) in another attribute. Such a MRS with a monetary denominator is often referred to as marginal willingness to pay (mWTP) (e.g., Mariel et al., 2021). While other types of MRS (i.e., with non-monetary denominators), such as maximum acceptable risk (MAR) (e.g., Veldwijk et al., 2023) and marginal willingness-to-wait (mWTW) (e.g., Genie et al., 2020), are also used in the literature, mWTP is the most popular one.

In addition to the derivation of mWTP estimates, non-market valuation researchers may also be interested in the welfare implications of a policy intervention and, therefore, in calculating an aggregate welfare measure (i.e., capturing the welfare implications of a specified combination of attribute level changes, rather than the marginal change in a single attribute) (Mariel et al., 2021). Typically, such an aggregate welfare measure takes the form of the 'compensating variation', indicating the amount of money that would need to be given to or taken from an individual after a change of the status quo (e.g., the implementation of a policy intervention) to leave them at the their initial level of utility (e.g., Lancsar & Savage, 2004). This is the appropriate welfare measure in case the payment vehicle in the choice experiment takes the form of a (hypothetical) payment due by the individual (e.g., a change in taxes, premiums, user fees). If the payment vehicle in the choice experiment is the reallocation of existing public resources, the appropriate welfare measure may be the 'compensating tax reallocation' (Bergstrom et al., 2004). This indicates the amount of money that would need to be reallocated away from other governmental spending areas to fund the policy intervention under consideration, keeping an individual's disposable income constant.

The (marginal) welfare measures derived from choice models can be incorporated into economic evaluation frameworks, such as cost-benefit analysis (e.g., McIntosh, 2006). Besides, given that welfare analysis may not always be necessary (Dekker et al., 2024) or most informative to policymakers (Chandoevwit & Wasi, 2020), these measures may also enable a meaningful interpretation of choice model estimates and the derivation of a ranking of relative importance (Mariel et al., 2021).

Figure 1.1: Stylized example of a choice task in a Discrete Choice Experiment (DCE)

Please select your most preferred alternative		
Attribute	Alternative A	Alternative B
Attribute 1	Level	Level
Attribute 2	Level	Level
Attribute 3	Level	Level
	☑	☐

DCEs have been widely used in the health domain. Recent systematic reviews show DCEs have been applied, for instance, to study patients' and physicians' choice of treatment, individuals' vaccine uptake decisions, and medical career choices (Clark et al., 2014; De Bekker-Grob et al., 2012; Soekhai et al., 2019). For many of such choices, the single discreteness of the choice task in a DCE resembles the real-life choice environment. Even though to a lesser extent than previously mentioned fields of application, DCEs are also used to inform resource allocations and public policy decisions, such as the implementation of health-promoting policies (e.g., Dieteren et al., 2023; Lancsar et al., 2022; Pechey et al., 2014), the adoption of vaccines in the national immunization program (Luyten et al., 2022), or investments in the healthcare system (e.g., Erdem & Thompson, 2014). In the context of policy decisions, however, DCEs may resemble the natural choice environment of the decision-maker to a lesser extent. For instance, policymakers often adopt and implement several policy measures simultaneously to address a specific policy issue. In such an instance, policymakers make multiple discrete choices (i.e., whether to adopt a policy measure or not) rather than a single discrete choice. At the same time, policymakers typically decide on the proportion of the budget that is allocated to a policy issue. This is a continuous choice element, which is also not easily captured within the DCE choice task. As a result, respondents cannot express their preference for the amount of resources allocated to the policy issue. These potential disparities may increase the extent of hypothetical bias and limit the relevance of the method for eliciting preferences for public resource allocation and policy decisions.

In recent years, several preference-elicitation methods have been developed that allow respondents to choose combinations of alternatives. For example, the menu-based choice experiment expands the common pairwise-comparison DCE by allowing respondents to choose both alternatives in each choice task (Huynh et al., 2024). In other expansions of the DCE approach, such as the Basked-Based Choice Experiment (BBCE), Basket-and-Expenditure-Based Choice Experiment (BEBCE), and Volumetric Choice Experiment (VCE), respondents can indicate a quantity of preference for each of the alternatives in a choice task (e.g., Caputo & Lusk, 2022; Carson et al., 2022; Neill & Lahne, 2022; Pellegrini et al., 2022). Typically, these methods are framed in the context of private consumption rather than public policy decisions and do not incorporate a resource constraint, which does not acknowledge the practice of scarcity of resources. As an alternative method, Participatory Value Evaluation (PVE) has been developed for public policy questions and can be characterized as a constrained portfolio choice experiment[2]: in a single choice task, respondents are presented with a set of policy alternatives addressing a particular policy problem and are asked to compose a portfolio of their preferred policy alternatives, subject to a resource constraint. Each of the policy alternatives is described by a number of attributes, of which the levels are experimentally varied between respondents. PVE has been introduced in transportation and environmental economics (Mouter et al., 2021a; 2021b) and may also be valuable for use in the context of health policy questions, which will be further explored in this dissertation.

Valuation perspective

Different perspectives may be taken in the elicitation of preferences for public policy alternatives and public resource allocations, most commonly a consumer or a citizen perspective (also referred to as personal or socially inclusive personal perspective (e.g., Dolan et al., 2003)). Under the consumer perspective, respondents in a choice experiment are asked which alternative(s) they would prefer (for) themselves. Under the citizen perspective, respondents are asked which alternative(s) they would prefer for society, of which they themselves are a part, or which alternative(s) they would recommend the government to adopt. Several studies have documented differences in study results when framing the elicitation task in either of these perspectives (e.g., Alphonce et al., 2014; Mouter et al., 2017; Ovaskainen & Kniivilä, 2005; Özdemir et al., 2023; Russell et al., 2003). For instance, one study found respondents were more averse to vaccine side-effects when asked to choose which vaccine should be approved by the government than when asked to choose which vaccine they would take themselves (Özdemir et al., 2023). Another study found that respondents were willing to pay more for food safety improvements in restaurants when asked for their vote on food safety regulations than when asked for their consumption choices (Alphonce et al., 2014). Nyborg (2000) argues that individuals make use of personal well-being functions and thereby maximize their own well-being in their role as 'Homo Economicus' (i.e., considering themselves mostly as 'consumer'), while they make use of subjective social welfare functions and thereby (aim to) maximize societal welfare in their role as 'Homo Politicus' (i.e., considering themselves mostly as 'citizen'). This divergence between personal and social preferences may be due to a variety of cognitive and normative

2 For PVE, the derivation of (marginal) welfare measures, as described in Textbox 1, is conceptually similar to other preference-elicitation methods.

reasons and is referred to as the 'consumer-citizen duality' (e.g., Alphonce et al., 2014; Mouter et al., 2018). Consequently, the valuation perspective to be taken when eliciting preferences and values should be carefully considered (e.g., Costa-Font & Rovira, 2005; Dolan et al., 2003; Mouter et al., 2017). In health economics, choice experiments have been used to a limited extent to elicit preferences for public policy alternatives and public resource allocations, compared with their application in private choice settings. Also, in the limited instances of choice experiments in a public choice setting, the citizen perspective has been rarely applied. This dissertation will make use of this perspective in its elicitation of public preferences to inform health policy decisions.

Dissertation objectives and structure

The aim of this dissertation is to advance the literature on the elicitation of public preferences for health policies by addressing three objectives.

The first objective is to position a relatively new multi-attribute preference elicitation method, Participatory Value Evaluation (PVE), relative to other more commonly used methods in the health domain, like discrete choice experiments (DCEs). This will contribute to researchers' and policymakers' understanding of the potential advantages and disadvantages of PVE for eliciting public preferences.

The second objective is to examine the endogeneity of preferences elicited in DCEs and PVEs to design characteristics of the choice experiment. This will contribute to insight on the internal validity of DCEs and PVEs. Since the preferences elicited using DCE and PVE may be used to inform public policy decisions, it is important that the validity of these methods is scrutinized extensively and regularly.

The third objective is to explore public preferences for health policy alternatives from a citizen perspective using DCE and PVE. This will contribute to the limited literature on choice experiments adopting a citizen perspective and provide information for policy decisions in two specific and very relevant health policy areas, namely long-term care and prevention of onset and progression of skin cancer.

The structure of the dissertation is as follows:

Chapter 2 introduces PVE in the health domain by discussing the method and its previous applications. Since PVE is relatively new, it is rather unknown. By illustrating how the method can be adapted to the policy question at hand and positioning PVE conceptually relative to a few more established multi-attribute preference-elicitation methods in the health domain, this chapter aims to contribute to a well-informed selection of methods for preference elicitation and a research agenda for further development of PVE.

Chapter 3 reviews the literature on the impact of the presentation order of alternatives, attributes and choice sets on respondents' choices in a DCE. Ordering effects form an issue that is not always considered in the design of DCE applications. At the same time, methodological insights on this topic may not diffuse across the different domains of application. Therefore, this cross-domain review provides recommendations to mitigate ordering effects in future applications and suggests directions for further research on ordering effects in DCEs.

Chapter 4 makes use of the DCE method to elicit citizens' preferences for policy measures to prevent the onset and progression of skin cancer, in three European countries: Austria, the Netherlands, and Spain. Previous preference studies elicited consumer preferences with respect to individual prevention measures, while little is known yet about public preferences for collective skin cancer prevention policies. In a sequence of twelve choice tasks, each consisting of two policy packages, respondents were asked to select their most preferred policy package.

Chapter 5 makes use of the PVE method to elicit citizens' preferences for policy action regarding long-term care (LTC) for older people in the Netherlands in 2040. Most studies, thus far, have focused on eliciting individuals' preferences regarding their current or future care recipience or caregiving situation. At the same time, the few studies approaching the topic from a citizen perspective took an attitudinal approach, which arguably does not capture the trade-offs that policymakers are facing. This study, instead, took a preference-based approach to elicit preferences. In a constrained portfolio choice experiment, respondents were asked to compose a portfolio of their most preferred policy measures for LTC in 2040, subject to a resource constraint.

Chapter 6 examines whether expenditure preferences and consequentiality perceptions of respondents in a PVE are sensitive to the payment vehicle and the priming of opportunity costs. Many stated preference studies adopting a citizen perspective include a tax increase as the payment vehicle to fund newly provided public goods or services, while reallocation of existing resources may be more realistic. Also, previous research found respondents to neglect opportunity costs in hypothetical choice situations, which may be reduced by priming these opportunity costs. Using three versions of a PVE survey, differing in their payment vehicle and inclusion of an opportunity cost priming question, this study tests the impact of both design variations on respondents' preferences and perceptions of the study's consequentiality.

Together, Chapters 2 addresses the first objective of this dissertation, Chapters 3 and 6 the second objective, and Chapters 4 and 5 the third objective. While all chapters are of scientific relevance, Chapters 4 and 5 also have clear societal relevance by using state-of-the-art applications of DCE and PVE to elicit preferences for societally relevant policy issues and provide policymakers with directions for publicly supported policy action. This may increase the legitimacy of public policy processes and help policymakers to generate public support for their policy decisions (e.g., Bryson et al., 2013; Nabatchi, 2012; Yang, 2016). The structure of the dissertation highlights the connections between the chapters. First, Chapter 2 compares the two methods of focus in this dissertation, DCE and PVE, conceptually with one another and with other preference-elicitation methods. This informed the selection of preference-elicitation methods for the policy questions addressed in Chapters 4 and 5.[3] After Chapter 2, the dissertation focuses on DCE first, in Chapters 3 and 4, and afterwards on PVE, in Chapters 5 and 6. Each of the two components contains a combination of interconnected chapters, one with a methodological focus and one with an applied, policy-relevant focus. For the DCE component, the findings of the systematic literature review on ordering effects in Chapter 3 and its resulting recommendations for future DCEs have been incorporated in the design and analysis of the DCE application on skin cancer prevention in Chapter 4, to minimize the potential bias resulting from ordering effects in this study. For the PVE component, the application on long-term care in Chapter 5 makes use of the payment vehicle most common to previous applications of PVE (and other preference-elicitation methods applied to questions in public resource allocations): a tax increase. To scrutinize the impact of the payment vehicle and of opportunity cost priming on

3 Practical considerations played a role in the selection of preference-elicitation methods, too. For example, Chapter 4 was originally designed as a PVE. However, the design of this PVE initially came which mutually exclusive policy alternatives (i.e., alternatives that could not be chosen together in a portfolio), which the software platform did not accommodate at the time. Therefore, DCE was used as the preference-elicitation method, instead.

respondents' expenditure preferences and consequentiality perceptions, Chapter 6 compares the version of the PVE application presented in Chapter 5 with two other versions, which included either an alternative payment vehicle or an opportunity cost priming question.

Finally, in **Chapter 7**, the results of the different studies are summarized and discussed. This chapter puts the findings in a broader perspective, addresses main strengths and limitations of the conducted research, and formulates implications for policy and directions for further research.

Chapter 2

Participatory Value Evaluation (PVE): A new preference-elicitation method for decision-making in healthcare

Based on:

Boxebeld, S., Mouter, N. and Van Exel, J. (2024). Participatory Value Evaluation (PVE): A New Preference-Elicitation Method for Decision Making in Healthcare. *Applied Health Economics and Health Policy*, *22*, 145 – 154

Abstract

Participatory Value Evaluation (PVE) has recently been introduced in the field of health as a new method to elicit stated preferences for public policies. PVE is a method in which respondents in a choice experiment are presented with various policy options and their attributes, and are asked to compose their portfolio of preference given a public-resource constraint. This paper aims to illustrate PVE's potential for informing healthcare decision-making and to position it relative to established preference-elicitation methods. We first describe PVE and its theoretical background. Next, by means of a narrative review of the eight existing PVE applications within and outside the health domain, we illustrate the different implementations of the main features of the method. We then compare PVE to several established preference-elicitation methods in terms of the structure and nature of the choice tasks presented to respondents. The portfolio-based choice task in a PVE requires respondents to consider a set of policy alternatives in relation to each other and to make trade-offs subject to one or more constraints, which more closely resembles decision-making by policymakers. When using a flexible budget constraint, respondents can trade-off their private income with public expenditures. Relative to other methods, PVE may be cognitively more demanding and is less efficient, however, it seems a promising complementary method for the preference-based assessment of health policies. Further research into the feasibility and validity of the method is required before researchers and policymakers can fully appreciate the advantages and disadvantages of PVE as a preference-elicitation method.

Introduction

Over the past decades, the use of preference-elicitation methods such as Discrete Choice Experiments (DCE) and Best-Worst Scaling (BWS) has rapidly expanded, including in the health field (Haghani et al., 2021a). One of the main purposes of employing such methods is the preference-based assessment of health-policy alternatives to inform governmental decision-making in the authorization of new pharmaceuticals and the public funding of treatments (Marsh et al., 2020; Van Til & IJzerman, 2014; Whichello et al., 2020a; Whitty et al., 2014a). In this way, governmental decisions may be better aligned with public preferences and decision-makers are provided with additional perspectives from citizens (Van Til & IJzerman, 2014).

In addition to the methods commonly used for this purpose, Participatory Value Evaluation (PVE) has been introduced in the fields of transportation (Mouter et al., 2021b) and environmental sciences (Mouter et al., 2021c; 2021d). PVE is a method in which respondents in a choice experiment are presented with various policy projects and their characteristics and effects, and are asked to compose their preference portfolio given a public-resource constraint (Dekker et al., 2019). Respondents seem to find PVE a relevant, credible and legitimate method (Juschten & Omann, 2023) that increased their awareness about the policy issue in question and may be valuable for policymakers (Juschten & Omann, 2023; Mouter et al., 2021a; 2021d; 2022; Mulderij et al., 2021; Rotteveel et al., 2022).

Given the use of PVE for incorporating public preferences in resource-allocation decisions, one may compare the method to a variety of participatory and deliberative methods, such as Participatory Budgeting, referendums and opinion polls. Mouter et al. (2021a) have provided a conceptual comparison of PVE with such methods. PVE has also been compared conceptually with Willingness to Assign (WTAS)/Willingness to Allocate Public Budget (WTAPB) Experiments (Mouter, 2021; Mouter et al., 2021c), in which respondents allocate a public budget for several collective goods or services (Costa-Font & Rovira, 2005) without any connection between public and private resource capacities. Finally, PVE has been compared both conceptually and empirically with the economic-evaluation framework of Cost-Benefit Analysis (CBA) (Mouter et al., 2021b). PVE has not yet been compared with other multi-attribute preference-elicitation methods. Such a comparison is straightforward as, from the modelling perspective, PVE essentially forms an extension of existing choice-modelling approaches (Dekker et al., 2019). A comparison also provides a better understanding of PVE compared to established preference-elicitation methods.

Now that PVE has been applied in the context of health (Mouter et al., 2021a; 2022; Mulderij et al., 2021; Rotteveel et al., 2022), it seems appropriate to discuss the method more specifically and in relation to established methods for eliciting health preferences. To do so, this paper will first introduce PVE in more detail and discuss its theoretical background. Next, the main features of published applications will be discussed. This is not a systematic literature review, as only eight PVE-applications have been published so far, but illustrates how the PVE-design can be adapted to the policy question at hand. Finally, PVE is positioned relative to established preference-elicitation methods, with the aim of helping researchers and policymakers understand the comparative (dis) advantages of PVE and contributing to a better-informed selection of methods for preference-based assessments of health policy alternatives in future.

Participatory Value Evaluation: The method and its theoretical background

Policy setting

Policymakers are typically faced with multiple-decision problems when allocating scarce resources, such as a public budget. Not only do they need to decide on the amount of the budget to spend on a particular purpose, but also on the budget allocation to specific goods or services, and how much to spend on each good or service. These decisions take the form of both discrete choices (i.e., whether to allocate resources towards a specific good or service) and continuous choices (i.e., the amount of the budget spent in total and on each selected good or service). PVE has been developed as a method to elicit citizens' preferences towards each of these decision problems simultaneously.

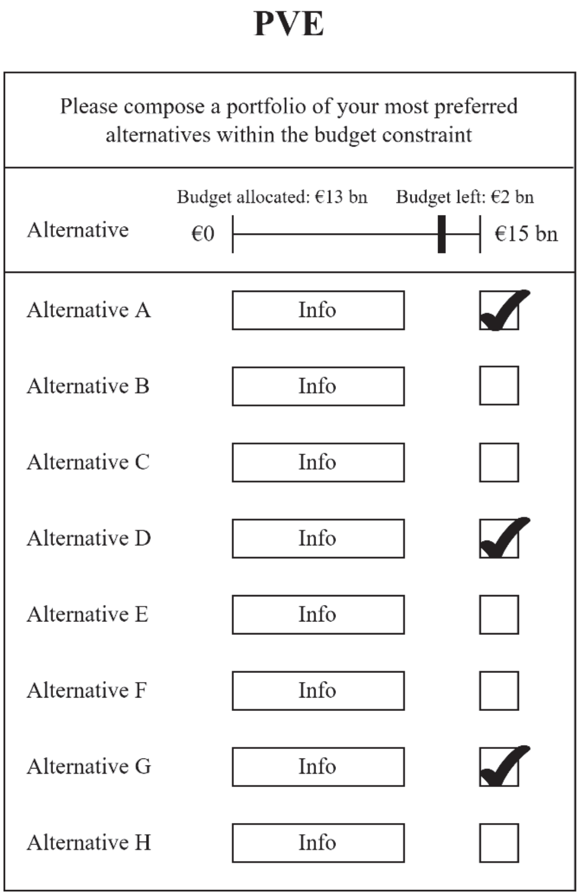
Choice task

PVE assesses the desirability of different policy options and their attributes by means of a choice experiment. Respondents are presented with a specific policy problem faced by a policymaker, a set of policy alternatives that address this problem and a (set of) constraint(s).[1] See Figure 2.1 for a stylized example of a PVE choice task. Each policy alternative is described by a set of attributes, specifying its estimated impact on several relevant outcomes. Respondents are asked to select a portfolio of policies according

1 As an illustration of the PVE development process, Juschten & Omann (2023) suggest seven development steps and describe the methods they used for knowledge creation within each step.

to their preferences by comparing and trading-off the attribute levels of the policy alternatives on offer, respecting the specified constraint(s). These constraints can, for example, take the form of a maximum budget and/or a target level on a relevant outcome (e.g., a minimum increase in a desired outcome or a minimum decrease in an undesired outcome). PVE thus combines a portfolio-based choice task with the allocation of public resources, all assembled within a single framework embedded in Random Utility Theory (RUT)(Dekker et al., 2019; Mouter et al., 2021a; 2021c).

Figure 2.1. Stylized example of the choice task of Participatory Value Evaluation (PVE)



An interesting feature of a PVE is that the budget constraint can be either fixed or flexible. In case of a fixed budget, respondents can only select policies within the given budget constraint. In case of a flexible budget, respondents may decide to raise or lower the budget, but then also need to accept that taxes (or premiums and tariffs) issued to finance the policy will change upwards or downwards accordingly. Thus, in a flexible-budget PVE, respondents do not only select a portfolio from a set of policy alternatives but simultaneously also trade-off public and private spending capacities.

Experimental design

While the set of policy alternatives is constant, the levels of the attributes are randomized across respondents so that the effect of these levels on respondents' choices can be estimated (Dekker et al., 2019). Ideally, the experimental design should include all combinations of attribute levels, as the PVE then captures respondents' trade-offs between all possible combinations (i.e., a full-factorial design). However, such a design is typically unfeasible for the analyst to construct in practice due to the exponential growth in the number of possible combinations (i.e., profiles) when increasing the number of alternatives, attributes or levels. Therefore, a 'min-max correlation' design can be constructed using an algorithm, in which the correlation between attribute levels is minimized within a reasonable number of profiles. This algorithm is explained in the Appendix of the article by Mouter et al. (2021a).

Data analysis and outcomes

Under RUT, the utility of each choice alternative can be divided into a deterministic component (i.e., the aggregate of the utilities attached to its attribute levels and, if applicable, its label) and a stochastic component captured in the error term of the utility function (e.g., Baltas & Doyle, 2001). In a PVE framework, an individual's utility is affected by both the utility of the choice alternatives as well as the utility of private consumption and any remaining (non-allocated) public budget. The PVE choice model can be econometrically estimated using the Multiple Discrete-Continuous Extreme Value (MDCEV) model, which is an established choice model for the estimation of both discrete and continuous choices (Bhat, 2008). Dekker et al. (2019) have proposed extensions to the MDCEV-model for the analysis of PVE data.[2] An alternative choice model that can be used for PVE is the portfolio choice model (Bahamonde-Birke & Mouter, 2024), which is more useful in the absence of a resource constraint.

2 These extensions include the non-linear utility impact that the two outside goods may have and the connection of public and private spending capacities through the tax system. The rationale for and formalization of these extensions are described by Dekker et al. (2024).

Dekker et al. (2019) show that PVE's embeddedness in RUT makes it possible to estimate and aggregate individual utility functions and implement these into the social-welfare function, yielding welfare estimates that can be used as inputs into economic-evaluation frameworks like Cost-Utility Analysis (CUA) or Cost-Benefit Analysis (CBA). As such, the link between public and private budget constraints through the tax system allows one to align PVE with the Kaldor Hicks welfare-economics framework (Dekker et al., 2019; Mouter et al., 2021c) to evaluate the (re)allocation of scarce resources. Thereby, it becomes possible to derive willingness-to-pay estimates from PVE data. However, most existing PVE applications estimate direct utility functions in preference space (Dekker et al., 2019), since there is often no need for a monetary valuation, as the PVE is already framed in the context of the application and the results can therefore directly inform policymakers (Dekker et al., 2024). Thus, the analysis of PVE data typically yields preference parameters that capture the marginal utility that respondents attach to a policy alternative or one level-increase of an attribute. These preference parameters can be used to calculate the marginal rate of substitution (MRS) between attributes, the probability that a portfolio of policy alternatives results in an improvement of social welfare, and the optimal composition of a portfolio given a specific constraint (Dekker et al., 2019; Mouter et al., 2021c; Mulderij et al., 2021).

The main features of published applications

To provide a better understanding of how PVE can be used, the main features of published studies applying PVE published up until March 2023 (either in peer-reviewed journals or in online working paper repositories) are discussed below. Given the distinct design features of these applications, the discussion centres around the variety of choices one can make to adapt specific core elements (i.e., the constraint(s) and the alternatives) in the design of a PVE to the policy question at hand.

Type of constraint

The type of constraint is a distinct design feature that varies between existing applications. Most studies use a monetary constraint, typically in the form of a maximum public budget that can be allocated towards a range of policy alternatives to be selected by the respondent. For example, Mulderij et al. (2021) conducted a PVE to elicit citizens' preferences regarding the public funding of interventions promoting healthy body weight among people with low incomes. Respondents were asked to select their preferred portfolio of policies considering a maximum public budget that was not sufficient to fund all projects. They were informed that any surplus budget would be shifted to next year and used for the same policy purpose (Mulderij et al., 2021). Alternatively, a monetary constraint may also take the form of a minimum rather than a maximum. In a study on citizens' preferences for disinvestment in healthcare, Rotteveel et al. (2022) asked respondents to select a portfolio of treatments for which the government should discontinue reimbursement, so that a minimum saving of €100 million could be achieved.

A constraint can also take a form other than monetary. For example, in two different PVE-applications regarding citizens' preferences for COVID-19 lockdown restrictions, it was considered that the pressure on the healthcare system was the most important constraint for policymakers. Therefore, in the PVE-application by Mouter et al. (2021a) on the relaxation of COVID-19 lockdown restrictions, respondents could select a portfolio of restrictions they preferred to be relaxed while respecting the constraint of a maximum of 50% additional pressure on the healthcare system. Similarly, in one of the scenarios of a PVE-application regarding public preferences for the introduction of COVID-19 lockdown restrictions under different scenarios by Mouter et al. (2022), the constraint was that respondents were required to select a portfolio of policy alternatives resulting in a risk reduction of at least 30%. It should be noted that the link with the Kaldor-Hicks framework is lost when a non-monetary constraint is selected, as respondents no longer trade-off their private income with public-resource allocations.

Fixed or flexible constraint

Another design feature is the choice of a fixed or flexible constraint. Most existing PVE-applications have included a fixed constraint. This may be desirable in cases where the level of the constraint is predetermined and policymakers need to adopt and implement policies within that constraint. A flexible constraint may be more appropriate if the goal of the PVE is to elicit citizens' preferences towards both a set of policy alternatives and the trade-off between public expenditure and private spending capacity. Two PVE-applications, on citizens' preferences for flood-protection programs and for urban-mobility investments applied such a flexible constraint (Dekker et al., 2019; Mouter et al., 2021c). In one of the versions of both experiments, respondents were allowed to select a portfolio of projects with a total expenditure that was either lower or higher than the target budget, in which case the related tax would be lowered or raised accordingly. This allowed the studies to elicit public preferences for the policy alternatives and the level of governmental expenditure on the policy issue simultaneously.

Number of constraints

In a PVE, one or multiple constraints can be implemented. Most existing applications have included a single constraint. However, an application on public preferences for CO_2-emission reduction policies required respondents to consider two constraints when selecting policy options: The target level for CO_2-reduction and the available budget (Van Beek et al., 2022). The potential to include multiple constraints in a PVE is an advantage if policymakers must consider (all) those constraints in the actual policy context. Disadvantages are that it may increase the cognitive burden on respondents and it complicates the model estimation.

Labelled or unlabelled alternatives

The policy alternatives in a PVE are described by a range of attribute levels and may come with or without labels. Most existing PVE applications are labelled, meaning that respondents are informed about the actual policy alternatives represented by the attribute levels, such as policies promoting a healthy body weight (Mulderij et al., 2021), lockdown restrictions (Mouter et al., 2021a; 2022), or climate policies (Hössinger et al., 2023; Van Beek et al., 2022). The application by Rotteveel et al. (2022) on disinvestment in healthcare, however, employed unlabelled alternatives because the authors anticipated that labels could influence respondents' preferences, when their study was focused on the importance of the attributes of healthcare interventions. Like in a labelled DCE, the inclusion of labels for the alternatives in a PVE limits the generalizability of the preference estimates for attribute levels, since respondents may incorporate other factors in their decision-making. However, the inclusion of labels adds to the realism of the choice task (De Bekker-Grob et al., 2010; Kruijshaar et al., 2009).

Overview of published applications

All in all, this discussion of the distinct design characteristics of existing PVE applications shows that there is considerable room within the PVE-framework to adapt and tailor the design to the relevant features of the policy question at hand. This concerns especially the constraint (i.e., fixed or flexible, monetary or another type, single or multiple) as well as the presentation of policy alternatives (labelled or unlabelled). Table 2.1 presents a summary overview of these characteristics and their implementation in the PVE applications published so far, four in the health domain and four in other domains. An overview of other characteristics of these eight studies (e.g., the number of respondents, the estimated choice model) is provided in Table A2.1 in Appendix 2A.

Table 2.1. An overview of the design characteristics of published PVE applications

Study	Topic	Fixed or flexible constraint	Type of constraint	Single or multiple constraints	Presentation of policy alternatives
Health					
Mulderij et al. (2021)	Policies promoting a healthy body weight	Fixed	Budget	Single	Labelled
Mouter et al. (2021a)	COVID-19 lockdown policies	Fixed	Max. pressure on healthcare system	Single	Labelled
Rotteveel et al. (2022)	Disinvestment of healthcare interventions	Fixed	Min. expenditure savings	Single	Unlabelled
Mouter et al. (2022)	COVID-19 restrictions under different scenarios	Fixed	Max. pressure on healthcare system	Single	Labelled
Other domains					
Dekker et al. (2019)	Urban mobility investments	Fixed/ Flexible	Budget	Single	Labelled
Mouter et al. (2021c)	Flood protection programs	Flexible	Budget	Single	Labelled
Van Beek et al. (2022)	Reduction of CO_2 emission	Fixed	Budget and min. CO_2-emission reduction	Multiple	Labelled
Hössinger et al. (2023)	Reduction of CO_2 emission in transport	Flexible	CO_2 emission reduction target	Single	Labelled

Abbreviation: PVE=Participatory Value Evaluation

Table 2.2. An overview of the various characteristics regarding the structure and focus of the choice tasks in the included preference-elicitation methods

Method	Number of choice sets	Type of choice task	Focus of choice task	Embedded in RUT	Constraint
DCE	Multiple	1 discrete choice	Attribute levels (and alternatives)	Yes	No
BWS-1	Multiple	2 discrete choices	Attributes	Yes	No
BWS-2	Multiple	2 discrete choices	Attribute levels	Yes	No
SW	Single	Ranking and point allocation	Attribute levels	No	No
PVE	Single	Continuous and discrete choices	Alternatives (and attribute levels)	Yes	Yes

Abbreviations: DCE=Discrete Choice Experiment, BWS-1=Best-Worst Scaling Case 1, BWS-2=Best-Worst Scaling Case 2, SW=Swing Weighting, PVE=Participatory Value Evaluation, RUT=Random Utility Theory

Position of PVE relative to other preference-elicitation methods

In the domain of health, a wide range of preference-elicitation methods is used (Soekhai et al., 2019b). To obtain a better view on the position of PVE relative to other methods, in this section the PVE method is compared to a selection of established methods. This selection is based on the final recommendations of the PREFER consortium (2022), in which eleven preference-elicitation methods are recommended based on an appraisal of methods by stakeholders and experts (Whichello et al., 2020b). Of these, five were explored in-depth by the PREFER consortium: The Discrete Choice Experiment (DCE) or Best-Worst Scaling Case 3 (BWS-3), Best-Worst Scaling Case 1 (BWS-1), Best-Worst Scaling Case 2 (BWS-2), the (Probabilistic) Threshold Technique (TT), and Swing Weighting (SW). All of these are included in the comparison with PVE[3], except for TT, since this is not a multi-attribute method (Hauber & Coulter, 2020) and therefore the least related to PVE. Table 2.2 presents an overview of the similarities and differences in the structure and nature of the choice tasks of the four remaining preference-elicitation methods and PVE.[4]

Number of choice tasks

While SW and PVE present all attribute levels (and all alternatives in the case of PVE) in a single choice task, DCE and both types of BWS involve multiple choice tasks. An advantage of the former is that it is probably closer to the reality of the policymaker, who faces all choice options at once rather than in multiple choice tasks. On the other hand, an advantage of multiple choice sets is that this is more efficient as multiple choices are observed for every respondent and, therefore, a smaller number of respondents is required. As another potential advantage of multiple choice tasks, the cognitive burden imposed on respondents may be lower, given that these choice tasks typically offer only two[5] rather than all policy alternatives simultaneously.

---

3 *DCE*: See Lancsar and Louviere (2008) and Mühlbacher and Johnson (2016) for introductions into the DCE-method, and De Bekker-Grob et al. (2012), Clark et al. (2014), and Soekhai et al. (2019a) for systematic reviews of DCE applications. *BWS*: See Flynn et al. (2007) for an introduction into BWS-2, Mühlbacher et al. (2016) for a survey of all three cases of BWS, methodological issues and the applied BWS literature, and Cheung et al. (2016) and Hollin et al. (2022) for extensive reviews of BWS applications. *SW*: See Edwards and Barron (1994) and Srivastava et al. (1995) for early discussions and comparisons of various ranking methods including SW, Tervonen et al. (2017) for a description of the SW-method and a conceptual comparison with DCE, and Whichello et al. (2023) for an empirical comparison with DCE.

4 Figure A2.1, included in Appendix 2B, provides stylized examples of the choice tasks of all five compared preference-elicitation methods.

5 In a systematic review of DCE applications in health, 83% of the 301 identified studies between 2013 and 2017

Type of choice task

The methods present respondents with different types of choice tasks. In DCE, respondents need to make one discrete choice per choice task, for their most preferred alternative. In BWS-1 and BWS-2, respondents need to make two discrete choices per choice task, for the most and least preferred attribute or attribute level respectively. In SW, respondents do not make discrete choices, but are asked to first rank level improvements in each attribute from most to least desired, and then assign points to weigh the importance of each attribute level improvement. In PVE, finally, respondents make multiple discrete choices by selecting policies in their portfolio and simultaneously make a continuous choice by determining the extent of allocated resources. This portfolio-based choice task allows respondents to evaluate all the alternatives on offer in relation to each other. This may lead them to select combinations of portfolios that are not necessarily in line with their ranking of the individual alternatives, as synergies between projects and distributional effects may be considered (Bahamonde-Birke & Mouter, 2024; Mouter et al., 2021b; 2021c).

Focus of choice task

The methods focus on different aspects of the decision problem. In DCE, respondents choose between two or more alternatives described by a number of attribute levels. Commonly, alternatives are unlabelled in a DCE (De Bekker-Grob et al., 2010) and, therefore, respondents base their choices on the attribute levels only. As such, the focus of the choice task is on the attribute levels. If the alternatives are labelled, there is an additional focus since respondents are also informed about the labels of the policy alternatives and can, therefore, incorporate factors other than the included attributes and levels in their decision-making. In BWS-1, respondents are presented with a single alternative (in the context of this method often referred to as 'object') described by a set of attributes and are asked to select their most- and least-preferred attribute. The attributes are presented without levels so there is an exclusive focus on attributes. In BWS-2, respondents are also presented with a single alternative described by a number of attributes, however, the attributes are presented with levels, and respondents need to select their most- and least-preferred attribute levels. The focus of BWS-2 is, therefore, on attribute levels. In SW, the focus is also on attribute levels since respondents need to rank and weight improvements in various attribute levels. Finally, in a PVE, the focus

were found to include two alternatives per choice set (excluding any opt-out or *status quo* alternative) (Soekhai et al., 2019a).

is predominantly on the alternatives since respondents compose portfolios of labelled alternatives. In addition, there is a secondary focus on attribute levels since these are also included to describe the impact of the alternatives on various outcomes.

Theoretical foundations

Four of the five methods are embedded in random-utility theory, only SW is not. Therefore, welfare estimates can be derived from DCE, BWS-1, BWS-2, and PVE, but not from SW. For DCE, BWS-1 and BWS-2, this is straightforward (Lancsar & Savage, 2004), but it requires a more elaborate procedure for PVE (Lloyd-Smith, 2018). The resulting welfare estimates can be used as inputs in other economic methods for policy evaluation, such as cost-effectiveness analysis (CEA), cost-benefit analysis (CBA), or cost-utility analysis (CUA) (e.g., McIntosh, 2006).

Inclusion of a constraint

Only the PVE-design includes a constraint. Thus, PVE is the only method that forces respondents to explicitly incorporate the constraint(s) faced by policymakers in their actual decision-making where resources are scarce and the allocation of (collective) resources is therefore constrained. In cases of a flexible budget, PVE also allows respondents to trade-off public and private expenditures.

Discussion

PVE is a new preference-elicitation method for the preference-based assessment of policy alternatives. This paper introduces PVE in the health policy domain, discusses its theoretical background and the main features of recently published practical applications, and positions it relative to the established methods DCE, BWS-1, BWS-2, and SW. We find that PVE comes with three (potential) advantages and two (potential) disadvantages relative to established methods.

Potential advantages

A first advantage of PVE is that its portfolio-based choice task allows respondents to evaluate policy alternatives in relation to each other while also considering synergies between alternatives and distributional consequences. For example, in two recent PVE applications on investments in transport projects and flood-protection programs, a substantial number of respondents selected a portfolio of projects in different parts of the region or country under consideration and explained that they considered spatial fairness in their portfolio choice (Mouter et al., 2021b; 2021c). In the health domain, such considerations may play a role in, for example, the distribution of healthcare services across regions or health outcomes across population subgroups. Unlike most other preference-elicitation methods, such distributional considerations as well as synergies between projects can be explicitly captured by the PVE framework (Bahamonde-Birke & Mouter, 2024). A potential second advantage is that it forces respondents to make their decisions within the constraint(s) that policymakers face. As stressed in the literature applying portfolio theory to economic evaluation and resource allocation in health, healthcare budgets can be considered fixed in the short run and to be spent on a portfolio of goods and services. The choice set of policymakers is, therefore, constrained by the public budget, rendering opportunity costs important (Bridges, 2004; Bridges et al., 2002; Sendi et al., 2003). Other preference-elicitation methods typically do not incorporate budget constraints and opportunity costs explicitly. Previous research has shown that a substantial share of respondents in these studies either discount the scarcity of resources (e.g., Ding et al., 2005) or even ignore the cost attribute entirely (e.g., Erdem et al., 2015; Koetse, 2017; Sever et al., 2019b), which may reduce the external validity of the findings.

These characteristics of the PVE choice task mean that it reflects actual policy decisions more closely than the other methods discussed, which may contribute to the involvement of respondents in the study and the acceptance and support of its findings. Respondents in the PVE studies discussed indicated that they appreciated the method for presenting them with the dilemmas policymakers actually face, increasing their awareness, and as a means for voicing their opinion (Juschten & Omann, 2023; Mouter et al., 2021a; 2021d; 2022; Mulderij et al., 2021; Rotteveel et al., 2022). A third advantage of PVE is its capability to simultaneously elicit public preferences for policy alternatives and the trade-off between public and private expenditure in the respective policy area. This may be especially useful in the context of deciding on the reimbursement of new treatments in the context of increasing healthcare expenditures.

Potential disadvantages

A first disadvantage of PVE is that it is less efficient than a preference-elicitation method that uses multiple choice tasks to elicit preferences (i.e., DCE, BWS-1, BWS-2). Since respondents are only presented with a single choice task in a PVE, and there is only experimental variation in attribute levels between respondents and not within respondents, the method requires larger samples of respondents to accomplish an estimation of similar accuracy. Secondly, because of its single choice task presenting all alternatives and attribute levels at once, PVE may impose a larger cognitive burden on respondents than methods containing multiple choice tasks. The amount of information presented to respondents and the complexity of the choice task may limit the inclusiveness of the method (Juschten & Omann, 2023). On the other hand, the single choice task in PVE may also prevent respondent fatigue and boredom that is sometimes observed in methods with multiple choice tasks, such as DCE (e.g., Savage & Waldman, 2008; Swait & Adamowicz, 2001; Weng et al., 2021). This risk of cognitive overload requires close attention to PVE design and consideration of the feasibility of using PVE across all population subgroups (e.g., older individuals, people on the lower end of the cognitive ability distribution) and warrants further study.

Discussion of limitations and directions for future research

Two reflections should be made regarding the selection of methods in this paper for comparison with PVE. Firstly, we compared PVE only with a selection of frequently used multi-attribute preference-elicitation methods. Other preference-elicitation methods such as the Volumetric Choice Experiment (VCE) (Carson et al., 2022; Chalak et al., 2023), Constant-Sum Paired Comparisons (Skedgel & Regier, 2015; Skedgel et al., 2015), the Basked-Based Choice Experiment (BBCE) (Caputo & Lusk, 2022) and Basked and Expenditure Based Choice Experiment (BEBCE) (Neill & Lahne, 2022) are more comparable to PVE as they ask respondents to make continuous (and discrete) choices. These methods have not been included in this study, however, as they have not (yet) or rarely been applied in the health domain. Further research should compare PVE with these as well as a wider range of other methods, such as frameworks that only evaluate policy alternatives without eliciting preferences themselves (e.g., CEA, CUA) as well as methods that are not multi-attribute in nature (e.g., (Probabilistic) Threshold Technique) (Hauber & Coulter, 2020) or that scored worse in the appraisal of preference-elicitation methods by Whichello et al. (2020b), like Contingent Valuation (CV) (e.g., Diener et al., 1998; Smith & Sach, 2010) or Person Trade-Off (PTO) (e.g., Green, 2001; Nord, 1995). One could also envisage positioning PVE relative to Multi-Criteria Decision Analysis (MCDA), which is a framework often used to support decision-making in healthcare (e.g., Hansen & Devlin, 2019; Marsh et al., 2016; Thokala et al., 2016). MCDA has not been included in this paper as it is not considered to be an elicitation method itself, but instead a decision-making framework that incorporates preference-elicitation methods as its choice task (The PREFER consortium, 2022).

Another limitation worth mentioning is that PVE is still a relatively novel method. The literature on the method is growing, including in the healthcare field, but is still limited. For example, while PVE might seem closer to the reality of policymakers than other preference-elicitation methods due to its constraint(s) and portfolio-based choice task in a single choice set, it has not yet been studied empirically whether respondents - or policymakers - in fact experience this. Even though PVE may be expected to impose a larger cognitive burden on respondents relative to some of the other methods, this has not yet been empirically examined. Therefore, further research is warranted to empirically assess the feasibility and face validity of PVE as well as the extent to which the method actually reflects the reality of political decision-making, including in comparison with more established preference-elicitation methods. Information on these aspects would allow researchers and policymakers to make better-informed choices for preference-elicitation methods. Also, additional applications of PVE to policy problems in health are needed to further explore its usefulness and implications for health-policy decision-making.

Conclusion

PVE seems a promising complementary method for eliciting preferences and involving citizens or patients in healthcare decision-making, but there is still room to further explore the method. PVE differs from the other preference-elicitation methods in its inclusion of an explicit resource constraint and its ability to simultaneously elicit preferences for policy alternatives and trade-off public and private spending, while also considering synergies between alternatives and distributional effects. This may come at the expense of the efficiency of the method and the understandability of the choice task for a broad set of respondents. These findings suggest that researchers and policymakers interested in the preference-based assessment of health-policy alternatives should trade-off the advantages and disadvantages of PVE against each other in their selection of a preference-elicitation method for a policy dilemma at hand. In a context in which a portfolio of multiple policy alternatives can be selected within a constraint and in which both public and private resources can be allocated, PVE seems to add value. Further research is required, nevertheless, into the feasibility and validity of PVE.

Appendix 2A: Details of existing applications

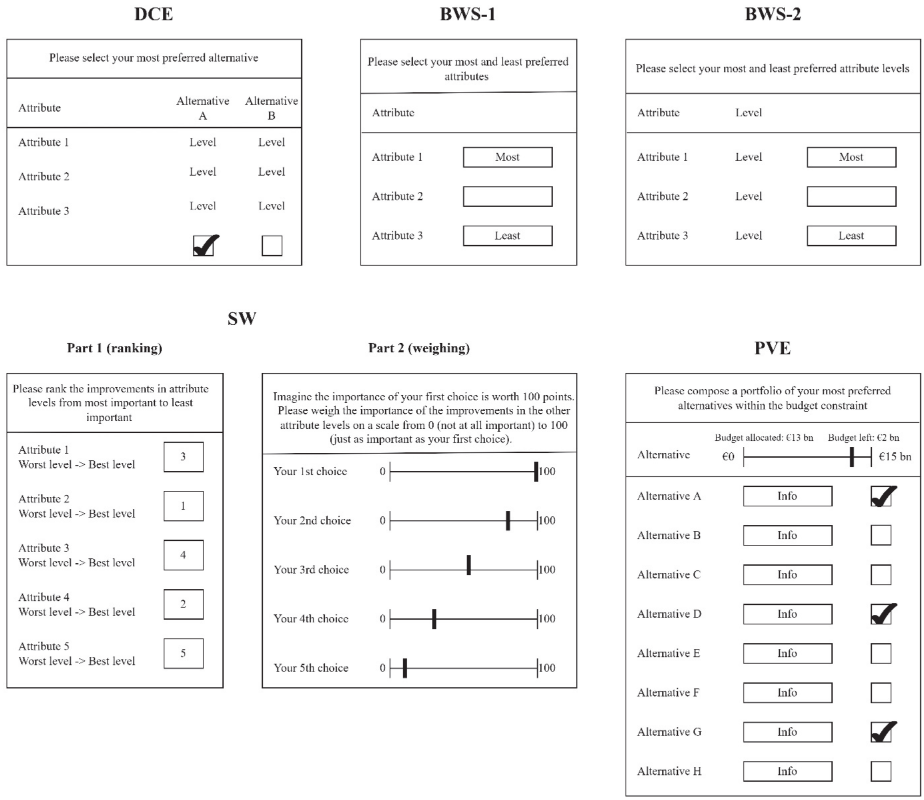
Table A2.1. An overview of the published PVE-applications

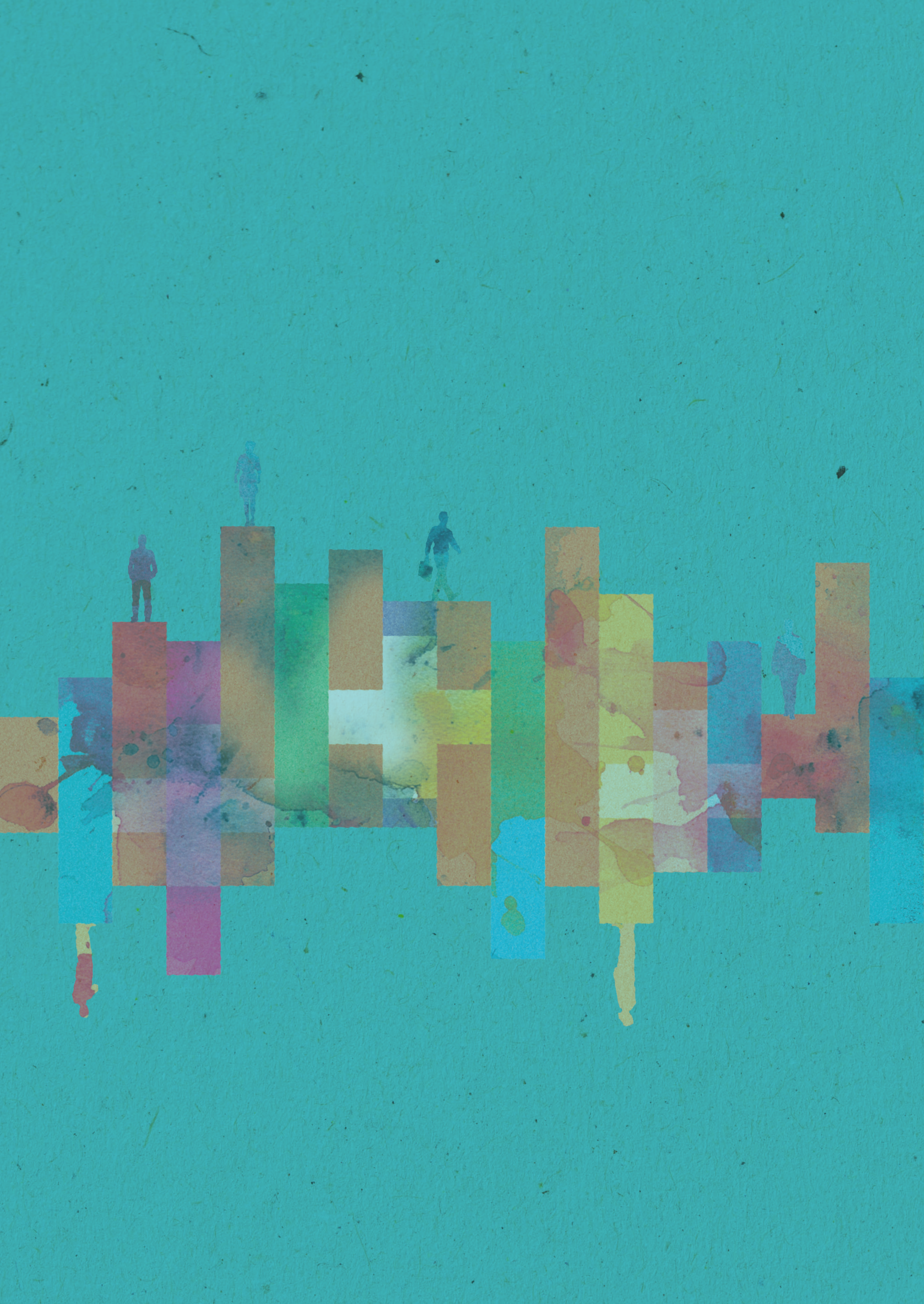
Study	Topic	N policy alternatives	N attributes	N respondents	Choice model
Health					
Mulderij et al. (2021)	Policies promoting a healthy body weight	8	7	1,053	MDCEV
Mouter et al. (2021a)	COVID-19 lockdown policies	8	6	29,651[a]	MDCEV-PVE
Rotteveel et al. (2022)	Disinvestment of healthcare interventions	8	7	1,143	PCM
Mouter et al. (2022)	COVID-19 restrictions under different scenarios	9 – 14[b]	1	2,011	PCM
Other domains					
Dekker et al. (2019)	Urban mobility investments	16	7	2,498	MDCEV-PVE
Mouter et al. (2021c)	Flood protection programs	14	6	2,900[c]	MDCEV-PVE
Van Beek et al. (2022)	Reduction of CO_2 emission	10	2	10,810[d]	-[e]
Hössinger et al. (2023)	Reduction of CO_2 emission in transport	11	16[f]	1,650	MNL

Abbreviations: MDCEV(-PVE)=Multiple Discrete-Continuous Extreme Value Model (for Participatory Value Evaluation), MNL=Multinomial Logit Model, PCM=Portfolio Choice Model. a) This consist of 3,358 respondents recruited from an online panel and 26,293 respondents who filled out the openly accessible online PVE. b) The PVE consisted of four different pandemic scenarios, with varying number of policy alternatives presented to respondents. c) This consists of 1,855 respondents who received a fixed-budget PVE, and 1,045 respondents who received a flexible-budget PVE. d) This consists of 2,163 respondents recruited from an online panel and 8,647 respondents who filled out the openly accessible online PVE. e) The data of this PVE were not analysed using a choice model, but only using descriptive statistics. f) In this PVE, a distinction is made between 5 "main effects", presented as ratio attributes, and 11 "further effects", presented on a uniform relative scale.

Appendix 2B: Stylized examples of all five compared preference elicitation methods

Figure A2.1. Stylized examples of the choice tasks of the Discrete Choice Experiment (DCE), Best-Worst Scaling Case 1 (BWS-1), Best-Worst Scaling Case 2 (BWS-2), Swing Weighting (SW) and Participatory Value Evaluation (PVE)





Chapter 3

Ordering effects in discrete choice experiments: A systematic literature review across domains

Based on:

Boxebeld, S. (2024). Ordering effects in discrete choice experiments: A systematic literature review across domains. *Journal of Choice Modelling*, 51, 100489

Abstract

Discrete choice experiments (DCEs) are increasingly used in several scientific domains. Since their results may be used to inform governmental decision-making, it is important that the validity of the method is continuously scrutinized. An often-studied design artefact is the impact of the presentation order of alternatives, attributes, and choice sets on the results of a DCE. No systematic review of the literature on ordering effects existed until now, and many applied studies using a DCE do not explicitly consider the role of ordering effects. I conducted a systematic review of the literature on ordering effects in this study. Using a three-step snowball sampling strategy, 85 studies were identified across various scientific domains. The majority of included studies documented statistically significant ordering effects. Alternative and attribute ordering effects are primarily caused by lexicographic behaviours, while choice set ordering effects seem to be caused by respondent learning, fatigue, or anchoring. Although ordering effects may not always occur, the majority of studies that did find statistically significant effects warrants the use of mitigation methods. An overview of potential mitigation methods for the applied DCE literature is presented, including randomization of presentation orders, advance disclosure of DCE core elements, and inclusion of alternative-specific constants (ASCs), attribute level overlap, and an instructional choice set (ICS). Finally, several directions for future methodological research on this topic are provided, particularly regarding heterogeneity in ordering effects by study design traits and respondent characteristics, and interactions between ordering effects. Insights in these aspects would further our understanding of respondents' processing of DCEs.

Introduction

The past decades have seen a rapid increase in the use of discrete choice experiments (DCEs) across scientific domains (Haghani et al., 2021a). The results of DCEs may be used to inform governmental decision-making regarding, for example, the design of more targeted policies or interventions. Therefore, it is crucial that the validity of the method is scrutinized on a regular basis. There has been much attention in the methodological literature for the internal validity of DCEs. A range of studies found the preferences elicited in choice experiments to be influenced by design artefacts, such as the number of attributes (Gao & Schroeder, 2009; Louviere et al., 2008), the framing of trade-offs (Rolfe & Brouwer, 2012), the use of words or graphics to present attribute levels (DeLong et al., 2021; Shr et al., 2019; Veldwijk et al., 2015), and the framing of attributes (Kragt & Bennett, 2012; Veldwijk et al., 2016b).

Another often-studied design artefact in the general survey methodology literature is the effect of the order of presentation of survey elements on respondents' answers. The order of questions in a survey has been found to affect responses (e.g., Thau et al., 2021; Van de Walle & Van Ryzin, 2011), and the presentation order of response options to a question likewise affects the results (e.g., Garbarski et al., 2016; Krosnick & Alwin, 1987). Statistically significant ordering effects have also been found for stated preference elicitation methods other than DCEs, such as Contingent Valuation (Boyle et al., 1993; Powe & Bateman, 2003) or rating-based conjoint analysis (DeSarbo et al., 2004; Ryan et al., 1998). Based on these findings, therefore, it is reasonable to expect ordering effects to be present in DCEs as well. Two early studies found statistically significant choice set ordering effects and significant alternative, attribute, and choice set ordering effects, respectively (Bradley & Daly, 1994; Chrzan, 1994). If not mitigated or accounted for, such ordering effects may bias the preferences estimated in a DCE.

Since the existence and magnitude of ordering effects may arguably depend on the design and topic of the choice experiment in question, many studies have since examined this. No systematic review of studies on this topic is available to date, however.[1] To address this knowledge gap, this study conducts a cross-domain systematic literature review. Even though many studies have methodological relevance

1 The Appendix of Czajkowski et al. (2014) provides a review that was not systematic and did only include studies on choice set ordering effects published up until 2012. Relative to that study, this study conducts a more rigorous and comprehensive review by also including studies on alternative and attribute ordering effects, also those published after 2012.

across domains, it may happen that important insights do not diffuse to other domains as a result of the somewhat fragmented research landscape of choice modelling (Haghani et al., 2021a). For example, it seems that many studies do not vary the order of attributes, an issue mostly examined in the health domain. By bridging the insights from the literature in various domains, this study aims to contribute to the applied DCE literature by providing recommendations for the mitigation of ordering effects and to the methodological literature with directions for future research on heterogeneity in and interactions between ordering effects.

Theoretical Framework

Before going into the methodology of this review and the empirical findings of the included studies, a concise theoretical framework is presented in this Section to provide the reader with a theoretical underpinning of the different types of ordering effects. The theoretical mechanisms are described below and summarized in Table 3.1.

In case of alternative ordering effects, the presentation order of alternatives within the choice set influences respondents' preferences. The theoretical mechanism underlying this type of ordering effect relates to lexicographic behaviour. A DCE choice task typically consists of a matrix with the alternatives presented in columns and the attributes in rows. Given that people read from left to right in most languages, they tend to process the alternatives sequentially from left to right.[2] As such, the alternative ordering effect is also called left-right bias[3], left-to-right bias, or position bias. It is not clear, however, which position in the matrix results in a higher choice probability, ceteris paribus, as theoretical arguments could be provided for different directions. On the one hand, the left-hand alternative is likely to be examined first by respondents, which may result in a higher probability of the left-hand alternative being selected if

2 Notably, several languages in which people read from right to left are actively in use, such as Arabic, Fula, and Hebrew. This is potentially relevant, given that there is evidence that attention shifting operates in opposite directions for those who read from left to right as compared to those who read from right to left (e.g., Smith & Elias, 2013; Spalek and Hammad, 2005). Since no clear pattern has been found in the direction of alternative ordering effects, it is unclear whether and how alternative ordering effects vary by respondents' language background.

3 The names 'left-right bias' and 'left-to-right bias' imply that the alternatives are presented in the columns of the DCE matrix, which is usually applicable. In rare cases of the alternatives being presented in rows instead of columns (e.g., Bennett et al., 2004; Boto-García et al., 2022; Rolfe & Windle, 2005), one would have to speak of 'top-bottom bias' or 'top-to-bottom bias', instead.

the first impression is more influential on one's choice or if a respondent immediately chooses the first examined alternative without further processing the remainder of the choice set. On the other hand, if respondents do process the entire choice task and the last examined alternative is more influential on one's choice, the right-hand alternative might be selected more often.

For attribute ordering effects, the underlying mechanism is very similar to that of alternative ordering effects; since attributes are typically presented in rows in a choice experiment matrix and most people read from the top to the bottom, attributes are likely to be processed sequentially by respondents. Again, theoretically, it is not clear whether this results, ceteris paribus, in the attribute positioned in the top row, in the bottom row, or in between, being more important in respondents' choices. Apart from the impact of attribute order on preference estimates, it can be argued that it may affect the error variance, too. Especially in case there is a natural grouping of attributes (e.g., the benefits of a transport project or medical treatment on the one hand and its risks or costs on the other hand; dimensions concerning physical health on the one hand and dimensions concerning mental health on the other hand in the context of valuing health states), the cognitive burden for respondents and error variance of the estimated model may be lower in case the attributes are clustered together in the choice task rather than presented in an entirely random order (Heidenreich et al., 2021; Norman et al., 2016a). A similar argument may be provided if there is a logical ordering of attributes (e.g., context variables, that vary between choice sets but not between alternatives, presented before variables that vary between alternatives), in which case adhering to this logical ordering in the choice task may result in a lower error variance than in case this logical ordering is ignored.

The choice set ordering effect, also called sequence effect, refers to the impact of the position of a choice task in the sequence of choice tasks presented to a respondent within a choice experiment. In contrast with alternative and attribute ordering effects, choice set ordering effects are not considered to be the result of lexicographic behaviour per se; after all, respondents are presented with only one choice set at a time. Instead, the existence of choice set ordering effects is linked to processes that may take place in respondents over the course of completing the choice experiment. A first proposed mechanism is commonly referred to as the learning effect; respondents gain experience with the choice setting of the experiment (i.e., institutional learning) and may learn about their preferences regarding the topic in question (i.e., value learning) as they go through the sequence of choice tasks. Learning is expected to result in a decreasing error variance; as respondents become more familiar with the choice

3

environment and the choice context, they learn about their choices and their choices consequently become more deterministic (e.g. Czajkowski et al., 2014). Also, better-defined preferences resulting from value learning may result in a decreasing probability of choosing the status quo/opt-out over the sequence of choice sets (e.g. Weng et al., 2021).

As a second mechanism, respondents may become tired or lose interest after a series of choice tasks because of the cognitive burden imposed on them in the processing of the choice tasks and lose concentration, denoted as the fatigue effect. Contrarily to the learning effect, the presence of fatigue is expected to result in an increasing error variance and/or an increasing probability of choosing the status quo/opt-out over the sequence of choice sets; as respondents become tired and, therefore, less focused on/devoted to the task, they may make more random choices in the final choice sets (e.g. Bradley and Daly, 1994; Savage and Waldman, 2008) or resort more often to the status quo/opt-out (e.g. Swait and Adamowicz, 2001). Finally, a third proposed mechanism underlying ordering effects is focused on the impact on preference parameters rather than the error variance. This mechanism is called the anchoring effect and suggests that respondents' choices in later choice sets may be affected by the attribute levels presented in earlier choice sets.[4] For example, if the first choice set contains an alternative with a low level for the price attribute, this may lead to a lower willingness-to-pay (i.e., increased cost sensitivity) and higher probability of choosing the status quo/opt-out in case a later choice set contains more expensive alternatives (e.g. Day and Pinto-Prades, 2010; Scheufele and Bennett, 2012).[5]

---

4 The anchoring effect may also be the explanation of what is called 'starting point bias'. This refers to respondents' choices in later choice sets being affected by the first attribute levels they see, so either in the first choice set or in an instructional choice set. Essentially, this can be seen as a narrower version (only focused on the difference between the first choice set and the rest of the choice experiment) of the choice set ordering effect (Lades et al., 2022).

5 Some studies, therefore, also speak of 'strategic learning' (e.g., Scheufele & Bennett, 2012); by going through the sequence of choice sets, respondents see varying attribute levels in the different alternatives that are presented to them. As they learn about the full attribute level range, respondents may develop the strategic tendency to choose the status quo in later choice sets because they have seen more attractive alternatives (e.g., cheaper, more effective, better quality) in previous choice sets. This paper uses the term 'anchoring' to distinguish this theoretical mechanism more clearly from other types of learning.

Table 3.1. Overview of the theoretical mechanisms underlying ordering effects and their expected results.

Type of ordering effect	Theoretical mechanism	Expected result
Alternative ordering effects	Lexicographic behaviour	• Preferences influenced by alternative order (direction unknown)
Attribute ordering effects	Lexicographic behaviour	• Preferences influenced by attribute order (direction unknown) • Error variance increases by deviation from any natural deterministic ordering/grouping of attributes
Choice set ordering effects	Learning	• Error variance decreases over the sequence of choice sets presented to a respondent • Probability of status quo/opt-out being chosen decreases over the sequence of choice sets presented to a respondent
	Fatigue	• Error variance increases over the sequence of choice sets presented to a respondent • Probability of status quo/opt-out being chosen increases over the sequence of choice sets presented to a respondent
	Anchoring	• Preferences influenced by attribute levels in previous choice sets presented to a respondent • Probability of status quo/opt-out being chosen increases over the sequence of choice sets presented to a respondent

3

Methodology

This study is, where applicable, reported according to the Preferred Reporting Items for Systematic Review and Meta-analysis (PRISMA) guidelines (Page et al., 2021). See Appendix 3B for the PRISMA checklist.

Inclusion criteria

In the literature search and selection process described in the following section, three inclusion criteria were used to determine the eligibility of a study for this review. Firstly, a study needs to empirically examine the effect of the ordering of alternatives/attributes/choice sets, by varying the order of these elements between/within respondents and/or including a term in the equation capturing the order effect. Secondly, the outcome measure of a study needs to be a preference (or scale) estimate. This may take the form of marginal utility estimates, derived measures of preference such as Willingness-to-Pay (WTP), Marginal Rate of Substitution (MRS), or Marginal Acceptable Risk (MAR), or of a scale parameter. Thirdly, a study needs to apply the discrete choice experiment (DCE) method, sometimes also called stated choice experiment or choice-based conjoint analysis. A DCE is defined here as a survey-based stated preference elicitation method in which respondents are required to make a single discrete choice for their most preferred option in a choice set[6] when faced with at least two alternatives (one of which may be an opt-out or status quo option). Also, the respondent needs to answer at least two choice sets sequentially, and each alternative is described by at least two attributes with varying levels.

Only if all these three substantive criteria as well the criterion that the paper was written in English were met, it was included in the systematic literature review. Thus, various types of studies have been excluded from this review. For example, many studies failed the first inclusion criterion by having only mentioned ordering effects as a potential limitation to their studies. Some studies have, alternatively, mentioned an attempt to minimize any bias resulting from ordering effects by varying the order of the alternatives/attributes/choice sets within their study, without empirically testing its effect. Other studies have been excluded for examining the existence of choice set ordering effects using two occurrences of the same choice set at different points in the sequence of choice sets (e.g. Carlsson et al., 2012; Segovia and Palma, 2021).[7] Besides, some studies have not fulfilled the second inclusion criterion by testing for ordering effects on other outcome measures, such as self-reported measures of choice certainty (e.g. Olsen et al., 2011) or visual attention measures (e.g. Ryan et al., 2018; Selivanova and Krabbe, 2018).[8] Finally, some studies have not fulfilled the third inclusion criterion by examining ordering effects using methods other than a DCE. This holds for both studies that employed DCE-resembling methods with a single choice set (e.g. Oppewal et al., 2015) as well as studies that used other methods, such as Contingent Valuation (e.g.

6 An exception is Best-Worst Scaling (BWS) Case 3, also called Multiprofile-Case BWS or Best-Worst DCE. In BWS Case 3, respondents choose both their most preferred and least preferred alternative for each choice set. This is conceptually closer to the traditional type of DCE than the other cases of BWS. The choice tasks for both methods are namely highly similar and the convergent validity of both methods seems to be high (Xie et al., 2014). However, recent research has suggested that respondents make use of different decision rules for choosing the most and least preferred alternatives in BWS Case 3 (Geržinič et al., 2021). As such, it may still be questioned to what extent the occurrence of ordering effects differs between DCE and BWS Case 3. Since only two (otherwise eligible) studies applying BWS Case 3 were identified in the search process of this study (Marsh & Phillips, 2012; Mulhern et al., 2017), their inclusion/exclusion would not change the overall findings of this review and these studies have been included in the final study sample.

7 A difference in choices by a respondent in two occurrences of the same choice task does not necessarily identify a choice set ordering effect for two reasons. Firstly, a respondent may recognize the repetition of the choice set and be annoyed by this. This may induce protest behaviour and, resultingly, unreliable responses. Secondly, there is a chance that a respondent is completely indifferent between two alternatives and randomly selects one of the alternatives. If a choice set is repeated, a respondent may then choose the other alternative without this being the result of an ordering effect. Therefore, studies aiming to identify choice set ordering effects using repeated choice sets have been excluded from the literature review.

8 Notably, even if the ordering of alternatives/attributes would affect attention fixation in a choice experiment, this does not necessarily result in an effect on choices, as shown by Meißner et al. (2016).

Boyle et al., 1993; Clark and Friesen, 2008; Powe and Bateman, 2003), Factorial Surveys and Conjoint Analysis types other than Choice-Based Conjoint Analysis (e.g. Auspurg and Jäckle, 2017; DeSarbo et al., 2004; Ryan et al., 1998), Best-Worst Scaling Case 1 or 2 (e.g. Campbell and Erdem, 2015; Nguyen et al., 2022), Time Trade-Off (e.g. Craig et al., 2015; Pinto-Prades et al., 2019), and other types of survey experiments (Atalay et al., 2012; Dayan and Bar-Hillel, 2011; Mantonakis et al., 2009).

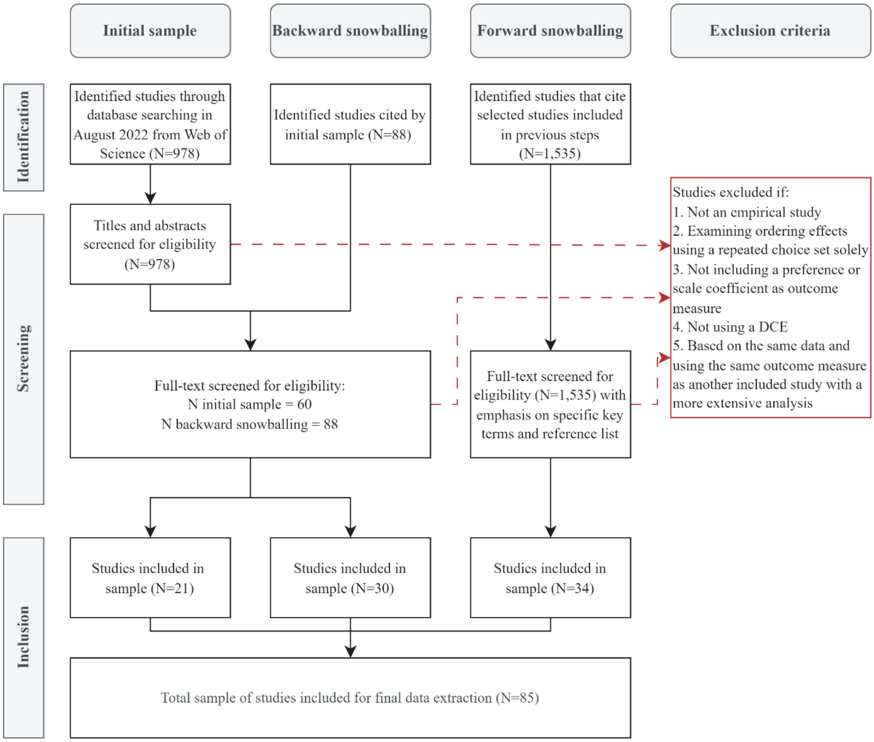
Database selection

As the database for use in the first step of the literature search process, Web of Science was selected. This choice for Web of Science as single database is consistent with other cross-domain literature reviews in the context of DCEs (Haghani et al., 2021a; 2021b; 2022; Mahieu et al., 2017). Relative to those reviews, this study has minimized the potential bias regarding the non-identification of potentially relevant studies resulting from the use of a single database by employing a structured backward and forward snowball sampling procedure. As database for the forward snowballing, Google Scholar was used. Unlike Web of Science and comparable databases, Google Scholar also includes many working papers and conference papers, doctoral dissertations, book chapters and unpublished work. As such, Google Scholar generally finds substantially more citations than other databases (Martín-Martín et al., 2018a; 2018b), which was deemed useful in the forward snowballing phase to expand the study sample and reduce the impact of publication bias. It was not used as the database in the first identification step, however, for its more limited data extraction functionalities (De Winter et al., 2014). This was considered to be a limitation for the first step in particular, given the abstract review in this step and the impossibility of exporting abstracts from Google Scholar.

Literature search and selection process

The literature search and selection process consisted of three sequential snowball-sampling steps (Wohlin et al., 2022), which are graphically depicted in Figure 3.1.

Figure 3.1. Graphical representation of the literature search and selection process



First step

In the first step of the selection process, a search in the Web of Science Core Collection has been performed on titles, abstracts and keywords indexed up until August 25, 2022, using the follow query:

- TS=((“discrete choice experiment” OR “choice experiment” OR “choice based conjoint” OR (“conjoint analysis” AND “choice*”)) AND (((“ordering” OR “order” OR “sequencing” OR “sequence” OR “positioning” OR “position” OR “left-right” OR “left-to-right”) AND (“effect” OR “effects” OR “bias” OR “anomalies” OR “anomaly” OR “attributes” OR “attribute” OR “choice sets” OR “choice set” OR “alternatives” OR “alternative”)) OR (“learning” OR “fatigue” OR “anchoring”)))

This search yielded 978 studies, of which data was extracted from Web of Science. All abstracts were screened for their suitability for this study using the inclusion criteria described in the previous section. Also, in case a study did not explicitly report in the abstract to have estimated any ordering effects, but instead only reported to have randomized the order of alternatives/attributes/choice sets, the study was included by abstract. In total, the abstract screening yielded 60 studies for a full-text screening, while 918 studies were excluded from the review sample. After the full-text screening, 21 of the 60 studies were included in the study sample.[9]

Second step

As the second step of the selection process, all 21 studies included under the first step were screened for references to potentially relevant previous studies (i.e., backward snowballing). All 88 studies identified as potentially relevant were screened on their full text to determine their eligibility. This yielded an additional 30 studies admissible to the review sample, resulting in a provisional sample of 51 included studies.

Third step

Finally, on the 26th and 27th of September 2022, in the third step of the selection process, Google Scholar was used to compose a list of all studies that have cited one or more of the studies included in the provisional sample[10] (i.e., forward snowballing). The resulting 1535 studies have been screened for the words 'order', 'position', 'sequence', 'learn', 'fatigue', 'anchor', and the in-text reference to the study/studies included in the provisional sample. If the search results suggested a study to be potentially eligible, the study's full text was screened and its eligibility examined using the inclusion criteria.

---

9 Two pairs of studies (i.e. four of the studies meeting the inclusion criteria) were based on a single dataset each. In order to avoid duplicates, the most recent study of each pair (McNair et al., 2012; Nguyen et al., 2021) was included in the study, since this expands the analysis of the previous study (McNair et al., 2011; Nguyen et al., 2015) in both cases. Additionally, three other studies (identified in a later step) were also based on a single dataset; Meyerhoff et al. (2015) and Oehlmann et al. (2017) have both examined the impact of choice set position on the scale parameter. Since Oehlmann et al. (2017) expand the previous analysis from Meyerhoff et al. (2015) by also examining the probability of choosing the status quo over the sequence of choices, the former is included while the latter is excluded. The third study based on the same dataset, Mariel and Meyerhoff (2016), has examined the impact of choice set position on status-quo choices and preference parameters. Since the other two studies do not examine the impact of choice set order on preference parameters, Mariel and Meyerhoff (2016) adds another perspective and is also included. Similarly, the study by Lundhede et al. (2009) uses data from both Ladenburg and Olsen (2008) and Jacobsen and Thorsen (2010). However, since the latter two studies focus on the impact on preference estimates and Lundhede et al. (2009) on error variance, all studies are included.

10 Not all studies identified in the first two steps were included in the base set of studies for the forward snowballing, since for some studies, the examination of ordering effects was not the main focus. Forward snowballing for these studies would constitute a substantial additional workload, while arguably yielding very few additional admissible studies for the total sample for final data extraction. Therefore, 18 of the 51 studies identified in the first and second step have been excluded from the forward snowballing sampling (see Appendix 3D).

34 of these studies have been considered eligible to the review sample. All in all, this yielded a total sample of 85 studies.

Data extraction

For all 85 studies in the total sample, data was extracted in a systematic manner with a precomposed (unregistered) form. This form is presented with a filled-out example for one of the included studies in Table A3.3 in Appendix 3C.

Descriptive statistics of included studies

Table 3.2 provides a summary of some descriptive statistics of the included studies. The vast majority of studies is on choice set ordering effects (57 studies, relative to 22 studies for attribute ordering effects and 10 studies for alternative ordering effects). Also, the distribution of included studies is skewed towards more recent publication periods for the total sample as well as for each separate type of ordering effect. Finally, the majority of studies in the overall sample makes use of applications in the environmental domain, which is driven by the subsample of studies on choice set ordering effects. For the subsamples on alternative and attribute ordering effects, the literature is predominantly stemming from the health domain.

Table 3.2. Overview of descriptive statistics of included studies.

Characteristic	Total N (%)	Alternatives N (%)	Attributes N (%)	Choice sets N (%)
Domain				
Environment	39 (46%)	1 (10%)	5 (23%)	33 (57%)
Health	22 (26%)	5 (50%)	12 (55%)	7 (12%)
Marketing	12 (14%)	3 (30%)	4 (18%)	7 (12%)
Transportation	13 (15%)	1 (10%)	1 (5%)	11 (19%)
Publication period				
1990 – 1999	5 (6%)	1 (10%)	4 (18%)	2 (4%)
2000 – 2009	20 (24%)	2 (20%)	6 (27%)	12 (21%)
2010 – 2019	48 (56%)	5 (50%)	9 (41%)	35 (61%)
2020 -	12 (14%)	2 (20%)	3 (14%)	8 (14%)
Total N	85	10	22	57

Please note that some included studies are counted multiple times if they examine more than one type of ordering effect or include multiple datasets of applications in different domains. Due to rounding, percentages may not add up to 100.

Results

Below, the results are presented in a narrative synthesis for each of the three types of ordering effects separately. In the synthesis, studies are grouped together based on their outcome of focus (e.g., marginal utility estimates, error variance, probability of choosing the status quo/opt-out option), their scope (e.g., only examined the existence of an ordering effect, or also tested the effectiveness of a mitigation method) and their findings (e.g., an increasing or decreasing error variance over the sequence, no significant ordering effects, etc.). An overview of all included studies and their findings can be found in Table A3.1 in Appendix 3A, while an overview of some of their design characteristics can be found in Table A3.4 in Appendix 3D.

Alternative ordering effects

The firstly discussed ordering effect is the alternative ordering effect. Ten identified studies have empirically examined alternative ordering effects. While a majority of the studies (n = 7) found statistically significant effects, the evidence regarding the direction of these effects is mixed. One study found that the left-hand alternative has a significantly higher choice probability (Sandorf et al., 2018), two studies found the right-hand alternative to be chosen significantly more often (Gerstenblüth et al., 2022; Krucien et al., 2017b), while a few other studies found significant but inconsistent ordering effects (Chrzan, 1994; Van der Waerden et al., 2006). Three studies found no statistically significant alternative ordering effects (Koç and Van Kippersluis, 2017; Ryan and Bate, 2001; Zhao et al., 2022).

One study has combined choice data with eye-tracking data to examine whether any disparities in visual attention also translate into disparities in choices (Meißner et al., 2016). The study conducted three DCEs, all with at least three alternatives on offer, and found that the middle alternative(s) receive(s) significantly more visual attention in each study. However, in only one of the three studies, this resulted in the middle alternative(s) being chosen significantly more often. In the other two studies, no statistically significant disparities in choice probability between alternatives were documented (Meißner et al., 2016). Finally, Sandorf et al. (2018) examined alternative ordering effects for both a traditional DCE display as well as a transposed matrix, in which alternatives were presented in the rows instead of the columns. They found alternative ordering effects to exist in the traditional matrix display, but not in the transposed display. This suggests that the layout of the choice task may influence the extent to which alternative ordering effects are present (Sandorf et al., 2018).[11]

Attribute ordering effects

The secondly discussed ordering effect is the attribute ordering effect. In the systematic review, 22 studies on attribute ordering effects have been identified, which all experimentally varied the order of (a subset of) attributes within a choice set. A majority of studies (N = 16) has documented statistically significant ordering effects, albeit the nature of the found effect differs by study. One study found an attribute to be significantly more important when presented first, which applied to only some of the attributes (Keshavarzian and Wu, 2021). Three studies found an attribute to be significantly more important when presented last (Glenk, 2007; Kjær et al., 2006; Scott and Vick, 1999). Four studies reported significant effects of the attribute ordering on the error variance (Boyle and Özdemir, 2009; Heidenreich et al., 2021; Krucien et al., 2017a; Mulhern et al., 2019). Interestingly, two of these studies found that the error variance was lower when the cost attribute was positioned as the first attribute (Boyle and Özdemir, 2009; Krucien et al., 2017a). Another study claims that the scale parameters of DCE versions with the cost attribute positioned as first or last are nearly identical, but has not reported any testing of the difference in scale (Kjær et al., 2006). Several studies found either other significant types of attribute ordering effects (Tseng and Lii, 2006) or significant but inconsistent effects (Chrzan, 1994; Kumar and Gaeth, 1991; Logar et al., 2020; Mulhern et al., 2016; Norman et al., 2016a; Soliño et al., 2017; Tsuchiya et al., 2019), some of which are discussed in more detail below. Finally, six studies documented no statistically significant ordering effects (Berchi et al., 2006, 2016; Farrar and Ryan, 1999; Mulhern et al., 2017; Ohdoko and Yoshida, 2012; Sjöstrand, 2001).

While most studies varied the order of attributes between respondents, one study also varied the order of attributes within respondents (between choice sets) for one of their subsamples (Mulhern et al., 2017). Even though the expectation was that this would result in a reduced choice consistency/increased error variance by increasing the cognitive burden for respondents, the authors did not find any significant effect on error variance (Mulhern et al., 2017). Some studies did not randomize all attributes but focused on the impact of the position of one or two attributes, particularly the cost attribute (e.g. Boyle and Özdemir, 2009; Glenk, 2007; Kjær et al., 2006). Whereas

11 It should be noted, though, that Damman et al. (2012) have also used a transposed matrix and they did find statistically significant ordering effects. They did not compare this with a regular matrix display, however.

Glenk (2007) and Kjær et al. (2006) found that respondents' price sensitivity increases (and thus WTP decreases) when the cost attribute was positioned as the last attribute, Boyle and Özdemir (2009) did not find any significant effect on preference estimates. Another study on patients' preferences for insomnia treatments divided its attributes in benefits and risks, grouped these into two blocks and varied the order of presentation of the benefit and risk blocks (Heidenreich et al., 2021). No significant differences were found between the two versions. However, a third version in which the order of attributes was completely randomized (i.e., no use of blocks) resulted in a significantly higher error variance relative to the two versions with a deterministic order (Heidenreich et al., 2021). One study that found a significant attribute ordering effect on error variance also involved eye-tracking (Krucien et al., 2017a). It found the cost attribute to be one of the least visually processed attributes when presented last, and the most visually processed attribute when presented first (Krucien et al., 2017a).

An early study suggested attribute ordering effects to be only present for goods unfamiliar to respondents (Kumar and Gaeth, 1991). However, later studies that did find significant ordering effects for familiar goods (e.g. Keshavarzian and Wu, 2021; Scott and Vick, 1999; Tseng and Lii, 2006) indicate that this suggestion was probably misguided. Finally, a recent study found significant attribute ordering effects on WTP-estimates only when attribute non-attendance was accounted for, suggesting that the presentation order of attributes may affect attribute non-attendance (Logar et al., 2020). Their suggestion is in line with an earlier study that asked respondents to rank a list of attributes in order of importance prior to its choice experiment (Tseng and Lii, 2006), as this study suggested respondents to incorporate more attributes in their choices when less important attributes were presented first.

Choice set ordering effects

Finally, in this review, 57 studies were identified that examined choice set ordering effects. Most studies (N = 40) found (some) significant effects, although these effects relate to different outcomes. Some studies reported the error variance to increase over the sequence of choice sets (e.g. Bradley and Daly, 1994; Maddala et al., 2003; Savage and Waldman, 2008), while some other studies found the error variance to decrease over the sequence (e.g. Czajkowski et al., 2014; Oehlmann et al., 2017; Uggeldahl et al., 2016). Several studies documented both patterns within the same choice experiment, with typically a decrease in error variance in the initial choice tasks and an increase in the final choice tasks (e.g. Balcombe et al., 2015; Campbell et al., 2015; Meyerhoff and Glenk, 2015). Besides, some studies have focused on the probability of the status quo/

opt-out alternative being selected over the sequence of choice tasks, again resulting in mixed findings. Two studies found the status quo/opt-out alternative to be less likely selected in choice sets later in the sequence (Nguyen et al., 2021; Scheufele and Bennett, 2012), while more studies found the opposite (e.g. Boxall et al., 2009; Petrolia et al., 2018; Oehlmann et al., 2017; Swait and Adamowicz, 2001; Weng et al., 2021). Furthermore, some studies have focused on the effect of choice task sequence on attribute importance (e.g. Cao et al., 2018; Crastes dit Sourd et al., 2020), with as most prevalent finding that willingness-to-pay for an alternative decreased, and thus cost sensitivity increased, if a previous choice set contained a more attractive alternative (better quality against a similar or lower price or similar quality against a lower price) (e.g. Day and Pinto-Prades, 2010; Groeneveld, 2010; Ladenburg and Olsen, 2008; McNair et al., 2012; Scheufele and Bennett, 2012). Finally, some studies found no significant choice set ordering effects (e.g. Brouwer et al., 2010; Dardanoni and Guerriero, 2021; Hensher and Collins, 2011; Hole, 2004; Lanz and Provins, 2013; Oppewal et al., 2010).

The documented findings lend themselves to different suggestions as to which mechanisms are driving the ordering effects. Even though a decreasing error variance is typically interpreted as a result of learning, this does not necessarily need to be the case (Czajkowski et al., 2014; Oehlmann et al., 2017). In contrast, it may also be the consequence of respondents making more use of non-compensatory choice heuristics over the sequence of choice sets (e.g., due to fatigue) (Swait and Adamowicz, 2001). Besides, it should be noted that not all studies examining learning and fatigue have varied the order of choice sets between respondents. In the absence of random variation, learning and fatigue patterns may be confounded with variation in utility balance (i.e., choice task difficulty) between choice sets (Abate et al., 2018).

Some studies have also included additional experimental elements or performed additional analyses to examine heterogeneity in ordering effects. For example, in an attempt to provide more clarity on the proposed mechanisms of learning and fatigue, some of the included studies have also elicited respondents' self-reported choice certainty regarding each of their choices, to mixed results. Brouwer et al. (2010) found a significant increase in self-reported choice certainty over the sequence of choice sets, but no significant effect on preference or scale parameters. Uggeldahl et al. (2016) reported a significantly decreasing self-reported level of choice certainty over the sequence, together with a significantly decreasing error variance. Finally, Logar and Brouwer (2017) reported no significant effect of choice set order on self-reported choice certainty, but they did find a statistically significant decrease in error variance over the sequence of choice sets in at least one of their two subsamples. As another topic, to examine the influence of survey administration mode, Savage and Waldman (2008) compared a choice experiment administered online and on paper. They found a significant increase in error variance only in the online survey administration mode and suggested fatigue to be more prevalent in online choice experiments.

Four studies have examined heterogeneity in the prevalence of choice set ordering effects between respondents. Campbell et al. (2015) have employed a probabilistic decision process model, in which respondents are probabilistically assigned to classes on the basis of their learning or fatigue patterns. They found that only a minority of respondents in their sample (about 10%) showed inconsistent preferences or error variance between different phases of the choice sequence (Campbell et al., 2015), but they have not further examined the characteristics of these respondents. Nguyen et al. (2021) distinguished between strategic and non-strategic respondents, in which strategic respondents did not believe in the payment consequentiality of the choice experiment. They found a sharp decrease in willingness-to-pay for policy alternatives over the choice sequence for strategic respondents, but a rather stable choice pattern for non-strategic respondents. Regarding starting point bias, two studies found this effect to be only significant among women (Ladenburg and Olsen, 2008; Ladenburg, 2013). On the contrary, another study documented men to be more susceptible to starting point bias (Bechtold and Abdulai, 2012). Additionally, Ladenburg (2013) found starting point bias to be significant only among respondents with less experience with the topic (those without children in a choice experiment on lunch programs in kindergartens).

In order to reduce, or perhaps even offset, choice set ordering effects, some studies have tested the effectiveness of various potential mitigation methods. For instance, Day et al. (2012) suggested that advance disclosure of the set-up of the choice experiment (e.g., informing respondents about the number of choice sets and that attribute levels will randomly vary) mitigates ordering effects by inducing institutional learning and anchoring prior to the start of the choice experiment. Another way to induce learning (and anchoring) before the start of the preference elicitation task is including an instructional choice set (ICS), also called 'warm-up task'. In an ICS, respondents are presented with an exemplary choice set, enabling them to become familiar with the choice environment and topic before they start making choices in the 'actual' choice sets. An ICS has been found to significantly affect the preference structure (Abate et al., 2018) but not the scale parameters (Abate et al., 2018; Meyerhoff and Glenk, 2015). Finally, two studies have examined the potential mitigating impact of attribute level overlap, i.e., when some of the attributes in a choice set have the same level for all the alternatives on offer, on choice set ordering effects. While Maddala et al. (2003) found no significant effect, a more recent study by Jonker et al. (2018a) did find a significant effect of attribute level overlap on error variance in the first choice tasks; attribute level overlap is suggested to take away learning effects at the start of the choice experiment.

Conclusion and Discussion

Summary of main findings

In this study, the literature on alternative, attribute and choice set ordering effects in choice experiments has been systematically reviewed. Regarding alternative ordering effects, seven of the ten identified studies found a statistically significant ordering effect, but with mixed findings regarding the direction of this effect. Additional findings include that the order of presentation of alternatives may significantly affect visual attention without necessarily affecting choices (Meißner et al., 2016) and that any alternative ordering effect may be reduced in a transposed matrix display with the alternatives presented in rows instead of columns (Sandorf et al., 2018). Regarding attribute ordering effects, again a majority of the 22 included studies found significant ordering effects. Most of the studies focused on the impact of the presentation order of attributes on preferences, but some studies have focused on the impact on error variance. Some interesting supplementary findings include that the complete randomization of attributes may increase error variance in case there is a natural grouping of attributes (Heidenreich et al., 2021) and that the presentation order of attributes may affect attribute non-attendance (Logar et al., 2020).

Finally, most studies included in this review were on choice set ordering effects, the majority of which reported statistically significant effects. Their results provide evidence for learning, fatigue, and anchoring as the underlying mechanisms, sometimes together in a single study. The findings include a dynamic error variance over the sequence of choice sets, a varying probability of the status quo/opt-out alternative being selected over the sequence of choices, and heterogeneous price sensitivity depending on the price levels presented in earlier choice sets. One study found ordering effects to be prevalent for only a minority of respondents (Campbell et al., 2015), while other studies suggest the effects to vary by gender (Bechtold and Abdulai, 2012; Ladenburg and Olsen, 2008; Ladenburg, 2013), experience with the topic (Ladenburg, 2013), and believe in the payment consequentiality of the choice experiment (Nguyen et al., 2021). All in all, the results are mixed for the three types of ordering effects; the substantial number of studies with null findings prevent us from drawing any definitive conclusions regarding the existence of ordering effects. Given that the majority of studies do find significant ordering effects, however, the applied literature is recommended to take mitigative measures. The final section of this paper provides both an overview of such mitigative measures as well as suggestions for further methodological research.

Limitations of this study

Despite the rigorous approach of this review, it may not qualify as systematic in all of its aspects (Haddaway et al., 2020). This particularly applies to the absence of a pre-registered protocol and the fact that the literature search and selection process and data extraction were conducted entirely by a single researcher. Even though this is not without precedence, the latter aspect has resulted in a lack of cross-validation (Haddaway et al., 2020; Stoll et al., 2019). Nevertheless, in several cases of doubt regarding the inclusion of particular studies, colleagues with DCE expertise were consulted. Also, the systematic review process is documented in a transparent manner to facilitate reproduction by other researchers.

Besides, this study comes with several other limitations. Firstly, it may well be possible that not all studies that have empirically examined one of the three ordering effects have been included in this review. Particularly, it may be that the choice to exclude the references to some of the studies included in the first and second step from the review in the third step, as described in the section 'Literature search and selection process', may have contributed to the exclusion of relevant studies. It should be emphasized, however, that the snowball sampling strategy used already led to a very extensive literature scanning process in its current form. Expanding the third step would have been possible, but would have led to a substantial increase in the number of studies to be assessed, arguably against an only modest gain in eligible studies. Secondly, this study focuses on alternative, attribute and choice set ordering effects exclusively, while there are also other types of ordering effects in stated choice surveys that have been studied to a smaller extent. For instance, some studies have examined the effect of positioning supplementary (attitudinal) questions before or after the choice experiment (Cai et al., 2011; Liebe et al., 2016) or of the order of elicitation methods in case of combining a choice experiment with another stated preference elicitation method like Contingent Valuation (e.g. Brouwer et al., 2017; Meldrum et al., 2020; Metcalfe et al., 2012).

Thirdly, like in all literature reviews, one should consider the possibility of publication bias. That is, researchers (and reviewers and editors of journals alike) prefer statistically significant results and, as such, studies that found significant ordering effects may have been published disproportionally often, while studies with null findings may have ended up 'in the file drawer' (Franco et al., 2014; Stanley, 2005). Relatedly, p-hacking is a well-known phenomenon of authors specifying their data or adjusting their analyses until their results have become statistically significant (Head et al., 2015). Considering the risk of publication bias, the search engine Google Scholar has been included deliberately in the search strategy, resulting in the inclusion of several unpublished working papers. Besides, several studies included in this review have examined ordering effects only as a side-issue, while the focus in their studies was on the topic of the choice experimental task (e.g. Dardanoni and Guerriero, 2021; Hole, 2004; Krucien et al., 2017b; Mulhern et al., 2019; Ohdoko and Yoshida, 2012). This arguably provides more room for null findings to be published. Also, several included studies that did focus on ordering effects and reported null findings have nevertheless been published in high-quality scientific journals (e.g. Farrar and Ryan, 1999; Mulhern et al., 2017). Nevertheless, we have not focused on the existence of publication bias and p-hacking in the ordering effects literature and, therefore, do not rule out the possibility of these mechanisms affecting the study findings. Future research could provide more insights into this in the context of ordering effects using meta-regression analytical methods like funnel plots and meta-significance testing (Stanley, 2005).

Implications for research

The applied literature on choice modelling can profit from the insights of this study. While it cannot be concluded that ordering effects are always present, given the heterogeneity in findings and the limitations of this review outlined above, it is at least safe to conclude that ordering effects may seriously bias the estimates of a choice experiment if not adjusted for. Yet, many recent studies in the applied choice modelling literature still do not report whether the order of alternatives, attributes and choice sets was varied between respondents. Other studies report that the order was fixed without justifying this, and some even justify the fixed order by referring to one of the included studies with null findings. Given the mixed evidence and the number of studies in this review finding significant ordering effects, justification based on a single study seems misguided.

In practical terms, for the sake of transparency, it is recommendable for DCE studies to report whether the order of alternatives, attributes and choice sets in their study was varied between respondents and to argue why this was (not) done. It is advisable to randomize the order of these elements in a choice experiment and, in most cases, there does not seem to be any harm in doing so, while it may prevent biased estimates due to ordering effects. There are cases, however, when a randomized order may not be feasible. As Norman et al. (2016a) point out, there may be cases in which there is a natural ordering of attributes,[12] in which case presenting the attributes in that order may help increase the acceptability of the choice experiment and reduce the cognitive burden on respondents. Also, there may be groups of variables that are more logically grouped together, such as the benefits and the risks or costs of an intervention (Heidenreich et al., 2021). In such instances, the order of attributes can be randomized within groups (rather than entirely randomly) and the order of groups can be randomized, too, between respondents. Secondly, regarding alternative ordering effects, studies should include alternative-specific constants (ASCs) for all alternatives (minus one), including but not limited to any status quo/opt-out alternative. If the ASC is not included in the model, any alternative ordering effects would be captured in the preference parameters and thereby potentially bias these parameters and resulting welfare estimates. Therefore, it is recommendable to include the ASC in the estimated choice model as a simple way to adjust for any alternative ordering effects. Its coefficient, however, should not be interpreted for any purpose other than the analysis of alternative ordering effects (at least in case of an unlabelled choice experiment) and should not be considered in any post-hoc derivations. Finally, several complementary measures to mitigate choice set ordering effects have been proposed and future studies can consider implementing. These measures include advance disclosure of the specific elements of the choice tasks (e.g., the full attribute level ranges and the number of choice tasks) (Day et al., 2012), the presentation of an instructional choice set (Abate et al., 2018), and the incorporation of attribute level overlap (Jonker et al., 2018a). Table 3.3 provides an overview of the various mitigation methods identified in this review.

12 A similar argument can be provided for the ordering of alternatives.

Table 3.3. Overview of the methods mitigating ordering effects identified in the literature review

Ordering effect	Mitigation method	Remarks
Alternative ordering effects	Alternative-Specific Constants (ASCs)	Include for N-1 alternatives (including the opt-out/status quo alternative)
	Randomizing the order of alternatives between respondents	
Attribute ordering effects	Randomizing the order of attributes between respondents	Be mindful of any natural ordering or grouping of attributes
Choice set ordering effects	Randomizing the order of choice sets between respondents	
	Advance disclosure of core elements of the choice tasks	Core elements may include the number of choice sets and an overview of all attributes and (randomly varying) levels
	Instructional Choice Set (ICS)	Randomize the attribute levels in the ICS to prevent the introduction of starting point bias
	Attribute level overlap	

Finally, in selecting the number of alternatives, attributes, and choice sets for their DCE design, researchers may consider the potential for ordering effects. The impact of respondent fatigue, for example, is arguably larger with a larger number of alternatives, attributes, and choice sets. Therefore, some argue for the use of a single binary choice format to reduce complications from inconsistent choice behaviour over the sequence of choice sets and strategic voting (Johnston et al., 2017; Mariel et al., 2021). This is not the only relevant consideration in such design choices, however; for example, in some choice contexts, offering more than two alternatives per choice set is more compatible with the real-life choice situation (e.g., food choice, mode of transportation choice, etc.). Also, sometimes the potential sample size is limited due to scarcity of data collection resources or a small study population (e.g., patients with a rare disease), so that multiple choice sets per respondent are required to accomplish sufficient statistical power. Thus, researchers should trade-off criteria related to statistical efficiency, incentive compatibility, and realism in selecting the number of alternatives and attributes per choice set and the number of choice sets per respondent[13] (Johnston et al., 2017; Mariel et al., 2021), and are strongly encouraged to mitigate ordering effects by adopting (some of) the methods presented in Table 3.3.

13 The focus of this review was not on the convergent validity of designs with varying numbers of alternatives, attributes, or choice sets. The reader is referred to, for instance, Bech et al. (2011), Caussade et al. (2005), Dellaert et al. (2012), Hensher (2006), Meyerhoff et al. (2015), Oehlmann et al. (2017), Weng et al. (2021), and Zhang & Adamowicz (2011), which have all assessed the impact of varying one or more of these three design dimensions.

With respect to the methodological literature on ordering effects, there are several ways in which future research may provide new insights. Firstly, future studies may further explore the extent to which there is heterogeneity in the effects of presentation order on respondents' choices, both with respect to study design traits and respondents' characteristics. For example, regarding the role of the study design in explaining the mixed findings of the included studies, the finding by Sandorf et al. (2018) that the matrix display of alternatives and attributes influences the existence of alternative ordering effects deserves further inquiry. Moreover, there is reason to believe that the complexity of the choice experiment design (e.g., the number of choice sets per respondent, the number of alternatives and attributes per choice set, the psychological distance between the topic of the study and respondents) may affect ordering effects. For example, Meyerhoff et al. (2015) find that the magnitude of the rise in error variance over the sequence of choice sets increases in the number of alternatives per choice set. Even though some of the design characteristics of included studies are listed in Table A3.1 in Appendix 3A and in Table A3.4 in Appendix 3D, this paper cannot provide conclusions on the role of these characteristics in the heterogeneity of results. A meta-analysis based on this systematic literature review could provide further insights.

Regarding respondent heterogeneity in ordering effects, Campbell et al. (2015) found that only a subset of their respondents was prone to learning and fatigue. Unfortunately, they have not yet related this to respondents' background characteristics. This would make an interesting avenue for further research. For example, it may be an interesting avenue for further research to examine to what extent respondents' perceptions of the consequentiality of the choice experiment influence the presence of ordering effects. Consequentiality is deemed to be crucial for the elicitation of truthful preferences (e.g. Vossler et al., 2012); respondents who believe their answers to the DCE have consequences in terms of a payment they have to make (i.e., payment consequentiality) or in terms of any policy actions implemented by the government (i.e., policy consequentiality) have more incentive to make well-considered choices in the DCE. Arguably, this might provide less room for ordering effects. Nguyen et al. (2021) found significant differences in choice set ordering effects by respondents' perceptions of the payment consequentiality of the DCE, and it would be interesting to extend this to policy consequentiality and alternative and attribute ordering effects. Furthermore, Ladenburg (2013) found that starting point bias is present only for respondents with less experience with/knowledge of the topic of the choice task. A question that arises is whether such an 'experience/knowledge' effect is also present for alternative and attribute ordering effects. For instance, since respondents' background knowledge has been found to affect attribute (non-)attendance (Sandorf et al., 2017) and the presentation order of attributes may in turn affect attribute non-attendance (Logar et al., 2020), one could expect a moderating role of respondents' knowledge/experience in attribute ordering effects.

Finally, future research could examine whether there are any interactions between the different types of ordering effects. For instance, alternative and attribute ordering effects may be more prevalent towards the end of the choice sequence, when respondents learn about their preferences and the choice setting or, alternatively, when fatigue kicks in and respondents start making use of simplifying choice heuristics. Swait and Adamowicz (2001) found indeed that respondents seem to switch to a simpler choice strategy over the sequence of choice sets. Likewise, studies by Meißner et al. (2016) and Orquin et al. (2013) using choice experiments and Li et al. (2016) using another type of preference elicitation task found a decrease in the average number of visual fixations over the sequence of choice sets. Combined, these findings give reason to suspect that an interaction between choice set ordering effects and alternative or attribute ordering effects may indeed be in place. Information on each of these aspects would help us to better understand respondents' processing of choice experiments and may contribute to the improvement of choice experimental designs and the validity of their findings.

Appendix 3A –Overview of studies

Table A3.1. Overview of studies included in the systematic literature review.

Study	Type	Domain	Exp. Treatment	Effect
Chrzan (1994)*	AL	M	5 different orderings of alternatives	Sign. but inconsistent effects on marginal utilities
Damman et al. (2012)	AL	H	2 versions with alternatives ordered either alphabetically or by performance	Sign. higher probability of best-performing alternative being chosen when ranked in alphabetical order (rather than ranking by performance)
Gerstenblüth et al. (2022)	AL, CS	H	2 opposite orderings of alternatives, randomized order of choice sets	Sign. higher probability of right-hand alternative to be selected. No sign. choice set ordering effect
Koç & Van Kippersluis (2017)	AL	H	-	No sign. effect
Krucien et al. (2017b)	AL, CS	H	-	Sign. higher probability of right alternative to be selected. Sign. decrease in error variance over sequence of choice sets
Meißner et al. (2016)	AL	M	Randomization of alternatives	Sign. higher probability of middle alternative to be selected in laptop study, but no sign. effects in other studies
Ryan & Bate (2001)	AL	H	2 different orderings of alternatives	No sign. effect
Sandorf et al. (2018)	AL	M	-	Sign. higher probability for left alternative to be chosen, but only in regular display (not in transposed display)
Van der Waerden et al. (2006)	AL	T	6 different orderings of alternatives, 10 orderings of choice sets	Yes, sign. effect of alternative order, but inconsistent between trip purpose
Zhao et al. (2022)	AL	E	Random position of alternatives	No sign. effect
Berchi et al. (2006)	AT	H	2 different orderings of attributes	No sign. effect
Berchi et al. (2016)	AT	H	Randomized order of attributes	No sign. effect
Boyle & Özdemir (2009)	AT	E	2 versions; one with cost attribute as last attribute and other with cost attribute as the first attribute	Sign. lower error variance (in unforced choice) when cost attribute is placed first. No sign. effect on preference coefficients and welfare estimates

Table A3.1. Overview of studies included in the systematic literature review. *(Continued)*

Study	Type	Domain	Exp. Treatment	Effect
Chrzan (1994)*	AT	M	2 different orderings of attributes	Sign. but inconsistent effects on marginal utilities
Farrar & Ryan (1999)	AT	H	2 versions, upper two and lower two attributes switched	No sign. effect
Glenk (2007)	AT	E	2 versions, one with cost attribute as first attribute and one with cost attribute as last attribute	Sign. higher sensitivity to attribute (either cost or anao) if presented as the bottom attribute
Heidenreich et al. (2021)	AT	H	3 orderings: benefits before risks, risks before benefits, entirely random	No sign. effect on preference estimates, but sign. higher error variance for entirely random order
Keshavarzian & Wu (2021)	AT	M	2 versions with groups of destination and airline attributes reversed	Sign. increase in marginal utility when group of attributes is positioned first, but only for some of the attributes
Kjær et al. (2006)	AT	H	Cost attribute either first or last	Sign. lower WTP when price attribute as last
Krucien et al. (2017)	AT	H	2 versions with opposite attribute ordering	Sign. lower error variance when cost attribute is placed first
Kumar & Gaeth (1991)	AT	M	3 different orderings of attributes; 2 versions varying between respondents in opposite order, 1 version varying within respondents	Sign. differences between versions for burglar alarm choice, but no sign. differences for tv choice
Logar et al. (2020)	AT	E	2 (opposite) attribute orders	No sign. effect. However, when accommodating ANA, sign. diff. in WTP estimates
Mulhern et al. (2016)	AT	H	3 different orderings of attributes	Sign. but inconsistent differences between 3 versions
Mulhern et al. (2017)	AT	H	1 version with fixed order, 1 with randomized order between respondents, 1 with randomized order within respondents	No sign. effect
Mulhern et al. (2019)	AT	H	Attributes grouped into 2 blocks, blocks presented in opposite orderings to both halves of the sample	Sign. differences in scale between different orderings
Norman et al. (2016a)	AT	H	Randomized attribute order, except life expectancy always as last attribute	No sign. effect, except for two of the 30 additional order of presentation coefficients
Ohdoko & Yoshida (2012)	AT	E	4 different orders of the 4 non-price attributes	No sign. effect

Table A3.1. Overview of studies included in the systematic literature review. *(Continued)*

Study	Type	Domain	Exp. Treatment	Effect
Scott & Vick (1999)	AT	H	2 versions with opposite attribute ordering	The attribute 'being able to talk to the doctor' was significantly more important when presented as the last attribute than as the first attribute.
Sjöstrand (2001)	AT	T	4 different orderings of attributes	No sign. effect
Soliño et al. (2017)	AT	E	Price attribute presented as first, third, or last of six attributes	No sign. differences in price sensitivity between different orderings, but significant differences for some of the other attribute coefficients, resulting in differences in WTP. Only when cost attribute is positioned last, all attribute coefficients are significant
Tseng & Lii (2006)	AT	M	3 different orderings of attributes	Sign. differences in preferences between different versions
Tsuchiya et al. (2019)	AT	H	5 different orderings of attributes	Sign. but inconsistent effects on marginal utilities
Abate et al. (2018)	CS	E	-	Sign. decrease in error variance initially, and sign. increase later in choice sequence, and then decrease
Achtnicht (2012)	CS	T	-	No sign. effect
Arentze et al. (2003)	CS	T	2 opposite orderings of blocks of choice sets, randomized order of choice sets within blocks	No sign. effect
Balcombe et al. (2015)	CS	E	-	Sign. decrease in error variance initially, and sign. increase in error variance in second half of sequence
Bansback et al. (2014)	CS	H	-	Sign. higher error variance in later groups of choice sets in the sequence
Bech et al. (2011)	CS	H	random blocking of choice sets	Sign. higher error variance in last block of 4 choice sets for respondents with 16 choice sets
Bechtold & Abdulai (2012)	CS	E	2 versions, one with high cost levels in first choice set and other with low cost levels in first choice set	Sign. higher WTP when presented with high cost levels in first choice set

Table A3.1. Overview of studies included in the systematic literature review. *(Continued)*

Study	Type	Domain	Exp. Treatment	Effect
Björklund & Swärdh (2017)	CS	T	randomized order of choice sets	No sign. difference in preference coefficients between first and second block of four choice sets
Boxall et al. (2009)	CS	E	8 different blocks (woodland) and 4 blocks (forest management)	Sign. increase in probability of SQ being chosen over sequence of choice sets
Bradley & Daly (1994)	CS	T	-	Sign. increase in error variance over sequence of choice sets
Brouwer et al. (2010)	CS	E	Randomized order of choice sets	No sign. difference in error variance between choice sets
Campbell et al. (2015)	CS	E	randomized order of choice sets	Sign. decrease in error variance in the beginning of the sequence, sign. increase towards the end of the sequence, and sign. but inconsistent effects of choice set order on preference estimates. However, all effects sign. only for a subset of respondents.
Cao et al. (2018)	CS	E	-	Sign. differences in preference parameters and WTP between choice sets
Carlsson & Martinsson (2001)	CS	E	Randomized order of two blocks of 8 choice sets each	Sign. higher error variance in second block of 8 choice tasks
Caussade et al. (2005)	CS	T	-	Sign. decrease in error variance initially, and sign. increase in error variance eventually
Chrzan (1994)*	CS	M	Different starting point each time, but same relative sequence of choice sets	Sign. but inconsistent effects on preference coefficients
Crastes dit Sourd et al. (2020)	CS	T, E	Randomized order of choice sets	Sign. and inconsistent differences in preference coefficients between first choice set and subsequent ones in all four data sets, depending on model specification
Czajkowski et al. (2012)	CS	E	Randomization of choice set order	Sign. decrease in error variance in first choice tasks
Dardanoni & Guerriero (2021)	CS	E	-	No sign. effect
Day & Pinto-Prades (2010)	CS	H	6 orderings of choice sets	Sign. decrease in probability of alternative being chosen if cheaper or better alternative was presented in previous choice task

Table A3.1. Overview of studies included in the systematic literature review. *(Continued)*

Study	Type	Domain	Exp. Treatment	Effect
Day et al. (2012)	CS	E	One version presenting extreme levels in first choice set, second version intermediate levels first	Sign. increase in probability of SQ being chosen after first choice set, and in further sequence only for subsample without advance disclosure. Sign. and inconsistent effect of choice set order on preference coefficients for one of the attributes, but only in subsample without advance disclosure
Groeneveld (2010)	CS	E	1 version with low levels of cost attribute in first choice set and other version with high levels first	Sign. higher WTP when presented with high cost levels first
Hensher & Collins (2010)	CS	T	Randomized order of choice sets	No sign. effect
Hess et al. (2012)	CS	T	Randomized order of choice sets	In some data sets, sign. decrease in error variance over the sequence of choices sets, in some data sets no sign. effect. At most minimal impact on WTP
Hildebrand et al. (2023)	CS	M	4 different blocks of choice sets	No sign. effect
Hole (2004)	CS	T	-	No sign. effect
Jacobsen & Thorsen (2010)	CS	E	3 orderings; in 1 highest cost levels in first choice set, in 2 cheapest cost levels in first choice set, in 3 entirely randomized	Sign. higher WTP when presented with highest cost levels in first choice set
Jarvis (2011)	CS	E	-	No sign. effect
Jiang et al. (2022)	CS	E	6 different blocks of choice sets	No sign. effect
Jonker et al. (2018a)	CS	H	Randomized order of choice sets	Sign. higher error variance in first block of three choice sets relative to later blocks
Koppelman & Sethi (2005)	CS	T	-	Sign. higher error variance in second choice set relative to first choice set. Also, sign. increase in error variance over the sequence, but only for subsample of existing train users (who were also presented with a different design than non-users).
Ladenburg & Olsen (2008)	CS	E	2 versions, with one version having higher price levels in ICS than other version	Sign. higher WTP when presented with highest cost levels in ICS. Effect only prevalent among women

Table A3.1. Overview of studies included in the systematic literature review. *(Continued)*

Study	Type	Domain	Exp. Treatment	Effect
Ladenburg (2013)	CS	E	2 versions, one with high cost levels in ICS and other with low cost levels	Sign. higher WTP when high cost levels presented in ICS, but only among women and only for respondents without children
Lanz & Provins (2013)	CS	E	randomized order of blocks and choice sets within blocks	No sign. effect
Lanz & Provins (2015)	CS	E	randomized order of blocks	Sign. increase in probability of SQ being chosen in last choice set relative to first choice set
Logar & Brouwer (2017)	CS	E	-	Sign. lower error variance in last choice set relative to first, at least in one of the two subsamples
Lundhede et al. (2009)	CS	E	3 different blocks (new motorways), 4 different blocks (national parks	Sign. increase in error variance over sequence of choice sets in new motorways study, no sign. effect in national parks study
Maddala et al. (2003)	CS	H	-	Sign. higher error variance in second half of choice sets
Mariel & Meyerhoff (2016)	CS	E	Randomized order of choice sets	Sign. increase in probability of SQ being chosen over sequence of choice set. Also sign. but inconsistent effect of choice set order on preference coefficients for two of the attributes
Marsh & Phillips (2012)	CS	E	-	Sign. decrease in error variance over the sequence of choice sets
McNair et al. (2012)	CS	E	-	Sign. decrease in probability of alternative being chosen if cheaper alternative was presented in previous choice set
Meyerhoff & Glenk (2015)	CS	E	5 versions, 1 without ICS and 4 with ICS with high or low levels for price attribute and for water quality attribute	Sign. decrease in error variance after the first choice task. Sign. increase in error variance towards the end of the sequence under some of the treatments. Sign. lower WTP when presented with an ICS with either levels indicating high quality and low price or low quality and high price relative to subsample without ICS
Mokas et al. (2021)	CS	E	randomized order of choice sets	No sign. effect
Nguyen et al. (2021)	CS	E	Randomization of choice set order	Sign. higher probability of SQ being chosen in second half of the sequence. Sign. decrease in WTP, but only for strategic respondents

Table A3.1. Overview of studies included in the systematic literature review. *(Continued)*

Study	Type	Domain	Exp. Treatment	Effect
Oehlmann et al. (2017)	CS	E	Randomized order of choice sets	Sign. increase in probability of choosing SQ over the sequence of choice sets. Sign. decrease in error variance over sequence of choice sets
Oppewal et al. (2010)	CS	M	-	No sign. effect
Petrolia et al. (2018)	CS	E	6 different blocks of choice sets	Sign. higher probability of the SQ being chosen in the third and fourth choice set relative to the first
Sælensminde (2001)	CS	T	randomized order of choice sets	No sign. difference in error variance for most choice sets relative to the first choice set. Sign. higher error variance for fourth and fifth choice set (so halfway the sequence) relative to first choice set
Savage & Waldman (2008)	CS	M	-	Sign. higher error variance in second half of DCE, but only for online survey mode
Scheufele & Bennett (2012)	CS	E	4 different versions of choice set order	Sign. decrease in WTP over sequence of choice sets. Sign. lower probability of SQ being selected in first choice set
Swait & Adamowicz (2001)	CS	M	Different blockings of choice sets	Yes, sign. increase in probability of SQ being chosen and simpler decision heuristics being employed in second half of the DCE
Uggeldahl et al. (2016)	CS	E	Randomized order of choice sets	Sign. decrease in error variance over the sequence of choice sets
Weng et al. (2021)	CS	E	-	Sign. increase in probability of SQ being chosen over sequence of choice sets in designs with 2 or 3 alternatives + SQ. No sign. ordering effect in design with 1 alternative + SQ
Zhang & Adamowicz (2011)	CS	E	8 different blocks	No sign. effect

*Note: *) This study included data from three datasets, used for the analysis of a different type of ordering effect each. Abbreviations: AL=Alternatives, AT=Attributes, CS=Choice Sets, E=Environment, Exp.=Experimental, H=Health, ICS=Instructional Choice Set, M=Marketing, Sign.=Significant(ly) (at the 95% level), T=Transportation*

Appendix 3B – PRISMA checklist

Table A3.2: PRISMA 2020 reporting checklist

Section and Topic	Item #	Checklist item	Reported in section(s) #
TITLE			
Title	1	Identify the report as a systematic review.	Introduction
ABSTRACT			
Abstract	2	See the PRISMA 2020 for Abstracts checklist.	Abstract
INTRODUCTION			
Rationale	3	Describe the rationale for the review in the context of existing knowledge.	Introduction
Objectives	4	Provide an explicit statement of the objective(s) or question(s) the review addresses.	Introduction
METHODS			
Eligibility criteria	5	Specify the inclusion and exclusion criteria for the review and how studies were grouped for the syntheses.	Inclusion criteria, Figure 3.1
Information sources	6	Specify all databases, registers, websites, organisations, reference lists and other sources searched or consulted to identify studies. Specify the date when each source was last searched or consulted.	Database selection, Literature search and selection process
Search strategy	7	Present the full search strategies for all databases, registers and websites, including any filters and limits used.	Literature search and selection process
Selection process	8	Specify the methods used to decide whether a study met the inclusion criteria of the review, including how many reviewers screened each record and each report retrieved, whether they worked independently, and if applicable, details of automation tools used in the process.	Limitations of this study
Data collection process	9	Specify the methods used to collect data from reports, including how many reviewers collected data from each report, whether they worked independently, any processes for obtaining or confirming data from study investigators, and if applicable, details of automation tools used in the process.	Limitations of this study

Table A3.2: PRISMA 2020 reporting checklist *(Continued)*

Section and Topic	Item #	Checklist item	Reported in section(s) #
Data items	10a	List and define all outcomes for which data were sought. Specify whether all results that were compatible with each outcome domain in each study were sought (e.g. for all measures, time points, analyses), and if not, the methods used to decide which results to collect.	Table A3.3
	10b	List and define all other variables for which data were sought (e.g. participant and intervention characteristics, funding sources). Describe any assumptions made about any missing or unclear information.	Table A3.3
Study risk of bias assessment	11	Specify the methods used to assess risk of bias in the included studies, including details of the tool(s) used, how many reviewers assessed each study and whether they worked independently, and if applicable, details of automation tools used in the process.	Limitations of this study, NA
Effect measures	12	Specify for each outcome the effect measure(s) (e.g. risk ratio, mean difference) used in the synthesis or presentation of results.	Inclusion criteria, Results
Synthesis methods	13a	Describe the processes used to decide which studies were eligible for each synthesis (e.g. tabulating the study intervention characteristics and comparing against the planned groups for each synthesis (item #5)).	Results
	13b	Describe any methods required to prepare the data for presentation or synthesis, such as handling of missing summary statistics, or data conversions.	NA
	13c	Describe any methods used to tabulate or visually display results of individual studies and syntheses.	Results
	13d	Describe any methods used to synthesize results and provide a rationale for the choice(s). If meta-analysis was performed, describe the model(s), method(s) to identify the presence and extent of statistical heterogeneity, and software package(s) used.	Results
	13e	Describe any methods used to explore possible causes of heterogeneity among study results (e.g. subgroup analysis, meta-regression).	NA
	13f	Describe any sensitivity analyses conducted to assess robustness of the synthesized results.	NA
Reporting bias assessment	14	Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting biases).	NA
Certainty assessment	15	Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome.	NA

Table A3.2: PRISMA 2020 reporting checklist *(Continued)*

Section and Topic	Item #	Checklist item	Reported in section(s) #
RESULTS			
Study selection	16a	Describe the results of the search and selection process, from the number of records identified in the search to the number of studies included in the review, ideally using a flow diagram.	Literature search and selection process, Figure 3.1
	16b	Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were excluded.	Inclusion criteria
Study characteristics	17	Cite each included study and present its characteristics.	Table A3.1, Table A3.4
Risk of bias in studies	18	Present assessments of risk of bias for each included study.	NA
Results of individual studies	19	For all outcomes, present, for each study: (a) summary statistics for each group (where appropriate) and (b) an effect estimate and its precision (e.g. confidence/credible interval), ideally using structured tables or plots.	Table A3.4/NA
Results of syntheses	20a	For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies.	NA
	20b	Present results of all statistical syntheses conducted. If meta-analysis was done, present for each the summary estimate and its precision (e.g. confidence/credible interval) and measures of statistical heterogeneity. If comparing groups, describe the direction of the effect.	NA
	20c	Present results of all investigations of possible causes of heterogeneity among study results.	NA (see Implications for research)
	20d	Present results of all sensitivity analyses conducted to assess the robustness of the synthesized results.	NA
Reporting biases	21	Present assessments of risk of bias due to missing results (arising from reporting biases) for each synthesis assessed.	NA
Certainty of evidence	22	Present assessments of certainty (or confidence) in the body of evidence for each outcome assessed.	NA

Table A3.2: PRISMA 2020 reporting checklist *(Continued)*

Section and Topic	Item #	Checklist item	Reported in section(s) #
DISCUSSION			
Discussion	23a	Provide a general interpretation of the results in the context of other evidence.	Summary of results
	23b	Discuss any limitations of the evidence included in the review.	Limitations of this study
	23c	Discuss any limitations of the review processes used.	Limitations of this study
	23d	Discuss implications of the results for practice, policy, and future research.	Implications for research
OTHER INFORMATION			
Registration and protocol	24a	Provide registration information for the review, including register name and registration number, or state that the review was not registered.	Literature search and selection process (unregistered)
	24b	Indicate where the review protocol can be accessed, or state that a protocol was not prepared.	Literature search and selection process (unregistered)
	24c	Describe and explain any amendments to information provided at registration or in the protocol.	NA
Support	25	Describe sources of financial or non-financial support for the review, and the role of the funders or sponsors in the review.	NA
Competing interests	26	Declare any competing interests of review authors.	See Competing interests
Availability of data, code and other materials	27	Report which of the following are publicly available and where they can be found: template data collection forms; data extracted from included studies; data used for all analyses; analytic code; any other materials used in the review.	Table A3.1, Table A3.3, Table A3.4

Please note that the checklist has been adapted for this study to refer to sections instead of pages.
Adapted from: Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71. doi: 10.1136/bmj.n71

3

Appendix 3C: Data extraction form

Table A3.3. The form used for data extraction with a filled-out example

Study reference	Boyle & Özdemir (2009)
Type of ordering effect	Attributes
Domain of application (topic)	Environment (farmland conservation programs)
Type of preferences (target sample)	Policy preferences (general population)
Country of study	United States (Maine)
Sample size	329 (for analysis of ordering effects, 697 respondents used for other analyses)
Number of choice sets per respondent	4
Number of attributes	5
Number of alternatives per choice set	2 (with dual-response status quo alternative)
Labelled or unlabelled alternatives	Unlabelled
Survey administration mode	Paper (mail)
Experimental treatment	2 different versions of the DCE to which respondents were randomized – one with the cost attribute as the last attribute, and one with the cost attribute as the first attribute
Significant effect (at the 95% level)	▸ No significant effect on preference coefficients and welfare estimates ▸ Significantly smaller error variance (in unforced choice) when cost attribute is placed first
Any other remarks	This study also tested for the effect of two other design characteristics: the number of choice alternatives and the inclusion of a status quo alternative.

Appendix 3D: Design characteristics of included studies

Table A3.4: Overview of design characteristics of the included studies

Study	Type	Topic	Sample size	N choice sets[@]	N attributes	N alternatives per choice set	Labelled	Survey administration mode
Chrzan (1994)	AL	fashion accessories brands	300	16	NS	5 + OO	no	paper
Damman et al. (2012)	AL	home care providers	438	NS	2-3	5	no	online
Gerstenblüth et al. (2022)	AL, CS	cigarette packages and health	97	12	3	2	no	online, monitored
Koç & Van Kippersluis (2017)	AL	dinner choice	2869	18	6	2	no	online
Krucien et al. (2017)	AL, CS	health state valuation	293	32	6	2	no	paper
Meißner et al. (2016)	AL	coffee, beach vacation and laptop choice	60, 35, 70	12, 8, 20	6	3 + OO, 5, 4	no	on computer in lab
Ryan & Bate (2001)	AL	rheumatology care	189	8	6	2	no	paper
Sandorf et al. (2018)*	AL	coffee machines	518	8	5	2 + SQ	no	on computer in lab
Van der Waerden et al. (2006)	AL	transport modes	965	9	6	3	yes	paper
Zhao et al. (2022)	AL	neighbourhood public spaces	276	2	7	2	no	online
Berchi et al. (2006)	AT	colorectal cancer screening	294	3	6	2	no	Either on paper (mail) or online (NS)
Berchi et al. (2016)	AT	medication prescription	188	12	7	2	no	in person interviews, online
Boyle & Özdemir (2009)*	AT	farmland conservation	329[#]	4	5	2 + SQ (DR)	no	paper
Chrzan (1994)	AT	fashion services	876	16	10	2 + OO	no	paper
Farrar & Ryan (1999)	AT	clinical service development programs	216	NS	5	2	no	paper
Glenk (2007)	AT	ecosystem services	310	4	4	2 + SQ	no	in person interview
Heidenreich et al. (2021)	AT	insomnia treatments	156	12	7	2 + OO (DR)	no	online
Keshavarzian & Wu (2021)	AT	holiday destinations	415	2	13	5	yes	online

Table A3.4: Overview of design characteristics of the included studies *(Continued)*

Study	Type	Topic	Sample size	N choice sets[@]	N attributes	N alternatives per choice set	Labelled	Survey administration mode
Kjær et al. (2006)	AT	psoriasis treatments	444	9	6	2	no	paper
Krucien et al. (2017)*	AT	health and lifestyle programs	58	12	7	2 + OO	no	on computer in lab
Kumar & Gaeth (1991)	AT	televisions and burglar alarms	120	23	4	4	yes	NS
Logar et al. (2020)	AT	degraded river restoration	NS	6	6	2 + SQ	no	in person interviews
Mulhern et al. (2016)	AT	health state valuation	456	7	5	2	no	in person interviews
Mulhern et al. (2017)	AT	health state valuation	1073	10	6	3[+]	no	online
Mulhern et al. (2019)	AT	health state valuation	975	15	13	2	no	online
Norman et al. (2016a)	AT	health state valuation	2053	16	12	2	no	online
Ohdoko & Yoshida (2012)	AT	biodiversity of forest ecosystems	957	16	3-5	2	no	online
Scott & Vick (1999)*	AT	general practitioner choice	639	7-8	5	2	no	paper
Sjöstrand (2001)	AT	public transport quality	1030	6	4	2	no	paper
Soliño et al. (2017)	AT	recreational hunting	557	8	6	3 + SQ	no	NS
Tseng & Lii (2006)	AT	student apartments	279	NS	12	3	no	NS
Tsuchiya et al. (2019)	AT	health state valuation	1300	10	6	2	no	online
Abate et al. (2018)	CS	food safety of chicken breast filets	1168	8	5	2 + SQ	no	online
Achtnicht (2012)	CS	car fuel characteristics and emissions	598	6	6	7	no	in person interviews
Arentze et al. (2003)	CS	transport modes for work	346	16	3-5	2-3	yes	NS
Balcombe et al. (2015)*	CS	food choice	40	12	5	2 + SQ	no	online
Bansback et al. (2014)	CS	health state valuation	1799	15	6	2	no	online
Bech et al. (2011)	CS	dentist choice	1053	4 - 16	6	2 + OO	no	online

Table A3.4: Overview of design characteristics of the included studies *(Continued)*

Study	Type	Topic	Sample size	N choice sets[@]	N attributes	N alternatives per choice set	Labelled	Survey administration mode
Bechtold & Abdulai (2012)	CS	functional dairy foods	1309	21	3	2 + SQ	no	paper
Björklund & Swärdh (2017)	CS	comfort and crowdedness in public transport	2003	8	4	2 + OO	no	online
Boxall et al. (2009)*	CS	woodland improvement, forest management	519, 192	8, 16	6, 5-7	2 + SQ	no	paper (mail)
Bradley & Daly (1994)	CS	train services	243	10-16	4	2	yes	in person interview
Brouwer et al. (2010)	CS	water supply security	300	5	3	2 + SQ	no	in person interviews
Campbell et al. (2015)	CS	endangered fish species conservation	624	16	6	2 + SQ	no	in person interviews
Cao et al. (2018)	CS	egg purchase	1554	8	5	2 + OO	no	online
Carlsson & Martinsson (2001)*	CS	WWF environmental projects	35	16	3	2	no	NS
Caussade et al. (2005)*	CS	route choice	403	6 - 15	3 - 6	2 - 4 + SQ	no	NS
Chrzan (1994)	CS	mail-order purchase clubs	605	8	5	2 + OO	no	paper
Crastes dit Sourd et al. (2020)	CS	value of travel time, toll road choice, forest quality	NS	8	2, 4, 5	2, 2 + SQ, 3 + SQ	no	in person interviews, online
Czajkowski et al. (2012)	CS	forest protection & recreation infrastructure	1001	26	4	3 + SQ	no	in person interviews
Dardanoni & Guerriero (2021)	CS	WWF environmental projects	366	7	3	2 + SQ	no	in person interviews, paper
Day & Pinto-Prades (2010)	CS	drug treatments for hypothetical disease	500	3	2[^]	2	no[%]	in person interview
Day et al. (2012)	CS	tap water quality	704	16	3	1 + SQ	no	in person interviews
Dekker et al. (2014)	CS	flood risk protection	477	9	4	2 + SQ	no	online

Table A3.4: Overview of design characteristics of the included studies *(Continued)*

Study	Type	Topic	Sample size	N choice sets[@]	N attributes	N alternatives per choice set	Labelled	Survey administration mode
Groeneveld (2010)	CS	sustainability improvement of cutter fishing	NS	8	4	2 + SQ	no	online
Hensher & Collins (2010)*	CS	route choice	378	16	5	2 + SQ	no	in person interview
Hess et al. (2012)	CS	route choice, mode choice	3819, 397, 2197, 442, 304	8-16	2-6	2-3	no	in person interviews, online
Hildebrand et al. (2023)	CS	sod production	22	12	6	2 + OO	no	online
Hole (2004)*	CS	park and ride services	255	9	2 + L	2 + OO	yes	paper, online
Jacobsen & Thorsen (2010)*	CS	national parks	952	8	5	2 + OO	no	paper
Jarvis (2011)	CS	recreational fishing	906	8-24	8	2 + OO	no	mail
Jiang et al. (2022)	CS[$]	butterfly conservation	440	6	4	1 + OO	no	online
Jonker et al. (2018a)	CS	health state valuation	2731	21	5	2	no	online
Koppelman & Sethi (2005)	CS	intercity travel	> 1400[€]	8-9	4 + L	6-8	yes	paper
Ladenburg & Olsen (2008)	CS	nature impact of new motorways	579	6	5	2 + SQ	no	online
Ladenburg (2013)	CS	lunch programs for kindergartens	54	8	4	2 + OO	no	paper
Lanz & Provins (2013)	CS	local environmental improvements	106	12	2	2 + SQ	no	in person interviews
Lanz & Provins (2015)	CS	water, sewerage and environmental services	1517	12	4	2 + SQ	no	in person interviews, online
Logar & Brouwer (2017)*	CS	water quality improvement	500	6	4	2 + OO	no	online
Lundhede et al. (2009)*	CS	nature impact of new motorways, national parks	595, 636	6, 8	5	2 + SQ	no	online, paper
Maddala et al. (2003)*	CS	HIV testing	353	11	6	2	no	mail
Mariel & Meyerhoff (2016)	CS	land use change	1661	6-24	4-7	2-4 + SQ	no	online

Table A3.4: Overview of design characteristics of the included studies *(Continued)*

Study	Type	Topic	Sample size	N choice sets[@]	N attributes	N alternatives per choice set	Labelled	Survey administration mode
Marsh & Phillips (2012)	CS	river water quality	505	6	6	4 + SQ	no	online
McNair et al. (2012)	CS	electricity wires	582	4	5 + L	1-2 + SQ	yes	online
Meyerhoff & Glenk (2015)	CS	water quality improvement	753	12	6	2 + SQ	no	online
Mokas et al. (2021)	CS	value of urban greenery	180	12	4	2 + SQ	no	on computer in lab
Nguyen et al. (2021)	CS	cyclone warning services	1014	6	4	1 + SQ	no	in person interviews
Oehlmann et al. (2017)*	CS	land use change	1661	6-24	4-7	2-4 + SQ	no	online
Oppewal et al. (2010)*	CS	DVD recorder purchase	406	12	8	2 + OO	no	online
Petrolia et al. (2018)	CS	ecosystem service delivery	145, 134, 117	4	5	2 + SQ	no	online
Sælensminde (2001)	CS	value of travel time	508	9	3	2	no	in person interviews
Savage & Waldman (2008)	CS	internet access choice	682	8	5	2	no	online, paper
Scheufele & Bennett (2012)	CS	nature preservation	NS	16	3	1 + SQ	no	online
Swait & Adamowicz (2001)*	CS	orange juice	280	16	5	3 + OO	no	NS
Uggeldahl et al. (2016)	CS	food choice	190	12	4	2 + SQ	no	on computer in lab
Weng et al. (2021)	CS	wetland restoration	1827	8	5	1-3 + SQ	no	online
Zhang & Adamowicz (2011)*	CS	drinking water treatment	366	4	5	1-2 + SQ	no	online

*Abbreviations: AL=Alternatives, AT=Attributes, CS=Choice Sets, DR=dual-response, NS=Not Specified, OO=Opt-out, SQ=Status Quo. Notes: *) Study not included in the base set for the forward snowballing. @) Number of choice sets per respondent, excluding instructional choice sets and repeated choice sets (see footnote 7 in the paper). #) Subsample used for analysis of ordering effects. +) One of the alternatives is constant (i.e., immediate death). ^) Actually 4 alternatives, but 2 of these were provided in the introduction and stayed constant over the course of the DCE. The other 2 alternatives were presented in the DCE matrix. %) The alternatives were unlabelled in the choice sets but labelled in the introduction. $) Only the subsample with the unincentivized choice experiment. €) Not reported more precisely.*

Chapter 4

Public preferences for skin cancer prevention policies: A discrete choice experiment in three European countries

Based on:
Boxebeld, S., Mouter, N. and Van Exel, J. (2025). Public preferences for skin cancer prevention policies: a discrete choice experiment in three European countries. *Social Science & Medicine*, *378*, 118155

Abstract

Objective

In many countries, the incidence of skin cancer is growing rapidly, resulting in a substantive health and economic burden. While the wide range of available skin cancer prevention policies may have large individual and societal benefits, many countries still lack a policy strategy, and little is known about public preferences for collective prevention policy measures. We elicited these preferences using a discrete choice experiment (DCE) in Austria, the Netherlands, and Spain to inform policy action.

Methods

Respondents were asked to choose twelve times between two packages of different prevention policies. Each package was described by its estimated effectiveness and costs. Before and after the DCE, respondents were asked for their support for any policy action. We quota-sampled adult citizens in each of the countries from an online panel (N=2,442). The choice data were analysed using multinomial logit (MNL) and mixed multinomial logit (MMNL) models.

Results

Almost all attributes significantly influenced respondents' choices, with the tax attribute being most influential in each country. Among the six policy measures, information campaigns and a price reduction of sunscreen were the most preferred policy measures, and the prohibition of solar bed sales and solaria the least preferred. Preference structures were largely consistent across the countries. Finally, most respondents supported policy action, particularly after the DCE.

Conclusions

Citizens in the three countries recommended their governments to take policy action against the increasing incidence of skin cancer. The results provide policymakers with directions for publicly supported policy action, which should be complemented with additional information on preference heterogeneity, citizens' argumentation, and policies' relative (cost-)effectiveness. The suggestion that preferences for policy action adapted over the course of completing the DCE survey should be further examined.

Introduction

In many countries, the incidence of skin cancer is high relative to other cancer types and, moreover, increasing rapidly (Hu et al., 2022; Leiter et al., 2020). For instance, skin cancer accounts for approximately a third of all cancer diagnoses worldwide (Roky et al., 2025). The global age-standardized incidence rate of non-melanoma skin cancer was estimated to have increased by about 46% between 1990 and 2019, and its number of new cases and deaths is predicted to grow by at least another 50% between 2020 and 2044 (Hu et al., 2022). As such, some experts speak of a skin cancer epidemic (e.g., Apalla et al., 2017; Asadi et al., 2023; Urban et al., 2020), which is supposedly caused by a combination of demographic developments (i.e., population ageing), ecological factors (e.g., ozone layer depletion, global warming), and behavioural trends (e.g., changes in clothing style and beauty norms) (e.g., Asadi et al., 2023; Chang et al., 2014; Watson et al., 2024).

The growing incidence of skin cancer is associated with increasing healthcare expenditures (e.g., Guy et al., 2015; Meertens et al., 2024; Noels et al., 2020). The global economic burden of skin cancer was estimated to amount to $715 billion international dollars (i.e., $80.90 international dollars per capita or 0.015% of total GDP) in the period 2020 – 2050 (Chen et al., 2023). It is estimated that the vast majority of skin cancer cases (around 90%) is attributable to excess ultraviolet radiation (UVR) exposure and, as such, preventable (e.g., Leiter et al., 2020; Teng et al., 2021; Yu et al., 2024). Therefore, the gains of prevention policies are likely substantial (Collins et al., 2024; Gordon & Rowell, 2015; Hirst et al., 2012; Køster et al., 2020; Shih et al., 2017) and include an improved population health and wellbeing and reduced (functional) morbidity, increased labour force productivity, and healthcare expenditure savings.

Therefore, investing in skin cancer prevention is paramount from a public health and economic perspective. A range of policy alternatives is available, including awareness campaigns, prohibition of solar beds or solar studios, screening programs, and free provision or price regulation of sunscreen, all varying in their effectiveness, costs, and restriction of individual freedoms. It remains unclear, though, which prevention policies are preferred by the public. A few studies have elicited user preferences for individual prevention methods such as sunscreen (Solky et al., 2007), screening programs (Houston et al., 2016), and mobile screening applications (Finch et al., 2015; Gaube et al., 2024; Haggenmüller et al., 2021; Sangers et al., 2021). However, no studies have elicited citizens' preferences for collective action.

It is important that citizens' preferences are incorporated in the policy development and implementation process for several reasons. This contributes to the legitimacy of policy interventions, which is important in democratic societies. Citizen involvement may also help policymakers in enacting specific policies and adapting their communication to different population segments. Finally, societal support is desirable for an effective implementation of health policies, as it contributes to adherence (e.g., Gustavsson & Lindblom, 2025; Salloum et al., 2017).

Therefore, this study aims to elicit preferences from a representative sample of the general population for various skin cancer prevention policies using a discrete choice experiment (DCE) in three countries: Austria, the Netherlands, and Spain. Using EU-wide data from the European Cancer Information System (ECIS) (European Commission, 2023) on the incidence of melanoma, the most severe type of skin cancer, we selected one EU country with a relatively high incidence (the Netherlands), one with a relatively low incidence (Spain), and one around the EU average (Austria). The aim of this study is to provide insight into between-country similarities and differences in public preferences for skin cancer prevention policies, not to explain them.

Methodology

Set-Up of the DCE

We used DCE as the stated preference elicitation method for its ability to capture the trade-offs that respondents make between different policy measures and their characteristics and effects. As such, the method is generally highly appreciated by stakeholders and experts (Whichello et al., 2020) and has been widely applied in the health domain (e.g., Soekhai et al., 2019). One of the potential uses of DCE is the elicitation of citizens' preferences towards health policies, such as preventive interventions. DCE applications with this purpose have, for example, elicited citizens' preferences for policies promoting a healthy diet (Dieteren et al., 2023), reducing and preventing obesity (Lancsar et al., 2022), stimulating the uptake of a COVID-19 vaccine (Mouter et al., 2022), and limiting the consumption of alcohol (Pechey et al., 2014).

An important step in the conduct of a DCE is the selection of policy alternatives, attributes and levels. This selection is based on a review of the scientific literature and existing practices of skin cancer prevention, expert consultation, think-aloud pre-testing, and pilot studies and is described in more detail in Appendix 4B. The six selected policy measures (see Table 4.1) are included as dichotomous attributes (Yes/No) in the choice tasks, so that each alternative in a choice task is a policy package consisting of one or more policy measures.

The policy packages differed in the policies they contain and in their estimated effects. Three effect attributes were included in the DCE, capturing the impact of a policy package on the (1) yearly number of new cases of skin cancer, (2) the yearly number of deaths due to skin cancer, and (3) a tax increase. Since skin cancer typically develops over a long period of accumulating excess exposure to UVR, the policy packages are expected to affect the number of new cases and deaths only in twenty years. On the contrary, the tax increase is effective immediately; the policy packages namely require public investments upon their implementation (and enforcement), while the revenues in the form of averted healthcare expenditures or increased workforce productivity are uncertain and expected to be realized in the long run. The levels for all three effect attributes are presented textually as well as graphically (using bars) to enhance respondents' understanding of the attribute levels. An overview of all attributes and levels is presented in Table 4.1.

4

Table 4.1. Overview of attributes and levels in the DCE

Attribute	Levels			
	1	2	3	4
Policy measures				
Information campaigns	No	Yes		
Prohibition of the sale of solar beds for home use	No	Yes		
Prohibition of solar studios	No	Yes		
30% reduction of the price of sunscreen	No	Yes		
Free provision of sunscreen in public areas	No	Yes		
Free provision of an app for skin cancer detection	No	Yes		
Effects of the measures				
Number of new cases per year[1]	-5%	-10%	-15%	-20%
Number of deaths per year[1,]	-10%	-15%	-20%	-25%
Costs (tax increase)[2]	€36 per year (€3 per month)	€72 per year (€6 per month)	€108 per year (€9 per month)	€144 per year (€12 per month)

Notes: 1) For each country, a status quo in twenty years from now in the absence of any measure was determined (see Appendix 4B) and the percentages were therefore expressed in absolute numbers that differed between countries. 2) The costs in this table were presented in Austria and the Netherlands, which had similar price levels, and were adjusted to match the price level in Spain using OECD data (OECD, 2023), so that respondents in Spain were presented with prices between €30 - €120 per year.

All in all, each choice task included two policy packages described by nine attributes. In each choice task, respondents were asked to choose one of the two policy packages. We opted for a forced choice (i.e., not offering an opt-out or status quo alternative) to elicit respondents' trade-offs, given that the question to respondents was which policies to prevent skin cancer they preferred the government to implement, not whether they preferred policies to be implemented. We asked respondents whether they would recommend the government to implement any (additional) skin cancer policies separately, both before and after the DCE.[1] At the top of each choice task screen, respondents were informed about the estimated number of new skin cancer cases and deaths per year in twenty years under the status quo (i.e., when no policy package is implemented). In case of level overlap (i.e., both policy packages containing the same level for a specific attribute), the background of the levels was coloured in grey to simplify the comparison of policy packages for respondents (e.g., Jonker et al., 2018; Jonker, 2024; Norman et al., 2016). To mitigate attribute ordering effects (Boxebeld, 2024), the order of the six policy measure attributes was randomized between respondents, while the order of the three effect attributes was fixed for all respondents, considering that presenting both effectiveness attributes first and the tax attribute next would be a more natural grouping of these attributes for respondents than presenting them in an entirely random order, and given limitations of the survey software. Similarly, the left-right position of the policy package in the choice task was randomized and an alternative-specific constant (ASC) was included in the choice models to capture any alternative ordering effects (Boxebeld, 2024). An example of a choice task is presented in Figure 4.1.

Apart from an introduction and the DCE choice tasks, the survey contained several additional questions: prior to the choice tasks, respondents were asked for their age, gender and educational attainment (as screening questions for the quota sampling) and after the choice tasks, they were asked to motivate their choices using open-ended questions. The survey instrument, including the DCE, was programmed in Sawtooth Lighthouse Studio v.9.14.2 (Sawtooth Software, n.d.).

1 The question presented before the DCE was: 'Would you recommend the government to take any policy measures to protect people against skin cancer?'. The question presented after the DCE was: 'Now that you have made a choice between policy packages twelve times, would you recommend the government to take any policy measures to protect people against skin cancer?'.

Figure 4.1. Example of a DCE choice task (translated to English)

	Package A	Package B
Information campaigns and school programs	No	Yes
Ban on the sales of solar beds for home use	Yes	Yes
Ban on solaria	No	No
30% price reduction of sunscreen	Yes	No
Free provision of sunscreen in public spaces	No	Yes
Free app for skin cancer detection	Yes	No
Effects of the policy measures Number of new cases per year	90.000 (-10.000)	85.000 (-15.000)
Number of deaths per year	1.760 (-440)	1.870 (-330)
Costs (tax increase)	€90 per year (€7.50 per month)	€60 per year (€5 per month)
I choose for:	Package A	Package B

Experimental Design

For the pilot studies, an efficient design was generated using Ngene v.1.2.1 (ChoiceMetrics, n.d.). The priors for the policy measure attributes were set at zero. The attributes regarding the effects of the policy measures were all dummy-coded. For reductions of 5%, 10%, 15%, and 20% in the number of new cases of skin cancer per year, the priors were set at 0.1, 0.2, 0.3 and 0.4, respectively. The same priors were used for a 10%, 15%, 20% or 25% reduction in the number of deaths due to skin cancer per year. Finally, the priors for the cost attribute were specified at -0.1, -0.2, -0.3 and -0.4 for a tax increase of €36, €72, €108, or €144 per year (i.e., €3, €6, €9, or €12 per month) (for AT and NL, or equivalent levels in ES). The coefficients resulting from the estimation of an MNL model on the pilot data in the Netherlands (N=151) were used as inputs for Bayesian priors in the generation of the final design for all three countries to eliminate between-country variation in results due to experimental design differences. The pilots in Austria (N=102) and Spain (N=101) were only used to check whether respondents correctly understood the survey. The final design was optimized for the Bayesian D-criterion for an MNL model (without interactions) using 1,000 Sobol draws. Two restrictions on possible combinations of attribute levels were imposed (see Appendix 4C) and 36 choice tasks were generated and grouped into three blocks. Respondents were randomly assigned to one of the three blocks of 12 choice tasks each. To minimize any bias from choice task ordering effects (Boxebeld, 2024), we randomized the order of choice tasks in the DCE sequence between respondents. Also, we presented respondents with two instructional choice sets (with fixed levels) to gradually build up the choice task complexity and disclosed the attribute level ranges and number of choice tasks in advance.

Data Collection

The data were collected in the three countries from online panels administered by Dynata (Dynata, n.d.), a worldwide-operating provider of survey services. Panel members were quota-sampled by the panel provider with the aim of obtaining samples representative for the country's adult population in terms of age, gender, and education level. Data collection took place between November 21 and December 11, 2023. Given the size of the choice task, survey access was restricted to computers only.[2] To exclude low-quality response patterns, a few data exclusion criteria were used (see Appendix 4D). After exclusion of 50 respondents (i.e., 2.0% of the initial sample)[3], a sample of 2,442 respondents remained for the analysis. The country-specific subsamples are described in terms of sociodemographic characteristics in Table A4.1 in Appendix 4A.

Model Specification and Estimation

The DCE data were analysed for the three countries separately using a Multinomial Logit (MNL) model. Under this model, embedded in Random Utility Theory, the utility derived from an alternative can be divided into a deterministic component and a stochastic component. The deterministic component consists of the sum of the utilities derived from the attribute levels of the alternative, while the stochastic component is captured in an error term.

---

2 The implications of this are unclear for this specific study, and studies on differences in DCE results by survey access mode resulted in different findings. For example, while Liebe et al. (2015) found differences between respondents who used a desktop/laptop or a tablet/smartphone in terms of price sensitivity, Vass & Boeri (2021) found no significant differences in preferences. DCE characteristics (e.g., the number of attributes) may play a moderating role in the impact of the survey access mode (Vass & Boeri, 2021).

3 The MNL results are robust to the inclusion of the respondents that were excluded from the main analyses, as well as to the exclusion of respondents who indicated to prefer no policy action regarding skin cancer prevention prior to the DCE (see Appendix 4G).

When comparing the three countries, there may be heterogeneity in preferences as well as in scale, because of which the beta coefficients cannot be compared directly. Therefore, relative measures were derived from the estimated choice models, as these relative measures can be compared between countries. For the MNL models, relative attribute importance was measured using both attribute-based normalization and profile-based normalization (Gonzalez, 2019). The effect attributes in the first estimated MNL models were dummy-coded, like in the experimental design, to check for linearity of the parameters. Based on the MNL estimates, we applied an attribute-based normalization. For each of the attributes, the greatest attribute importance (i.e., the difference in utility between the most and least preferred attribute level in a country) was derived. Next, the importance of the attribute with the greatest difference in utility between the most and least preferred attribute level was normalized to 1, and the importance of the other attributes was expressed relative to the tax attribute. Notably, in the attribute-based normalization, it is assumed that the importance of the attribute with the greatest importance is equal between countries, which may not be the case. Therefore, we also applied a profile-based normalization, for which the total difference in utility between the (theoretically) most and least preferred policy package was calculated (Gonzalez, 2019).

In addition, to accommodate random heterogeneity in preferences, Mixed Multinomial Logit Models (MMNL) were estimated. We allowed for random heterogeneity in all attributes, including the ASC, to avoid the misattribution of heterogeneity. MMNL models are continuous mixture models, in which the choice probabilities do not come with a closed-form solution. Therefore, the choice probabilities were approximated using simulation based on 5,000 Sobol draws. The panel structure of the data was accounted for, so that random preference heterogeneity is allowed for between respondents, but not within respondents. Given that the coefficients of the dummy-coded tax attribute in the initial MNL models showed a reasonable degree of linearity (see Table 4.2), we the tax attribute is treated continuously in the MMNL models. This facilitates the calculation of welfare estimates and unifies the estimates choice models with economic theory (Mariel et al., 2021). The coefficients of the two dummy-coded effectiveness attributes in Table 4.2 show a lack of linearity. To account for this non-linearity while simultaneously allowing these two variables to be included in a continuous fashion, which facilitates model convergence, these were Box-Cox transformed (e.g., Tuhkanen et al., 2016). The resulting utility function of the MMNL model takes the form:

$$U_{itj} = ASC_j + \beta'_i X_{itj} + \frac{\delta'_i {N_{itj}}^{\lambda} - 1}{\lambda} + \rho_i tax_{itj} + e_{itj}$$

in which U_{itj} represents the utility that a respondent i derives from choosing alternative j in choice task t, ASC is an alternative-specific constant estimated for one of the two alternatives in a choice task to capture any alternative ordering effects (Boxebeld, 2024), and e_{itj} is a stochastic error term. Furthermore, X_{itj} is a vector of the policy-specific attributes that characterize alternative j, and β'_i is a vector of taste coefficients corresponding to the policy-specific attributes. N_{itj} is a vector of the two effectiveness attributes (i.e., number of new skin cancer cases; number of skin cancer deaths), δ'_i is a vector of taste coefficients corresponding to the effectiveness attributes, and λ is the non-linear transformation parameter to be estimated. Finally, tax_{itj} is the tax attribute level of j and ρ_i is the taste coefficient for the tax attribute.

To interpret and compare the MMNL estimates across countries, we computed the marginal rate of substitution (MRS) between each of the policy-specific and effectiveness attributes and the tax increase attribute. We take the (negative) ratio of the unconditional distributions for both parameters, which takes the following form for the policy-specific attributes:

$$MRS_i = -\frac{\beta_i}{p_i}$$

The standard errors have been computed using the Delta method (Bliemer & Rose, 2013). Since the effectiveness attributes are included non-linearly in the MMNL models, the MRS distribution between these attributes and the tax attribute is not constant either but varying by the level of the effectiveness attribute. To obtain the MRSs for these attributes, we worked out the partial derivates of the utility function including the estimated transformation parameter λ and the unconditional distribution of the δ for the attribute in question, with respect to the attribute levels included in the DCE. Then, the ratio was taken between the resulting distribution and the unconditional distribution for the tax attribute parameter, yielding a MRS distribution that is specific to a particular value of the attribute:

$$MRS_{N,i} = -\frac{\delta_i N^{\lambda-1}}{p_i}$$

The distribution of the random parameters is specified as normal for the ASC_i and β_i parameters:

$$\beta_i = \mu + \sigma\zeta_i$$

in which μ and σ are the mean and standard deviation of the random parameter, and ζ_i is a vector of standard normal draws for i. For the effectiveness attributes, we expected a direction of preference (i.e., respondents were expected to derive positive utility from reductions in the number of new skin cancer cases and the number of skin cancer deaths), because of which we constrained the distribution of their parameters. That is, we assumed a log-normal distribution:

$$\delta_i = e^{(\mu_N + \sigma_N\zeta_{N,i})}$$

For the tax attribute, we expected respondents to derive negative utility from a tax increase. Assuming a negative log-normal distribution (i.e., without shifting the distribution) may result in 'exploding implicit prices', however (Crastes dit Sourd, 2024). This potential issue was mitigated by 'mu-shifting' the point mass of the distribution of the tax attribute away from zero (Crastes dit Sourd, 2024):

$$\rho_i = -e^{(\mu_{tax})} - e^{(\mu_{tax} + \sigma_{tax}\zeta_{tax,i})}$$

All models were estimated in R v.4.4.0, with the choice modelling package Apollo v.0.3.0 (Hess & Palma, 2019) and using the BGW algorithm (Bunch et al., 1993).

Results

The results from the MNL model, in which respondents had to choose one of the two policy packages in each of the twelve choice tasks presented to them, are presented in Table 4.2. All policy measures were significantly and positively associated with the utility respondents derived from a policy package, except for both types of prohibition in Austria, which were not significantly associated with derived utility at the 95% level. With respect to the effect attributes, the reductions in number of new skin cancer cases and skin cancer-attributable deaths were significantly and positively associated with the utility derived from a policy package. The only exception was the attribute level of a 15% reduction in skin cancer deaths in Austria and the Netherlands. The tax increase attribute was significantly and negatively associated with the utility derived from a policy package for all levels in each country. Finally, the significant ASC parameters suggest left-right bias in each country (i.e., a higher choice probability for the left-hand alternative, ceteris paribus) (Boxebeld, 2024).

Table 4.2. Multinomial logit (MNL) model estimates with dummy-coded effect attributes

Attribute level	AT		NL		ES	
	Coeff. (Rob. SE)	p-value	Coeff. (Rob. SE)	p-value	Coeff. (Rob. SE)	p-value
Policy attributes						
Information campaigns	0.3326 (0.0370)	< 0.0001	0.1862 (0.0363)	< 0.0001	0.2993 (0.0352)	< 0.0001
Prohibition of sale tanning beds	-0.0179 (0.0367)	0.6266	0.0775 (0.0368)	0.0352	0.0891 (0.0342)	0.0091
Prohibition of solaria	0.0727 (0.0377)	0.0539	0.0819 (0.0400)	0.0404	0.1038 (0.0333)	0.0018
Price sunscreen 30% lower	0.2810 (0.0354)	< 0.0001	0.3169 (0.0363)	< 0.0001	0.3693 (0.0339)	< 0.0001
Free provision sunscreen in public areas	0.1065 (0.0400)	0.0078	0.1279 (0.0396)	0.0013	0.1284 (0.0356)	< 0.0001
Free skin cancer detection app	0.1869 (0.0313)	< 0.0001	0.1229 (0.0302)	< 0.0001	0.1916 (0.0275)	< 0.0001
Effect attributes						
Effect on N new cases of skin cancer per year						
-5% (Ref.)	-	-	-	-	-	-
-10%	0.0818 (0.0381)	0.0159	0.2003 (0.0405)	< 0.0001	0.1681 (0.0360)	< 0.0001
-15%	0.1897 (0.0409)	< 0.0001	0.3424 (0.0433)	< 0.0001	0.3232 (0.0381)	< 0.0001
-20%	0.4020 (0.0417)	< 0.0001	0.5948 (0.0438)	< 0.0001	0.4078 (0.0388)	< 0.0001
Effect on N deaths due to skin cancer per year						
-10% (Ref.)	-	-	-	-	-	-
-15%	0.0419 (0.0409)	0.1528	-0.0249 (0.0440)	0.2853	0.1632 (0.0417)	< 0.0001
-20%	0.1290 (0.0482)	0.0037	0.1879 (0.0482)	< 0.0001	0.1699 (0.0474)	< 0.0001
-25%	0.2023 (0.0415)	< 0.0001	0.3082 (0.0466)	< 0.0001	0.4400 (0.0451)	< 0.0001

Table 4.2. Multinomial logit (MNL) model estimates with dummy-coded effect attributes *(Continued)*

Attribute level	AT		NL		ES	
	Coeff. (Rob. SE)	p-value	Coeff. (Rob. SE)	p-value	Coeff. (Rob. SE)	p-value
*Additional tax**						
€36 per year (Ref.)	-	-	-	-	-	-
€72 per year	-0.3315 (0.0435)	< 0.0001	-0.4825 (0.0462)	< 0.0001	-0.3492 (0.0417)	< 0.0001
€108 per year	-0.8131 (0.0650)	< 0.0001	-1.0932 (0.0683)	< 0.0001	-0.8400 (0.0616)	< 0.0001
€144 per year	-1.2075 (0.0768)	< 0.0001	-1.6108 (0.0854)	< 0.0001	-1.0633 (0.0727)	< 0.0001
ASC						
ASC right-hand alternative	-0.0913 (0.0286)	0.0014	-0.1278 (0.0289)	< 0.0001	-0.0806 (0.0289)	0.0053
Model summary statistics						
N respondents	793		787		862	
LL (final)	-6080.74		-5803.96		-6647.37	
AIC	12193.49		11639.93		13326.75	
BIC	12308.06		11754.38		13442.66	

*P-tests are two-sided for the policy attributes and one-sided for the effect attributes. Notes: *) The presented levels for the cost attribute are for Austria and the Netherlands and were adapted for Spain, as explained in the note to Table 4.1. Abbreviations: ASC=Alternative-Specific Constant, AT=Austria, Coeff.=Coefficient, ES=Spain, LL=Log-likelihood, NL=The Netherlands, Rob. SE=Robust Standard Error.*

The relative attribute importance is presented in Figure 4.2. As can be observed from the attribute-based normalization in Panel A, the tax attribute was the most important in respondents' choices in all three countries, and the difference in importance between the tax attribute and the other attributes was large. In the profile-based normalization in Panel B, the importance of each attribute is expressed as the proportion of the overall difference in utility between the most and least preferred policy package in a country accounted for by that attribute. Here, we do see differences between countries in the importance of the tax attribute, with the greatest importance in the Netherlands and the lowest in Spain. Regarding the two 'effectiveness attributes', the reduction of new cases was more important in respondents' choices than the reduction in deaths in both Austria and the Netherlands. In Spain, these two attributes were of similar importance. With respect to the policy measures, the preference structures of the three countries were rather similar. On average, lowering the price of sunscreen and information campaigns were more influential in respondents' choices than both types of prohibition, free sunscreen in public areas, and the free provision of a detection app. The most striking difference between countries is that the policy measures of information campaigns and the free provision of a detection app were less influential in respondents' choices in the Netherlands relative to Austria and Spain.

Figure 4.2. Relative importance of the attributes by country

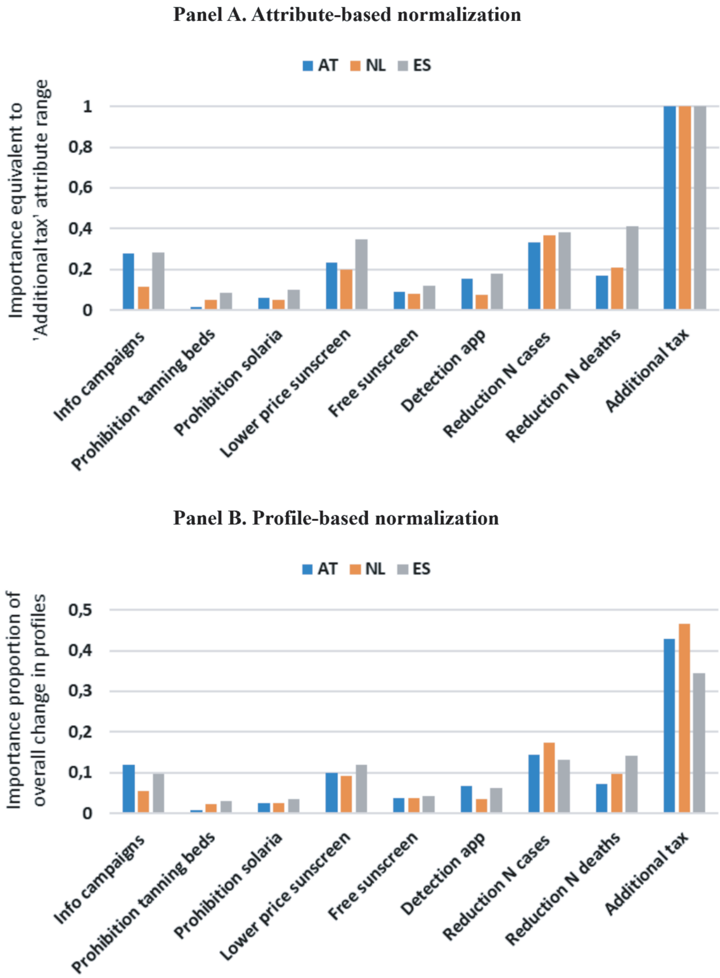


Table 4.3. Mixed multinomial logit (MMNL) model estimates

Attribute level	AT		NL		ES	
	Coeff. (Rob. SE)	**p-value**	**Coeff. (Rob. SE)**	**p-value**	**Coeff. (Rob. SE)**	**p-value**
Policy attributes*						
Mean						
Information campaigns	0.6685 (0.0660)	< 0.0001	0.4124 (0.0636)	< 0.0001	0.5199 (0.0576)	< 0.0001
Prohibition of sale tanning beds	0.0437 (0.0602)	0.4684	0.2417 (0.0598)	< 0.0001	0.2048 (0.0521)	< 0.0001
Prohibition of solaria	0.0292 (0.0653)	0.6549	0.0435 (0.0691)	0.5289	0.1453 (0.0535)	0.0066
Price sunscreen 30% lower	0.5072 (0.0616)	< 0.0001	0.6222 (0.0652)	< 0.0001	0.5890 (0.0559)	< 0.0001
Free provision sunscreen in public areas	0.1801 (0.0609)	0.0031	0.2597 (0.0640)	< 0.0001	0.2426 (0.0529)	< 0.0001
Free skin cancer detection app	0.4310 (0.0576)	< 0.0001	0.3569 (0.0540)	< 0.0001	0.3779 (0.0461)	< 0.0001
SD						
Information campaigns	0.8971 (0.0858)	< 0.0001	0.8252 (0.0835)	< 0.0001	0.8943 (0.0716)	< 0.0001
Prohibition of sale tanning beds	0.8127 (0.0863)	< 0.0001	0.6607 (0.1038)	< 0.0001	0.5668 (0.0874)	< 0.0001
Prohibition of solaria	1.017 (0.0903)	< 0.0001	1.0549 (0.0987)	< 0.0001	0.6832 (0.0818)	< 0.0001
Price sunscreen 30% lower	-0.6079 (0.1045)	< 0.0001	0.7106 (0.1015)	< 0.0001	0.4425 (0.1102)	< 0.0001
Free provision sunscreen in public areas	0.6187 (0.1029)	< 0.0001	0.5977 (0.1046)	< 0.0001	0.4576 (0.0989)	< 0.0001
Free skin cancer detection app	0.6768 (0.0854)	< 0.0001	-0.5456 (0.0962)	< 0.0001	-0.3804 (0.0945)	< 0.0001
Effect attributes*						
Mean						
Effect on N new cases of skin cancer per year #	-5.7920 (1.0871)	< 0.0001	-3.0533 (0.6755)	< 0.0001	-3.7025 (0.8167)	< 0.0001
λ (transf. par.) N new cases	1.7109 (0.4090)	< 0.0001	0.9432 (0.2724)	0.0005	0.9041 (0.3112)	0.0037

Table 4.3. Mixed multinomial logit (MMNL) model estimates *(Continued)*

Attribute level	AT		NL		ES	
	Coeff. (Rob. SE)	p-value	Coeff. (Rob. SE)	p-value	Coeff. (Rob. SE)	p-value
Effect on N deaths due to skin cancer per year #	-4.3020 (1.7571)	0.0072	-7.6417 (1.0429)	< 0.0001	-5.3104 (1.1973)	< 0.0001
λ (transf. par.) N deaths	0.6206 (0.6282)	0.3233	2.1073 (0.2988)	< 0.0001	1.0963 (0.3929)	0.0053
Additional tax	-5.6066 (0.1338)	< 0.0001	-5.1789 (0.1071)	< 0.0001	-5.5866 (0.1334)	< 0.0001
SD						
Effect on N new cases of skin cancer per year #	1.5839 (0.1601)	< 0.0001	1.2166 (0.1039)	< 0.0001	1.4089 (0.1231)	< 0.0001
Effect on N deaths due to skin cancer per year #	2.4965 (0.2010)	< 0.0001	2.2615 (0.1236)	< 0.0001	2.8487 (0.1862)	< 0.0001
Additional tax	2.7586 (0.1462)	< 0.0001	2.7735 (0.1201)	< 0.0001	2.5403 (0.1299)	< 0.0001
ASC*						
Mean						
Right-hand alternative	-0.1395 (0.0428)	0.0011	-0.1879 (0.0407)	< 0.0001	-0.1109 (0.0389)	0.0043
SD						
Right-hand alternative	0.6026 (0.0663)	< 0.0001	-0.4192 (0.0762)	< 0.0001	-0.6390 (0.0657)	< 0.0001
Model summary statistics						
N respondents	793		787		862	
LL (final)	-5508.69		-5152.81		-6061.49	
AIC	11061.38		10349.61		12166.97	
BIC	11218.91		10506.98		12326.34	

*P-tests are one-sided for the means of the effect attributes, and two-sided for all other coefficients. Notes: *) The random coefficients for the ASC and Policy attributes are specified to be normally distributed, those for the effects on the N of cases and N of deaths are specified to be positively lognormally distributed, and for the additional tax is specified to be mu-shifted and negatively lognormally distributed. #) This variable was Box-Cox transformed. Abbreviations: ASC=Alternative-Specific Constant, AT=Austria, Coeff.=Coefficient, ES=Spain, LL=Log-likelihood, NL=The Netherlands, Rob. SE=Robust Standard Error, SD=Standard Deviation.*

The results of the MMNL models, which accommodate random heterogeneity in preferences, are presented in Table 4.3. Starting values for the MMNL models were taken from the corresponding MNL models (see Appendix 4E). The results show there was significant heterogeneity in preferences for all the attributes in each country. Preference heterogeneity seems relatively stronger for the two types of prohibition, particularly in Austria and the Netherlands.

From the results, we derived the MRSs. The median, mean and standard error of the mean for the MRSs between the policy-specific attributes and the tax increase are presented in Table 4.4. The MRSs can be interpreted as the yearly increase in taxes respondents are willing to accept for the adoption of a particular policy measure. For instance, the median value of €12.77 for information campaigns in the Netherlands indicates that the median respondent in the Netherlands is willing to accept a tax increase of €12.77 per year (i.e., a bit over €1 per month) if this results in the implementation of an information campaign. The much lower median values relative to the mean values indicate that the distributions of the MRSs for all attributes in all three countries are right-skewed.

Table 4.4. MRS estimates for the policy measures

Attribute	AT			NL			ES		
	Median	Mean	Rob. SE	Median	Mean	Rob. SE	Median	Mean	Rob. SE
Information campaigns	41.20	90.99	14.30	12.77	36.60	6.55	29.43	69.37	11.78
Prohibition of sale tanning beds	0.60	5.94	8.14	6.09	21.45	5.54	8.90	27.32	7.33
Prohibition of solaria	0.32	3.97	8.93	0.34	3.86	6.16	4.76	19.39	7.64
Price sunscreen 30% lower	33.47	69.03	10.71	27.43	55.21	7.44	50.90	78.58	11.26
Free provision sunscreen in public areas	6.06	24.51	8.33	7.33	23.04	5.82	13.05	32.37	7.58
Free skin cancer detection app	24.09	58.65	10.96	13.15	31.67	6.10	28.42	50.42	9.39

The estimates relate to the marginal rate of substitution (MRS) between each policy-specific attribute and the tax increase attribute. Abbreviations: AT=Austria, ES=Spain, NL=The Netherlands, Rob. SE=Robust Standard Error

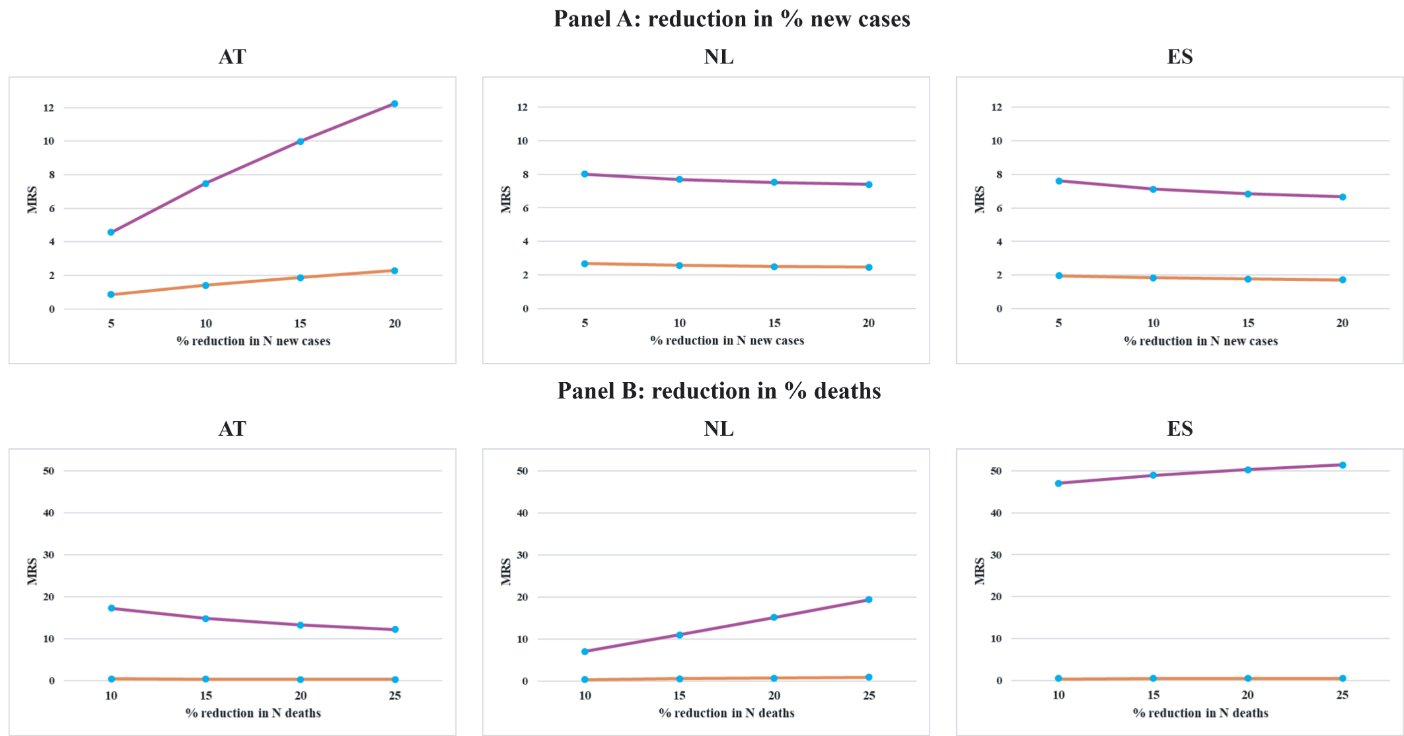
The observations that arise when comparing the MRS estimates roughly correspond with the findings from the MNL models; from the six policy measures, information campaigns and a price reduction in sunscreen were most valued across the three countries, followed by a free skin cancer detection app. The prohibition of tanning bed sales and of solaria were least valued. Also, some differences between countries arise. Respondents in the Netherlands derived less value from information campaigns and a skin cancer detection app than those in Austria and Spain, in line with the relative attribute importance measures presented before. Also, respondents in Spain were least averse towards both types of prohibition, while respondents in Austria were most averse.

The MRS estimates for both effectiveness attributes by country are plotted in Figure 4.3. It can be observed that the value of the MRS increases in the level of the effectiveness attributes for the reduction in new cases in Austria and the reduction in deaths in the Netherlands and Spain. In contrast, it decreases for the reduction in new cases in the Netherlands and Spain and for reduction in deaths in Austria. Similar to the policy-specific attributes, the median values are generally much lower relative to the mean values, indicating that the distributions of the MRSs for both effectiveness attributes in each of the countries are right-skewed. For reductions in the number of new cases, the MRS estimates are very similar for the Netherlands and Spain, while the mean MRS estimates in Austria are lower for lower values of the attribute and higher for higher values of the attribute. For reductions in deaths, the MRS estimates are rather similar for Austria and the Netherlands, although with opposite trends. While the median MRS estimates in Spain are similar to those in the other two countries, the mean MRS estimates are much higher. This indicates a substantially higher degree of skewness in the MRS distributions for this attribute in Spain compared with the other countries.

Both before and after the choice tasks, respondents were asked whether they would recommend the government to adopt any policy measures to protect people against skin cancer.[4] In Figure 4.4, the results are graphically presented. Prior to the DCE, most respondents are in favour of taking any policy action, ranging from 63.2% in Austria and 71.0% in the Netherlands to 83.1% in Spain.

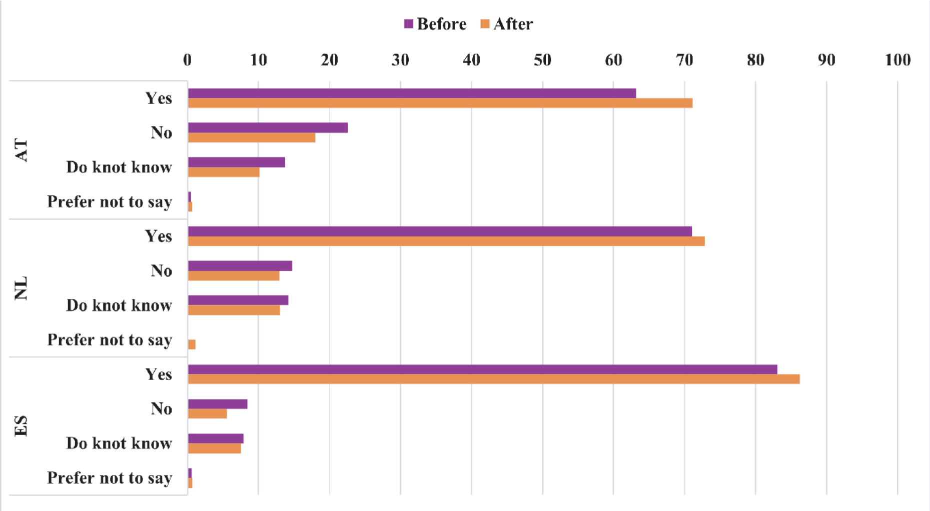
4 After the first time that respondents were asked this question, they were informed about the DCE design, asked to indicate for each of the included policy measures whether they thought the measure had already been in force in the year of data collection (i.e., 2023), they were presented with two instructional choice sets, and they completed the sequence of twelve choice tasks. Also, they were asked whether they themselves or anyone in their immediate surroundings had been diagnosed with skin cancer, and whether they had an occupation in which they were working outdoors (occasionally or frequently). Finally, before they were asked the question regarding their support for any policy action for the second time, respondents were asked to motivate their choices in the DCE, using two open-ended questions.

Figure 4.3. MRS estimates for the effectiveness attributes



The orange lines indicate the median MRS values, the purple lines indicate the mean MRS values. Abbreviations: AT=Austria, ES=Spain, NL=The Netherlands

Figure 4.4. Respondents' preferences for any policy action before and after the DCE



Abbreviations: AT=Austria, ES=Spain, NL=The Netherlands

These differences between countries are statistically significant at the 95% level in a logistic regression, also after adjusting for country sample composition differences in terms of age, gender, and education level (see Table A4.8 in Appendix 4F). After the DCE, the shares of respondents in favour of taking any policy measures have increased with 7.9%-point in Austria, 1.8%-point in the Netherlands, and 3.1%-point in Spain, reducing the difference in support between highest (i.e. Spain) and lowest (i.e., Austria) from 19.9%-point to 15.1%-point. This suggests that respondents adapted their preferences, based on their considerations of the policies and their effects while completing the DCE survey, in favour of taking policy action in all three countries, although this difference was not statistically significantly in the Netherlands.[5]

5 According to a McNemar's Test for each country, the differences in proportions of people answering 'Yes' (as opposed to any of the other answer options) before and after the DCE are statistically significant at the 95% level for Austria (McNemar's Chi-sq 29.84; p-value < 0.0001) and Spain (McNemar's Chi-sq 8.19; p-value 0.0042), but not for the Netherlands (McNemar's Chi-sq 1.34; p-value 0.2466). After the DCE, the differences in support for policy action between Austria and the Netherlands are no longer statistically significant, while respondents in Spain again show a significantly higher level of support (see Table A4.9 in Appendix 4F).

Conclusion and Discussion

This study has examined public preferences for policies targeted at the prevention of skin cancer and differences in these preferences between three European countries with a varying incidence of (melanoma) skin cancer: Austria, the Netherlands and Spain. To our knowledge, it is the first study that examines preferences for collective skin cancer prevention measures, rather than for individual prevention measures. Its findings can be categorized into three overall findings.

Firstly, the results from the choice models suggest that the policy measures, the effects on the number of new skin cancer cases and deaths, and the tax increase all played a role in respondents' choices in the three countries, except for the two types of prohibition policies in Austria. Furthermore, the tax attribute was the most influential attribute in each country, providing negative utility. Secondly, (almost) all policies were supported on average, and the preference structure was similar for the three countries. Respondents in the Netherlands valued information campaigns and the free provision of a skin cancer detection app less than respondents in Austria and Spain. Lowering the price of sunscreen was highly valued by respondents in all three countries, while both types of prohibition were less valued, particularly in Austria. This corresponds with previous studies that examined public preferences for preventive health interventions, which found that encouraging and less intrusive interventions receive more public support than discouraging and more intrusive interventions (Diepeveen et al., 2013; Dieteren et al., 2023; Mouter et al., 2022). The extent to which this is the case may vary by country and should also be considered in relation to (respondents' preferences towards) the effectiveness and costs of policy measures.

Finally, we find that the majority of respondents in each of the countries recommended the government to take policy measures to protect people against skin cancer. Public support for policy action was highest in Spain and lowest in Austria, both when asked before and after the DCE. However, the level of public support increased after the DCE, particularly in Spain and in Austria, so that the difference in public support between countries also decreased. This finding of policy support adapting over the course of the DCE survey provides an additional interesting insight[6], that deserves further inquiry in future studies.

6 Previous studies found that participation in a deliberation with others on the study topic (e.g., Jiang et al., 2023; Reckers-Droog et al., 2020) and information treatments in a DCE (e.g., Needham et al., 2018; Vanermen et al., 2021) may result in respondents adapting their attitudes and preferences. Also, some studies that used a DCE

Policy implications

Policy action is generally supported by a large majority of respondents in all three included countries, while a minority (i.e., 18.0 - 22.6% in Austria, 13.0% - 14.7% in the Netherlands, and 5.6 – 8.5% in Spain) would not recommend the government to take any policy action. As such, the governments of these countries are recommended to take policy action regarding this topic. When considering the implementation of preventive policies, governments are recommended to take measures that minimally increase the tax burden, since this is the most important (and disliked) attribute in respondents' preferences. This could be realized by means of implementing less expensive policies, or perhaps by reallocation of existing public resources rather than increasing the tax level.[7] At the same time, provided that the underlying assumption of fully compensatory decision-making holds, the MRS estimates show the extent to which respondents are willing to accept a tax increase for any specific measure and thereby indicate how much the government could spend on these policy actions while maintaining public support.

On average, almost all policy measures receive public support, but to varying extents. The two types of prohibition, the most intrusive policies, were the least supported policy measures. Governments are therefore recommended not to take these policies first. Dieteren et al. (2023) found a similar result in their DCE on policy measures promoting a healthy diet and suggested that implementing (less intrusive) policy measures may eventually raise support for more intrusive measures, referring to the stated preference literature surrounding tobacco and alcohol policies (Dieteren et al., 2023). Policies that are particularly recommended to be adopted (first) are lowering the price of sunscreen and information campaigns, as these policies were most preferred by respondents. While information campaigns may be generic and tailored towards everyone, their (cost-)effectiveness may be particularly high when targeted to groups with the highest risk of developing skin cancer or the greatest potential benefits of prevention, such as people with an outdoor occupation and children (Kasparian et al., 2009). Finally, governments from countries for which no studies on preferences for collective skin cancer prevention policies are available yet may take away from this study that, across the three countries of study, there was broad support for less intrusive prevention

including an opt-out or status quo alternative (i.e., an unforced choice setting) found a change in the probability of choosing the opt-out or status quo alternative over the sequence of choice tasks (Boxebeld, 2024). These results, although investigated using different study approaches, relate to our findings.

7 The latter would require that the respondents' willingness to allocate public budget to skin cancer prevention policies is higher than their willingness to do so for alternative public spending purposes, which is a condition that could be examined in future studies.

policies. Nevertheless, the relationships between respondents' preferences and individual, institutional, cultural and other contextual characteristics remain unclear and, therefore, one should be cautious when extrapolating the results. Also, respondents in this study were informed about the specific mechanism through which policies would be financed (i.e., increasing taxes). Applicability and support for such mechanisms may vary across countries, which also should be considered when extrapolating the findings. For context-specific evidence about policy support for skin cancer prevention policies, conducting a study like this locally is strongly recommended.

Limitations and recommendations for future research

While the study has examined between-country differences in public preferences for skin cancer prevention policies, it has not attempted to explain these differences or to assess within-country (i.e., between-respondent) preference heterogeneity. Many individual- and country-level characteristics may contribute to preference heterogeneity within and between countries (e.g., Kasparian et al., 2009). Even though examining the role of such characteristics in public preferences is beyond the scope of our paper, it seems valuable to further explore preference heterogeneity regarding this topic. Also, we have excluded several policy measures from this DCE based on the pre-testing, such as the implementation of population-based screening programs and shading policies (see Appendix 4B). Future choice experiments may examine citizens' preferences towards these and perhaps other policy options, too.

Furthermore, in the DCE, we have presented respondents with a forced choice setting only. Future research may examine which factors influence respondents' choices for an opt-out or status quo alternative. Besides, since preferences may be endogenous to design characteristics of the DCE, future studies may examine the robustness of findings to design changes. For example, future studies may position the tax attribute in between the policy-specific attributes and the effectiveness attributes or change the specification of the payment vehicle or the visual presentation of attribute levels to examine the impact of these design traits on the importance of the tax attribute in respondents' choices.

Also, future studies may examine the robustness of the results to the analytical decisions made. For instance, due to limits to the available computational capacity, the simulation of the value of the log-likelihood function for the MMNL models is based on 5,000 Sobol draws. Following recommendations from recent research comparing simulation noise under different types of draws (Czajkowski & Budziński, 2019) and given the rather large number of random parameters in our MMNL models, we would ideally have used a larger number of (shuffled or scrambled) draws. Furthermore, to assure model convergence, we assumed uncorrelated random parameters in the estimation of our MMNL models, like most applied DCE studies in health economics. However, it has been recommended to allow for correlation between random parameters in an MMNL model (Mariel et al., 2021). Inclusion of all potential correlation patterns would substantially raise the number of parameters and complicate the model estimation. Finally, the estimates are based on the assumption of respondents employing fully compensatory decision heuristics. Previous studies have shown that respondents may not attend all attributes (Gonçalves et al., 2022) and, therefore, this assumption may not hold in practice. Even though attribute non-attendance (ANA) could be accounted for in the modelling, different methods of doing so are available (Gonçalves et al., 2022) and may lead to different results. Also, some studies argue it is difficult, or perhaps impossible, to disentangle the sources of attribute non-attendance (e.g., heuristics or true preferences)(Heidenreich et al., 2017), putting the analyst at risk of imposing rather than revealing preferences. For these reasons, we have not attempted to incorporate ANA in our models and acknowledge the potential bias resulting from this.

Furthermore, as applicable to all stated preference research, hypothetical bias may compromise the external validity of study findings (Haghani et al., 2021a; 2021b). To mitigate hypothetical bias, we have implemented a form of a consequentiality script in the introduction by stating that the results will be shared with the national ministry of health and national cancer foundation of the respective country. Nevertheless, we cannot exclude the possibility of hypothetical bias influencing the results. As another dimension of external validity, the study's results are time- and place-specific. For instance, stated preferences may be affected by respondents' psychological distance to the study topic (Veldwijk et al., 2019). Arguably, the psychological distance to the topic of study may be larger at the end of the year (when UV exposure is lowest), when data was collected, than in the summer (when UV exposure is highest). Besides, a variety of survey modes and sampling methods is available, with varying advantages and disadvantages (Mariel et al., 2021). The choice for online data collection may affect the data quality and representativeness of the study sample, even though its influence may be limited in practice (e.g., Determann et al., 2017). Also, we hope that this study in three countries inspires future research to examine citizens' preferences in other countries too, since preferences may depend on cultural, institutional, and other factors that differ between countries.

Finally, respondents in this study were asked to choose the most preferred policy package in each choice task of two packages, limiting the room for respondents to indicate their preferences towards particular combinations of policy measures. One of the respondents indicated that they would have liked to have the opportunity to compose a policy package of their preference, instead of choosing between two predetermined packages. To meet such demands, further research may make use of alternative preference elicitation methods to elicit citizens' preferences for skin cancer prevention policies. For example, Participatory Value Evaluation (PVE) seems a useful method in this context. Respondents in a PVE are asked to compose their most preferred policy package (called 'portfolio') from a set of policy measures, subject to a resource constraint. This allows them to express their preference towards particular combinations of policy measures and the extent to which resources are allocated to this policy area (Boxebeld et al., 2024; Mouter et al., 2021).

Conclusion

This study explored public preferences for collective skin cancer prevention policies in three European countries. It provided governments with directions for publicly supported policy action to address the rising incidence of skin cancer and, with it, its increasing societal burden. The results suggested a large majority of citizens to support policy action against skin cancer. Less intrusive policy measures, such as reducing the price of sunscreen and information campaigns, are favoured over more intrusive policy measures, such as the prohibition of solar bed sales and solaria. Also, while the study's results can inform governments with directions for policy action that are publicly supported, these should be complemented with additional information on the relative effects of the different policy measures, the relation between preferences and individual, institutional, cultural and other contextual factors, and citizens' argumentation, to form a more complete understanding of public support for collective skin cancer prevention policies.

Appendix 4A: Descriptive sample statistics

Table A4.1. Descriptive statistics of the study sample for selected sociodemographic characteristics (after exclusion of respondents of suspected low quality)

Socio-demographic characteristic	Country		
	AT	NL	ES
Gender			
Man	368 (46.4%)	373 (47.4%)	425 (49.3%)
Woman	423 (53.3%)	411 (52.2%)	435 (50.5%)
Non-binary	2 (0.3%)	2 (0.3%)	2 (0.2%)
Do not know	-	1 (0.1%)	-
Age			
18 – 34	211 (26.6%)	203 (25.8%)	192 (22.3%)
35 – 64	403 (50.8%)	421 (53.5%)	493 (57.2%)
65+	178 (22.4%)	163 (20.7%)	177 (20.5%)
Prefer not to say	1 (0.1%)	-	-
Education level			
Education other than university (of applied sciences)	567 (71.5%)	560 (71.2%)	543 (63.0%)
University (of applied sciences)	225 (28.4%)	226 (28.7%)	316 (36.7%)
Do not know	-	-	1 (0.1%)
Prefer not to say	1 (0.1%)	1 (0.1%)	2 (0.2%)
Total N	793	787	862

Appendix 4B: Selection of attributes and levels

Selection process

Our selection of policy measures, effect attributes, and levels was informed by reviewing the scientific and 'grey' literature on (the evaluation of) existing policies and initiatives in other countries, for which we used Web of Science, Google, and Google Scholar. Also, we have been in contact with seven experts on skin cancer in each of the three countries of study (four by digital interviews and three by email). A concept survey was pre-tested using a convenience sample of three lay persons of different ages and education levels and about twenty colleagues (two of which one-on-one and the others in two group sessions). This resulted in three changes to the DCE, discussed below, yielding the eventual survey instrument used for the final data collection. This survey was also piloted in each of the countries (N=102 for Austria, N=151 for the Netherlands, and N=101 for Spain), but this did not result in any changes.

Policy measures

When it comes to primary prevention, we distinguish between three types, for each of which we included at least one policy measure in the DCE. Firstly, there is the type of prevention measure that aims to affect individuals' knowledge of and attitude towards the health behaviour. In our case, this includes interventions that make people more aware of the risk of excess UV exposure, the importance of prevention, and the possibilities to protect oneself. Two of such policy interventions are information campaigns and educational programs in schools. Information campaigns may target the general population or specific groups and make use of traditional media (e.g., television, radio, newspaper advertisements), billboards, and social media channels. Educational programs in schools aim to learn children from a young age onwards about appropriate sun protection behaviour. Both types of interventions have been implemented in various countries, such as Australia, where these have demonstrated to be (cost-)effective (e.g., Shih et al., 2009; Sinclair & Foley, 2009). We initially included both as separate policy measures in the DCE. As the cognitive burden of the DCE on respondents seemed high when pre-testing the survey, we eventually combined both interventions into a single policy measure in the DCE.

Secondly, another type of prevention measure aims to support people in changing their health behaviour by facilitating healthy choices and reducing the barriers for adopting individual protection measures. The two policy measures included in our DCE of this type both aim to stimulate the use of sunscreen, as sunscreen is (conditional on appropriate use) a highly effective individual protection measure. One of the two policy measures stimulates sunscreen use by reducing the price of sunscreen by 30%. This percentage was inspired by Dieteren et al. (2023), who included a 30% reduction in the price of vegetables and fruits in their DCE on preferences for policies stimulating a healthy diet. We considered this percentage sufficiently substantial to be meaningful to respondents, without overshooting. The other policy measure stimulates sunscreen use by providing free sunscreen in public areas (e.g., beaches, parks, schoolgrounds and sports facilities). On top of these two policy measures, we initially also included a policy intervention raising the amount of shadow in public areas by means of planting trees and placing shade sails in public spaces like parks, schoolgrounds and beaches. This type of intervention allows individuals to find shelter against the sun while still being outdoors and has been implemented in *inter alia* the Canadian province of Toronto (Holman et al., 2018). During the pretesting, however, respondents explained to choose policy packages including this measure for reasons totally unrelated to the topic of the DCE (e.g., favouring more trees in public areas for environmental or aesthetic reasons). While such spillovers, or positive externalities, may be an argument to implement this policy, it was a reason for us to exclude this intervention to reduce the role that such other considerations would play in respondents' choice behaviour.

Finally, a third type of primary prevention measure restricts individuals' room for making unhealthy choices. Of course, it is difficult to reduce exposure to UV radiation from the sun, but indoor tanning can be regulated more easily. Exposure to UV radiation from indoor tanning devices is considered an important risk factor for the development of skin cancer. In our DCE, the two legal bans on indoor tanning belong to this type of policy measure. We included a legal ban on the sale of solar beds for home use, as we considered the banning of their use to be unfeasible in terms of law enforcement. Also, we included a legal ban on solar studios. Legal bans on indoor tanning are in place for youths in several countries and for everyone in Australia and Brazil, and have been found effective in reducing tanning (e.g., Carpenter et al., 2023; Morais, 2022).

Besides primary prevention policies, a range of secondary prevention measures is available. Originally, we included two types of secondary prevention measures: a population-based screening campaign and the free provision of a skin cancer detection app. Under the first policy measure, either all adults or individuals of high-risk groups are periodically invited for a comprehensive or targeted screening of one's skin. Under the second policy measure, everyone with a smartphone can download an application for free and use it to scan a suspicious spot on their skin. The spot is then examined using artificial intelligence, after which the user may be referred to see a doctor in case of suspicion. While the two measures can co-exist, the respondents in the pretesting were confused by the presence of both policy interventions. Therefore, and to reduce the cognitive burden of the DCE, we decided to drop one of the two secondary prevention measures. We removed the population-based screening program, since this is much more expensive than the detection app and is facing a low take-up in the German states where it has been implemented already.

Effect attributes

Apart from the six policy-specific attributes, the DCE tasks also included three attributes to capture the effects of the policy measures. First, the reduction in number of new skin cancer cases is included. Any results of policy action are not to be expected in the short term, given that much of skin cancer risk is about the cumulative exposure to UVR. Therefore, respondents are told that the reduction in number of skin cancer cases is realized in twenty years from now. Avoiding the use of percentages, which may be difficult for respondents to process, the number of new skin cancer cases per year is expressed both as an absolute number and relative to a status quo (see next subsection). This attribute reflects the effectiveness of primary prevention policies (i.e., policies preventing the onset of skin cancer), such as both types of prohibition and both measures aimed to increase sunscreen use. However, it does not capture the effectiveness of secondary prevention policies (i.e., policies preventing the progression of the disease by stimulating early detection and treatment), such as the free provision of a skin cancer detection app. To capture the effectiveness of these secondary prevention efforts as well as primary prevention measures, the yearly number of deaths resulting from skin cancer in twenty years from now is included as a second effect attribute. This attribute is similarly expressed both as an absolute number as well as relative to a status quo. To put the number of skin cancer deaths in perspective, the survey also informs respondents about the yearly number of deaths resulting from traffic accidents in their country. Finally, a third attribute was included to capture the opportunity costs of implementing and enforcing skin cancer policies. This attribute presents a tax increase, which is expressed as a uniform increase (i.e., equal for every adult citizen) of the income tax and presented both per month and per year. Unlike the two effectiveness attributes, the tax attribute applies to the current situation; respondents are told that a tax increase would be necessary to cover the immediate costs of implementing the policy measures, while any benefits accrue in the long run only. In the DCE, we chose to exclude any second-order effects (e.g., the effects of health gains on population happiness and labour force productivity).

In the choice task, the levels for all three effect attributes are presented textually as well as graphically to enhance respondents' understanding of the attribute levels. We follow Pechey et al., (2014) in using bars to visualize the changes to the status quo. In contrast with their study, we avoid 'traffic light' colours, given that the most prevalent colour blindness concerns the colours red and green (Jonker et al., 2018) and given that these colours may be distortionary in respondents' preferences. Instead, following the literature on colour coding in DCEs, we make use of the colour purple as a presumably more neutral colour (Himmler et al., 2021; Jonker et al., 2018a; 2018b; Jonker, 2024). This colour indicated changes to the status quo. Yellow was considered a suitable colour to combine with purple, taking into account the most prevalent types of colour blindness (e.g., Nichols, n.d.). We have not pretested different visualizations than the one used in the final data collection, which was not commented on during the pretesting phase, and we acknowledge that our choice of visualization might have influenced respondents' choices. Given the variety of attribute level visualizations used in DCEs and the limited evidence of their impact on response, future research may examine (or synthesize evidence about) the influence of these visualizations on respondents' preferences.

Status quo

Since the two effectiveness attributes were presented both in absolute numbers as well as relative to a scenario without any additional policy action, country-specific status quos had to be estimated in terms of the number of yearly new cases of skin cancer and deaths resulting from skin cancer in twenty years from now. The status quo estimations are presented below and based on data presented in Table A4.2.

Austria

Number of new cases in twenty years:

It is estimated that the number of new cases of melanoma will rise to approximately 2,500 per year by 2030, which is an increase of about 30.1% from 2020. In the absence of longer-term projections, we assume that the growth rate stays the same after 2030. Following this assumption, we expect the number of new cases of melanoma to amount to approximately 3,600 in twenty years from now (2,551 * (1.301^0.1)^13). In the absence of any projections for the incidence and the lack of any precise estimates for the current incidence of non-melanoma skin cancer, the 30,000 cases per year of BCC/PCC (the most important forms of non-melanoma skin cancer) mentioned in Hautnah (2016) as the approximate current incidence is taken as the starting point. A growth rate equal to that of melanoma is assumed, which means that we come to an incidence of about

42,200 (30,000 * (1.301^0.1)^20) new cases of non-melanoma skin cancer per year in twenty years from now. Together with the projection for melanoma, this results in an expected incidence of ±46,000 ≈ 50,000 new skin cancer cases per year in twenty years from now (for the expected incidence, we round the projections to numbers that are easy to process for respondents).

Number of deaths in twenty years:

The number of deaths resulting from melanoma is expected to rise to approximately 540 in 2030, which is an increase of about 26.6% from 2020. In the absence of longer-term projections, we assume that the growth rate stays the same after 2030. Following this assumption, we expect the number of deaths resulting from melanoma to amount to approximately 740 (543 * (1.266^0.1)^13) in twenty years from now. For non-melanoma skin cancer, no information is available regarding both the current and future mortality. Therefore, it is assumed that the percentage of skin cancer deaths resulting from non-melanoma skin cancer in Austria now and in the future is equal to the current percentage in the Netherlands (±15%). The Netherlands is used as reference point, for the current incidence of non-melanoma and the current number of deaths resulting from melanoma in Austria are both about half the size of those numbers in the Netherlands. Therefore, we estimate the number of deaths resulting from non-melanoma skin cancer to amount to ±130 (740 / 85 * 15) in twenty years. Together with the projection for melanoma, this results in an expected number of skin cancer deaths of 870 ≈ 1,000 in twenty years from now.

Spain

Number of new cases in twenty years:

No projections were available regarding the future number of new cases of skin cancer. Therefore, the future incidence was estimated by means of projections for the future number of deaths. It is projected that the yearly number of deaths resulting from melanoma skin cancer will increase by approximately 25.6% (1,326 / 1,056 * 100%) over the following twenty years. Taking a constant mortality rate as a reference point in the absence of any policy changes, it is assumed that the yearly number of new cases of melanoma skin cancer will increase by the same percentage and amount to approximately 10,100 (1.256 * 8,049) in twenty years from now. For non-melanoma skin cancer, both registration of current incidence as well as projection of future incidence is lacking. Instead, we base the expected number of new cases of non-melanoma skin cancer in twenty years on the incidence rates for BCC, SCC and melanoma skin cancer estimated in the meta-analysis by Tejera-Vaquerizo et al. (2016). In this study, the estimated overall incidence rate for BCC and SCC together (the vast majority of non-melanoma skin cancer cases are of one of these two types) amounts to 151.21 cases per 100,000 person-years. For melanoma, this amounts to 8.82. Therefore, we estimate that the historical percentage of total new cases of skin cancer attributable to non-melanoma skin cancer was ± 94.5% (151.21 / (151.21 + 8.82) * 100%). This is expected to be lower in twenty years given that the projected increase in melanoma deaths (25.6%) is larger than the projected increase in non-melanoma skin cancer deaths (14.3%). Therefore, we assume that the percentage of total new cases of skin cancer attributable to non-melanoma skin cancer in twenty years will be 90%. Given this assumption, we project the future number of new cases of non-melanoma skin cancer to amount to ± 90,900 (10,100 / 10 * 90). As such, the total number of new cases of skin cancer in twenty years is estimated to be around 101,000 ≈ 100,000.

Number of deaths in twenty years:

The number of deaths in twenty years is derived from the projections of García-Souto et al. (2021) and of Sendín-Martin et al. (2021). The first study projects the number of melanoma deaths to amount to 6,632 between 2039 and 2043. Since we do not have any information on the distribution of deaths over these five years, we take the average per year as the predicted number of deaths in twenty years: 1,326 (6,632/5). The second study projects the number of deaths resulting from non-melanoma skin cancer between 2040 and 2044 to amount to 4,440. By again taking the average per year, we come to a predicted number of non-melanoma skin cancer deaths in twenty years of 888 (4,440/5). Together, this results in a predicted number of deaths in twenty years of ±2,200.

The Netherlands

Number of new cases in twenty years:

For the number of new cases in the Netherlands in twenty years from now, we make use of predictions for the three most prevalent forms of skin cancer by Schreuder et al. (2019). For BCC, they estimate that the number of new cases will amount to 58,902 in 2027, which is an increase of 22.7% from 2017. In the absence of longer-term projections, we assume that the growth rate stays the same after 2027. Following this assumption, we expect the number of new cases of BCC to amount to approximately 81,700 in twenty years from now (58,902 * (1.227^0.1)^16). For PCC, the projection for the number of new cases is 21,318 in 2027, an increase of 73.1% from 2017. Assuming an equal growth rate after 2027, the predicted number of new PCC cases in twenty years from now amounts to ±51,300 (21,318 * (1.731^0.1)^16). Finally, the number of new cases of melanoma is expected to rise to 9,496 by 2027, an increase of 53.2% from 2017. Using the same procedure as for BCC and PCC, the predicted number of new melanoma cases in twenty years from now is approximately 18,800 (9,496 * (1.532^0.1)^16). Taken together, this amounts to about 152,000 ≈ 150,000 new cases of skin cancer per year in twenty years from now.

Number of deaths in twenty years:

Van Niekerk et al. (2021) expect the number of melanoma-related deaths will double between 2025 and 2045. In their projection, there will be about 2,200 deaths per year resulting from melanoma in about twenty years. For non-melanoma types of skin cancer, no projections for future mortality were available. Therefore, this has been derived from the projection for melanoma. According to data from Statistics Netherlands, melanoma-related deaths account for about 85% of the total number of skin cancer deaths. Assuming this will increase to about 90% in twenty years from now, given that the incidence of melanoma is currently rising faster than other types of skin cancer (Schreuder et al., 2019), the total number of skin cancer deaths are estimated at approximately 2,500 per year in twenty years from now.

Table A4.2. Overview of data retrieved on the mortality and incidence of skin cancer in Austria, Spain and the Netherlands.

Outcome measure	Country		
	Austria	**Spain**	**The Netherlands**
Number of deaths			
Melanoma skin cancer	Eurostat (2023a):	Eurostat (2023a):	Eurostat (2023a):
	2017: 401	2017: 992	2017: 798
	2018: 360	2018: 982	2018: 798
	2019: 373	2019: 1,036	2019: 753
	2020: 396	2020: 1,079	2020: 814
	2021: 397	2021: 1,056	2021: 793
	Hackl et al. (2015) 429 in 2020, 543 in 2030	García-Souto et al. (2021): 6,632 in 2039 – 2043	Van Niekerk et al. (2021): 2,200 in 2045
Non-melanoma skin cancer	-	Statistics Spain (2023):	Statistics Netherlands (2023):
		2017: 689	2017: 129
		2018: 668	2018: 108
		2019: 694	2019: 133
		2020: 738	2020: 150
		2021: 777	2021: 133
			2022: 163
		Sendín-Martin et al. (2021): 4,440 in 2040 – 2044	
Traffic accidents (for reference)	Eurostat (2023b):	Eurostat (2023b):	Eurostat (2023b):
	2017: 414	2017: 1,830	2017: 535
	2018: 409	2018: 1,806	2018: 598
	2019: 416	2019: 1,755	2019: 586
	2020: 344	2020: 1,370	2020: 515
	2021: 362	2021: 1,533	2021: 509
Number of new cases			
Melanoma skin cancer	Hackl et al. (2015) 1,961 in 2020, 2,551 in 2030	REDECAN (2023): 8,049	Schreuder et al. (2019): 9,496 in 2027
Non-melanoma skin cancer	Hautnah (2016): > 30,000 currently	-	Schreuder et al. (2019): > 70.000 currently
			Schreuder et al. (2019): 58,902 of BCC and 21,318 of PCC in 2027
			Schreuder et al. (2022): 63,900 of BCC in 2029

Appendix 4C: Experimental design

Design restrictions

In the generation of the experimental design, two restrictions were imposed: firstly, candidate choice tasks were rejected from the final design in case they contained a legal ban on solar studios, but not on the sale of solar beds for home use. The underlying rationale is that it may be perceived as insensible to prohibit solar studios, in which people can engage in indoor tanning in a more controlled setting, while allowing for the sale of solar beds for indoor tanning in one's own home environment. Secondly, candidate choice tasks including policy packages with a 'no' as level for every policy-specific attribute (i.e., no additional policy action will be implemented) were also rejected from the final design, since we did not include zero levels for the effect attributes.

Ngene syntax

```
design
;alts=altA, altB
;rows=36
;block=3
;eff=(mnl, d, mean)
;bdraws=sobol(1000)
;reject:
altA.solarium > altA.tanbeds,
altB.solarium > altB.tanbeds,
altA.info = 1 and altA.tanbeds = 1 and altA.solarium = 1 and altA.pricesunscr = 1 and altA.
freesunscr = 1 and altA.app = 1,
altB.info = 1 and altB.tanbeds = 1 and altB.solarium = 1 and altB.pricesunscr = 1 and altB.
freesunscr = 1 and altB.app = 1

;model:
U(altA)= b1.dummy[(u, 0.01, 0.21)] * info[1,0]
+ b2.dummy[(n, 0.18, 0.09)] * tanbeds[1,0]
+ b3.dummy[(u, -0.11, 0.17)] * solarium[1,0]
+ b4.dummy[(n, 0.27, 0.14)] * pricesunscr[1,0]
+ b5.dummy[(n, 0.26, 0.13)] * freesunscr[1,0]
+ b6.dummy[(n, 0.13, 0.07)] * app[1,0]
+ b7.dummy[(n, 0.17, 0.09)|(n, 0.35, 0.18)|(n, 0.35, 0.18)] * cases[1,2,3,0]
```

+ b8.dummy[(n, 0.22, 0.11)|(n, 0.30, 0.15)|(n, 0.35, 0.18)] * deaths[1,2,3,0]
+ b9.dummy[(n, -0.34, 0.17)|(n, -0.73, 0.37)|(n, -1.10, 0.56)] * costs[1,2,3,0]
/

U(altB)= b1.dummy * info
+ b2.dummy * tanbeds
+ b3.dummy * solarium
+ b4.dummy * pricesunscr
+ b5.dummy * freesunscr
+ b6.dummy * app
+ b7.dummy * cases
+ b8.dummy * deaths
+ b9.dummy * costs

$

Table A4.3. Experimental design matrix

Block	Choice situation	Concept	Info	Tanbeds	Solarium	Pricesunscr	Freesunscr	App	Cases	Deaths	Costs
1	1	1	2	1	1	2	2	2	4	3	3
1	1	2	1	2	2	1	1	1	2	2	2
1	2	1	1	1	1	1	2	1	1	2	1
1	2	2	1	2	1	2	1	2	2	3	2
1	3	1	1	2	1	1	1	2	1	2	2
1	3	2	2	2	2	2	2	1	4	4	4
1	4	1	2	2	2	1	2	2	3	1	3
1	4	2	1	2	1	2	1	1	1	3	2
1	5	1	1	2	1	1	2	2	4	2	2
1	5	2	1	2	1	1	2	1	1	4	3
1	6	1	1	2	2	2	2	2	2	4	1
1	6	2	2	1	1	2	2	2	2	1	2
1	7	1	1	2	2	1	1	2	1	3	2
1	7	2	2	1	1	1	1	1	4	2	3
1	8	1	1	1	1	2	2	1	4	2	2
1	8	2	2	2	2	1	2	2	2	4	2
1	9	1	1	2	1	1	2	1	1	1	3
1	9	2	1	1	1	1	1	1	2	3	2

Table A4.3. Experimental design matrix *(Continued)*

Block	Choice situation	Concept	Info	Tanbeds	Solarium	Pricesunscr	Freesunscr	App	Cases	Deaths	Costs
1	10	1	1	1	1	2	2	2	4	1	1
1	10	2	2	1	1	2	2	2	3	2	2
1	11	1	2	2	1	2	1	1	2	3	4
1	11	2	2	1	1	2	1	2	3	3	1
1	12	1	2	2	2	2	2	1	1	4	2
1	12	2	2	2	2	1	2	1	2	1	3
2	1	1	2	2	2	2	1	2	1	4	2
2	1	2	1	2	1	1	2	1	2	1	1
2	2	1	2	2	1	1	2	1	3	3	3
2	2	2	1	2	2	1	1	2	2	4	2
2	3	1	2	2	2	1	2	2	4	4	4
2	3	2	2	1	1	2	1	1	1	1	2
2	4	1	1	2	1	1	1	2	3	1	4
2	4	2	1	1	1	2	2	1	2	3	2
2	5	1	2	2	2	1	2	1	3	2	4
2	5	2	2	2	1	1	1	2	2	4	1
2	6	1	1	2	1	1	2	1	4	4	2
2	6	2	2	2	2	2	1	2	1	3	3
2	7	1	2	2	2	2	2	1	4	1	2
2	7	2	1	1	1	2	2	1	3	4	3
2	8	1	2	1	1	1	1	2	4	2	4
2	8	2	2	2	2	2	1	2	2	2	4
2	9	1	1	1	1	1	1	2	1	2	1
2	9	2	2	1	1	2	1	1	3	4	2
2	10	1	2	1	1	1	1	2	2	1	2
2	10	2	2	1	1	1	1	2	1	1	1
2	11	1	2	2	2	2	1	1	3	2	2
2	11	2	2	2	1	2	2	1	4	2	4
2	12	1	1	1	1	2	2	2	2	1	4
2	12	2	1	2	2	1	2	2	4	4	2
3	1	1	1	1	1	1	2	1	1	4	2
3	1	2	2	2	1	2	1	1	3	2	3
3	2	1	2	2	2	1	2	2	2	2	3
3	2	2	1	2	2	1	1	1	3	3	4
3	3	1	2	2	1	2	2	1	2	1	1

Table A4.3. Experimental design matrix *(Continued)*

Block	Choice situation	Concept	Info	Tanbeds	Solarium	Pricesunscr	Freesunscr	App	Cases	Deaths	Costs
3	3	2	1	1	1	1	2	2	4	2	2
3	4	1	1	2	2	2	1	1	1	4	3
3	4	2	2	2	1	1	2	2	3	3	1
3	5	1	1	2	1	2	2	1	1	1	4
3	5	2	2	1	1	2	1	1	3	3	3
3	6	1	1	2	1	2	2	2	1	2	3
3	6	2	2	1	1	2	2	2	1	1	2
3	7	1	1	1	1	2	2	1	3	4	1
3	7	2	1	2	2	1	2	1	4	3	1
3	8	1	1	2	2	1	1	2	4	1	3
3	8	2	1	2	2	1	1	2	1	2	4
3	9	1	1	2	2	1	1	2	1	3	1
3	9	2	1	1	1	2	2	1	1	3	4
3	10	1	1	2	1	1	2	2	4	2	2
3	10	2	2	2	2	1	1	2	4	2	1
3	11	1	1	1	1	2	2	2	4	1	1
3	11	2	2	2	1	1	1	1	4	4	2
3	12	1	1	1	1	2	2	1	4	3	3
3	12	2	1	1	1	2	2	2	2	3	4

Appendix 4D: Data quality check

Addressing potential data quality concerns, respondents were excluded from the sample of analysis in case they were considered to have exerted low-quality response patterns according to a combination of criteria. First, a statistical criterion related to root likelihood (RLH) has been used to exclude respondents, as this is considered a well-performing data quality criterion in terms of specificity and sensitivity (Jonker et al., 2022). For every respondent, the average individual root likelihood (RLH) was calculated using the DCE software Sawtooth Lighthouse Studio. This RLH score expresses the performance of a sample-level MNL model in explaining every individual's choices. Respondents were excluded from the sample of analysis in case their average individual RLH score was equal to or below 0.5 (e.g., Gregor et al., 2018). The rationale underlying this threshold value is that, for each choice task, the probability of correctly explaining an individual's choice by pure chance is 0.5 in case of two alternatives. If the average RLH score is below this, the sample-level model performs poorly for this individual and the individual is considered to have made random (i.e., low-quality) choices. This applied to only 3 respondents in Austria, 1 in the Netherlands and 1 in Spain.

Apart from this statistical criterion, respondents were also excluded in case they fulfilled at least two out of the following three criteria: (A) straightlining (i.e., always choosing the left-hand alternative or always the right-hand alternative for every choice task in the sequence), (B) a survey completion time of less than 4 minutes (i.e., about 30–40% of the median response time reported in Table A4.4), and (C) providing nonsensical answers to the open-ended motivation questions. All in all, as presented in Table A4.5, between 0.9% and 2.7% of respondents were removed from the sample of analysis.

Table A4.6 reports the descriptive statistics of the full study sample for selected sociodemographic characteristics (i.e., before exclusion of respondents). When comparing this table with Table A4.1 in Appendix 4A, presenting the descriptive statistics of the sample used for the main analyses (i.e., after exclusion of respondents according to Table A4.5), the differences are small. Comparing the full study sample in Table A4.6 with the sample quotas, we find that women were slightly oversampled in Austria and the Netherlands and slightly undersampled in Spain. Also, we find that younger respondents were slightly undersampled in Spain, while middle-aged respondents were oversampled in Spain but undersampled in Austria and the Netherlands. Finally, respondents with a university (of applied sciences) degree were undersampled in Austria and the

Netherlands, but slightly oversampled in Spain. These patterns generally persisted after exclusion of respondents based on quality criteria (see Table A4.1).

Table A4.4. Survey completion time statistics

Statistic	Austria	The Netherlands	Spain
Mean (SD)	24.607 (102.866)	21.103 (107.365)	23.630 (112.700)
Median	11.8	10.333	11.167

Table A4.5. Exclusion of respondents

	Austria	The Netherlands	Spain
Total completes	815	807	870
Quality criteria			
RLH ≤ 0.500	3	1	1
Straightlining (always A or always B)	23	8	32
Of which provided nonsensical answers to the open-ended motivation questions:	9	2	4
Response time < 4 minutes (240 seconds)	16	39	18
Of which provided nonsensical answers to the open-ended motivation questions:	8	19	1
Straightlining AND Response time < 4 minutes (240 seconds)	2	1	2
Total excluded respondents based on quality criteria (duplicates removed)	22 (2.7%)	20 (2.3%)	8 (0.9%)

Table A4.6. Descriptive statistics of the study sample for selected sociodemographic characteristics (before exclusion of respondents of suspected low quality)

Socio-demographic characteristic	Country					
	AT		NL		ES	
	Collected	Target (%)	Collected	Target (%)	Collected	Target (%)
Gender						
Man	381 (46.7%)	48.2	387 (48.0%)	49.0	431 (49.5%)	48.6
Woman	432 (53.0%)	51.8	417 (51.7%)	51.0	437 (50.2%)	51.4
Non-binary	2 (0.2%)	-	2 (0.2%)	-	2 (0.2%)	-
Do not know	-	-	1 (0.1%)	-	-	-
Age						
18 – 34	219 (26.9%)	26.4	216 (26.8%)	26.0	194 (22.3%)	23.5
35 – 64	414 (50.8%)	52.0	426 (52.8%)	53.9	497 (57.1%)	53.9
65+	181 (22.2%)	21.5	165 (20.4%)	20.1	179 (20.6%)	22.7

Table A4.6. Descriptive statistics of the study sample for selected sociodemographic characteristics (before exclusion of respondents of suspected low quality) *(Continued)*

Socio-demographic characteristic	Country					
	AT		NL		ES	
	Collected	Target (%)	Collected	Target (%)	Collected	Target (%)
Prefer not to say	1 (0.1%)	-	-	-	-	-
Education level						
Education other than university (of applied sciences)	582 (71.4%)	65.0	568 (70.4%)	65.0	549 (63.1%)	65.0
University (of applied sciences)	232 (28.5%)	35.0	238 (29.5%)	35.0	318 (36.6%)	35.0
Do not know	-	-	-	-	1 (0.1%)	-
Prefer not to say	1 (0.1%)	-	1 (0.1%)	-	2 (0.2%)	-
Total N	815		807		870	

Appendix 4E: MNL model estimates with non-linear continuous effectiveness attributes

Table A4.7. Multinomial logit (MNL) model estimates with non-linear continuous effectiveness attributes

Attribute level	AT		NL		ES	
	Coeff. (Rob. SE)	p-value	Coeff. (Rob. SE)	p-value	Coeff. (Rob. SE)	p-value
Policy attributes						
Information campaigns	0.3248 (0.0365)	< 0.0001	0.1748 (0.0357)	< 0.0001	0.2637 (0.0347)	< 0.0001
Prohibition of sale tanning beds	-0.0192 (0.0356)	0.5902	0.0945 (0.0356)	0.0079	0.0778 (0.0336)	0.0208
Prohibition of solaria	0.0706 (0.0375)	0.0598	0.0762 (0.0394)	0.0528	0.1167 (0.0330)	0.0004
Price sunscreen 30% lower	0.2837 (0.0338)	< 0.0001	0.3424 (0.0348)	< 0.0001	0.3349 (0.0330)	< 0.0001
Free provision sunscreen in public areas	0.0920 (0.0369)	0.0127	0.1291 (0.0367)	0.0004	0.1261 (0.0335)	0.0002
Free detection app	0.1863 (0.0313)	< 0.0001	0.1327 (0.0298)	< 0.0001	0.1844 (0.0278)	< 0.0001
Effect attributes						
Effect on N new cases	0.0009 (0.0018)	0.3012	0.0219 (0.0214)	0.1539	0.1294 (0.1269)	0.1539

Table A4.7. Multinomial logit (MNL) model estimates with non-linear continuous effectiveness attributes *(Continued)*

Attribute level	AT		NL		ES	
	Coeff. (Rob. SE)	p-value	Coeff. (Rob. SE)	p-value	Coeff. (Rob. SE)	p-value
$\lambda_{new\ cases}$	2.3119 (0.7238)	0.0014	1.2340 (0.3905)	0.0016	0.3332 (0.4190)	0.4265
Effect on N deaths	0.0022 (0.0085)	0.3969	0.0000 (0.0002)	0.3458	0.0005 (0.0019)	0.3896
λ_{deaths}	1.6759 (1.3206)	0.2044	3.0622 (0.8473)	0.0003	2.3769 (1.2126)	0.0500
Additional tax	-0.0115 (0.0007)	< 0.0001	-0.0155 (0.0008)	< 0.0001	-0.0120 (0.0008)	< 0.0001
ASC						
ASC right-hand alternative	-0.0856 (0.0266)	0.0013	-0.1094 (0.0266)	< 0.0001	-0.0790 (0.0279)	0.0046
Model summary statistics						
N respondents	793		787		862	
LL (final)	-6084.54		-5810.29		-6658.20	
AIC	12193.08		11644.58		13340.41	
BIC	12279.01		11730.42		13427.34	

*P-tests are two-sided for the policy attributes, ASC, and lambda parameters and one-sided for the effect attributes. Notes: *) The presented levels for the cost attribute are for Austria and the Netherlands and were adapted for Spain, as explained in the note to Table 4.1. Abbreviations: ASC=Alternative-Specific Constant, AT=Austria, Coeff.=Coefficient, ES=Spain, LL=Log-likelihood, NL=The Netherlands, Rob. SE=Robust Standard Error.*

Appendix 4F: Additional analyses of preferences for any policy action

Table A4.8. Logistic regressions on country differences in support for any policy action before the DCE.

	(1)		(2)	
	Coefficient (rob. SE)	P-value	Coefficient (rob. SE)	P-value
Country (*base: Austria*)				
The Netherlands	0.357 (0.108)	0.001	0.369 (0.110)	0.001
Spain	1.050 (0.117)	< 0.001	1.063 (0.119)	< 0.001

Table A4.8. Logistic regressions on country differences in support for any policy action before the DCE. *(Continued)*

	(1)		(2)	
	Coefficient (rob. SE)	**P-value**	**Coefficient (rob. SE)**	**P-value**
Adjusted for sociodemographics?	No		Yes	

The outcome variable is binary and based on the question 'Would you recommend the government to take any policy measures to protect people against skin cancer?', with values 1 (answer option 'Yes, I would recommend the government to take policy measures to protect people against skin cancer') and 0 (answer options 'No, I would recommend the government not to take policy measures to protect people against skin cancer', 'Do now know', and 'Prefer not to say'). Selected sociodemographics include age, gender, and education level. Abbreviation: rob. SE=robust standard error.

Table A4.9. Logistic regressions on country differences in support for any policy action after the DCE.

	(1)		(2)	
	Coefficient (rob. SE)	**P-value**	**Coefficient (rob. SE)**	**P-value**
Country *(base: Austria)*				
The Netherlands	0.084 (0.112)	0.456	0.098 (0.114)	0.391
Spain	0.930 (0.126)	< 0.001	0.957 (0.129)	< 0.001
Adjusted for sociodemographics?	No		Yes	

The outcome variable is binary and based on the question 'Now that you have made a choice between policy packages twelve times, would you recommend the government to take any policy measures to protect people against skin cancer?', with values 1 (answer option 'Yes, I would recommend the government to take policy measures to protect people against skin cancer') and 0 (answer options 'No, I would recommend the government not to take policy measures to protect people against skin cancer', 'Do now know', and 'Prefer not to say'). Selected sociodemographics include age, gender, and education level. Abbreviation: rob. SE=robust standard error.

Appendix 4G: Sensitivity analyses

Table A4.10. Multinomial logit (MNL) model estimates with non-linear continuous effectiveness attributes, including respondents excluded from the main analyses

Attribute level	**AT**		**NL**		**ES**	
	Coeff. (Rob. SE)	**p-value**	**Coeff. (Rob. SE)**	**p-value**	**Coeff. (Rob. SE)**	**p-value**
Policy attributes						
Information campaigns	0.3205 (0.0357)	< 0.0001	0.1660 (0.0352)	< 0.0001	0.2654 (0.0344)	< 0.0001
Prohibition of sale tanning beds	-0.0187 (0.0347)	0.5895	0.0951 (0.0348)	0.0064	0.0723 (0.0334)	0.0306
Prohibition of solaria	0.0674 (0.0363)	0.0632	0.0800 (0.0386)	0.0381	0.1167 (0.0326)	0.0003

Table A4.10. Multinomial logit (MNL) model estimates with non-linear continuous effectiveness attributes, including respondents excluded from the main analyses *(Continued)*

Attribute level	AT		NL		ES	
	Coeff. (Rob. SE)	p-value	Coeff. (Rob. SE)	p-value	Coeff. (Rob. SE)	p-value
Price sunscreen 30% lower	0.2711 (0.0332)	< 0.0001	0.3436 (0.0343)	< 0.0001	0.3243 (0.0329)	< 0.0001
Free provision sunscreen in public areas	0.0935 (0.0360)	0.0094	0.1329 (0.0361)	0.0002	0.11778 (0.0333)	0.0004
Free detection app	0.1793 (0.0305)	< 0.0001	0.1244 (0.0295)	< 0.0001	0.1820 (0.0275)	< 0.0001
Effect attributes						
Effect on N new cases	0.0007 (0.0015)	0.3138	0.0185 (0.0190)	0.1649	0.1563 (0.1518)	0.1515
$\lambda_{new\ cases}$	2.3810 (0.7755)	0.0021	1.2870 (0.4071)	0.0016	0.2520 (0.4175)	0.5462
Effect on N deaths	0.0048 (0.0169)	0.3889	0.0000 (0.0001)	0.3426	0.0006 (0.0021)	0.3874
λ_{deaths}	1.4079 (1.2318)	0.2531	3.1855 (0.8273)	0.0001	2.3257 (1.1894)	0.0505
Additional tax	-0.0112 (0.0007)	< 0.0001	-0.0152 (0.0008)	< 0.0001	-0.0117 (0.0008)	< 0.0001
ASC						
ASC right-hand alternative	-0.0960 (0.0277)	0.0005	-0.1105 (0.0266)	< 0.0001	-0.0803 (0.0285)	0.0049
Model summary statistics						
N respondents	815		807		870	
LL (final)	-6280.77		-5983.14		-6733.21	
AIC	12585.54		11990.27		13490.41	
BIC	12671.79		12076.41		13577.45	

*P-tests are two-sided for the policy attributes, ASC, and lambda parameters and one-sided for the effect attributes. Notes: *) The presented levels for the cost attribute are for Austria and the Netherlands and were adapted for Spain, as explained in the note to Table 4.1. Abbreviations: ASC=Alternative-Specific Constant, AT=Austria, Coeff.=Coefficient, ES=Spain, LL=Log-likelihood, NL=The Netherlands, Rob. SE=Robust Standard Error.*

Table A4.11. Multinomial logit (MNL) model estimates with non-linear continuous effectiveness attributes on sample from main analyses, excluding respondents who recommended not to take any policy action prior to the choice tasks

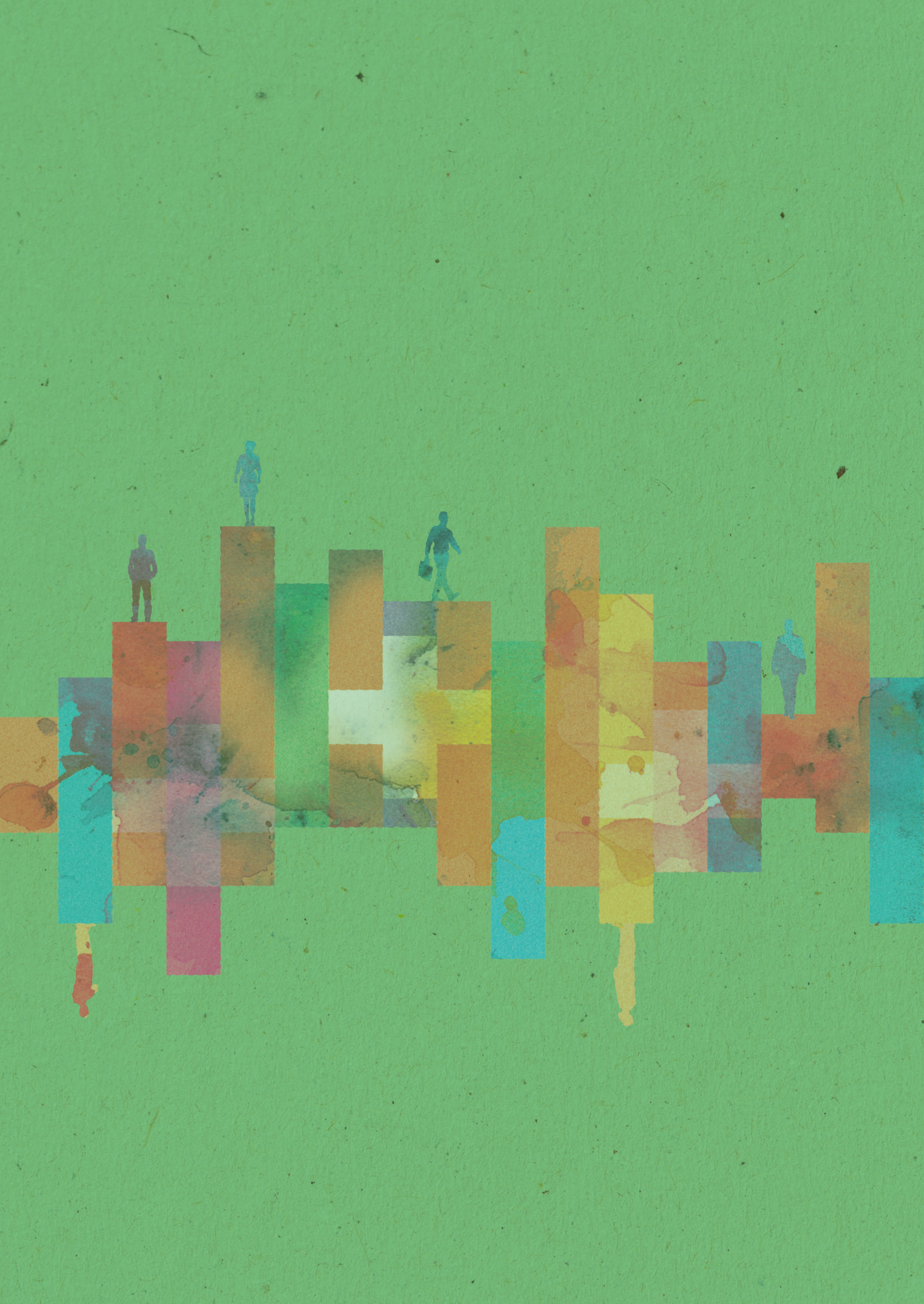
Attribute level	AT		NL		ES	
	Coeff. (Rob. SE)	p-value	Coeff. (Rob. SE)	p-value	Coeff. (Rob. SE)	p-value
Policy attributes						

Table A4.11. Multinomial logit (MNL) model estimates with non-linear continuous effectiveness attributes on sample from main analyses, excluding respondents who recommended not to take any policy action prior to the choice tasks *(Continued)*

Attribute level	AT		NL		ES	
	Coeff. (Rob. SE)	p-value	Coeff. (Rob. SE)	p-value	Coeff. (Rob. SE)	p-value
Information campaigns	0.3383 (0.0410)	< 0.0001	0.2057 (0.0385)	< 0.0001	0.2820 (0.0365)	< 0.0001
Prohibition of sale tanning beds	0.0240 (0.0395)	0.5432	0.1342 (0.0385)	0.0005	0.0849 (0.0350)	0.0152
Prohibition of solaria	0.1219 (0.0409)	0.0029	0.0910 (0.0432)	0.0351	0.1264 (0.0342)	0.0002
Price sunscreen 30% lower	0.3073 (0.0390)	< 0.0001	0.3800 (0.0381)	< 0.0001	0.3252 (0.0346)	< 0.0001
Free provision sunscreen in public areas	0.1260 (0.0417)	0.0025	0.1419 (0.0402)	0.0004	0.1404 (0.0353)	0.0004
Free detection app	0.1811 (0.0352)	< 0.0001	0.1570 (0.0317)	< 0.0001	0.2044 (0.0289)	< 0.0001
Effect attributes						
Effect on N new cases	0.0355 (0.0369)	0.1681	0.1524 (0.0542)	0.0025	0.2502 (0.0798)	0.0009
$\lambda_{new\ cases}$	2.3366 (0.9916)	0.0184	1.3470 (0.3804)	0.0004	0.1578 (0.4423)	0.7212
Effect on N deaths	0.0599 (0.0606)	0.1615	0.0284 (0.0195)	0.0724	0.0641 (0.0590)	0.1388
λ_{deaths}	1.3477 (1.0724)	0.2089	2.6086 (0.6303)	< 0.0001	1.8872 (0.9135)	0.0389
Additional tax	-0.3506 (0.0276)	< 0.0001	-0.5315 (0.0299)	< 0.0001	-0.3370 (0.0246)	< 0.0001
ASC						
ASC right-hand alternative	-0.0796 (0.0306)	0.0094	-0.0970 (0.0293)	0.0009	-0.0792 (0.0294)	0.0071
Model summary statistics						
N respondents	614		671		789	
LL (final)	-4768.38		-4959.70		-6101.05	
AIC	9560.76		9943.40		12226.11	
BIC	9643.62		10027.32		12311.98	

*P-tests are two-sided for the policy attributes, ASC, and lambda parameters and one-sided for the effect attributes. Notes: *) The presented levels for the cost attribute are for Austria and the Netherlands and were adapted for Spain, as explained in the note to Table 4.1. Abbreviations: ASC=Alternative-Specific Constant, AT=Austria, Coeff.=Coefficient, ES=Spain, LL=Log-likelihood, NL=The Netherlands, Rob. SE=Robust Standard Error.*

4



Chapter 5

Trade-offs in long-term care in an ageing society: A constrained portfolio choice experiment

Based on:

Boxebeld, S., Mouter, N. and Van Exel, J. (2025). Trade-offs in long-term care in an ageing society: A constrained portfolio choice experiment. *Journal of the Economics of Ageing [article in press]*

Abstract

Many countries face rapidly ageing populations, putting the sustainability of their long-term care (LTC) system under pressure. This urges governments to make challenging trade-offs between the affordability of the system and the quality and accessibility of care. To inform allocation decisions, this study elicited citizens' preferences for LTC policy in the Netherlands in 2040. We conducted a constrained portfolio choice experiment, in which 997 respondents composed a portfolio of their preferred policies for LTC for older people from seven policy alternatives, subject to a budget constraint. The policy alternatives embedded different types of care and support and were characterized by their estimated effects on meeting the demand for nursing care, informal care needs, and governmental expenditure. Choices were analysed using a Multiple Discrete Continuous Extreme Value (MDCEV) choice model and a Latent Class Cluster Analysis (LCCA). The results suggest a preference for distributing resources towards multiple policies over investing heavily in one or two particularly. The top-ranked portfolios all included both nursing care and social care policies, and most respondents chose portfolios constituting a substantial increase in public LTC expenditure. This study finds that most citizens in the Netherlands support increases in public expenditure on LTC for older people. Preference heterogeneity was identified regarding the preferred level of expenditure and the policy alternatives to be adopted, particularly for supportive care technologies and compulsory social service for young adults. Policymakers may use these results to support the selection of a portfolio of LTC policies that aligns with public preferences.

Introduction

The populations of many countries are ageing rapidly and predicted to continue ageing in the next decades (Eurostat, n.d.; WHO, 2024). Partially because of this demographic development, many of these countries are faced with substantial increases in their expenditures on long-term care (LTC) for older people (e.g., Breyer & Lorenz, 2021). Also, due to population ageing, the caregiving tasks for a growing number of older people will have to be borne and financed by a relatively small group of (potential) caregivers and taxpayers. Hence, the sustainability of LTC is under pressure in many ageing societies, regarding the availability of both financial resources and personnel (Mosca et al., 2017; Swartz et al., 2012). At the same time, there is substantial heterogeneity among countries when it comes to the public funding and delivery of LTC for older people (Swartz, 2013), as many different policy options exist with varying costs and benefits.

For policymakers involved in (re)designing the LTC system of the future, important trade-offs can be identified. There is a tension between quality and access to LTC for older people and the affordability of the care system (e.g., Da Roit, 2012). On the one hand, a collective LTC system with a comprehensive coverage guarantees a certain degree of access to care and thus horizontal equity. Also, a comprehensive provision of formal care is likely to reduce the provision of informal care (e.g., Hollingsworth et al., 2022; Miyakawi et al., 2020). Since the provision of informal care, for many caregivers, is associated with a substantial burden in terms of a reduced health and wellbeing (e.g., Bom et al., 2019) and economic opportunity costs (e.g., Schmitz & Westphal, 2017), comprehensive collective care provision may mitigate this burden (Hollingsworth et al., 2022; Løken et al., 2017; Miyawaki et al., 2020). On the other hand, comprehensive public provision of care requires large governmental expenditures, which may result in intergenerational inequities and an unsustainable care system in terms of financial and personnel requirements in the long run (e.g., Mosca et al., 2017; Swartz, 2013). Ultimately, this also comes down to the normative question of the extent to which LTC for older people is an individual or a collective responsibility and who should provide for this care and bear the associated burden (e.g., Hoefman et al., 2017; Janus & Koslowski, 2020; Read et al., 2021; Wittenberg et al, 2024).

At the same time, several studies find that changes in the availability of a certain type of LTC may be compensated by changes in the use of other types of long-term or medical care, suggesting substitution to take place to some extent between formal and informal LTC (e.g., Arora & Wolf, 2018; Hollingsworth et al., 2022; Miyawaki et al., 2020; Mommaerts, 2025), long-term and medical care (e.g., Bakx et al., 2020; Costa-

Font et al., 2018; Moura, 2022), and different types of formal LTC (e.g., Guo et al., 2015; Kattenberg & Bakx, 2021). Thus, policymakers face the challenging task of balancing all these different aspects in (re)designing a sustainable LTC system for the future (Da Roi, 2012; Mosca et al., 2017; Swartz, 2013).

In decisions about changes to the LTC system, it seems important for governments to incorporate citizens' preferences, since they are stakeholders in the system either as care recipient, caregiver, and/or taxpayer. In this study, we asked a sample of 997 adult citizens from the Netherlands to compose a portfolio of their preferred policy alternatives for LTC for older people in 2040, subject to a budget constraint. They were informed about the alternatives' estimated effect on the fulfilment of nursing care demand, the need for informal care, and governmental expenditure on LTC. We analysed their portfolio choices using a multiple discrete-continuous extreme value (MDCEV) choice model. We then used the resulting estimates as inputs into an optimal portfolio analysis, forecasting the expected utility of each portfolio and ranking the portfolios accordingly.

This study contributes to the existing literature on public preferences for LTC in two important ways. Firstly, many previous studies have elicited respondents' preferences regarding their own situation, such as characteristics of formal LTC (insurance) (e.g., Brau & Bruni, 2008; Chandoevwit & Wasi, 2020; De Bresser et al., 2022; 2024; Kaambwa et al., 2021; Lehnert et al., 2018; Milte et al., 2020) and willingness to provide, receive, or pay for informal care (e.g., De Jong et al., 2022; Hoefman et al., 2017; 2019; Mentzakis et al., 2011). Several other studies have asked respondents to make choices for hypothetical individuals (e.g., Amilon et al., 2020; Nieboer et al., 2010; Santos-Eggimann & Meylan, 2017). This study, instead, asks respondents to advise the government on what it should invest in and, therefore, about the LTC system for older people in 2040. This study thus elicits citizens' socially inclusive personal preferences, on a macro-level rather than on the individual level. Respondents' preferences may be different when reasoning from their personal perspective compared to reasoning from a socially inclusive personal perspective, as perhaps a broader set of factors would be included when making trade-offs from the latter perspective (e.g., Mouter et al., 2017; Nyborg, 2000). Both perspectives thus seem useful[1], yet, only few studies have adopted a

1 It is beyond the scope of this paper to provide an overview of the advantages and disadvantages of taking a socially inclusive personal perspective in preference-elicitation rather than a personal perspective, but several studies have discussed this issue in a variety of contexts (e.g., Costa-Font & Rovira, 2005; Dolan et al., 2003; Mouter et al., 2017; 2018; Nyborg, 2000; Russell et al., 2003). The national guidelines for economic evaluation in the health domain in the Netherlands prescribe taking a societal perspective (Versteegh et al., 2016).

socially inclusive personal perspective (e.g., Janus & Koslowski, 2020; Milte et al., 2024; Patterson & Reyes, 2024). Also, these studies typically took an attitudinal approach (i.e., asking respondents about their opinion) rather than a preference-based approach (e.g., asking respondents to make choices between different policy options in a choice experiment). Building on those studies, this study takes a preference-based approach, asking respondents to make choices between different LTC policies for older people. In addition, given the focus on the societal level, a broad sample of the adult population is included in this study (as the potential (tax)payers of the system, but also future care recipients and/or caregivers). Many previous studies are either based on samples consisting of older or middle-aged respondents (e.g., Lehnert et al., 2018; Nieboer et al., 2010; Santos-Eggimann & Meylan, 2017), while especially younger generations also have a stake in this policy dilemma.

Secondly, the study extends our knowledge on public preferences for LTC by eliciting preferences not only for policy alternatives, but also for the height of the overall public expenditure on LTC. In this Participatory Value Evaluation (PVE), respondents are presented with a constrained portfolio choice experiment. They are asked to select a portfolio of policy alternatives of their preference, requiring them to trade-off their private expenditure capacity with the level of public spending (Mouter et al., 2021b) (see section 2.2 'Constrained portfolio choice experiment').[2] This should approximate the situation of policymakers more closely than other preference elicitation methods, and provides additional insights into the preferred level of public spending on LTC.

We find that, overall, respondents derived positive utility from all policy alternatives. Also, the estimated effects of policy alternatives on the fulfilment of nursing care needs and reductions in need for informal caregiving played a significant role in respondents' choices. An optimal portfolio analysis underlines that respondents care about both effects, as each of the highest-ranked portfolios contains policy alternatives that affect both types of outcomes. Most respondents composed portfolios that would constitute substantial expenditure increases on LTC, which is both in line with previous studies (e.g., Amilon et al., 2020; Boxebeld et al., 2024; Milte et al., 2024) as well as partially at odds with recent policy developments in the Netherlands (e.g., Maarse &

2 This is particularly relevant given that several previous studies found that a substantial share of the public tends to believe LTC for older people should be funded publicly rather than privately (e.g., Janus & Koslowski, 2020; Patterson & Reyes, 2024; Simmons et al., 2024). Many people are, however, insufficiently aware of the opportunity costs of increased public expenditure (e.g., Cohen-Blankshtain & Sulitzeanu-Kenan, 2021; Persson & Tinghög, 2020), which may bias their preferences. Therefore, we decided to include an explicit opportunity cost of increased public expenditure in the choice task, in the form of a tax increase.

Jeurissen, 2016). We discuss the implications of these findings, and the additional research required, to inform policy action in LTC in ways that are aligned with public preferences. Also, we discuss how policymakers may use the findings of heterogeneous preferences by respondent characteristics to broaden the support base for particular policy measures.

The remainder of this paper is organized as follows: Section 2 describes the institutional setting, the study's methodology, survey design, and the estimation approach. Section 3 reports the results of the analyses, while Section 4 discusses these findings in the light of previous research, this study's limitations, and policy developments.

Methods

Institutional setting

The Netherlands is characterized by a universal and comprehensive LTC coverage, in which no private LTC insurance exists (Bakx et al., 2023). Also, LTC expenditure in the Netherlands as a proportion of gross domestic product (GDP) is the highest among OECD countries (OECD, n.d.). The comprehensive coverage and resulting large share of public expenditure on LTC makes the Netherlands vulnerable to population ageing (Bakx et al., 2023). In an attempt to curve this expenditure increase, a number of policy reforms have focused on promoting ageing-in-place. The most recent major reform took effect in 2015 and restricted access to institutional care, widened the availability of home-based care, and put greater emphasis on informal care (Maarse & Jeurissen, 2016). Nevertheless, the government has increased investments in institutional care since then to improve quality of care (Bakx et al., 2023) and to address the increasing LTC demand. More information on the institutional LTC setting of the Netherlands can be found, for instance, in Bakx et al. (2020; 2023), Bär et al. (2022), Bergeot & Tenand (2023), and Tenand et al. (2023).

Constrained portfolio choice experiment

PVE is a novel preference-elicitation method that can be characterised as a constrained portfolio choice experiment. After having been introduced in transportation (Mouter et al., 2021a) and environmental economics (Mouter et al., 2021b), the method is now applied in health economics, too (Boxebeld et al., 2024). In a PVE, respondents are faced with a policy question and presented with a single choice task, consisting of several policy alternatives that are all described by a set of attributes with randomly varying levels. They are asked to compose their preferred portfolio of policy alternatives to address the policy question, subject to a resource constraint (e.g., a public budget). The premise of PVE is that its portfolio-based choice task allows respondents to incorporate synergies between policy alternatives in their choices. Also, the resource constraint forces respondents to acknowledge the scarcity of resources that policymakers face in the context of specific policy issues. However, it is possible to deviate from a fixed constraint (e.g., a fixed budget) in the design of a PVE and allow respondents to choose for an adjustment of the public expenditure level on a policy area. In such a flexible-budget PVE, respondents trade-off the level of public expenditure with their private spending capacity (Boxebeld et al., 2024; Mouter et al., 2021b). This makes PVE a suitable method for the policy area of LTC for older people, in which multiple policy alternatives (e.g., different care arrangements) can be implemented simultaneously and both public and private resources can be allocated.

Choice task design

The selection of policy alternatives, attributes, levels and a resource constraint for this choice experiment was informed by a review of the literature and interviews with stakeholders and policy experts. An extensive description of these selections is provided in Appendix 5B. In addition, we conducted three rounds of pre-testing the design and a pilot study.

Three attributes were included in the choice task: 1) the effect of the policy alternatives on the percentage of older people in need of nursing care who receive this in 2040; 2) the costs of the policy alternatives, presented as a uniform increase of the tax burden for all adult citizens in 2040; and 3) the impact on the average amount of informal care required in 2040 (in hours per week per person). Table 5.1 provides a list of the policy alternatives with level ranges for each of the three attributes.

The status quo for the choice task presented to respondents concerned the scenario in which the supply of formal LTC services for older people in 2040 is maintained at current levels, while the demand is expected to increase substantially due to the projected population ageing. Respondents were informed that in this status quo, 65% of older people in need of nursing care would receive this in 2040 (compared to 95% now), while the population of 16 years and older would have to provide 12 hours of informal care per person per week (compared to 2 hours per week now) (see Appendix 5B). Implementing policy alternatives could mitigate these consequences; most policy alternatives either increased the capacity of nursing care or reduced the provision of informal care, while all policy alternatives required additional governmental investment (see Table 5.1). The resource constraint concerned this additional government investment: the chosen portfolio could not exceed an expenditure increase of €105 per adult per month (see Appendix 5B). On an aggregate level, this correspondents to an additional spending of €20 billion per year, which would approximately double current public expenditure on LTC for older people. According to recent government estimates (Rijksoverheid, 2023), this is the expenditure increase required for LTC for older people in 2040 given the expected rise in demand and income growth in the absence of any policy change, while maintaining accessibility at current levels. Increasing expenditures for LTC beyond this level is considered unrealistic.

Table 5.1. Overview of included policy alternatives, attributes and levels.

Policy alternative	**Attributes**		
	Fulfilment of nursing care needs (%)	**Costs (€ per adult per month)**	**Informal care provision (average N hours/week per adult)**
Increase capacity of nursing homes (by 10,000 places)	2, 4, 6	10, 15, 20	-1, -2
Increase capacity of nursing care at home (by 10,000 places)	2, 4, 6	5, 10, 15, 20	0
Increase use of supportive care technologies	2, 4, 6	5, 10, 15, 20	0
Introduce care homes (per 10,000 places)	0	10, 15, 20	-1, -2. -3
Increase capacity of social care at home (by 10,000 places)	0	5, 10, 15, 20	-1, -2, -3
Provide respite care to informal caregivers (by 3 months)	0	5, 10, 15	-1, -2. -3
Introduce compulsory social service for young adults (by 3 months)	0	5, 10, 15	-1, -2, -3

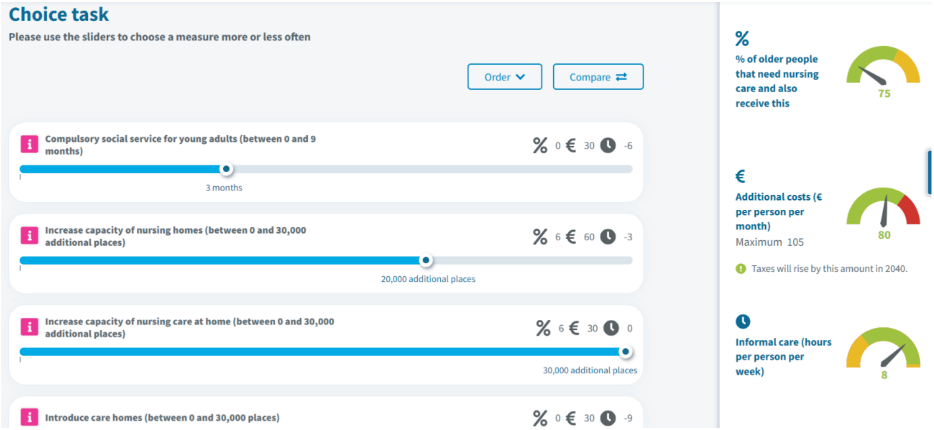
Overview of attribute level ranges by policy alternative. The distinctions between nursing and social care and between nursing homes and care homes are explained in Appendix 5B. See Appendix 5C for full descriptions of the attributes and policy alternatives as presented to respondents.

In the choice task, respondents thus faced clear trade-offs: they could increase the capacity of nursing care, so that more older people will receive the nursing care they need. Alternatively, they could reduce the required amount of informal care and alleviate the associated burden on informal caregivers. To a certain extent, trade-offs needed to be made between the policy alternatives fulfilling either of these needs, as both types of policy alternatives potentially exhausted the resource constraint. Finally, respondents could choose not to increase governmental spending on LTC, but this would result in waiting lists and welfare losses for older people with unfulfilled care needs as well as a substantial informal caregiving burden for the population at large. As such, each choice came with clear opportunity costs.

Respondents could choose each of the policy alternatives, which were presented in a random order to mitigate ordering effects (Boxebeld, 2024), between zero and three times by moving sliders. The attribute levels of each alternative were presented in the main screen, while supplementary information for each of the policy alternatives could be accessed via pop-up screens (see Appendix 5C). The total effects of their choices on the three attributes were presented in a dashboard on the right of the choice task screen (see Figure 5.1 for an exemplary (translated) choice task screen).

Due to the single choice task in a PVE, there is only experimental variation between (i.e., not within) respondents. A 'min-max correlation' design was generated using an algorithm that aims to minimize the maximum level of the correlation between different versions, resulting in 57 different versions of the choice task (i.e., different combinations of attribute levels). This design was created using the Python package PortChoice (Hernandez, n.d.). To force respondents to make trade-offs between the policy alternatives and the attributes, in each version the total costs of implementing all policy alternatives three times exceeded the maximum budget (€105 per person per month).

Figure 5.1. Exemplary screenshot of (part of) the choice task



Survey instrument

The survey, embedding the PVE choice task, was programmed in the software platform Wevaluate (Populytics, n.d.). Prior to the choice task, respondents were asked for their informed consent and were informed about the study objective, the policy question, and the choice task design. Also, at the start of the survey, a few screen-out questions for the quota sampling were presented, regarding the respondent's age group, gender, and education level. To induce value learning (i.e., familiarize respondents with the topic), we presented respondents with a few normative questions about the distribution of responsibilities for LTC prior to the choice task (see Appendix 5E). To induce institutional learning (i.e., familiarize respondents with the choice environment), the choice task was introduced in an instructional video.

After completing the choice task, respondents were asked to motivate their choices using an open-ended question and to provide information about their informal care experience and attitudes and sociodemographic and socioeconomic characteristics.

Data Collection and Sample Description

Respondents were recruited from an online panel (Dynata, 2022) and quote-sampled to be representative of the adult population of the Netherlands in terms of age, sex, and educational attainment. Data collection took place between June 18 and June 25, 2024 and resulted in 997 completed surveys. Descriptive statistics of the sample are presented and related to population-level statistics in Table A5.1 in Appendix 5A. The sample is roughly representative of the population in terms of gender and education level. In terms of age, older (65+) respondents are somewhat underrepresented.

All complete responses were included in the main analyses. Additionally, we conducted a sensitivity analysis, reported in Appendix 5F, in which responses of suspected low-quality (N=58, 5.8% of the total sample) were excluded. Its results support the robustness of the main results.

Estimation approach

First, we estimated a Multiple Discrete-Continuous Extreme Value (MDCEV) choice model. This model accounts for the discrete choices (i.e., whether policy alternatives are included in the chosen portfolio) and the continuous choices (i.e., how often a policy alternative is chosen) that respondents were facing in the choice task, as well as for the constraint (i.e., the budget restriction). The random utility function of the MDCEV model takes the following form (Bhat, 2008):

$$U(x) = \sum_{k=1}^{K} \frac{\gamma_k}{\alpha_k} \left[e^{\sigma(\beta' z_k + \varepsilon_k)}\right] \cdot \left\{\left(\frac{x_k}{\gamma_k} + 1\right)^{\alpha_k} - 1\right\}$$

where *U(x)* is the utility function with respect to consumption quantity vector *x*, consisting of *K* elements (i.e., the policy alternatives (*k*) in the choice task). The baseline marginal utility of each good (i.e., the marginal utility of each policy alternative at zero 'consumption') is represented by $e^{\sigma(\beta' z_k+\varepsilon_k)}$, in which β captures the marginal utility with respect to z_k, which is an attribute of the policy alternative k (or characteristic of the respondent), σ is a scale parameter, and ε_k is a stochastic error term. The translation parameter, γ_k (with $\gamma_k > 0$), allows for corner solutions (i.e., zero 'consumption' of a good). Finally, α_k (with $0 \leq \alpha_k \leq 1$), is a satiation parameter that allows for decreasing marginal utility of consumption (Bhat, 2008). More information on the MDCEV model is presented in Bhat (2008).

We specified the MDCEV model without outside good. Given the potential confounding between the translation and satiation parameters, joint estimation of γ and α is problematic. As a solution, different sub-utility functions (i.e., profiles) can be used (Bhat, 2008), of which we used the α-γ profile.[3] The MDCEV model allows us to examine the marginal utility respondents attached to the policy alternatives and attribute levels in their portfolio choices. Besides, using the preference estimates from the MDCEV model as inputs, we computed the expected utility of each portfolio (i.e., each possible combinations of policy alternatives and attribute levels), averaging over 1,000 repetitions with random draws for the stochastic error term (ε_k). Enumerating over all portfolios, this resulted in a ranking of portfolios with respect to their expected utility. The highest-ranked portfolios are most likely to maximize the expected utility of society given respondents' preferences and the present budget constraint (Dekker et al., 2024). The optimal portfolio analysis is presented as an alternative to the monetary valuation of policy alternatives and attribute level changes. In doing so, we follow Dekker et al. (2024), who adopt a social welfare function approach instead of a consumer surplus approach.[4] Moreover, Chandoevwit and Wasi (2020) argue that an analysis of demand, resembling the optimal portfolio computation presented here, better suits the needs of policy makers than marginal rates of substitution and willingness to pay estimates.

To explore preference heterogeneity, we examined the choice shares by respondents' characteristics. In addition, we conducted a Latent Class Cluster Analysis (LCCA). LCCA uses a factor model in which respondents are assigned to clusters based on simple indicators (i.e., their choices for the policy alternatives in the case of this study). Clusters are formed to maximize preference homogeneity within clusters and

3 The use of the α-γ profile is convenient given our use of the procedure by Pinjari and Bhat (2021), based on this profile, as implemented in Apollo.

4 They argue: "*The reason for doing so is that the PVE survey is already framed in the application context and the attractiveness of public sector projects can directly be quantified and compared in terms of citizens' cardinal utility without the need for monetary valuation.*" (Dekker et al., 2024, p. 2).

preference heterogeneity between clusters (Molin et al., 2016; Vermunt & Magidson, 2002). Since the clusters are latent, the number of clusters is unknown and should be determined by the analyst. A key criterion for this decision is the balance between model fit and model parsimony, which we assessed using the Bayesian Information Criterion (BIC). This criterion was supplemented by criteria regarding the interpretability and communicability of the model and the probabilistic cluster sizes (e.g., Molin et al., 2016). We estimated a three-step LCCA model (Vermunt, 2010). First, a set of models with up to ten clusters was estimated and the preferred model was identified. Next, all respondents were assigned with cluster membership probabilities and, finally, the association between these probabilities and respondents' characteristics (i.e., covariates) was examined (Vermunt, 2010).[5] The LCCA was based on a subsample of the data (N=928), whereby respondents who answered 'do not know' (for education level) or 'prefer not to say' (for all other covariates) were excluded from this analysis.[6]

The MDCEV model has been estimated using the BGW algorithm (Bunch et al., 1993) in the package Apollo (Hess & Palma, 2019) version 0.3.2. in R version 4.2.1. (R Core Team, 2022). The LCCA models were estimated using Latent GOLD 5.1 (Vermunt & Magidson, 2016).

Results

Mean preferences

Descriptive results

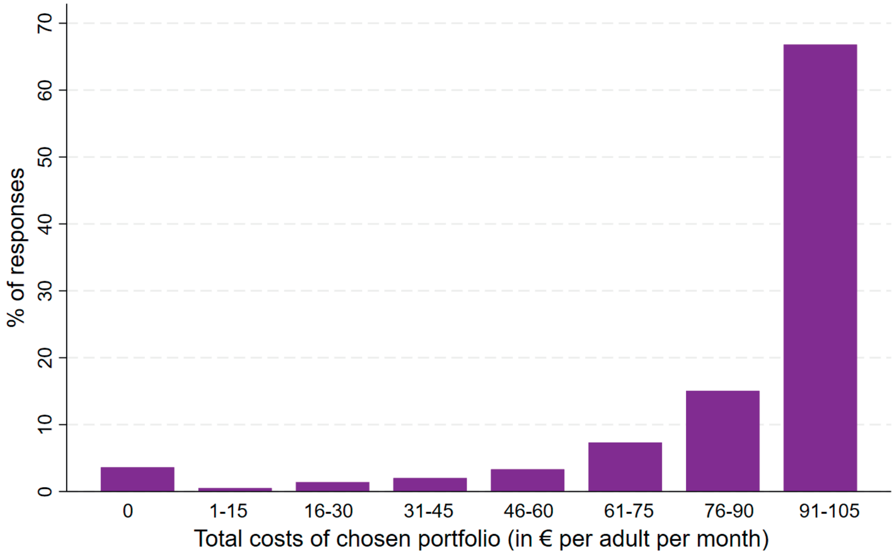
The descriptive analysis of the expenditure patterns resulting from respondents' portfolio choices shows that most respondents chose to increase the budget by (almost) the maximum amount that was possible and thus exhausted the resource constraint (nearly) entirely. As can be seen in Figure 5.2, two out of three respondents (N=666) chose portfolios that would constitute a public expenditure increase (through taxation)

5 To correct for the underestimation of the association between covariates and class membership probabilities that is typical to a three-step LCCA approach, a maximum likelihood-based correction method was applied. All LCCA models were estimated using 500 sets of random starting values and 500 iterations per set.

6 Including separate dummies for these answer categories in the models or imputing the most likely alternative answer given the other answers for that respondent were considered as alternative approaches. The first approach yielded many unidentified parameters given the small number of respondents choosing these answer options. The second approach was deemed undesirable because it requires strong assumptions, as we had limited other information on respondents.

of between €91 and €105 per adult per month. Almost half of this group (N=326) chose to exhaust the resource constraint entirely.[7] Less than 4% of respondents (N=36) did not choose any policy alternative and thus did not increase public expenditure on LTC for older people at all. On average, respondents' portfolio choices resulted in a public expenditure increase of about €89 (standard deviation: €24).

Figure 5.2. Distribution of the total costs of respondents' chosen portfolios



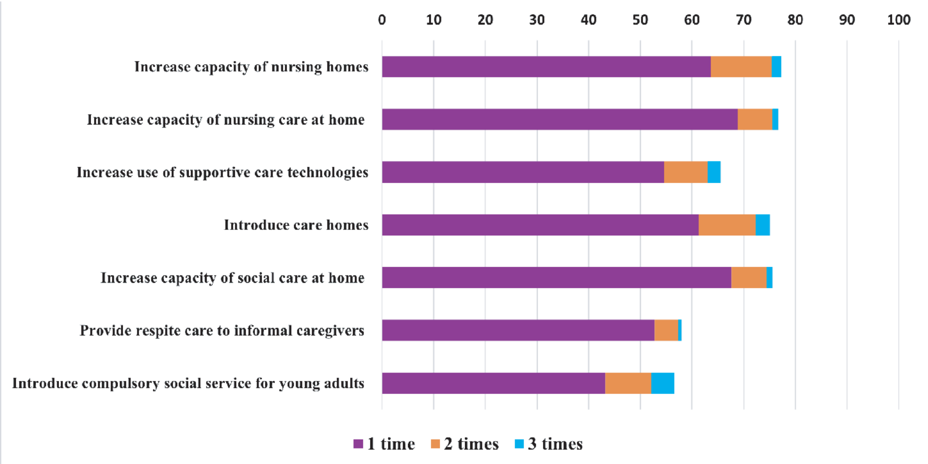
In this Figure, the different cost outcomes have been clustered together in groups of 15 for the sake of clearness of graphical display. In the choice experiment, a large number of cost outcomes was possible, depending on the design version and respondents' choices (see the histogram and Kernel density plot in Figure A5.1 in Appendix 5D).

The choice shares of the policy alternatives, presented in Figure 5.3, show that each alternative is included at least once in more than half of respondents' portfolios. Particularly two policy alternatives regarding nursing care (i.e., increasing the capacity of nursing homes, increasing the capacity of nursing care at home) and two policy alternatives regarding formal social care (i.e., introducing care homes, increasing the capacity of social care at home) were often chosen (by 75–77% of respondents). Two policy alternatives aimed at alleviating the burden on informal caregivers, namely providing respite care to informal caregivers and introducing compulsory social service

7 It was not possible/feasible for all respondents to come to the amount of €105 exactly, depending on the cost attribute levels of the design version they were presented with.

for young adults, were much less often included (by 57-58% of respondents). Regarding the number of times the policy alternatives are chosen (i.e., the intensive margin), Figure 5.3 shows that each policy alternative is typically chosen not more than once. Depending on the policy alternative, 5–14% of respondents chose that alternative two or three times. In all descriptive statistics, respondents' preferences for the policy alternatives and for the attribute levels are not yet disentangled. The MDCEV estimates presented below account for this.

Figure 5.3. Choice shares of the policy alternatives



The percentages of respondents by policy alternative who chose the alternative one, two or three times.

MDCEV estimates

Table 5.2 presents the estimated parameters of the MDCEV model. The policy alternative-specific parameters describe the relation between 'consuming' a policy alternative (i.e., choosing to allocate funding towards that policy alternative) and respondents' utility, independent of the attributes. All policy alternatives were significantly and positively associated with respondents' utility. The taste parameters indicate the association between the attribute levels and respondents' utility, independent of the policy alternatives. Both an increase in the fulfilment of nursing care needs as well as a reduction in the required amount of informal caregiving were significantly associated with respondents' utility. This suggests that respondents' choices are influenced by both attributes, and that respondents prefer to fund policy alternatives that increase the fulfilment of informal care and reduce the required amount of informal care provision.

Table 5.2. MDCEV estimates

Coefficient	Utility parameters (δ/β)	p-value	Translation parameters (γ)	p-value
Remaining budget	NA (fixed)		21.2201 (1.5074)	< 0.0001
Policy alternative-specific parameters				
Increase capacity of nursing homes	3.3404 (0.0740)	< 0.0001	0.8510 (0.0490)	< 0.0001
Increase capacity of nursing care at home	3.1650 (0.0719)	< 0.0001	0.7917 (0.0370)	< 0.0001
Increase use of supportive care technologies	2.7674 (0.0656)	< 0.0001	1.1136 (0.0473)	< 0.0001
Introduce care homes	3.2847 (0.0641)	< 0.0001	0.9368 (0.0423)	< 0.0001
Increase capacity of social care at home	3.3551 (0.0649)	< 0.0001	0.8108 (0.0359)	< 0.0001
Provide respite care to informal caregivers	2.6260 (0.0606)	< 0.0001	1.1024 (0.0380)	< 0.0001
Introduce compulsory social service for young adults	2.3545 (0.0653)	< 0.0001	1.3322 (0.0526)	< 0.0001
Taste parameters				
Additional 1% fulfilment of nursing care needs	0.0203 (0.0106)	0.0280		
Minus 1 hour of informal care provision	0.0748 (0.0184)	< 0.0001		
Scale parameter				
Scale (σ)	0.6112 (0.0101)	< 0.0001		
N	997			
LL(final)	-8929.33			
AIC	17894.65			
BIC	17982.94			

Robust standard errors in parentheses. P-values based on two-sided tests for the policy alternative-specific and scale parameters and one-sided tests for the taste parameters. Abbreviations: AIC=Akaike Information Criterion, BIC=Bayesian Information Criterion, LL(final)=Final log-likelihood, N=Number of observations (i.e., respondents).

5

Table 5.3. Optimal portfolio composition

Policy alternative	**Top 10 portfolios**									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Increase capacity of nursing homes	0	1	1	0	2	0	1	0	0	3
Increase capacity of nursing care at home	1	1	0	1	0	1	0	0	1	0
Increase use of supportive care technologies	0	1	1	1	2	1	2	1	2	0
Introduce care homes	2	0	2	0	0	1	1	3	0	0
Increase capacity of social care at home	1	1	0	0	0	1	0	1	1	3
Provide respite care to informal caregivers	3	2	3	3	2	3	3	0	3	0
Introduce compulsory social service for young adults	0	1	0	3	1	0	0	2	0	0

The top ten optimal portfolios within the budget constraint of €105 per adult per month of additional public expenditure. The bold numbers in black in the top row indicate the ranking of the portfolio, while the numbers in the other rows indicate the frequency of each policy alternative in each portfolio.

Optimal portfolio composition

Table 5.3 shows the ten portfolios with the highest expected utility. For example, portfolio 1 includes an increase in the capacity of nursing care at home by 10,000 places, the introduction of care homes with 20,000 places (i.e., two times 10,000 places), an increase in the capacity of social care at home by 10,000 places, and the provision of respite care to informal caregivers for a maximum of nine months (i.e., three times three months), while increase in nursing home capacity, increase in use of supportive care technologies and compulsory social service for young adults are not selected. Several patterns can be observed from the top ten portfolios. For example, each of these portfolios included at least one of the policy alternatives regarding nursing care and at least one regarding social care. Besides, all portfolios except portfolio 10 contained at least four of the seven policy alternatives. Additionally, increased use of supportive care technologies and provision of respite care to informal caregivers were included at least once in eight out of the ten highest-ranked portfolios. Finally, increasing the capacity of nursing homes and increasing the capacity of nursing care at home seemed strong substitutes: both policy alternatives were included four times while the other was not, and they were included together only once. All ten highest-ranked portfolios exhausted the resource constraint entirely.

Preference heterogeneity

Descriptive results

The choice shares at the extensive margin (i.e., whether a policy alternative is chosen or not) according to respondent characteristics are presented in Figures A5.2 and A5.3 in Appendix 5D. Considerable variation was found across policy alternatives, with more heterogeneity for increased use of supportive care technologies and introduction of compulsory social service for young adults. This heterogeneity was most pronounced between age groups: younger respondents included the increased use of supportive care technologies much more often than middle-aged and older respondents, while older respondents included the introduction of compulsory social service for young adults considerably more often than younger and middle-aged respondents. Because respondent characteristics may be correlated, these descriptive statistics should not be taken as more than a first indication of preference variation. To address this, we incorporated all respondent characteristics simultaneously as covariates in the LCCA discussed below.

LCCA estimates

After estimating ten cluster models, it became clear from the model fit statistics (presented in Table A5.3 and Figure A5.4 in Appendix 5D) that the BIC was minimized for models with between two and four clusters. These models were inspected more closely. Even though the model with three clusters comes with a slightly lower BIC value, the model with four clusters was considered more easily interpretable and communicable.

Table 5.4 provides the estimates of the four-cluster LCCA model in terms of the choice shares at the extensive margin. Graphical presentations of choice shares by clusters for both the extensive and intensive margin are presented in Figures A5.5 and A5.6 in Appendix 5D. From these results, it becomes clear that two of the clusters have rather uniform preferences across the various policy alternatives: Cluster 1 has choice shares of 81–83% for all policy alternatives and chooses each alternative once on average (between 0.87 and 1.04 times). This cluster, which is the cluster with the largest probabilistic share, thus spreads out the available resources and chooses a diverse portfolio. Cluster 4, on the other hand, has very low choice shares (<10% on the extensive margin and <0.19 on the intensive margin) for all policy alternatives. This cluster, with the smallest probabilistic share, thus seems to invest only few additional resources on LTC for older people.

In contrast with Clusters 1 and 4, Clusters 2 and 3 differentiated their portfolio choices over alternatives. Both clusters choose more often for the institutional and home-based nursing and social care policy alternatives and increasing the use of supportive care technologies than for providing respite care for informal caregivers and introducing compulsory social service for young adults. A difference is that Cluster 3 has higher choice shares for the institutional and home-based nursing and social care alternatives (88–100%) than Cluster 2 (53–67%). The same pattern applies to providing respite care to informal caregivers (44% for Cluster 3 versus 29% for Cluster 2), while the opposite holds for introducing compulsory social service for young adults (24% and 41% for Clusters 3 and 2, respectively). The expenditure patterns arising from the clusters' preferences[8] (presented in Figure A5.7 in Appendix 5D) show that the mean costs of respondents' portfolio choices are higher for Clusters 1 and 3 (€97 and €95,

8 The expenditure patterns by cluster and choice shares at the intensive margin are derived post-hoc from the LCCA with choice shares at the extensive margin as indicators (i.e., rather than from separate LCCAs with other indicators). For this aim, sample weights based on cluster membership probabilities were applied to the descriptive statistics.

respectively) than for the overall sample (€89). Cluster 2 has a mean portfolio cost of €81, while this amounts to only €7 in Cluster 4.

Table 5.4. Estimation results of the LCCA model with four clusters

Policy alternative	Overall mean	Cluster 1 (51%)	Cluster 2 (23%)	Cluster 3 (21%)	Cluster 4 (5%)
Increase capacity of nursing homes	77	81	67	95	1
Increase capacity of nursing care at home	77	83	57	97	1
Increase use of supportive care technologies	65	81	54	56	3
Introduce care homes	75	82	58	88	9
Increase capacity of social care at home	76	82	53	100	6
Provide respite care to informal caregivers	58	81	29	44	2
Introduce compulsory social service for young adults	57	82	41	24	8

Prediction of indicators (in % of respondents who included an alternative at least once in their portfolio). Probabilistic cluster shares between brackets.

Age, gender, and having work experience in healthcare were the only respondent characteristics significantly associated with cluster membership probabilities (at the 95% level). Education level, informal care provision, self-assessed health, housing situation and self-reported financial situation did not significantly vary with cluster membership probabilities. Older respondents (age 65+) were much more likely than younger and middle-aged respondents to belong to belong to Cluster 1. Younger respondents (1 –34 years) were more likely than the two older age groups to belong to Cluster 2, and respondents of middle age (35–64 years) were somewhat more likely to belong to Cluster 4. Men were more likely than women to belong to Cluster 2 and somewhat more likely to belong to Cluster 4, while women were more likely to belong to Cluster 3. Finally, people who have worked in healthcare were more likely to belong to Clusters 1 and 2 than respondents who have never worked in healthcare.

Table 5.5. Cluster membership probabilities by respondent characteristics

Policy alternative	Cluster 1 (51%) (ref.)	Cluster 2 (23%)	Cluster 3 (21%)	Cluster 4 (5%)	Wald-test score	p-value
Prediction of cluster membership probabilities						
Age					13.042	0.042
18–34 years (ref.)	0	0	0	0		
35–64 years	0	- 0.535 (0.380)	0.292 (0.434)	0.729 (0.565)		
65+ years	0	- 1.062 (0.437)	- 0.571 (0.524)	0.198 (0.640)		
Gender					8.467	0.037
Man (ref.)	0	0	0	0		
Woman	0	- 0.421 (0.314)	0.569 (0.305)	- 0.217 (0.376)		
Work experience in healthcare					8.284	0.041
No (ref.)	0	0	0	0		
Yes	0	- 0.045 (0.337)	- 0.969 (0.358)	- 0.119 (0.377)		
Probabilistic distribution for the covariate levels over the clusters (%)						
Age						
18–34 years	45	33	19	4		
35–64 years	49	19	26	6		
65+ years	66	15	16	4		
Gender						
Man	51	27	17	6		
Woman	52	17	27	4		
Work experience in healthcare						
No	50	19	22	4		
Yes	55	27	13	5		

Probabilistic shares (for the clusters) and robust standard errors (for the coefficients) in parentheses. Please note that the percentages may not sum up to 100 for each covariate level due to rounding to integers. Ref.: Reference category. Other (non-significant) covariates included in the reported model were education level, provision of informal care, self-reported health, housing status, and self-assessed financial situation.

Conclusion and Discussion

Summary and discussion of results

In this study, we examined the preferences of a broad sample of citizens in the Netherlands for LTC policies for older people in 2040. In a constrained portfolio choice experiment, respondents composed a portfolio of policy alternatives, subject to a budget constraint of €105 of additional expenditure per adult citizen per month. Four main findings emerge from the study results.

Firstly, on average, respondents derive positive utility from all policy alternatives, and each of the seven policy alternatives is chosen by more than half of the respondents. Policy alternatives regarding institutional and home-based nursing and social care were most preferred, while respite care and compulsory social service for young adults were least preferred. Policy alternatives were typically chosen only once in a portfolio. This suggests a preference for distributing public resources towards multiple policy alternatives over investing substantially in one or two particular policy alternatives.

Secondly, the attributes played a significant role in respondents' choice behaviour, since respondents derived positive utility from fulfilment of nursing care needs and reductions in need for informal caregiving. This is also reflected in the optimal portfolio analysis, as all ten highest-ranked portfolios contain policy alternatives affecting both these outcomes. In the optimal portfolios, increased use of supportive care technologies and provision of respite care to informal caregivers were often included.

Thirdly, most respondents chose for portfolios that would require substantial increases in expenditures. All ten highest-ranked portfolios completely exhausted the budget constraint. If taken as consequential, this would indicate a substantial average willingness to pay (WTP) for additional LTC services for older people. This goes against recent policy developments in the Netherlands, relying more strongly on participation of families and the community in providing informal care at home (Maarse & Jeurissen, 2016). At the same time, this finding corresponds with the results of a recent choice experiment in the Netherlands on resource allocation over different healthcare purposes, in which respondents allocated most additional resources to LTC (Boxebeld et al., 2024). It also corresponds with findings of several recent studies in other countries (using different research designs) documenting support for increasing expenditure to improve and expand (access to) LTC services (e.g., Amilon et al., 2020; Janus & Koslowski, 2020; Milte et al., 2024).

Fourthly, the results show the existence of preference heterogeneity in our sample. The LCCA results suggest this is associated with respondents' age, gender and work experience in healthcare. Heterogeneity in choice shares was most pronounced for increasing the use of supportive care technologies and the introduction of compulsory social service for young adults.

Limitations and recommendations for future research

In the choice task, some policy alternatives were designed to affect the fulfilment of nursing care needs only, while others only affected the required amount of informal care provision. This design choice clarifies and reinforces the trade-off between fulfilling nursing care needs and fulfilling social care needs (with the latter plausibly alleviating the burden on informal caregivers more strongly[9]) in the choice task, which arguably made the task somewhat easier for respondents and reflects the imperfect substitutability of formal and informal care. However, it does not fully capture the complex reality of LTC, as increasing the capacity of nursing care at home and increasing the use of supportive care technologies may also substitute informal care partially (e.g., Anderson & Wiener, 2015).

Besides, while this study used a budget as the choice task constraint, one may argue that another type of constraint is perhaps more relevant. For example, in many countries (including the Netherlands), staff shortages are a pressing constraint to the capacity of the care system (OECD, 2023). We considered implementing the personnel capacity as a second constraint, but decided not to do so because it was difficult to operationalize and would increase the cognitive burden for respondents. Also, given that many people exhaust the budget constraint (almost) entirely, it may be questioned to what extent respondents anchor on the budget constraint. To examine this, future research may experimentally vary the height of the budget constraint between respondents.[10] Besides, a uniform tax increase for all adults was included as the payment vehicle in the choice task. Future research may examine the robustness of the elicited preferences to the priming of the opportunity costs of increased public expenditure (e.g., Persson & Tinghög, 2020) and to a different payment vehicle, such

9 Given that informal caregivers more often fulfil social care needs than nursing care needs, we assumed that formal social care (i.e., the alternatives of introducing care homes and increasing the capacity of social care at home) would induce a stronger substitution effect than formal nursing care.

10 Two previous PVE applications, using split-samples, varied a fixed and flexible budget between respondents and found that most respondents in the flexible-budget version did not adjust the height of public expenditure (Dekker et al., 2024; Mouter et al., 2021b), but this finding may be application-specific. Also, another PVE study varied the height of the budget constraint within respondents in a sequence of PVE choice tasks (Bahamonde-Birke et al., 2024), finding a high level of consistency, but also variability due to the budget change (Bahamonde-Birke, 2024). No study, thus far, has varied the height of the budget constraint between respondents, however.

as an alternative tax specification or the reallocation of existing public resources away from other spending purposes (e.g., Andersson et al., 2023).

A final important limitation of the choice task design, generally applicable to stated preference research, is a potential lack of perceived consequentiality and unfamiliarity of respondents with the topic and choice environment, which may influence respondents' preferences and could give rise to hypothetical bias (Haghani et al., 2021). To address these aspects, we included a consequentiality script (Lewis et al., 2016), stating our intention to share the results with the Ministry of Health, Welfare and Sports. Additionally, we provided respondents with concise background information about the policy issue and the various policy alternatives and attributes, warm-up questions prior to the choice task[11], and an instructional video. Nonetheless, we cannot exclude (or test for) the possibility of hypothetical bias influencing our results. Therefore, we encourage future research to elicit respondents' preferences in different policy contexts[12] and using different choice task designs, including a variety of resource constraints and payment vehicles, to further investigate the validity of our results.

Regarding limitations to the modelling of the choice data, it would have been interesting to include all considered respondent characteristics as interaction terms in the MDCEV models to disentangle preference heterogeneity regarding the policy alternatives from preference heterogeneity regarding the attribute levels in more detail. We have only included respondents' age here as an example (see Table A5.6 in Appendix 5G), because most other respondent characteristics are considered latent variables, which cannot be directly included in an MDCEV model. Instead, one could estimate an integrated choice and latent variable (ICLV) MDCEV model. However, such a model is highly prone to specification issues, difficult to interpret, and may not be suitable to derive policy implications from (e.g., Campbell & Sandorf, 2020; Chorus & Kroesen, 2014), which we considered a disadvantage given our intention to contribute to the public debate. Another alternative is estimating a latent class MDCEV (LC-MDCEV) model with respondent characteristics as covariates. Despite our use of a search algorithm implemented in Apollo incorporating 500 different sets of starting values based on a

11 These warm-up questions, together with their answers, are documented in Appendix 5E. As explained in Boxebeld et al. (2024), these questions are used to induce value learning, but may have influenced respondents' choices in the choice task.

12 Given the importance of the institutional context, the results are to some extent context-specific. Many countries provide a less comprehensive LTC coverage than the Netherlands (Bakx et al., 2023), for example, while the populations of many countries will be aging more strongly than the population of the Netherlands (Eurostat, n.d.).

procedure by Bierlaire et al. (2010), the LC-MDCEV models did not converge for our data. Therefore, we estimated and presented the results of the LCCA model instead.

Policy implications

Concluding, policymakers can use the study results to align their policy decisions in LTC more closely with citizens' preferences. The results suggest there is broad public support for a substantial expenditure increase in LTC to address the projected challenges by 2040. Policymakers are recommended to implement a diverse portfolio of policy alternatives, providing in both nursing and social care needs. The policy alternatives regarding the increased use of supportive care technologies and the provision of respite care to informal caregivers are particularly encouraged, conditional on the policies' effectiveness and efficiency in practice. Besides, the study results may provide directions for governments to broaden the support base for specific policy alternatives. For instance, governments interested in the increased use of supportive care technologies may aim to better understand the reasons for the lower support among middle-aged and older respondents. To this aim, citizens' preferences and broader attitudes towards LTC for older people would need to be further explored, for instance by attending to the motivations underlying these preferences.

Appendix 5A: Descriptive sample statistics

Table A5.1. Study sample compared with the general population in terms of sociodemographic characteristics

Sociodemographic characteristic	Total sample (N=997)		LCCA sample (N=928)		Population [b]
	(N)	(%)	(N)	(%)	(%)
Age					
18–34	266	26.7	234	25.2	27.2
35–64	538	54.0	514	55.4	47.5
65+	191	19.2	180	19.4	25.3
Prefer not to say	2	0.2	-	-	-
Gender [a]					
Man	484	48.5	455	49.0	49.4
Woman	509	51.1	471	50.8	50.6
Non-binary	3	0.3	2	0.2	
Prefer not to say	1	0.1	-	-	-
Education level					
No university (of applied sciences)	647	64.9	599	64.5	64.0
University (of applied sciences)	349	35.0	329	35.5	35.4
Do not know	1	0.1	-	-	0.5

a) Respondents were asked for their gender identity, but the descriptive statistics for the general population are based on registered gender/sex, which is a different but rather strongly correlated concept.
b) Descriptive statistics for the general population aged 18 and older for June 2024 (for gender and age) and aged 15 and older for the second quartile of 2024 (for education level) were retrieved from Statistics Netherlands (n.d.$_a$; n.d.$_b$).

Appendix 5B: Selection of alternatives, attributes, levels and constraint

The selection of alternatives, attributes, levels and constraint is based on a review of the literature and five interviews to verify the findings from the review and inform the operationalization in the experiment. These interviews were held in March and April 2022 with a managing director of a large long-term care (LTC) supplier and board member of an LTC umbrella organization, policy officers from the Ministry of Health, Welfare and Sport of the Netherlands, and experts in the field of (long-term) care from the Council of Public Health & Society of the Netherlands, the Netherlands Institute for Social Research, and the Netherlands Scientific Council for Government Policy.

Policy alternatives

In the selection of policy alternatives, we considered that one may distinguish between nursing care and other types of caregiving tasks, from here labelled as 'social care'. Nursing care requires specialized training and is therefore typically supplied by adequately trained professionals, while social care may be provided by either formal or informal caregivers. For both types of care, we included several policy alternatives in the choice task that capture the relevant trade-offs.

For nursing care, we included three policy alternatives: increasing the capacity of nursing homes, increasing the capacity of nursing care at home, and increasing the use of supportive care technologies. Firstly, one may choose to increase the capacity of nursing homes. In nursing homes, care-dependent older people receive both nursing care as well as social care from professionals in an institutional setting. A major reform of the LTC system for older people in the Netherlands in 2015 has restricted the access to nursing homes considerably, so that only older people with severe physical or mental health conditions are admitted to nursing homes nowadays (Maarse & Jeurissen, 2016). Currently, there is place for 130,000 older people in nursing homes (which is about 0.7% of the total population of the Netherlands, or 7.5% of the population of age 75 and older). In the experiment, respondents could choose to increase the nursing home capacity by up to 30,000 places in steps of 10,000 places. Secondly, respondents could choose to increase the capacity of nursing care at home. Under this policy alternative, professional caregivers provide nursing care to care-dependent older people at their home. Also, for this alternative, respondents could choose to raise the capacity by up to 30,000 places in steps of 10,000 places.

Thirdly, there is the policy alternative of increasing the use of supportive care technologies. This policy alternative concerns technological innovations that assist professionals in their current tasks or complement current care by helping the care recipient with practical matters or providing surveillance when caregivers are absent (and not technologies that would completely replace professional caregiving, such as self-operating care robots). Such technological innovations are considered a promising way of enhancing the productivity of professional caregivers (Mosca et al, 2017). Potential threats relate to ethical dilemmas and privacy issues (e.g. Dickinson et al., 2021; Tian et al., 2025; Yew, 2021).

Social care involves tasks such as assistance with household tasks or companionship. For this type of LTC for older people, we included four policy alternatives. Firstly, there is the policy alternative of introducing care homes. In care homes, older people are residing in an institutional setting where professionals provide them with social care. A distinction with nursing homes is that the intensity of care is lower, as no nursing care is provided, and the admission threshold therefore also lower, than for nursing homes. In contrast with the institutional social care that existed before the 2015 LTC reform in the Netherlands, the funding of the housing component of care homes is private: residents pay rent, while the social care component is (partially) funded collectively. Respondents could choose to increase the capacity of this type of care by up to 30,000 places in total, in steps of 10,000 places. Secondly, there is the policy alternative of increasing the capacity of social care at home, in which professional caregivers provide social care to older people at their homes. Again, respondents could choose to raise the capacity by up to 30,000 places in terms of 10,000 places.

Thirdly, the policy alternative of providing respite care to informal caregivers was presented to respondents. In the case of respite care, professionals take over caregiving tasks from informal caregivers, to alleviate them from the burden of caregiving, giving them the opportunity to recover and increase the potential duration of their care provision. Unlike social care at home, respite care is provided only temporarily, after which informal caregivers are expected to step in again. Respondents could choose to implement this policy for a maximum duration of respite care of nine months, in steps of three months.

Finally, the policy alternative of a compulsory social service for young adults was presented. Similar to a military conscription, young adults would be obliged to provide a few months of (unpaid) community service, which may take the form of providing informal care to older people. An experiment took place in the Netherlands in recent years, during which young people could voluntarily provide such social service (Ministry of Education, Culture and Science of the Netherlands, 2024). The government considers extending the program to include compulsory social service, too. The effects of compulsory social service for a longer period are largely unknown. Some studies have evaluated compulsory social service programs of a shorter period for adolescents and/or young adults (e.g. for 40 hours in an entire year), which are in place in several countries. These studies came to mixed results, including null findings (e.g., Kim & Morgül, 2017), a higher probability of volunteering after having participated in such a program (e.g. Haski-Leventhal et al., 2010), and a negative experience of such a program, that may lead young people to be unmotivated for voluntary work in the long run and weaken their citizenship identity (Warburton & Smith, 2003). Respondents could choose to implement this policy for a maximum duration of compulsory community service of nine months, in steps of three months.

The policy alternatives included in the choice task facilitate respondents to express their preferences for the mix of institutionalized and home-based nursing and social care that should be provided to older people in the Netherlands in 2040 and, in case of social care, the mix between formal and informal caregivers. Respondents were informed that public investment comes with trade-offs that differ between the types of care. In case of nursing care, the trade-off relates to the relationship between public expenditure and the fulfilment of care needs; an increase of public expenditure increases the supply of professional nursing care and thus the fulfilment of nursing care needs, thereby reducing waiting lists and unmet care needs (and potentially improving the well-being of the care-dependent older population), but also results in a rise of the tax level. In case of social care, the trade-off relates to the relationship between public expenditure and informal care provision; an increase of public expenditure increases the supply of professional social care (or, in case of compulsory social service, provides an alternative to 'traditional' informal care), which likely reduces the need for informal care (e.g., Hollingsworth et al., 2022; Miyawaki et al., 2020). The provision of informal care may come with a substantial burden for caregivers in terms of a reduction in health and wellbeing (e.g., Bom et al., 2019; Stöckel & Bom, 2022) and economic opportunity costs in terms of productivity losses in the workplace (Keita Fakeye et al., 2023) and foregone formal employment and earnings (Heitmüller & Inglis, 2007; Schmitz & Westphal, 2017; Brimblecombe et al., 2020; Josten et al., 2024). Although a reduction in the required amount of informal caregiving may therefore be (perceived as) desirable, the public expenditure required for this results in a tax level increase in the choice experiment.

Attributes and levels

Three attributes were included in the design, capturing the different estimated effects of implementing the different policy alternatives. As a first attribute, the effect on fulfilment of nursing care needs was included. In the dashboard on the right of the choice task screen, respondents were presented with a meter showing the percentage of nursing care needs in 2040 that is fulfilled by the supply of nursing care (see Figure 1, main text). This percentage was 65% at the start and could go up by 2 to 6 %-point each time one of the policy alternatives related to nursing care was chosen (see Table 5.1). This attribute was not affected by selecting one of the four policy alternatives related to social care. Secondly, the effect on informal care provision was included as an attribute. This was operationalized as the average number of hours of informal care provision required per person (of 16 years and older) per week, which was again shown as a meter in the choice task. This number was 12 hours per adult per week at the start and decreased by 1 to 3 hours every time the respondent chose one of the four policy alternatives regarding social care or the policy alternative regarding the increase of the capacity of nursing homes.[13] The other two policy alternatives did not affect the provision of informal care.

Finally, there is the cost attribute. This concerns the public expenditure required to adopt the policy alternatives selected by the respondent and was also included as a meter in the choice task, starting at zero.[14] A choice for one of any of the policy alternatives raised the public expenditure on LTC for older people by €5 to €20 per adult per year. A key consideration in specifying the range of levels for each policy alternative is that we assumed that institution-based care arrangements (i.e. increasing the capacity of nursing homes, introducing care homes) are generally more costly than home-based care arrangements. This assumption is based on the finding by Krabbe-Alkemade et al. (2020) that a shift in the Dutch LTC system for older people from institutionalized and formal care to home-based and informal care led to a reduction in the growth of governmental LTC expenditure. Taking into account the uncertainty surrounding our assumption[15], the level ranges of institutional and home-based care policy options in the choice experiment were largely overlapping. Furthermore, the policy alternative concerning the compulsory social service was not costly in itself but may come with implementation and enforcement costs.

Constraint

In the design, a constraint was included. This constraint restricted participants' level of freedom in choosing a preferred portfolio and was related to public expenditure: that is, participants were faced with a limited public budget. Every choice for a policy alternative would constitute an increase of public expenditure, and respondents would therefore need to accept a tax increase. The additional expenditure allowed is capped at €105 per adult per year. On the country-level, this amount equals an expenditure increase of €20 billion per year, which is approximately the amount required to uphold

13 The underlying rationale is that, in contrast with the increased capacity of nursing care at home and the increased use of supportive care technologies, institutional nursing care (i.e., in nursing homes) also comes with social care for its residents and may therefore crowd out informal social care.

14 It should be noted that this attribute concerns the public expenditure on LTC exclusively. Moving to a broader evaluation perspective, e.g. including spillover effects to public expenditure in other domains or the wider economic impact of the policy alternatives, may lead to different conclusions. For instance, several studies suggest that not being admitted to institutional LTC leads to an increase in use of medical care (Bakx et al., 2020), while an expanded availability of publicly funded home-based care and nursing homes leads to a reduction in medical care use (Costa-Font et al., 2018; Moura, 2022).

15 An earlier study, for instance, resulted in deviating findings (Kok et al., 2015).

current levels of accessibility of the LTC system in 2040, taking into account the anticipated rise in care demands and price levels (Rijksoverheid, 2023). Going beyond this is considered unrealistic. Thus, participants needed to incorporate the financial constraint that policymakers face, but also to trade-off the level of public expenditure on LTC for older people with their own private spending capacity.

Status quo

Respondents were presented with a status quo (i.e., the starting point of the choice task). This status quo represented the baseline scenario with the current capacity of LTC. Respondents were told that there is currently capacity for 130,000 older people in nursing homes, 12,000 for nursing care at home, and 31,000 for social care at home. In 2040, to meet the same care capacity, a higher expenditure would be needed given price level increases. The National Institute of Public Health and the Environment of the Netherlands (RIVM, 2022) estimated that the total spending on LTC for older people amounted to approximately €18.64 billion in 2021. Based on a projected average yearly price level increase of 0.8% for the period between 2021 and 2030 and 0.4% between 2030 and 2040, they estimated that, to meet the current level of accessibility of LTC for older people in 2040, a total expenditure of approximately € 21 billion (i.e., (18,640,000,000 * 1.008^10) * 1.004^10 = 21,008,176,486) would be required.

Since the demand for LTC for older people will increase substantially, from about 180,000 older people in need of nursing care now to about 260,000 in 2040, the percentage of care needs fulfilled by formal care will decrease in the status quo scenario. It is estimated that, with the current capacity, the percentage of nursing care needs fulfilled by formal care will decline to 65% in 2040, while it is estimated to be around 95% at present.[16] Given this drop in formal care capacity relative to the demand, the required amount of informal care provision increases under the status quo scenario. The average number of hours per week per week spent on informal caregiving per citizen is estimated to amount to two currently[17], and this rises to twelve hours per week under the status quo scenario in 2040.

---

16 We base this on the estimates of 130,000 places in nursing homes, 12,000 places for extensive home-based care, 31,000 places for less extensive home-based care, and about 180,000 indications related to LTC for older people (i.e., people who qualify for receiving publicly provided LTC for older people) currently, and about 260,000 indications by 2040 (Hinkema et al., 2020).

17 According to Statistics Netherlands (2023), about 13% of the Dutch population of 16 years and older provides informal care, for about 13 hours per week. This equals to (13 * 13 / 100 = 1.69) approximately 2 hours per week on average for the entire 16+ population (numbers are rounded for respondents' convenience).

Appendix 5C: Pop-up screens in online choice experiment

Table A5.2. Information provided in the pop-up screens (translated to English)

Policy measure	Information in pop-up information screen
Increase capacity of nursing homes	In nursing homes, there is only room for older people in need who can no longer live independently at home and require nursing care. In addition to nursing care, residents of nursing homes also receive other care, such as meals and recreational activities. **Measure** If the government does not take any measures, there will be capacity for 130,000 older people in need in nursing homes. This means that there will not be enough space for all the older people who require nursing care in 2040. Consequently, there will be long waiting lists. Additionally, more informal care will be needed for the older people on the waiting list. Each time you choose this measure, capacity for 10,000 extra people will be added. **What is the effect of this measure?** By expanding the number of spots in nursing homes, more older people in need will receive the nursing care they require. As a result, the waiting lists will decrease, and less informal care will be needed.
Increase capacity of nursing care at home	With nursing care at home, older people in need receive nursing care in their own homes. This involves essential care provided by community nurses, who visit regularly (for example, daily). The older person in need is responsible for arranging other types of care, such as household help or recreational activities, possibly through informal caregivers or social care at home. **Measure** If the government does not take any measures, there will be space for about 12,000 older people in professional home-based nursing care. This means there will not be enough capacity for all the older individuals who require nursing care in 2040. Consequently, there will be long waiting lists. Each time you choose this measure, capacity for 10,000 extra people will be added. **What is the effect of this measure?** By expanding the number of spots in professional home-based nursing care, more older people in need will receive the nursing care they require. As a result, the waiting lists will decrease.

Table A5.2. Information provided in the pop-up screens (translated to English) *(Continued)*

Policy measure	Information in pop-up information screen
Increase use of supportive care technologies	Technology is increasingly being used in healthcare. In this choice task, we are only considering supportive technology; this does not refer to technology that replaces caregivers, but to technological tools that make the work of caregivers easier or complement the current care. **Measure** With this measure, the government provides funding to long-term care organizations to purchase supportive technology. The government also establishes regulations for the use of this technology, for example, to protect the privacy of older people, and ensures that these regulations are complied with. **What is the effect of this measure?** If the government promotes the use of supportive technology in long-term care for older people, the pressure on professional caregivers will decrease. As a result, they will be able to help more older people. This means that more older people will receive the nursing care they need. Consequently, the waiting lists will decrease.
Introduce care homes (*presented in the choice task as 'assisted living apartments'*)	In assisted living apartments, older people who need help with household tasks and personal care reside. They also receive meals and potentially recreational activities. However, they do not receive nursing care in assisted living apartments; for this, they would need to move to a nursing home. **Measure** In assisted living apartments, the residents have their own independent apartment, for which they pay rent. However, the government covers (or contributes to) the cost of the support and care that the older people receive there. Currently, there are only a small number of assisted living apartments available. By choosing this measure, more assisted living apartments will be created. Each time you select this measure, capacity for 10,000 additional people will be added. You can choose to create capacity for up to 30,000 extra people in total. **What is the effect of this measure?** By expanding the number of assisted living apartments, less informal care will be needed.

Table A5.2. Information provided in the pop-up screens (translated to English) *(Continued)*

Policy measure	Information in pop-up information screen
Increase capacity of social care at home	With social care at home, older people in need receive domestic help and other care in their own homes. This care is provided by community nurses, who visit regularly (for example, daily). However, this does not include nursing care; for this, older people would need to use nursing care at home or move to a nursing home. **Measure** If the government does not take any measures, there will be capacity for about 31,000 older people social care at home. This means there will not be enough capacity for all the older people who need domestic help and other social care by 2040. Consequently, there will be long waiting lists. Since older people on the waiting list still need this help and care, more informal care will be required. Each time you choose this measure, capacity for 10,000 additional people will be added. **What is the effect of this measure?** By expanding the number of spots in social care at home, less informal care will be needed.
Provide respite care to informal caregivers (*presented in the choice task as 'temporary replacement of informal caregivers'*)	Many people who provide informal care do so in addition to their paid work, studies, household tasks, etc. As a result, they often find it challenging to provide informal care. Sometimes, it is helpful for them to be able to temporarily step back from their caregiving duties. During this time, they are replaced by professional caregivers. This is known as respite care. **Measure** With this measure, the government makes it possible for people who provide long-term informal care to older people to temporarily hand over their caregiving duties to professional caregivers. This allows informal caregivers to maintain more balance in their own lives and often sustain their caregiving duties for a longer period. Each time you choose this measure, informal caregivers are granted the right to temporary replacement for an additional three months. **What is the effect of this measure?** By making it possible for informal caregivers to temporarily transfer their caregiving duties to professional caregivers, less informal care will be needed.

Table A5.2. Information provided in the pop-up screens (translated to English) *(Continued)*

Policy measure	Information in pop-up information screen
Introduce compulsory social service for young adults	In the Netherlands, some political parties have proposed introducing compulsory social service for young adults. Under this measure, all young adults (for example, from the age of 18) would be required to temporarily contribute to society full-time, such as by working in long-term care for older people. **Measure** Currently, young adults can choose to temporarily make a voluntary contribution to society, but this is not yet compulsory. By choosing this measure, it will become compulsory for young adults to temporarily make an unpaid contribution to society. Each time you select this measure, the compulsory social service will be extended by three months. **What is the effect of this measure?** By choosing this measure, more people will be available to provide care for older people. As a result, other informal caregivers will need to provide less informal care.

Appendix 5D: Additional results

Figure A5.1. Histogram and Kernel density plot of the distribution of the total costs of respondents' chosen portfolios

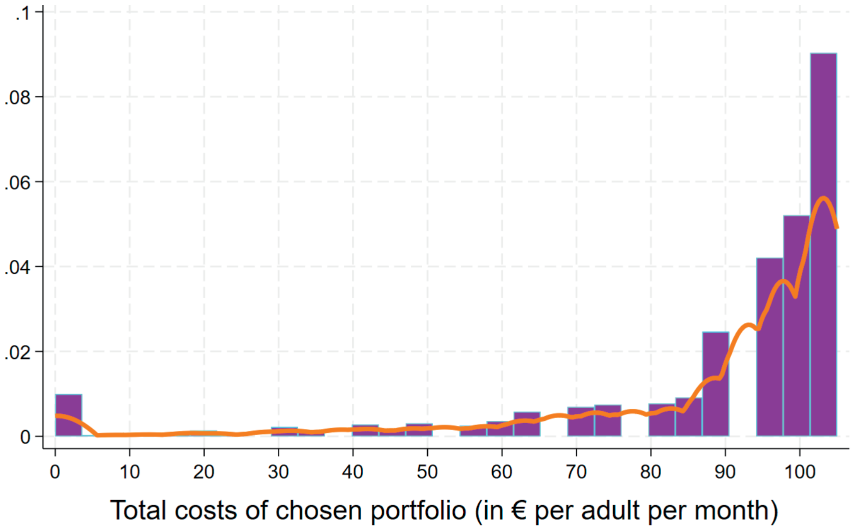


Figure A5.2. Choice shares of the policy alternatives (at the extensive margin) by selected respondent characteristics

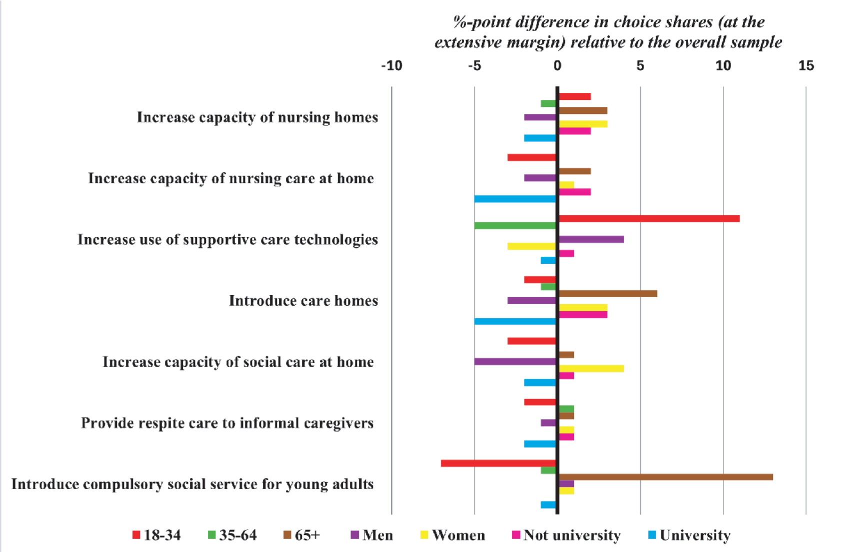


Figure A5.3. Choice shares of the policy alternatives (at the extensive margin) among selected subsamples of respondents

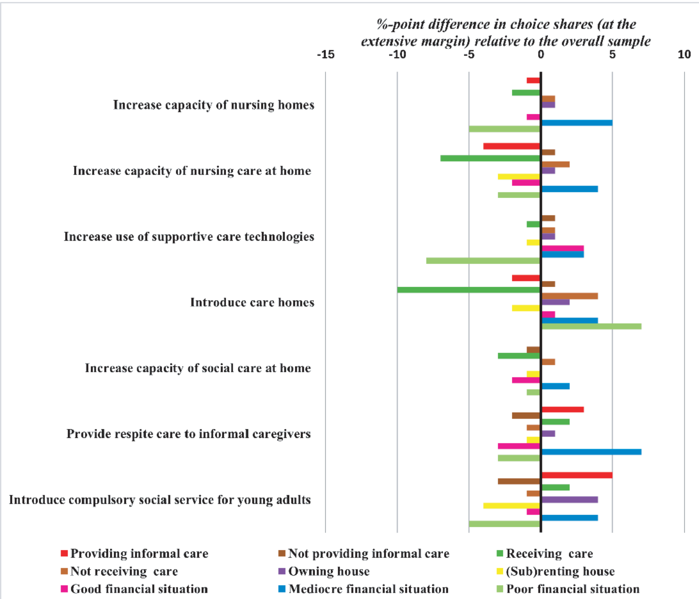
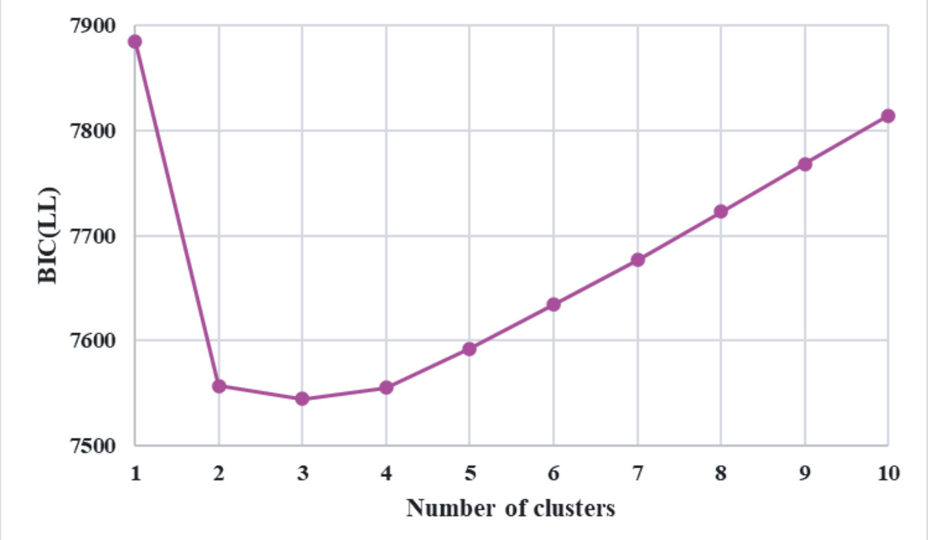


Table A5.3. Model fit statistics of the estimated LCCA models.

No. of clusters	No. of parameters	Log-Likelihood	BIC(LL)
1	7	-3918.17	7884.18
2	15	-3727.22	7556.94
3	23	-3693.77	7544.70
4	31	-3671.52	7554.86
5	39	-3662.82	7592.13
6	47	-3656.55	7634.24
7	55	-3650.58	7676.98
8	63	-3646.02	7722.52
9	71	-3641.59	7768.32
10	79	-3637.22	7814.26

BIC(LL): Bayesian Information Criterion (based on Log-Likelihood)

Figure A5.4. Trend in the balance between model fit and model parsimony over the number of clusters in the LCCA models



BIC(LL): Bayesian Information Criterion (based on Log-Likelihood)

Figure A5.5. Choice shares of the policy alternatives at the extensive margin by cluster

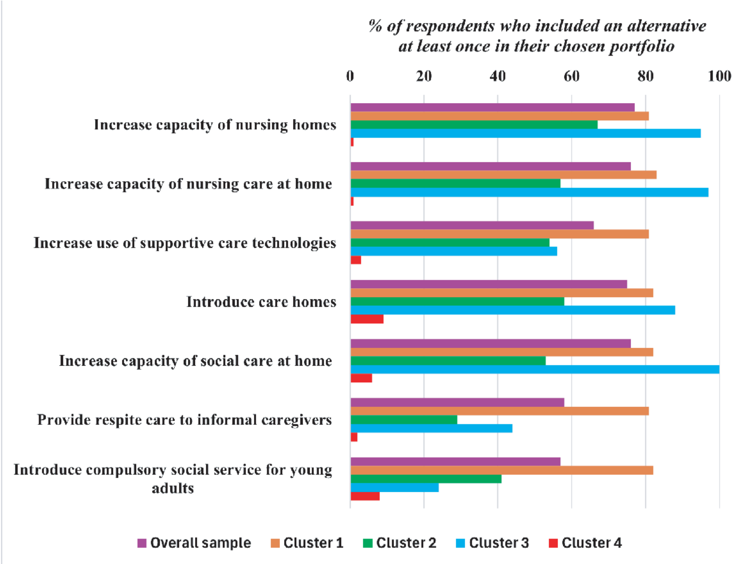
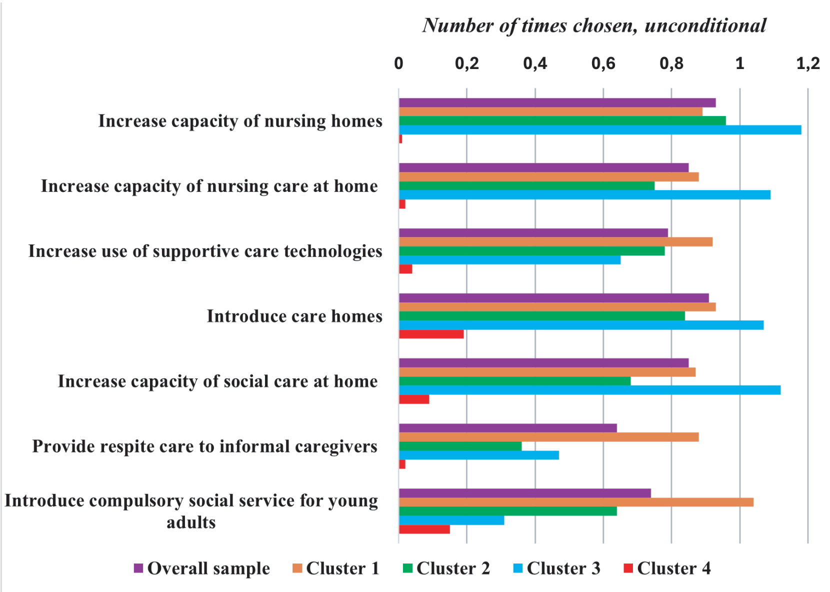
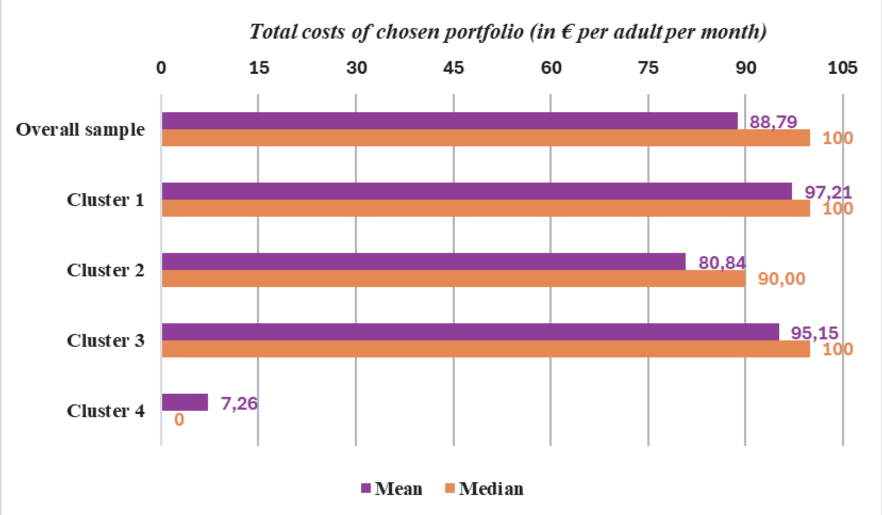


Figure A5.6. Choice shares of the policy alternatives at the intensive margin by cluster



5

Figure A5.7. Total costs of respondents' chosen portfolios by cluster



Appendix 5E: Warm-up questions

The following 'warm-up' questions were presented to respondents prior to the choice task, in an attempt to induce value learning (i.e., activate respondents to think about the topic and learn about their own preferences). Some of these questions are based on a study by Van Ooijen et al. (2017).

Questions

Who should bear the responsibility if someone at an older age needs care and support at home, the government or the individual and his/her family? Please give your opinion for each of the four types of care and support below [answer options: only the government; mostly the government; both equally; mostly the individual and his/her family; only the individual and his/her family; prefer not to say]:

- Necessary nursing and supportive care (e.g., help with getting dressed and bathroom visits, administering medication)
- Recreational activities under supervision (e.g., daytime activities, help with hobbies)
- Domestic help (e.g., preparing food, cleaning)
- Housing (e.g., providing a stairlift or custom bed)

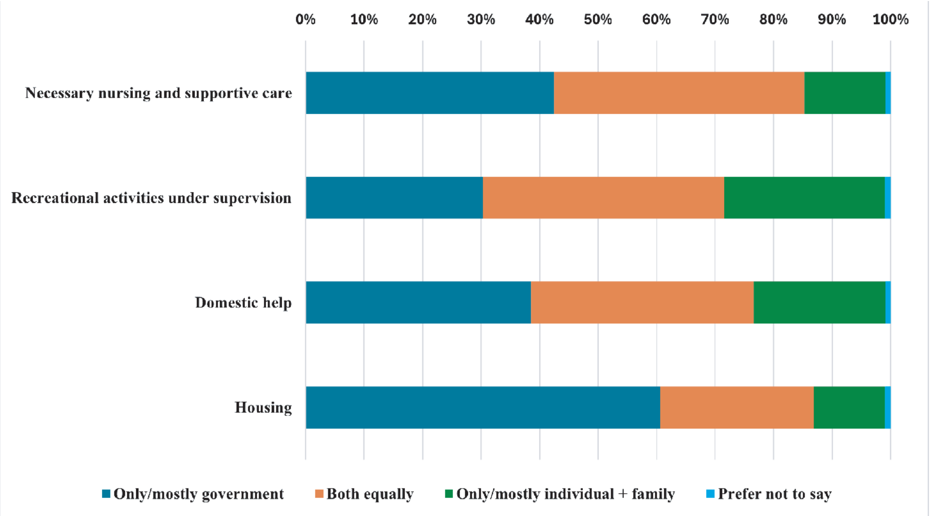
Do you think providing informal care within one's own environment should be a free choice or an obligation? [answer options: a free choice; an obligation; prefer not to say]

Do you currently need care or support from others yourself? (You may choose multiple answers) [answer options: yes, for necessary nursing and supportive care; yes, for recreational activities under supervision; yes, for domestic help; no; prefer not to say; other, namely:...]

Where would you prefer to receive care and support if you would be a care-needing older person? [answer option: within my own home as much as possible; within a care institution as much as possible; within my own home and within a care institution are equally preferred by me; prefer not to say]

Answers

Figure A5.8. The responses to the warm-up question about the distribution of responsibilities for long-term care for older people



Here, the answer options 'Only the government' and 'Only the individual and his/her family' have been merged with 'Mostly the government' and 'Mostly the individual and his/her family', respectively.

Figure A5.9. The responses to the warm-up question on whether informal care provision should be a free choice or an obligation

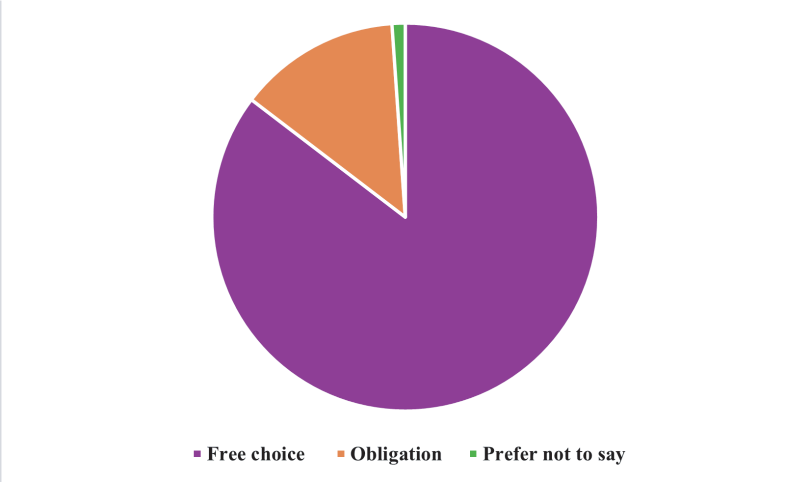
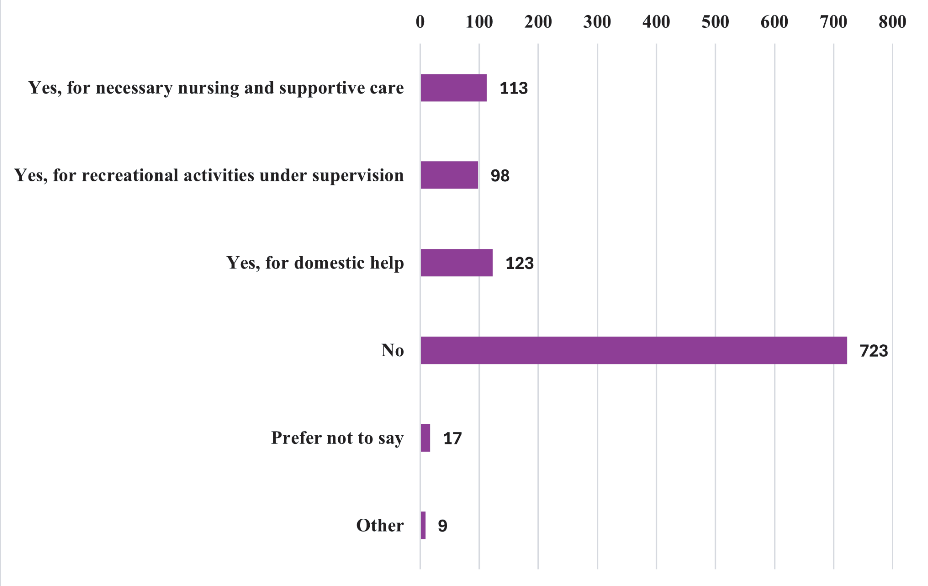
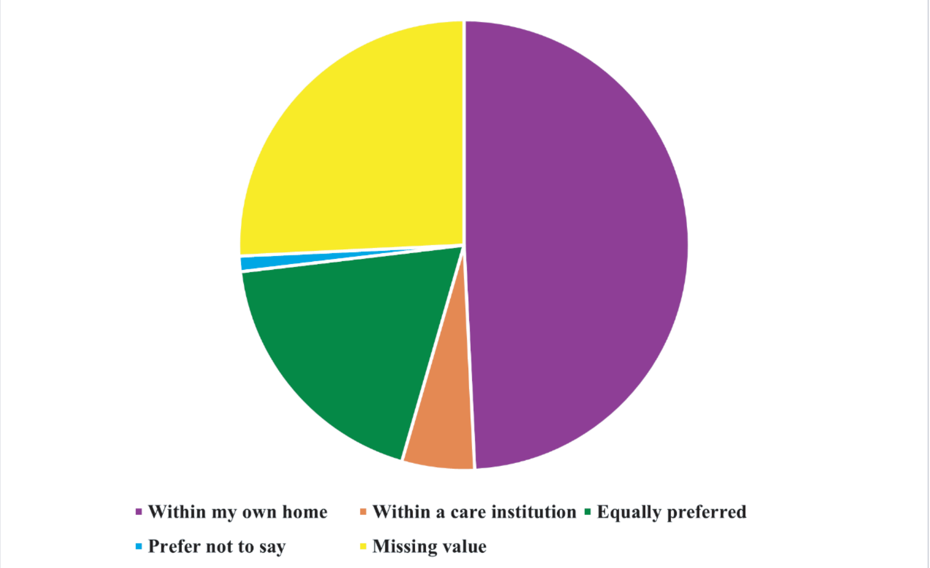


Figure A5.10. The responses to the warm-up question about care recipience



Respondents could choose multiple answer options. N=997

Figure A5.11. The responses to the warm-up question about preferences for receiving home-based or institutional care



Appendix 5F: Sensitivity analysis

Completion time statistics of total sample

Median duration: 8.967 minutes; **Mean** duration: 12.106 (standard deviation: 14.284)
Min. duration: 1.3 minutes; **Max.** duration: 231.05 minutes

Exclusion of suspected low-quality responses

As a sensitivity analysis, we repeat the analyses of the mean preferences (i.e., the descriptives, MDCEV model, and optimal portfolio analysis) from the paper on a restricted sample, from which responses of suspected low quality have been excluded. The 10% respondents with the shortest survey completion time (N=100), with the threshold value of completion time being 4 minutes and 13 seconds, were screened for their answers to the open-ended motivation question. In case they provided nonsensical answers to this question, they were considered to have provided responses of a low quality and were excluded from the sample for the sensitivity analysis. This applied to 58 respondents (i.e., 5.8% of the total sample used for the main analyses), yielding a restricted sample of 939 respondents for the sensitivity analysis.

Sensitivity analysis

Figure A5.12. Histogram and Kernel density plot of the distribution of the total costs of respondents' chosen portfolios, reduced sample

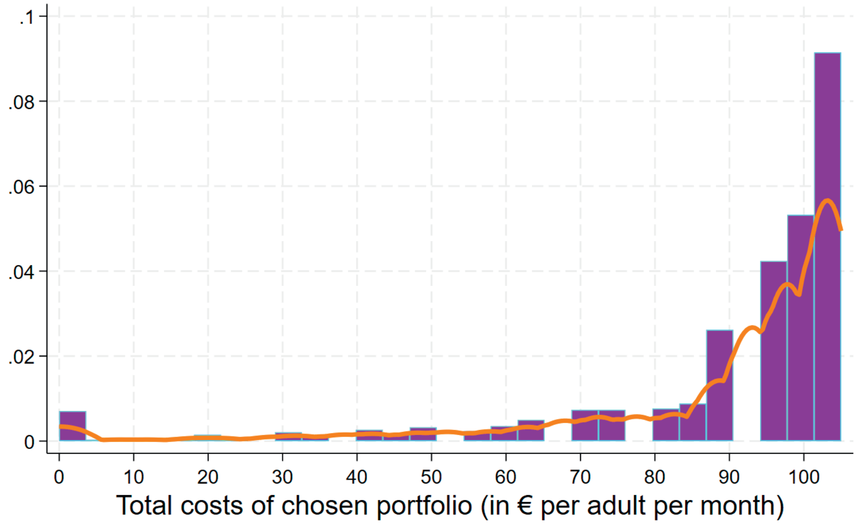


Figure A5.13. Distribution of the total costs of respondents' chosen portfolios, reduced sample

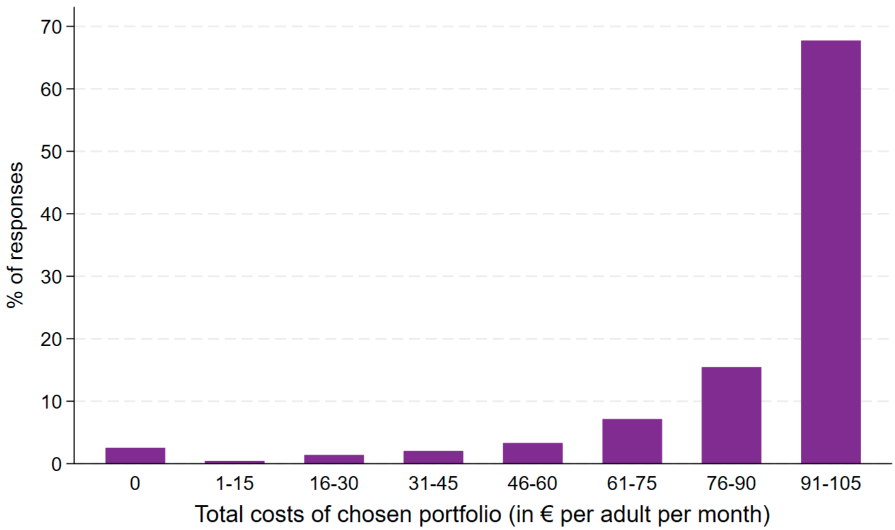
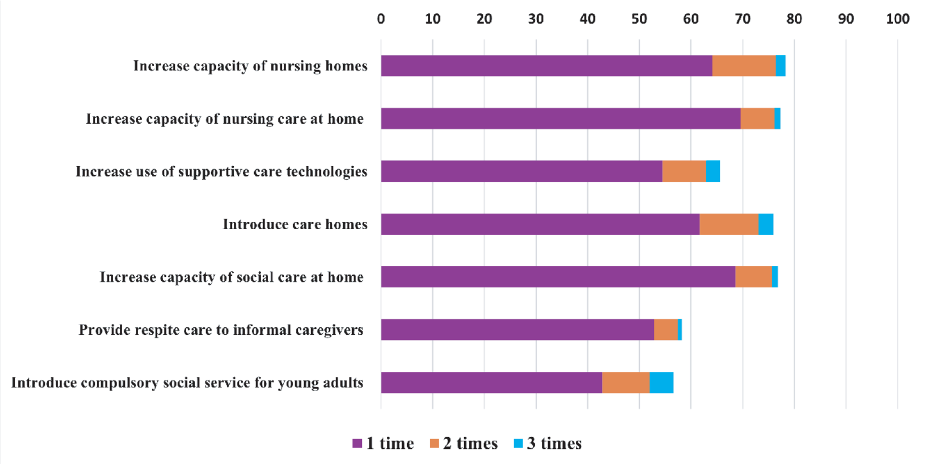


Figure A5.14. Choice shares of the policy alternatives, reduced sample



The percentages of respondents by policy alternative who chose the alternative one, two or three times.

Table A5.4. MDCEV estimates, reduced sample

Coefficient	Utility parameters (δ/β)	p-value	Translation parameters (γ)	p-value
Remaining budget	NA (fixed)		19.8385 (1.3713)	< 0.0001
Policy alternative-specific parameters				
Expansion of nursing home capacity	3.3432(0.0764)	< 0.0001	0.8517 (0.0425)	< 0.0001
Expansion of professional home-based nursing care	3.1616 (0.0739)	< 0.0001	0.7957 (0.0383)	< 0.0001
Deployment of supportive care technologies	2.7563 (0.0676)	< 0.0001	1.1398 (0.0495)	< 0.0001
Clustered homes with supportive care	3.2801 (0.0662)	< 0.0001	0.9431 (0.0441)	< 0.0001
Expansion of professional home-based supportive care	3.3605 (0.0670)	< 0.0001	0.8046 (0.0371)	< 0.0001
Respite care	2.6124 (0.0626)	< 0.0001	1.1108 (0.0394)	< 0.0001
Social conscription for young adults	2.3462 (0.0673)	< 0.0001	1.3488 (0.0549)	< 0.0001
Taste parameters				
Additional 1% fulfilment of nursing care needs	0.0196 (0.0109)	0.0360		
Minus 1 hour of informal care provision	0.0708 (0.0188)	< 0.0001		
Scale parameter				
Scale (σ)	0.6099 (0.0104)	< 0.0001		
N	939			

Table A5.4. MDCEV estimates, reduced sample *(Continued)*

Coefficient	Utility parameters (δ/β)	p-value	Translation parameters (γ)	p-value
LL(final)	-8441.33			
AIC	16918.66			
BIC	17005.86			

Robust standard errors in parentheses. P-values based on two-sided tests for the policy alternative-specific and scale parameters and one-sided tests for the taste parameters. Abbreviations: AIC=Akaike Information Criterion, BIC=Bayesian Information Criterion, LL(final)=Final log-likelihood, N=Number of observations (i.e., respondents).

Table A5.5. Optimal portfolio composition, reduced sample

Policy alternative	Top 10 portfolios									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Increase capacity of nursing homes	0	0	2	0	0	1	0	0	1	3
Increase capacity of nursing care at home	1	0	0	2	0	0	1	0	0	0
Increase use of supportive care technologies	0	0	0	0	2	2	1	1	0	0
Introduce care homes	2	1	0	0	2	1	1	0	1	0
Increase capacity of social care at home	1	1	2	2	1	0	1	1	2	3
Provide respite care to informal caregivers	3	3	1	2	0	3	3	3	2	0
Introduce compulsory social service for young adults	0	3	2	1	2	0	0	3	1	0

The top ten optimal portfolios within the budget constraint of €105 per adult per month of additional public expenditure. The bold numbers in black in the top row indicate the ranking of the portfolio, while the numbers in the other rows indicate the frequency of each policy alternative in each portfolio.

Appendix 5G: MDCEV with interactions

Table A5.6. MDCEV estimates with interactions for age

Coefficient	Utility parameters (δ/β)	p-value	Translation parameters (γ)	p-value
Remaining budget	NA (fixed)		21.5265 (1.6036)	< 0.0001
Policy alternative-specific parameters				
Increase capacity of nursing homes	3.2431 (0.1402)	< 0.0001	0.8791 (0.0439)	< 0.0001
* Age 35 – 64	0.0812 (0.1627)	0.6177		
* Age 65+	0.1169 (0.1927)	0.5441		
Increase capacity of nursing care at home	2.9787 (0.1354)	< 0.0001	0.8117 (0.0390)	< 0.0001
* Age 35 – 64	0.2037 (0.1555)	0.1902		
* Age 65+	0.2713 (0.1846)	0.1418		
Increase use of supportive care technologies	2.9107 (0.1210)	< 0.0001	1.0812 (0.0483)	< 0.0001
* Age 35 – 64	- 0.1992 (0.1423)	0.1616		
* Age 65+	- 0.1119 (0.1790)	0.5317		
Introduce care homes	3.1641 (0.1186)	< 0.0001	0.9396 (0.0439)	< 0.0001
* Age 35 – 64	0.0943 (0.1360)	0.4880		
* Age 65+	0.2330 (0.1634)	0.1538		
Increase capacity of social care at home	3.2798 (0.1221)	< 0.0001	0.8049 (0.0371)	< 0.0001
* Age 35 – 64	0.0868 (0.1385)	0.5311		
* Age 65+	0.1214 (0.1660)	0.4648		
Provide respite care to informal caregivers	2.5282 (0.1196)	< 0.0001	1.1034 (0.0393)	< 0.0001
* Age 35 – 64	0.0845 (0.1397)	0.5450		
* Age 65+	0.1724 (0.1709)	0.3131		

Table A5.6. MDCEV estimates with interactions for age *(Continued)*

Coefficient	Utility parameters (δ/β)	p-value	Translation parameters (γ)	p-value
Introduce compulsory social service for young adults	2.1702 (0.1313)	< 0.0001	1.3140 (0.0544)	< 0.0001
* Age 35 – 64	0.1446 (0.1532)	0.3453		
* Age 65+	0.4696 (0.1881)	0.0125		
Taste parameters				
Additional 1% fulfilment of nursing care needs	0.0295 (0.0220)	0.0893		
* Age 35 – 64	- 0.0170 (0.0266)	0.5231		
* Age 65+	0.0049 (0.0311)	0.8736		
Minus 1 hour of informal care provision	0.0696 (0.0403)	0.0421		
* Age 35 – 64	0.0133 (0.0478)	0.7805		
* Age 65+	- 0.0459 (0.0571)	0.4217		
Scale parameter				
Scale (σ)	0.6098 (0.0106)	< 0.0001		
N	928			
LL(final)	-8290.39			
AIC	16652.78			
BIC	16826.77			

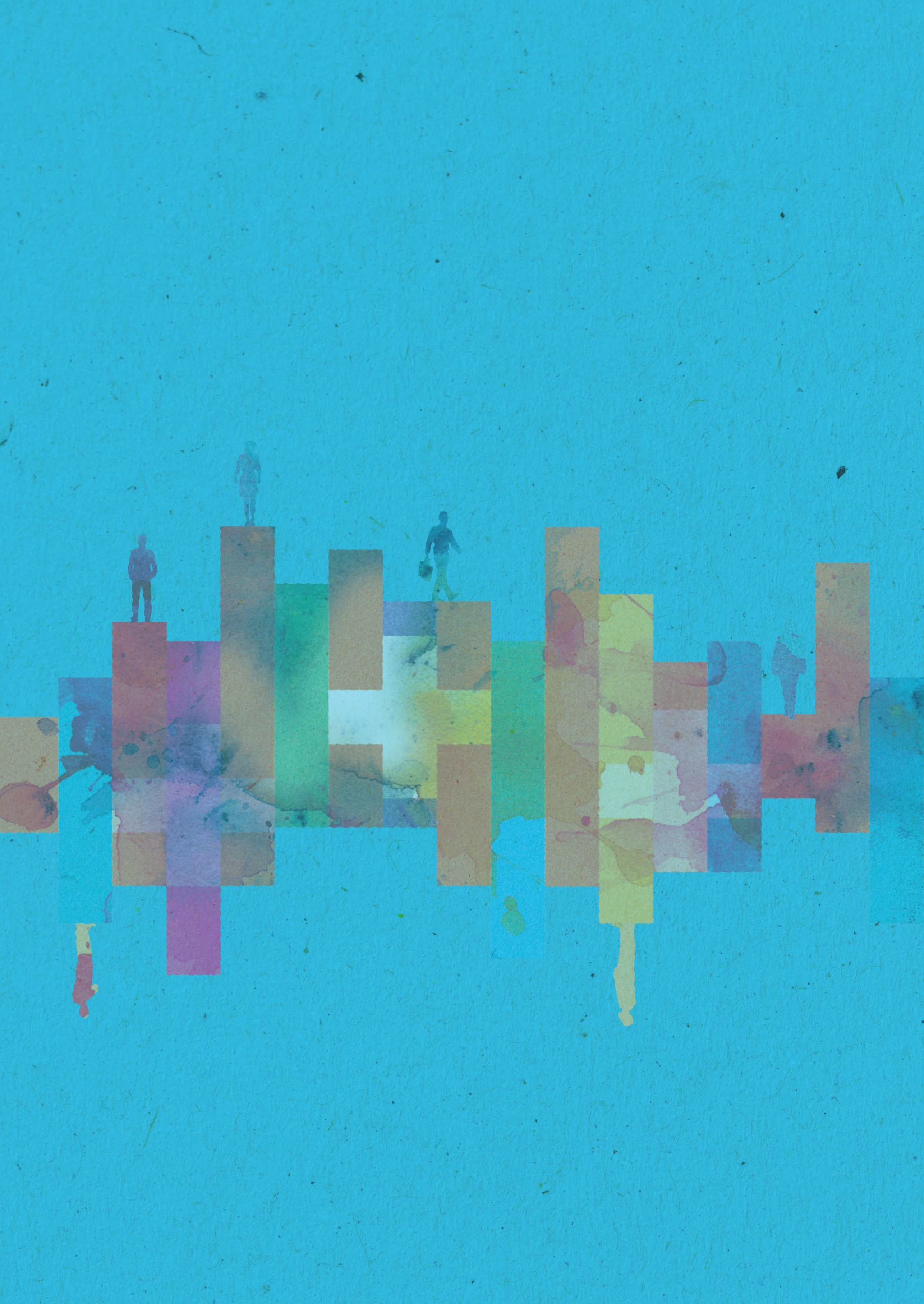
MDCEV model with interactions for age (reference category: age 18 – 34), estimated on the dataset that was also used for the LCCA. Robust standard errors in parentheses. P-values based on two-sided tests for the policy alternative-specific, interaction and scale parameters and one-sided tests for the taste parameters. Abbreviations: AIC=Akaike Information Criterion, BIC=Bayesian Information Criterion, LL(final)=Final log-likelihood, N=Number of observations (i.e., respondents).

Chapter 6

Do the choice of payment vehicle and priming of opportunity costs affect expenditure preferences and consequentiality perceptions in a choice experiment?

Based on:

Boxebeld, S., Mouter, N. and Van Exel, J. (2025). Do the choice of payment vehicle and priming of opportunity costs affect expenditure preferences and consequentiality perceptions in a choice experiment?



Chapter 7

General discussion

The aim of this dissertation was to advance the literature on the elicitation of public preferences for health policies. To meet this aim, the three main objectives were: 1) to position Participatory Value Evaluation (PVE) relative to other more commonly used multi-attribute preference-elicitation methods used in the health domain; 2) to examine the endogeneity of preferences elicited in a discrete choice experiment (DCE) or PVE to design characteristics of the choice experiment; and 3) to explore public preferences for health policy alternatives from a citizen perspective using DCE and PVE. In this Chapter, the main findings of the five studies addressing these three objectives (Chapters 2 – 6) are summarized, after which the strengths and limitations of the research carried out are discussed. Furthermore, this Chapter provides recommendations for future research and policy and concludes the dissertation.

Summary of main findings

The position of PVE (objective one)

PVE is a relatively new preference-elicitation method and, therefore, still unknown to many researchers and policymakers. To improve researchers' understanding of the method and its relative advantages and disadvantages in informing health policy decisions, **Chapter 2** introduced PVE in the health domain. By discussing the existing PVE literature, it illustrated the variation between applications in terms of design characteristics. This suggests that researchers have considerable flexibility to tailor a PVE design to the policy question at hand, particularly with respect to the number and nature of relevant resource constraints. A conceptual comparison with a selection of established multi-attribute preference-elicitation methods shed more light on the potential advantages and disadvantages of PVE as compared to other more commonly used methods in the health domain. On the one hand, the portfolio choice task in PVE allows respondents to evaluate synergies between alternatives and may better resemble the reality of policymakers. Also, the incorporation of an explicit resource constraint acknowledges scarcity of resources and enables researchers to elicit respondents' preferences for policy alternatives and the extent to which resources are allocated to the policy problem simultaneously. On the other hand, the single choice task in PVE may result in a higher cognitive burden on respondents and requires larger samples for achieving a similar level of statistical accuracy than the other methods. Finally, this Chapter underscored the need for more research on the feasibility and validity of PVE, also in relation to established preference-elicitation methods.

Effects of design characteristics (objective two)

To synthesize existing insights on ordering effects and promote the diffusion of these insights across the different domains of application of DCEs, **Chapter 3** conducted a cross-domain review of the literature on the impact of the presentation order of alternatives, attributes and choice sets on respondents' choices in a DCE. The majority of the 85 included studies found statistically significant ordering effects. Alternative and attribute ordering effects are caused by lexicographic behaviour, while choice set ordering effects are suggested to be the result of learning, fatigue, or anchoring. The review provided the applied literature with an overview of mitigation methods, including the randomization of presentation orders, advance disclosure of DCE core elements, and inclusion of alternative-specific constants (ASCs), attribute level overlap, and an Instructional Choice Set (ICS). Also, Chapter 3 provided the literature with several directions for methodological research to further our understanding of respondents' processing of DCEs, such as the study of heterogeneity in ordering effects by respondent and design characteristics.

Since conducting this research, two new studies have appeared that examine specific questions on ordering effects more closely. The first study (Rudolph et al., 2024) examined the effect of completely randomizing versus randomizing blocks of related attributes, relative to not applying any randomization in a choice experiment. They grouped the nine attributes in their choice experiment in five blocks and compared the results between complete attribute randomization, randomization of the five blocks, and partial randomization of the five blocks (i.e., with the "theoretically important" blocks being presented first and the order of the other blocks randomized). They found fewer and smaller differences in cognitive burden indications between the three versions than expected. Seeing that the perceived difficulty of the choice tasks was high across all versions, which may be due to the relatively high number of attributes and/or the difficult study topic (i.e., weapon exports), it remains unclear to what extent this result can be generalized. The second study (Dvorak et al., 2023) applied a Bayesian estimation of a choice model to account for preference updating over the course of the DCE sequence, with the aim to adjust for choice task ordering effects resulting from anchoring. They found that this cognitive model of choice behaviour better fitted the data, removed choice task ordering effects, and resulted in different willingness-to-pay (WTP) estimates, which suggests that such a modelling approach could form an effective method to mitigate choice set ordering effects. Nevertheless, as such a model is unlikely to mitigate choice task ordering effects due to fatigue, this method should be considered as a complementary mitigation method.

Furthermore, previous studies typically found respondents' stated preferences in a choice experiment to be endogenous to the specified payment vehicle (i.e., a tax increase or reallocation of existing public resources). In these studies, however, the outside good was typically undefined or limited to one or two goods. Therefore, in its examination of the impact of the payment vehicle on respondents' expenditure preferences, **Chapter 6** asked respondents to indicate their preferences towards the outside good, with the aim of reducing hypothetical bias and raising the realism of the choice task. Besides, previous studies tested the effect of an opportunity cost priming text in a single-alternative choice task. This Chapter tested the effect of an opportunity cost priming question on respondents' expenditure preferences in a multiple-alternative choice task. By asking respondents which private spending purposes they would economize on, we embedded the taxation payment vehicle in their household budgets. Finally, this Chapter examined the impact of the specified payment vehicle and the inclusion of an opportunity cost priming method on respondents' perception of the real-world impact (i.e., the consequentiality) of the choice experiment, which has not been studied before.

To examine the impact of the payment vehicle and opportunity cost priming on expenditure preferences and consequentiality perceptions, Chapter 6 fielded three versions of the PVE application on long-term care (LTC) reported in Chapter 5. The expenditure distributions were found to be highly consistent across the three versions, with no significant differences in both the overall sample and various restricted samples. Also, the observed differences between survey versions in consequentiality perceptions were mostly not significant and generally small. These counterintuitive findings lend themselves to various potential explanations, the plausibility of which arguably varies, and ask for a further investigation of respondents' perceptions of the substitution mechanisms implied by the different payment vehicles.

Public preferences for health policy alternatives (objective three)

While citizens' preferences towards individual skin cancer prevention measures have received attention in the academic literature, little is known regarding their preferences for collective skin cancer prevention policies. To explore public preferences for skin cancer prevention policies, **Chapter 4** fielded a discrete choice experiment in three European countries: Austria, the Netherlands, and Spain. The results showed that almost all attributes significantly influenced respondents' choices, with the tax attribute being most important in each country. Among the six included policy measures, a reduction in the price of sunscreen was most preferred, while prohibition of solar bed sales and of solaria were least preferred. This suggests that respondents favoured the least intrusive policy measures over the most intrusive ones. Respondents in the Netherlands valued information campaigns and the free provision of a skin cancer detection app less than respondents in the two other countries. Overall, though, the preference structure was rather similar across countries. Most respondents would recommend their government to take policy action regarding skin cancer prevention, and the levels of support were higher after the DCE than before.

To elicit public preferences for LTC for older people in the Netherlands in 2040, **Chapter 5** made use of the PVE method. In contrast with previous studies, this Chapter takes a citizen perspective and a preference-based approach in doing so. The study found that respondents derived positive utility from all seven policy alternatives. Each of the policy alternatives was chosen by more than half of the respondents, but typically only once, suggesting a preference for distribution resources over a variety of policy alternatives rather than substantially investing in one or two particular policies. Besides, respondents derived positive utility from increases in the fulfilment of nursing care needs and decreases in need for informal caregiving. This was also reflected in the optimal portfolio analysis, since the ten highest-ranked portfolios all contained policy alternatives affecting both outcomes. Respondents' portfolio choices would require a substantial increase in LTC expenditure, which is both in line with several previous studies as well as at odds with recent policy developments. Finally, there was preference heterogeneity in our sample, particularly regarding the increased use of supportive care technologies and the introduction of compulsory social service for young adults.

Strengths and limitations

This dissertation comes with several strengths and limitations. One of the strengths is the high societal relevance of both topics of application, future long-term care arrangements and skin cancer prevention. These applications aim to inform policy discussions and should therefore be well-aligned with the practice of policymakers. To this aim, the choice experiments have been developed based on consultation with experts and policymakers and extensive reviews of the scientific literature and existing policy measures and proposals, ensuring a sufficient embedding in the scientific literature as well as the policy practice. Another strength is the combination of these applied studies, with an emphasis on societal relevance, and methodological studies, with an emphasis on scientific relevance. The latter studies aimed to improve future preference-elicitations by informing researchers' choice of elicitation method, helping researchers to better understand and mitigate ordering effects, and considering the role of the payment vehicle and opportunity costs in the choice task.

Furthermore, advantage was taken from recent developments in different domains that make frequent use of choice experiments and choice models. Because the research landscape of choice modelling is somewhat fragmented (Haghani et al., 2021), new insights may not diffuse easily across different domains of application. This dissertation attempted to incorporate recent developments in choice modelling from different streams of the literature. A clear example is the application of PVE and MDCEV models, introduced in transportation and environmental economics, to health policy questions. But there are more examples; for example, while almost all mixed multinomial logit (MMNL) applications in health economics specify each random parameter to be normally distributed (Buckell et al., 2025), researchers in other domains have argued that this is often not a behaviourally realistic assumption for attributes for which we expect a direction of preference (e.g., treatment effectiveness, adverse side-effects and risks, costs) (e.g., Train & Sonnier, 2005). Chapter 4, therefore, carefully considered such alignment of behavioural expectations and model specification and, additionally, made use of a recent development from environmental economics to mitigate unrealistic WTP values (i.e., exploding implicit prices) (Crastes dit Sourd, 2024). Chapter 6 bridges streams of literature from transportation and environmental economics (i.e., on payment vehicles and consequentiality perceptions) and behavioural economics and marketing (i.e., on opportunity cost priming). And, finally, Chapter 3 reviews the literature on ordering effects across domains of application of choice experiments, from which designers of choice experiments in all domains may benefit. For instance, the insights resulting from previous research on attribute ordering effects, largely originating from the health domain, may also be of benefit to DCE applications in other domains.

Next to strengths, this dissertation also comes with limitations. One of these limitations relates to the second objective of the dissertation, on the positioning of PVE with respect to other multi-attribute preference-elicitation methods used in the health domain. Chapter 2 based this positioning on a conceptual comparison of methods, while it ideally would have been based on an empirical comparison as well, like has been performed for methods other than PVE (e.g., Himmler et al., 2021; Krucien et al., 2017; Soekhai et al., 2023; Veldwijk et al., 2024; Whichello et al., 2023; Whitty & Gonçalves, 2018). This could demonstrate, for example, whether PVE results in elicited preferences that are different from other methods, and whether the method results in a larger or smaller cognitive burden on respondents than other preference-elicitation methods. Another limitation is that, for both topics of application in the dissertation, the design of the choice tasks did not fully represent the reality of policymakers. Unfortunately, incompleteness of and uncertainty surrounding the available information on the effects of policy measures, for example on the relative effectiveness and costs of institutional versus home-based care, required us to make assumptions in the choice of attributes and levels. If these assumptions do not hold, the presented choice tasks may have been less accurate representations of reality. For this reason, both chapters emphasized that the results of these and other preference studies should be complemented with additional information on the relative effects of the different available policy measures. Besides, for complex policy questions like the ones addressed in Chapters 4 and 5, researchers typically face a trade-off between the complexity and completeness of the choice tasks presented to respondents. Presenting all possible policy measures and all relevant outcomes and constraints would arguably impose a large cognitive burden on respondents. To prevent this, only a selection is presented to respondents, but this also limits the realism -and therewith the face validity and external validity- of the choices presented.

Besides, the findings by Chapter 3 on ordering effects in DCEs may also have implications for PVE applications, but these have not been studied in this dissertation. The mechanism of lexicographic behaviour underlying alternative and attribute ordering effects is likely to be present among respondents in a PVE as well. As a precautionary measure, the order of the policy alternatives in the PVE presented in Chapter 5 and 6 is randomized between respondents. Given limitations to the survey software, the order of the attributes in the choice task (and the order of the dashboard meters in the choice task alike) could not be randomized, because of which attribute ordering effects may have played a role. Nevertheless, given that the findings in Chapter 3 suggested that particular design characteristics in a DCE choice task may influence ordering effects, and since the PVE choice task is different from a DCE choice task, this may result in different findings of ordering effects in PVE. Future research may examine this.

Finally, the optimal portfolio analysis in Chapters 5 and 6 assumed a certain decision rule in the aggregation of preferences. In these chapters, a preference ranking of portfolios was computed based on the MDCEV estimates. In the aggregation of preferences, a utilitarian social welfare function was used, maximizing utility of society as a whole. An important alternative would be to use a Rawlsian social welfare function, in which the utility of the worst-off member of society is maximized (Hindriks

& Myles, 2006). Such distributional considerations were not included in the context of this PVE choice task, as that would arguably further increase the cognitive burden on respondents. Further research may explore the feasibility of including such distributional considerations in the survey and the sensitivity of the obtained portfolio rankings to the adopted social welfare function. Besides, all respondents' preferences received equal weight in the aggregation of preferences, in the absence of any rationale to differentiate the weights. Dekker et al. (2024) show that it is possible to differentiate the weighting of respondents' preferences in the optimal portfolio analysis, and future PVE applications may additionally elicit public preferences for doing so.

Methodological considerations and directions for future research

Besides the previously mentioned suggestions for further research in relation to limitations of this dissertation, two methodological considerations - and directions for future research - are discussed more generally and elaborately below.

The validity of the citizen valuation perspective

First, as discussed in the introduction of the dissertation, the choice experiments in this dissertation make use of the citizen perspective, which is taken less frequently in stated preference research than the consumer perspective. It is important to take into account that framing the choice experiment in either perspective may influence respondents' preferences. In the context of policy preferences, it is unclear which of the two perspectives results in elicitation of preferences that are closer to real-life preferences (i.e., external validity), as we typically cannot compare these stated preferences with any observed behaviour. Nevertheless, if we decide to elicit preferences for public policy decisions and resource allocations, we have to take a perspective in the preference-elicitation task. Given the potential endogeneity of the elicited preferences to the perspective taken, as discussed in the introduction chapter, it is important to understand the relative advantages and disadvantages of the different perspectives.

The consumer perspective, on the one hand, is considered convenient for the analyst; if everyone considers their own interests exclusively, their estimated preferences represent the utilities they derive from how the good or service in question affects their own wellbeing. In theory, this makes preferences comparable across respondents, facilitating the aggregation of preferences. Even though this presumption is theoretically feasible, it is unclear to what extent it represents reality in the context of policy preferences; even when asked to consider their own individual interests exclusively, some respondents may still take into account (their perceptions of) societal interests, too. In that sense, the assumption of the purely self-interested utility-maximizing agent may not be universally applicable and relevant in all choice contexts. Or, as put even stronger by Bowles and Gintis (1993): *"…the self-interested behavior underlying neoclassical theory is artificially truncated: it depicts a charmingly Victorian but Utopian world…"* (Bowles & Gintis, 1993, p.83). Moreover, if respondents are determined to incorporate others' interests in their preferences, they are encouraged not to incorporate such altruistic considerations under the consumer perspective, even if these are part of respondents' 'true' preferences. This would violate the influential principle of consumer sovereignty (Beeson et al., 2025) and may, ultimately, result in respondents being annoyed by the consumer framing of a choice experiment (Nyborg, 2000), reducing its face validity and potentially resulting in protest behaviour.

The citizen perspective, on the other hand, results in stated preferences that are (in theory) less comparable across respondents. In fact, some respondents in a preference-elicitation task may still only consider the utility derived from their own consumption under the citizen perspective, while others may also or only consider others' interests and wellbeing and how this affects their utility. Effectively, when aggregating respondents' preferences, this may lead the analyst to compare apples and oranges (Nyborg, 2000). Additionally, when respondents are considering others' interests too, this leads to challenges in and of itself. For example, one runs the risk of inflated valuations, as respondents may consider only the benefits of goods and services to others (i.e., not the costs). In addition, respondents are often inaccurate in estimating the benefits for and preferences of other individuals, as was shown, for instance, regarding the provision of public goods and services (e.g., Gyrd-Hansen et al., 2016; Jung et al., 2020; Simonsen et al., 2021).

Future research should further examine to what extent and under which circumstances respondents inaccurately estimate the benefits for and preferences of other individuals. This could take the form of qualitative research on respondents' motivations underlying their choices, as well as explicitly eliciting preferences from a consumer and from a citizen perspective in a variety of decision contexts. Also, it could test whether information treatments affect the accuracy of respondents' estimates. At the same time, researchers may also examine to what extent framing a stated preference task from different perspectives induces respondent annoyance and protest behaviour, again using qualitative research methods (e.g., think aloud testing) or debriefing questions.

Ultimately, it is a hard choice between the lesser of two evils for researchers eliciting preferences for policy measures: framing the choice task from a consumer perspective, even though respondents are probably not exclusively self-interested, or framing it from a citizen perspective, even though respondents may inaccurately estimate others' preferences and benefits.

Understanding and capturing respondents' choice behaviour more accurately

A second methodological consideration, as also highlighted in Chapter 2, is that there is currently limited understanding of respondents' choice behaviour in PVE. This understanding is essential, nonetheless, for overseeing how researchers' decisions in the design of a PVE choice task would affect respondents' choices and whether the specified models accurately represent respondents' choice behaviour.

Several directions for future studies on respondents' choice behaviour in PVEs can be proposed, many of which heavily build upon previous research using other preference-elicitation methods, particularly DCE. Firstly, researchers could make use of qualitative research methods to better understand respondents' choice behaviour, for instance using interviews. This may take place during the completion of the choice task, by asking respondents to think out loud and explain their processing of the choice task, thinking, and choice process (e.g., Ryan et al., 2009; Whitty et al., 2014), or post-hoc, by asking respondents to motivate the choices they made and testing their understanding of key concepts in the choice task (e.g., Veldwijk et al., 2016). This could provide insight into whether respondents are processing the choice task in the intended manner (e.g., attending all alternatives and attributes), whether they understood everything, and how they came to their choices (i.e., internal validity).

Secondly, researchers may observe respondents' processing of the choice task by using tracking methods. For instance, eye-tracking may be used to obtain information on respondents' visual attention in a PVE choice task. Visual fixation time spent on different parts of the choice task and eye movements could provide insight into which aspects of choice scenarios respondents pay (most) attention to and how they process the choice task (Bansal et al., 2024), while pupil width may be used as a proxy for mental effort, providing insight into the cognitive burden on respondents (Genie et al., 2023). It should be noted that a challenge for the use of eye-tracking technologies is that they are expensive, limiting the sample sizes and prohibiting the estimation of a choice model on the collected data. This is arguably even more constraining in a PVE: as mentioned in

Chapter 2, due to the absence of within-respondent experimental variation, PVE is less efficient and, therefore, typically requires larger sample sizes than DCE or BWS. Given the impossibility for many eye-tracking studies to relate the visual attention data to respondents' choices in the experiment, it is often assumed that a larger share of visual attention for alternatives or attributes translates into higher choice probabilities (for alternatives) or greater importance (for attributes). As mentioned in Chapter 3, though, the findings by Meißner et al. (2016) suggest that this assumption may not always hold. Especially when applying eye-tracking to PVE, therefore, one should be very considerate and transparent regarding the assumptions made. To address this limitation, eye-tracking in a lab-based setting could be performed simultaneously to fielding the PVE in an online panel. In case the descriptive choice information (i.e., the choice shares and cost distribution) and other information (e.g., survey completion time, self-reported experiences) correspond between the two samples, one may combine the two datasets and relate the visual attention data to the choice data in this way (Bansal et al., 2024). Alternatively, one may employ mouse tracking as an alternative method to observe respondents' choices; in this method, respondents' mouse movements and clicking behaviour are collected alongside their choices (e.g., Nova & Guevara, 2023; Tanasache et al., 2023). It could be recorded, for example, when and how often respondents open the pop-up information screens for the alternatives in a PVE, and this may be linked to respondents' background characteristics and choice behaviour.

Thirdly, researchers may elicit additional information, next to respondents' choices, to understand their processing of a PVE survey. For example, supplementary survey questions could be included to measure respondents' stated choice certainty (e.g., Dekker et al., 2016; Regier et al., 2014; Rose et al., 2015), attribute attendance (e.g., Caputo et al., 2018; Hole et al., 2013; Scarpa et al., 2013), and broader experiences (i.e., perceptions of the choice task difficulty, topic relevance, etc.) (e.g., De Ruijter et al., 2025; Pearce et al., 2020). The resulting information, as well as meta data like survey completion time, may be used to understand the relation between these additional variables and respondents' choices, and to improve model accuracy by joint modelling of these variables and choices (e.g., Dekker et al., 2016; Rose et al., 2015) or by calibrating or reweighting responses (e.g., Penn & Hu, 2023; Regier et al., 2014). Alternatively, one may also regress the estimated error variance on these additional variables and respondents' background characteristics. The error variance (i.e., the scale) is often used as a proxy of the perceived complexity of the choice task (e.g., Bech et al., 2011; Dellaert et al., 2012) and may thus be used to explore the cognitive burden of completing a PVE choice task for different population segments.

Each of the research directions discussed above would improve our understanding of respondents' choice behaviour, which could inform researchers' decisions when designing a PVE choice task and analysing PVE data. The effect of design decisions in a PVE could also be studied and documented more elaborately. This may be done in the form of a meta-analysis of multiple PVE applications, where the dependent variable could be, for instance, the estimated error variance, survey completion time, or respondents' reported survey experiences. The independent variables could include various design dimensions in PVEs (e.g., Rolfe & Brouwer, 2012). For instance, some PVEs included all attribute levels of the policy alternatives in the choice task screen (like in Chapter 5), while others included most attribute levels only in the pop-up information screens (e.g., Mulderij et al., 2021). Such design choices may affect both the importance of the attribute levels in respondents' choices and the cognitive burden on respondents. Besides, future studies may use split-samples to empirically test the impact of a design variations on respondents' preferences in the context of a specific application. Chapter 6 provided an example of testing the impact of two design decisions (i.e., the choice of payment vehicle and the inclusion of an opportunity cost priming question) on respondents' choices. Many more empirical tests of the impact of design dimensions can be thought of, such as by varying the number of alternatives, the number of attributes, the number of constraints, the type of constraints (i.e., monetary or non-monetary, a minimum or maximum, as discussed in Chapter 2), and the lay-out of the choice task (e.g., Caussade et al., 2005; Dellaert et al., 2012; Hensher, 2006; Meyerhoff et al., 2015; Sandorf et al., 2018).

One dimension of PVE studies that should be further examined in particular is the height of the budget constraint. In the PVE on long-term care covered in Chapter 5, a third of the respondents fully exhausted the budget constraint, and another third of the sample almost exhausted the resource constraint. As shown in Chapter 6, similar observations arise from the samples with another payment vehicle and with an opportunity cost priming method. This may be due to respondents in this PVE anchoring on the budget constraint, which is undesirable in case the PVE is used to elicit respondents' preferences for the preferred level of public expenditure on the topic in question. Also, the budget constraint may mask unreasonably high levels of preferred expenditures. Two previous PVE studies examined the influence of the budget constraint using split-samples, in which one subsample of respondents could adjust the height of the budget and another subsample could not (Dekker et al., 2024; Mouter et al., 2021). While they found that the relative ranking of policy alternatives in terms of preference was the same across the two subsamples and that most respondents in the flexible-budget version did not adjust the height of the budget, the computed optimal portfolios differed between the two versions (Dekker et al., 2024; Mouter et al., 2021). Building on these studies, a future PVE application using split-samples might randomize the height of the budget constraint between respondents and thereby examine whether this influences respondents' choice behaviour and experiences.

Finally, improved insight into respondents' choice behaviour, along the lines previously suggested, may improve the modelling of PVE data. For instance, the data from several previous PVE applications, including the application on LTC presented in Chapters 5 and 6, have been modelled using the traditional MDCEV model or an adapted version for PVEs in which the alternatives have a discrete choice dimension only (i.e., the MDCEV-PVE model). Both models assume independently and identically distributed error terms for the different choice alternatives, following the Independence of Irrelevant Alternatives (IIA) assumption that is at the heart of traditional choice models and implies a uniform pattern of substitution between alternatives (McFadden, 2001).[1] As a result, while one of the aims of PVE was to allow respondents to incorporate synergies between policy alternatives in their choices, as discussed in Chapter 2, these synergies are currently neglected in the modelling of PVE data because of the IIA assumption.[2] The behavioural plausibility of this assumption may be questioned in practice (McFadden, 2001), however. In fact, two alternatives may be closer substitutes, in case they both satisfy the same needs of an individual. For instance, the optimal portfolio analysis for the PVE on LTC, presented in Chapter 5, suggested that increasing the capacity of nursing homes and increasing the capacity of nursing care at home were substitutes. Contrarily, alternatives may also be complements, if their joint consumption satisfies a particular need (e.g., Lattin & Mcalister, 1985).

To align choice models with respondents' choice behaviour, multiple models have been developed to accommodate such complementarity and substitution patterns in discrete and multiple discrete-continuous (MDC) choice models alike. Arguably the most

1 This does not mean that alternatives are chosen completely independently from each other in the MDCEV and MDCEV-PVE models. As these models come with a resource constraint (e.g., a budget), choices induce income effects across alternatives (Palma et al., 2023), given that 'consumption' of one alternative requires the allocation of budget and thereby crowds out 'consumption' of other alternatives. Given the IIA assumption, however, the cross-elasticities between all alternatives are assumed to be equal, meaning that increases in the consumption of one alternative reduces the consumption of all other alternatives to equal extents.

2 An exception is the PVE application by Bahamonde-Birke et al. (2024), making use of the Portfolio Choice Model (PCM) instead, which does accommodate for correlation between alternatives (Bahamonde-Birke & Mouter, 2024). However, as the PCM considers error terms at the portfolio level only and not also at the level of alternatives, it is computationally/statistically more expensive than the alternative models proposed here, which may be particularly problematic in case of larger choice sets (Bahamonde-Birke & Mouter, 2024).

popular of such models for MDC choices is the Multiple Discrete-Continuous Nested Extreme Value (MDCNEV) choice model (Pinjari & Bhat, 2010), in which alternatives are grouped together in nests, between which closer substitution patterns are expected. Another recent addition is the extended Multiple Discrete-Continuous (eMDC) choice Model (Palma & Hess, 2022; Palma et al., 2023), in which interactions between alternatives are estimated to capture complementarity and substitution patterns. Future research may apply these models to the data from previous PVE applications in which the policy alternatives contained a continuous choice dimension (i.e., the PVE in Chapter 5 of this dissertation and the PVE by Mulderij et al. (2021)) to examine whether this improves statistical model fit, provides additional insights into respondents' choice behaviour in terms of complementarity and substitution between alternatives, and ultimately results in more accurate policy implications. Likewise, the MDCEV-PVE model could be extended to allow for correlation between alternatives.

Besides, in the modelling of all PVE studies so far, including Chapters 5 and 6 of this dissertation, respondents were assumed to employ fully compensatory decision heuristics in their choices. This assumption may not hold, however, as respondents may make use of simplifying heuristics (e.g., Veldwijk et al., 2023), such as attribute non-attendance (ANA) (Gonçalves et al., 2022) or elimination/selection by aspect (EBA/SBA) (e.g., Erdem et al., 2014). Not accommodating for these heuristics in the modelling of respondents' choices may bias the model estimates.[3] Furthermore, the modelling in all PVE applications thus far has been based on the assumption that respondents are utility-maximizing in their choices. In the wider choice modelling literature, however, several models have been developed and applied to capture alternative decision rules, such as Random Regret Minimization (RRM) (e.g., Buckell et al., 2022; Chorus et al., 2014) or Decision Field Theory (DFT) (e.g., Hancock et al., 2018; Meester et al., 2023). Future research may examine whether models based on these alternative decision rules capture respondents' choice behaviour in a PVE more accurately than models based on utility-maximization, which may result in an improved model fit. At the same time, one should recognize that deviating from a utility-maximization framework also comes with disadvantages in terms of model interpretability and limited post-estimation possibilities (e.g., welfare analysis) (Hess et al., 2018).

3 At the same time, one should be mindful of the source of the non-compensatory choice behaviour. ANA, for example, could be the result of respondents' preferences rather than their use of simplifying heuristics (Heidenreich et al., 2017). In that case, correcting for ANA could imply the researcher to be imposing instead of revealing preferences, as pointed out in Chapter 4.

Implications for policy

The DCE and PVE applications presented in Chapters 4 and 5 are directly aimed at informing health policy questions. The elicited preferences provide policymakers with directions for publicly supported policy action. For example, in Chapter 4, given that a large majority of respondents in each of the three countries supported policies aimed at preventing skin cancer, governments are recommended to take policy action to protect their citizens against skin cancer. When doing so, they are advised to (first) adopt information campaigns and measures lowering the price of sunscreen and to minimize the impact on tax levels. The policy options prohibiting solar beds and solar studios was least preferred and, therefore, such more intrusive policies are not recommended as first to be adopted if public support is considered important. In Chapter 5, regarding future long-term care policies, governments are recommended to invest in a broad range of policy measures, encompassing both nursing care and social care policies, rather than to focus on investing in one or two specific policies. Also, in case they are interested in broadening the support base for particular policies, they may take advantage of the results of Chapter 5 in targeting specific subgroups of the population. For instance, they may want to better understand middle-aged and older citizens' motivations underlying their lower preference for the increased use of supportive care technologies, or the conditions under which a form of social service might be attractive and acceptable for younger adults.

More generally, Chapters 4 and 5 illustrate the potential of choice experiments to inform health policy questions. In case policymakers decide to issue or conduct a study to elicit public preferences for health policy alternatives, they may take advantage of the insights from Chapter 2. This Chapter compared a number of multi-attribute preference-elicitation methods conceptually and thereby provided guidance for researchers and policymakers in selecting a suitable preference-elicitation method for the policy question at hand. For instance, DCE or BWS seem more suitable when policymakers are interested in implementing only one policy alternative to address a particular policy question, while PVE is more suitable when policymakers expect to implement a combination of policy measures. Besides, the results of preference-elicitation methods are most valuable in combination with sufficient information on the relative effectiveness and costs of different policy measures. In the health domain, the study of these outcomes has become standard practice for curative interventions (i.e., new treatments), but is not as developed yet in other areas in health, such as long-term care. As such, if we aim to incorporate public preferences for these trade-offs in the public policy process, the advancement of our understanding of the trade-offs between policy measures should be a priority.

Furthermore, the limited generalizability of elicited preferences poses a challenge in the prioritization of future research needs. On the one hand, preferences may be endogenous to characteristics of the choice experiment design and data collection as well as the institutional and cultural context in which these methods are employed. Therefore, it is desirable to inform policy decisions with preferences elicited using a variety of elicitation methods and designs applied to different samples. On the other hand, as conducting preference-elicitation studies can be costly, this desirable situation is constrained by the scarcity of resources discussed in Chapter 1. This gap between the desired situation, in which preferences are elicited for every context separately using a variety of designs, and the scarcity of resources, which does not allow researchers and policymakers to do so, provides room and urgence for more attention to the transferability of preferences. The concept of transferability is most developed in the domain of environmental economics, where much attention is paid to the transferability of citizens' valuation of environmental public goods across contexts and study characteristics (e.g., Brouwer & Neverre, 2020; Johnston et al., 2018; Mattman et al., 2016; Rolfe et al., 2015). While transferability has not been as developed and used regarding stated preferences in the health domain, it has received more attention recently (e.g., Marsh et al., 2025; Veldwijk et al., 2025). An advantage of stated preference research in health is the wide diffusion and adoption of good practices, guidelines, and checklists (e.g., Bridges et al., 2011; Hauber et al., 2016; Johnson et al., 2013; Lancsar et al., 2017; Ride et al., 2024). This promotes a uniform standard of conduct and reporting of stated preference studies, which has been suggested to contribute to a higher quality of survey development relative to the domain of transportation in applications of DCEs to COVID-19 (Haghani et al., 2022). Nevertheless, there is still room for improvement in terms of a more complete study reporting to raise the transparency and reproducibility of stated preference research (e.g., Soekhai et al., 2019). Finally, an important condition for transferability of preferences is an improved understanding of the impact of methodological characteristics, such as design dimensions and data collection features, on the elicited preferences (Veldwijk et al., 2025). This dissertation has contributed, particularly with Chapters 3 and 6, to such an improved understanding. Nevertheless, there is still considerable scope for diffusion of methodological insights across the main domains of application and for additional research, particularly for PVE, as discussed in this Chapter. Researchers, policymakers, and other research funders are therefore encouraged to allocate resources to more methodological research on preference-elicitation methods. This could limit resource use for numerous applied studies in the long run and contributes to the validity of stated preference research.

Concluding remarks

This dissertation aimed to advance the literature on the elicitation of public preferences for health policy alternatives. Insights into these preferences could inform policy action in a way that promotes a close alignment between public preferences and collective resource allocation decisions. The applications of DCE and PVE in this dissertation illustrate how these methods can be of value. At the same time, however, each of the studies included in this dissertation also show the current limitations to the use of choice experiments. One may wonder, given these limitations, whether stated preferences are sufficiently reliable and valid to inform policy decisions. Even though the opinions on this are differing, I tend to hold the opinion that "some number" has become better than "no number" (Kling et al., 2012); over the past decades, we have witnessed many methodological improvements in choice modelling and choice experiments (e.g., Caputo & Scarpa, 2022; Johnston et al., 2017), and the consistency between stated preferences elicited using DCEs and revealed preferences is often found to be reasonably high on an aggregate level (e.g., De Bekker-Grob et al., 2020; Quaife et al., 2018; Zhang et al., 2025). Moreover, the counterfactual of no longer using stated preferences would imply that we have no information on valuations of many non-market goods. This puts us at the risk of making policy decisions based on implicit rather than explicit value trade-offs, which reduces the accountability and transparency of these decisions, or renders policymakers with a welfare-maximizing objective clueless as to which policies to adopt. Of course, this does not free stated preference researchers and policymakers from the task to continue improving preference-elicitation methods, to increase their internal and external validity and reliability and ultimately elicit preferences more accurately to inform public policy decisions. On the bright side, this means there is enough work for me and others to be done in the decades to come, for which this dissertation provides many directions.

References

Abate, T.G., Mørkbak, M.R. and Olsen, S.B. (2018). Inducing value and institutional learning effects in stated choice experiments using advanced disclosure and instructional choice set treatments. *Agricultural Economics*, *49*, 339 – 351

Achtnicht, M. (2012). German car buyers' willingness to pay to reduce CO_2 emissions. *Climatic Change*, *113*, 679 – 697

Ahtiainen, H., Pouta, E., Zawadzki, W. and Tienhaara, A. (2023). Cost vector effects in discrete choice experiments with positive status quo costs. *Journal of Choice Modelling*, *47*, 100401

Alphonce, R., Alfnes, F. and Sharma, A. (2014). Consumer vs. citizen willingness to pay for restaurant food safety. *Food Policy*, *49*, 160 – 166

Amilon, A., Ladenburg, J., Siren, A. and Østergaard, S.V. (2020). Willingness to pay for long-term home care services: Evidence from a stated preferences analysis. *Journal of the Economics of Ageing*, *17*, 100238

Anderson, W.L. and Wiener, J.M. (2015). The Impact of Assistive Technologies on Formal and Informal Home Care. *The Gerontologist*, *55*(3), 422 – 433

Andersson, M., Jonsson, L., Brundell-Freij, K. and Berdica, K. (2023). Who should pay? Public acceptance of different means for funding transport infrastructure. *Transportation*, *50*, 1425 – 1448

Apalla, Z., Lallas, A., Sotiriou, E., Lazaridou, E. and Ioannides, D. (2017). Epidemiological trends in skin cancer. *Dermatology Practical & Conceptual*, *7*(2), 1

Arentze, T., Borgers, A., Timmermans, H. and DelMistro, R. (2003). Transport stated choice responses: effects of task complexity, presentation format and literacy. *Transportation Research Part E: Logistics and Transportation Review*, *39*, 229 – 244

Arora, K. and Wolf, D.A. (2018). Does Paid Family Leave Reduce Nursing Home Use? The California Experience. *Journal of Policy Analysis and Management*, *37*(1), 38 – 62

Asadi, L.K., Khalili, A. and Wang, S.Q. (2023). The sociological basis of the skin cancer epidemic. *International Journal of Dermatology*, *62*, 169 – 176

Atalay, A.S., Bodur, H.O. and Rasolofoarison, D. (2012). Shining in the Center: Central Gaze Cascade Effect on Product Choice. *Journal of Consumer Research*, *39*(4), 848 – 866

Auspurg, K. and Jäckle, A. (2017). First Equals Most Important? Order Effects in Vignette-Based Measurement. *Sociological Methods & Research*, *46*(3), 490 – 539

Bahamonde-Birke, F.J. (2024). *Is it possible to collect more than one observation per individual in PVE-experiments?* Conference abstract, International Choice Modelling Conference 2024

Bahamonde-Birke, F.J. and Mouter, N. (2024). About positive and negative synergies of social projects: Treating correlation in participatory value evaluation. *Environment and Planning B: Urban Analytics and City Science*, *52*(3), 563 – 577

Bahamonde-Birke, F.J., Geigenmüller, I.M., Mouter, N., Van Lierop, D.S. and Ettema, D.F. (2024). How do I want the city council to spend our budget? Conceiving MaaS from a citizen's perspective … (as well as biking infrastructure and public transport). *Transport Policy*, *145*, 96 – 104

Bakx, P., Wouterse, B. and Van Doorslaer. (2023). *Long-term care in the Netherlands*. NBER Working Paper, No. 31823. Cambridge, MA: National Bureau of Economic Research

A

Bakx, P., Wouterse, B., Van Doorslaer, E. and Wong, A. (2020). Better off at home? Effects of nursing home eligibility on costs, hospitalizations and survival. *Journal of Health Economics*, *73*, 102354

Balcombe, K., Fraser, I. and McSorley, E. (2015). Visual Attention and Attribute Attendance in Multi-Attribute Choice Experiments. *Journal of Applied Econometrics*, *30*(3), 447 – 467

Baltas, G. and Doyle, P. (2001). Random utility models in marketing research: a survey. *Journal of Business Research*, *51*(2), 115 – 125

Bansal, P., Kim, E.-J. and Ozdemir, S. (2024). Discrete choice experiments with eye-tracking: How far have we come and ways forward. *Journal of Choice Modelling*, *51*, 100478

Bansback, N., Hole, A.R., Mulhern, B. and Tsuchiya, A. (2014). Testing a discrete choice experiment including duration to value health states for large descriptive systems: Addressing design and sampling issues. *Social Science & Medicine*, *114*, 38 – 48

Bär, M., Bakx, P., Wouterse, B. and Van Doorslaer, E. (2022). Estimating the health value added by nursing homes. *Journal of Economic Behavior and Organization*, *203*, 1 – 23

Barnes, L., Blumenau, J. and Lauderdale, B.E. (2022). Measuring Attitudes toward Public Spending Using a Multivariate Tax Summary Experiment. *American Journal of Political Science*, *66*(1), 205 – 221

Baumgartner, D. and Kolassa, J. (2023). Power considerations for Kolmogorov-Smirnov and Anderson-Darling two-sample tests. *Communications in Statistics – Simulation and Computation*, *52*(7), 3137 – 3145

Bech, M., Kjaer, T. and Lauridsen, J. (2011). Does the number of choice sets matter? Results from a web survey applying a discrete choice experiment. *Health Economics*, *20*(3), 273 – 286

Bechtold, K.-B. and Abdulai, A. (2012). *Willingness-To-Pay for Functional Dairy Products and the Influence of Starting Point Bias: Empirical Evidence for Germany*. AAEA Conference Paper. Available at: https://ageconsearch.umn.edu/record/124776

Beeson, M., Chilton, S., Metcalf, H. and Nielsen, J.S. (2025). Citizen Preferences and BCA: A Model of Willingness-to-Pay behind a Veil of Ignorance. *Journal of Benefit-Cost Analysis* [article in press]

Bennett, J., Van Bueren, M. and Whitten, S. (2004). Estimating society's willingness to pay to maintain viable rural communities. *Australian Journal of Agricultural and Resource Economics*, *48*(3), 487 – 512

Berchi, C., Degieux, P., Halhol, H., Danel, B., Bennani, M. and Philippe, C. (2016). Impact of falling reimbursement rates on physician preferences regarding drug therapy for osteoarthritis using a discrete choice experiment. *International Journal of Pharmacy Practice*, *24*, 114 – 122

Berchi, C., Dupuis, J.-M. and Launoy, G. (2006). The reasons of general practitioners for promoting colorectal cancer mass screening in France. *European Journal of Health Economics*, *7*, 91 – 98

Bergeot, J. and Tenand, M. (2023). Does informal care delay nursing home entry? Evidence from Dutch linked survey and administrative data. *Journal of Health Economics*, *92*, 102831

Bergstrom, J.C., Boyle, K.J. and Yabe, M. (2004). Trading Taxes vs. Paying Taxes to Value and Finance Public Environmental Goods. *Environmental & Resource Economics*, *28*, 533 – 549

Bhat, C.R. (2008). The multiple discrete-continuous extreme value (MDCEV) model: Role of utility function parameters, identification considerations, and model extensions. *Transportation Research Part B: Methodological*, *42*, 274 – 303

Bierlaire, M. Thémans, M. and Zufferey, N. (2010). A Heuristic for Nonlinear Global Optimization. *INFORMS Journal on Computing*, *22*(1), 59 – 70

Björklund, G. and Swärdh, J.-E. (2017). Estimating policy values for in-vehicle comfort and crowding reduction in local public transport. *Transportation Research Part A: Policy and Practice*, *106*, 453 – 472

Bliemer, M.C.J. and Rose, J.M. (2013). Confidence intervals of willingness-to-pay for random coefficient logit models. *Transportation Research Part B: Methodological*, *58*, 199 – 214

Bom, J., Bakx, P., Schut, F. and Van Doorslaer, E. (2019). The Impact of Informal Caregiving for Older Adults on the Health of Various Types of Caregivers: A Systematic Review. *The Gerontologist*, *59*(5), 629 – 642

Börger, T., Abate, T.G., Aanesen, M. and Zawojska, E. (2021). Payment and Policy Consequentiality in Dichotomous Choice Contingent Valuation: Experimental Design Effects on Self-Reported Perceptions. *Land Economics*, *97*(2), 407 – 424

Boto-García, D., Mariel, P., Baños Pino, J. and Alvarez, A. (2022). Tourists' willingness to pay for holiday trip characteristics: A Discrete Choice Experiment. *Tourism Economics*, *28*(2), 349 – 370

Bowles, S. and Gintis, H. (1993). The Revenge of Homo Economicus: Contested Exchange and the Revival of Political Economy. *Journal of Economic Perspectives*, *7*(1), 83 – 102

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

Boxebeld, S. (2024). Ordering Effects in Discrete Choice Experiments: A Systematic Literature Review Across Domains. *Journal of Choice Modelling*, *51*, 100489

Boxebeld, S., Geijsen, T., Tuit, C. et al. (2024a). Public preferences for the allocation of societal resources over different healthcare purposes. *Social Science & Medicine*, *341*, 116536

Boxebeld, S., Mouter, N. and Van Exel, J. (2024b). Participatory Value Evaluation: A New Preference-Elicitation Method for Decision Making in Healthcare. *Applied Health Economics and Health Policy*, *22*, 145 – 154

Boxebeld, S., Mouter, N. and Van Exel, J. (2025). Trade-offs in long-term care for older people in an ageing society: a constrained portfolio choice experiment. *Journal of the Economics of Ageing* [article in press]

Boyle, K.J. and Özdemir, S. (2009). Convergent Validity of Attribute-Based, Choice Questions in Stated-Preference Studies. *Environmental and Resource Economics*, *42*, 247 – 264

Boyle, K.J., Welsh, M.P. and Bishop, R.C. (1993). The Role of Question Order and Respondent Experience in Contingent-Valuation Studies. *Journal of Environmental Economics and Management*, *25*(1), 580 – 599

Bradley, M. and Daly, A. (1994). Use of the logit scaling approach to test for rank-order and fatigue effects in stated preference data. *Transportation*, *21*, 167 – 184

Brau, R. and Bruni, M.L. (2008). Eliciting the demand for long-term care coverage: a discrete choice modelling analysis. *Health Economics*, *17*(3), 411 – 433

Breyer, F. and Lorenz, N. (2021). The "red herring" after 20 years: ageing and health care expenditures. *European Journal of Health Economics*, *22*, 661 – 667

Bridges, J. (2004). Understanding the risks associated with resource allocation decisions in health: An illustration of the importance of portfolio theory. *Health, Risk & Society*, *6*(3), 257 – 275

Bridges, J.F.P., Stewart, M., King, M.T. and Van Gool, K. (2002). Adapting portfolio theory for the evaluation of multiple investments in health with a multiplicative extension for treatment synergies. *European Journal of Health Economics*, *3*, 47 – 53

Bridges, J.F.P., Hauber, A.B., Marshall, D. et al. (2011). Conjoint analysis applications in health – a checklist: A report of the ISPOR Good Research Practices for Conjoint Analysis Task Force. *Value in Health*, *14*(4), 403 – 413

Brimblecombe, N., Knapp, M., King, D., Stevens, M. and Cartagena Farias, J. (2020). The high cost of unpaid informal care by young people: health and economic impacts of providing unpaid care. *BMC Public Health*, *20*(1), 1115

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

Brouwer, R., Dekker, T., Rolfe, J. and Windle, J. (2010). Choice Certainty and Consistency in Repeated Choice Experiments. *Environmental and Resource Economics*, *46*, 93 – 109

Brouwer, R., Logar, I. and Sheremet, O. (2017). Choice Consistency and Preference Stability in Test-Retests of Discrete Choice Experiment and Open-Ended Willingness to Pay Elicitation Formats. *Environmental and Resource Economics*, *68*, 729 – 751

Bryson, J.M., Quick, K.S., Slotterback, C.S. and Crosby, B.C. (2013). Designing public participation processes. *Public Administration Review*, *73*(1), 23 – 34

Buckell, J., Vasavada, V., Wordsworth, S., Regier, D.A. and Quaife, M. (2022). Utility maximization versus regret minimization in health choice behavior: Evidence from four datasets. *Health Economics*, *31*(2), 363 – 3381

Buckell, J., Wreford, A., Quaife, M. and Hancock, T. (2025). A Break from the Norm? Parametric Representations of Preference Heterogeneity for Discrete Choice Models in Health. *Medical Decision Making* [article in press]

Bunch, D.S., Gay, D.M. and Welsch, R.E. (1993). Subroutines for Maximum Likelihood and Quasi- Likelihood Estimation of Parameters in Nonlinear Regression Models. *ACM Transactions on Mathematical Software*, *19*(1), 109 – 130

Cai, B., Cameron, T.A. and Gerdes, G.R. (2011). Distal order effects in stated preference surveys. *Ecological Economics*, *70*(6), 1101 – 1108

Campbell, D. and Erdem, S. (2015). Position Bias in Best-worst Scaling Surveys: A Case Study on Trust in Institutions. *American Journal of Agricultural Economics*, *97*(2), 526 - 545

Campbell, D. and Sandorf, E.D. (2020). The use of latent variable models in policy: A road fraught with peril? *Bio-based and Applied Economics*, *9*(3), 305 – 324

Campbell, D., Boeri, M., Doherty, E. and Hutchinson, W.G. (2015). Learning, fatigue and preference formation in discrete choice experiments. *Journal of Economic Behavior & Organization*, *119*, 345 – 363

Cao, Y., Cranfield, J. and Widowski, T. (2018). Position-Dependent order effects on the prediction of consumer preferences in repeated choice experiments. *Applied Economics*, *50*(3), 287 – 302

Caputo, V. and Lusk, J.L. (2022). The Basket-Based Choice Experiment: A Method for Food Demand Policy Analysis. *Food Policy*, *109*, 102252

Caputo, V. and Scarpa, R. (2022). Methodological advances in food choice experiments and modeling: Current practices, challenges, and future research directions. *Annual Review of Resource Economics*, *14*, 63 – 90

Caputo, V., Van Loo, E.J., Scarpa, R., Nayga Jr., R.M. and Verbeke, W. (2018). Comparing serial, and choice task stated and inferred attribute non-attendance methods in food choice experiments. *Journal of Agricultural Economics*, *69*(1), 35 – 57

Carlsson, F. and Martinsson, P. (2001). Do Hypothetical and Actual Marginal Willingness to Pay Differ in Choice Experiments? *Journal of Environmental Economics and Management*, *41*, 179 – 192

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

Carpenter, C.S., Churchill, B.F. and Marcus, M. (2023). Bad lighting: Effects of youth indoor tanning prohibitions. *Journal of Health Economics*, *88*, 102738

Carson, R.T. and Czajkowski, M. (2014). The discrete choice experiment approach to environmental contingent valuation. In: S. Hess and A. Daly (eds.). *Handbook of Choice Modelling*. Chapter 9. Edward Elgar Publishing

Carson, R.T. and Groves, T. (2007). Incentive and informational properties of preference questions. *Environmental and Resource Economics*, *37*, 181 – 210

Carson, R.T. and Hanemann, W.M. (2005). Contingent Valuation. In: K.-G. Mäler and J.R. Vincent (eds.). *Handbook of Environmental Economics*. Vol. 2, Chapter 17. Amsterdam, the Netherlands: Elsevier

Carson, R.T. and Louviere, J.J. (2011). A common nomenclature for stated preference elicitation approaches. *Environmental and Resource Economics*, *49*, 539 – 559

Carson, R.T., Eagle, T.C., Islam, T. and Louviere, J.J. (2022). Volumetric choice experiments (VCEs). *Journal of Choice Modelling*, *42*, 100343

Caussade, S., Ortúzar, J.D., Rizzi, L.I. and Hensher, D.A. (2005). Assessing the influence of design dimensions on stated choice experiment estimates. *Transportation Research Part B: Methodological*, *39*, 621 – 640

Chalak, A., Nakkash, R., Abu-Rmeileh, N.M.E. et al. (2023). Own-price and cross-price elasticities of demand for cigarettes and waterpipe tobacco in three Eastern Mediterranean countries: a volumetric choice experiment. *Tobacco Control*, *32*(1), 86 – 92

Chandoevwit, W. and Wasi, N. (2020). Incorporating discrete choice experiments into policy decisions: Case of designing public long-term care insurance. *Social Science & Medicine*, *258*, 113044

Chang, C., Murzaku, E.C., Penn, L. et al. (2014). More Skin, More Sun, More Tan, More Melanoma. *American Journal of Public Health*, *104*(11), e92 – e99

Chen, S., Cao, Z., Prettner, K. et al. (2023). Estimates and Projections of the Global Economic Cost of 29 Cancers in 204 Countries and Territories From 2020 to 2050. *JAMA Oncology*, *9*(4), 465 – 472

Cheung, K.L., Wijnen, B.F.M., Hollin, I.L. et al. (2016). Using Best-Worst Scaling to Investigate Preferences in Health Care. *PharmacoEconomics*, 34, 1195 – 1209

A

ChoiceMetrics. (n.d.). *Ngene 1.2.1*.

Chorus, C., Van Cranenburgh, S. and Dekker, T. (2014). Random regret minimization for consumer choice modeling: Assessment of empirical evidence. *Journal of Business Research*, *67*(11), 2428 – 2436

Chorus, C.G. and Kroesen, M. (2014). On the (im-)possibility of deriving transport policy implications from hybrid choice models. *Transport Policy*, *36*, 217 – 222

Chrzan, K. (1994). Three kinds of order effects in choice-based conjoint analysis. *Marketing Letters*, *5*, 165 - 172

Clark, J. and Friesen, L. (2008). The causes of order effects in contingent valuation surveys: An experimental investigation. *Journal of Environmental Economics and Management*, *56*(2), 195 – 206

Clark, M.D., Determann, D., Petrou, S., Moro, D. and De Bekker-Grob, E.W. (2014). Discrete choice experiments in health economics: a review of the literature. *PharmacoEconomics*, *32*, 883 – 902

Cohen-Blankshtain, G. and Sulitzeanu-Kenan, R. (2021). Foregone and predicted futures: challenges of opportunity cost neglect and impact bias for public participation in policymaking. *Journal of European Public Policy*, *28*(5), 677 – 697

Collins, L.G., Gage, R., Sinclair, C. and Lindsay, D. (2024). The Cost-Effectiveness of Primary Prevention Interventions for Skin Cancer: An Updated Systematic Review. *Applied Health Economics and Health Policy*, *22*, 685 – 700

Costa-Font, J. and Rovira, J. (2005). Eliciting preferences for collectively financed health programmes: the 'willingness to assign' approach. *Applied Economics*, *37*, 1571 – 1583

Costa-Font, J., Jimenez-Martin, S. and Vilaplana, C. (2018). Does long-term care subsidization reduce hospital admissions and utilizations? *Journal of Health Economics*, *58*, 43 – 66

Craig, B.M., Runge, S.K., Rand-Hendriksen, K., Ramos-Goñi, J.M. and Oppe, M. (2015). Learning and Satisficing: An Analysis of Sequence Effects in Health Valuation. *Value in Health*, *18*(2), 217 – 223

Crastes dit Sourd, R. (2024). A new empirical approach for mitigating exploding implicit prices in mixed multinomial logit models. *American Journal of Agricultural Economics*, *106*(1), 76 – 95

Crastes dit Sourd, R., Hess, S., Borejesson, M. and Hensher, D. (2020). *Preference Instability in stated choice surveys: more evidence*. Working paper. Available at: http://www.stephanehess.me.uk/papers/working_papers/Crastes_Hess_Borjesson_Hensher_2 020.pdf

Curtis, J.A. and McConnell, K.E. (2002). The citizen versus consumer hypothesis: Evidence from a contingent valuation survey. *Australian Journal of Agricultural and Resource Economics*, *46*(1), 69 – 83

Czajkowski, M. and Budziński, W. (2019). Simulation error in maximum likelihood estimation of discrete choice models. *Journal of Choice Modelling*, *31*, 73 – 85

Czajkowski, M., Giergiczny, M. and Greene, W.H. (2014). Learning and Fatigue Effects Revisited: Investigating the Effects of Accounting for Unobservable Preference and Scale Heterogeneity. *Land Economics*, *90*(2), 324 – 351

Da Roit, B. (2012). The Netherlands: the struggle between universalism and cost containment. *Health and Social Care in the Community*, *20*(3), 228 – 237

Damman, O.C., Hendriks, M., Rademakers, J., Spreeuwenberg, P., Delnoij, D.M.J. and Groenewegen, P.P. (2011). Consumers' interpretation and use of comparative information on the quality of health care: the effect of presentation approaches. *Health Expectations*, *15*, 197 – 211

Dardanoni, V. and Guerriero, C. (2021). Young people's willingness to pay for environmental protection. *Ecological Economics*, *179*, 106853

Day, B. and Prades, J.-L.P. (2010). Ordering anomalies in choice experiments. *Journal of Environmental Economics and Management*, *59*, 271 – 285

Day, B., Bateman, I.J., Carson, R.T. et al. (2012). Ordering effects and choice set awareness in repeat-response stated preference studies. *Journal of Environmental Economics and Management*, *63*, 73 – 91

Dayan, E. and Bar-Hillel, M. (2011). Nudge to nobesity II: Menu positions influence food orders. *Judgment and Decision Making*, *6*(4), 333 – 342

De Bakker, E. and Dagevos, H. (2012). Reducing meat consumption in today's consumer society: Questioning the citizen-consumer gap. *Journal of Agricultural and Environmental Ethics*, *25*, 877 – 894

De Bekker-Grob, E.W., Donkers, B., Bliemer, M.C.J., Veldwijk, J. and Swait, J.D. (2020). Can healthcare choice be predicted using stated preference data? *Social Science & Medicine*, *246*, 112736

De Bekker-Grob, E.W., Hol, L., Donkers, B. et al. (2010). Labeled versus Unlabeled Discrete Choice Experiments in Health Economics: An Application to Colorectal Cancer Screening. *Value in Health*, *13*(2), 315 – 323

De Bekker-Grob, E.W., Ryan, M. and Gerard, K. (2012). Discrete choice experiments in health economics: a review of the literature. *Health Economics*, *21*(2), 145 – 172

De Bresser, J., Knoef, M. and Van Ooijen, R. (2022). Preferences for in-kind and in-cash home care insurance. *Journal of Health Economics*, *84*, 102626

De Bresser, J., Knoef, M. and Van Ooijen, R. (2024). *The market for life care annuities: using housing wealth to manage longevity and long-term care risk*. Working paper

De Jong, L., Schmidt, T., Stahmeyer, J.T., Eberhard, S., Zeidler, J. and Damm, K. (2022). Willingness to provide informal care to older adults in Germany: a discrete choice experiment. *European Journal of Health Economics*, *24*, 425 – 436

De Ruijter, A., Van Exel, J. and Mouter, N. (2025). *"It should be relevant, reliable and feasible": Introducing FACE, an instrument for assessing the face validity of choice experiments*. SSRN Working paper. Available at: https://dx.doi.org/10.2139/ssrn.5153288

De Winter, J.C.F., Zadpoor, A.A. and Dodou, D. (2014). The expansion of Google Scholar versus Web of Science: a longitudinal study. *Scientometrics*, *98*, 1547 – 1565

Dekker, T. (2014). Indifference based value of time measures for Random Regret Minimisation models. *Journal of Choice Modelling*, *12*, 10 – 20

Dekker, T., Hess, S., Brouwer, R. and Hofkes, M. (2016). Decision uncertainty in multi-attribute stated preference studies. *Resource and Energy Economics*, *43*, 57 – 73

Dekker, T., Koster, P. and Brouwer, R. (2014). Changing with the Tide: Semiparametric Estimation of Preference Dynamics. *Land Economics*, *90*(4), 717 – 745

Dekker, T., Koster, P. and Mouter, N. (2019). *The economics of Participatory Value Evaluation*. Tinbergen Institute Discussion Paper, No. TI 2019-008/VIII. Amsterdam/Rotterdam, the Netherlands: Tinbergen Institute

Dekker, T., Koster, P. and Mouter, N. (2024). A micro-econometric framework for Participatory Value Evaluation. *Journal of Choice Modelling*, *52*, 100507

Dellaert, B.G.C., Donkers, B. and Van Soest, A. (2012). Complexity Effects in Choice Experiment- Based Models. *Journal of Marketing Research*, *49*(3), 424 – 434

DeLong, K.L., Syrengelas, K.G., Grebitus, C. and Nayga Jr., R.M. (2021). Visual versus Text Attribute Representation in Choice Experiments. *Journal of Behavioral and Experimental Economics*, *94*, 101729

DeSarbo, W.S., Lehmann, D.R. and Hollman, F.G. (2004). Modeling Dynamic Effects in Repeated- Measures Experiments Involving Preference/Choice: An Illustration Involving Stated Preference Analysis. *Applied Psychological Measurement*, *28*(3), 186 – 209

Determann, D., Lambooij, M.S., Steyerberg, E.W., De Bekker-Grob, E.W. and De Wit, G.A. (2017). Impact of Survey Administration Mode on the Results of a Health-Related Discrete Choice Experiment: Online and Paper Comparison. *Value in Health*, *20*(7), 953 – 960

Dickinson, H., Smith, C., Carey, N. and Carey, G. (2021). Exploring governance tensions of disruptive technologies: the case of care robots in Australia and New Zealand. *Policy and Society*, *40*(2), 232 – 249

Diener, A., O'Brien, B. and Gafni, A. (1998). Health care contingent valuation studies: a review and classification of the literature. *Health Economics*, *7*(4), 313 – 326

Diepeveen, S., Ling, T., Suhrcke, M., Roland, M. and Marteau, T.M. (2013). Public acceptability of government intervention to change health-related behaviours: a systematic review and narrative synthesis. *BMC Public Health*, *13*, 756

Dieteren, C., Bonfrer, I., Brouwer, W.B.F. and Van Exel, J. (2023). Public preferences for policies promoting a healthy diet: a discrete choice experiment. *European Journal of Health Economics*, *24*, 1429 – 1440

Ding, M., Grewal, R. and Liechty, J. (2005). Incentive-aligned conjoint analysis. *Journal of Marketing Research*, *42*(1), 67 – 82

DiSantostefano, R.L., Smith, I.P., Falahee, M. et al. (2024). Research priorities to increase confidence in and acceptance of health preference research: What questions should be prioritized now? *The Patient*, *17*, 179 – 190

Dolan, P., Olsen, J.A., Menzel, P. and Richardson, J. (2003). An inquiry into the different perspectives that can be used when eliciting preferences in health. *Health Economics*, *12*, 545 – 551

Dvorak, F., Glenk, K., Logar, I. and Meyerhoff, J. (2023). *Cognitive models of Bayesian anchoring in discrete choice experiments*. SSRN Working paper. Available at: https://dx.doi.org/10.2139/ssrn.4473059

Dynata. (2022). *Dynata 2022 Panel Book*. Available at: https://www.dynata.com/content/Dynata-2022-Panel-Book.pdf

Eckermann, S. and Pekarsky, B. (2014). Can the Real Opportunity Cost Stand Up: Displaced Services, the Straw Man Outside the Room. *PharmacoEconomics*, *32*, 319 – 325

Edwards, W. and Barron, F.H. (1994). SMARTS and SMARTER: Improved Simple Methods for Multiattribute Utility Measurement. *Organizational Behavior and Human Decision Processes*, *60*(3), 306 – 325

Engmann, S. and Cousineau, D. (2011). Comparing distributions: the two-sample Anderson-Darling test as an alternative to the Kolmogorov-Smirnoff test. *Journal of Applied Quantitative Methods*, *6*(3), 1 – 17

Erdem, S., Campbell, D. and Hole, A.R. (2015). Accounting for Attribute-Level Non-Attendance in a Health Choice Experiment: Does it Matter? *Health Economics*, *24*(7), 773 – 789

Erdem, S. and Thompson, C. (2014). Prioritising health service innovation investments using public preferences: a discrete choice experiment. *BMC Health Services Research*, *14*, 360

Erdem, S., Campbell, D. and Thompson, C. (2014). Elimination and selection by aspects in health choice experiments: Prioritising health service innovations. *Journal of Health Economics*, *38*, 10 – 22

European Commission. (2023). *Estimates of cancer incidence and mortality in 2022, for all countries*. Retrieved from: https://ecis.jrc.ec.europa.eu/explorer.php?$0-0$1-All$2- All$4-1,2$3-27$6-0,85$5-2022,2022$7-7$CEstByCountry$X0_8-3$X0_19- AE27$X0_20-No$CEstBySexByCountry$X1_8-3$X1_19-AE27$X1_-1-1$CEstByIndiByCountry$X2_8-3$X2_19-AE27$X2_20-No$CEstRelative$X3_8- 3$X3_9-AE27$X3_19- AE27$CEstByCountryTable$X4_19-AE27

Eurostat. (2023a). *Causes of death – deaths by country of residence and occurrence.* Retrieved from: https://ec.europa.eu/eurostat/databrowser/view/hlth_cd_aro__custom_7963734/default/table?lang=en

Eurostat. (2023b). *Persons killed in road accidents by sex*. Retrieved from: https://ec.europa.eu/eurostat/databrowser/view/tran_sf_roadse__custom_7963048/default /table?lang=en

Eurostat. (n.d.). *Ageing Europe – statistics on population developments*. Available at: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Ageing_Europe_- _statistics_on_population_developments

Farrar, S. and Ryan, M. (1999). Response-Ordering Effects: A Methodological Issue in Conjoint Analysis. *Health Economics*, *8*, 75 – 79

Finch, L., Youl, P., Marshall, A.L., Soyer, H.P., Baade, P. and Jande, M. (2015). User preferences for text message-delivered skin cancer prevention and early detection. *Journal of Telemedicine and Telecare*, *21*(4)

Flynn, T.N., Louviere, J.J., Peters, T.J. and Coast, J. (2007). Best-worst scaling: What it can do for health care research and how to do it. *Journal of Health Economics*. *26*(1), 171 – 189

Franco, A., Malhotra, N. and Simonovits, G. (2014). Publication bias in the social sciences: Unlocking the file drawer. *Science*, *345*(6203), 1502 – 1505

Franke, T.M., Ho, T. and Christie, C.A. (2012). The Chi-Square Test: Often Used and More Often Misinterpreted. *American Journal of Evaluation*, *33*(3), 448 – 458

Frederick, S., Novemsky, N., Wang, J., Dhar, R. and Nowlis, S. (2009). Opportunity cost neglect. *Journal of Consumer Research*, *36*(4), 553 – 561

Gao, Z. and Schroeder, T.C. (2009). Effects of Label Information on Consumer Willingness-to-Pay for Food Attributes. *American Journal of Agricultural Economics*, *91*(3), 795 – 809

A

Garbarski, D., Schaeffer, N.C. and Dykema, J. (2016). The effect of response option order on self-rated health: a replication study. *Quality of Life Research*, *25*, 2117 – 2121

García-Souto, F., Durán-Romero, A.-J and Pereyra-Rodríguez, J.-J. (2021). Melanoma mortality in Spain: predictions up to 2043. *International Journal of Dermatology*, *60*(7), 844 – 850

Gaube, S., Biebl, I., Engelmann, M.K.M., Kleine, A.-K. and Lermer, E. (2024). Comparing preferences for skin cancer screening: AI-enabled app vs dermatologist. *Social Science & Medicine*, *349*, 116871

Genie, M.G., Nicoló, A. and Pasini, G. (2020). The role of heterogeneity of patients' preferences in kidney transplantation. *Journal of Health Economics*, *72*, 102331

Genie, M.G., Ryan, M. and Krucien, N. (2023). Keeping an eye on cost: What can eye tracking tell us about attention to cost information in discrete choice experiments? *Health Economics*, *32*(5), 1101 – 1119

Gerstenblüth, M., Harris, J.E. and Triunfo, P. (2022). *Addiction and Rational Choice: Evidence from an Eye Tracking Experiment with Cigarette Packages*. medRxiv working paper. Available at: https://doi.org/10.1101/2022.07.07.22277382

Geržinič, N., Van Cranenburgh, S., Cats, O., Lancsar, E. and Chorus, C. (2021). Estimating decision rule differences between 'best' and 'worst' choices in a sequential best worst discrete choice experiment. *Journal of Choice Modelling*, *41*, 100307

Glenk, K. (2007). A split-sample experiment to test for effects of attribute order in choice experiments. In: J. Meyerhoff, N. Lienhoop and P. Elsasser (eds.). *Stated Preference Methods for Environmental Valuation: Applications from Austria and Germany*. Metropolis

Glenk, K., Meyerhoff, J., Akaichi, F. and Martin-Ortega, J. (2019). Revisiting cost vector effects in discrete choice experiments. *Resource and Energy Economics*, *57*, 135 – 155

Gonçalves, T., Lourenço-Gomes, L. and Costa Pinto, L.M. (2022). The role of attribute non-attendance on consumer decision-making: Theoretical insights and empirical evidence. *Economic Analysis and Policy*, *76*, 788 – 805

Gonzalez, J.M. (2019). A Guide to Measuring and Interpreting Attribute Importance. *The Patient*, *12*, 287 – 295

Gordon, L.G. and Rowell, D. (2015). Health system costs of skin cancer and cost-effectiveness of skin cancer prevention and screening: a systematic review. *European Journal of Cancer Prevention*, *24*, 141 – 149

Green, C. (2001). On the societal value of health care: what do we know about the person trade-off technique? *Health Economics*, *10*(3), 233 – 243

Gregor, J.C., Williamson, M., Dajnowiec, D., Sattin, B., Sabot, E. and Salh, B. (2018). Inflammatory bowel disease patients prioritize mucosal healing, symptom control, and pain when choosing therapies: results of a prospective cross-sectional willingness-to-pay study. *Patient Preference and Adherence*, *12*, 505 – 513

Groeneveld, R.A. (2010). Framing and Training to Induce Preference Learning in Choice Experiments. *Marine Resource Economics*, *25*(2), 233 – 245

Groothuis, P.A., Mohr, T.M., Whitehead, J.C. and Cockerill, K. (2017). Endogenous Consequentiality in Stated Preference Referendum Data: The Influence of the Randomly Assigned Tax Amount. *Land Economics*, *93*(2), 258 – 268

Guo, J., Konetzka, R.T. and Manning, W.G. (2015). The Causal Effects of Home Care Use on Institutional Long-Term Care Utilization and Expenditures. *Health Economics*, *24*(S1), 4 – 17

Gustavsson, E. and Lindblom, L. (2025). Justification of principles for healthcare priority setting: the relevance and roles of empirical studies exploring public values. *Journal of Medical Ethics*, *51*, 285 – 292

Guy, G.P., Machlin, S.R., Ekwueme, D.U. and Yabroff, K.R. (2015). Prevalence and Costs of Skin Cancer Treatment in the U.S., 2002–2006 and 2007–2011. *American Journal of Preventive Medicine*, *48*(2), 183 – 187

Gyrd-Hansen, D. (2013). Using the Stated Preference Technique for Eliciting Valuations: The Role of the Payment Vehicle. *PharmacoEconomics*, *31*, 853 – 861

Gyrd-Hansen, D., Kjær, T. and Nielsen, J.S. (2016). The value of mortality risk reductions. Pure altruism – a confounder? *Journal of Health Economics*, *49*, 184 – 192

Hackl, M., Hanika, A., Klotz, J., Leitner, B. and Zielonke, N. (2015). *Trends der Entwicklung von Krebserkrankungen in Österreich*. Statistics Austria and Austrian Ministry for Health

Haddaway, N.R., Bethel, A., Dicks, L.V. et al. (2020). Eight problems with literature reviews and how to fix them. *Nature Ecology & Evolution*, *4*, 1582 – 1589

Haggenmüller, S., Krieghoff-Henning, E., Jutzi, T. et al. (2021). Digital Natives' Preferences on Mobile Artificial Intelligence Apps for Skin Cancer Diagnostics: Survey Study. *JMIR mHealth and uHealth*, *9*(8), e22909

Haghani, M., Bliemer, M.C.J. and De Bekker-Grob, E.W. (2022). Applications of discrete choice experiments in COVID-19 research: Disparity in survey qualities between health and transport fields. *Journal of Choice Modelling*, *44*, 100371

Haghani, M., Bliemer, M.C.J. and Hensher, D.A. (2021a). The landscape of econometric discrete choice modelling research. *Journal of Choice Modelling*, *40*, 100303

Haghani, M., Bliemer, M.C.J., Rose, J.M., Oppewal, H. and Lancsar, E. (2021b). Hypothetical bias in stated choice experiments: Part I. Macro-scale analysis of literature and integrative synthesis of empirical evidence from applied economics, experimental psychology and neuroimaging. *Journal of Choice Modelling*, *41*, 100309

Haghani, M., Bliemer, M.C.J., Rose, J.M., Oppewal, H. and Lancsar, E. (2021c). Hypothetical bias in stated choice experiments: Part II. Conceptualisation of external validity, sources and explanations of bias and effectiveness of mitigation methods. *Journal of Choice Modelling*, *41*, 100322

Hancock, T.O., Hess, S. and Choudhury, C.F. (2018). Decision field theory: Improvements to current methodology and comparisons with standard choice modelling techniques. *Transportation Research Part B: Methodological*, *107*, 18 – 40

Hansen, P. and Devlin, N. (2019). Multi-Criteria Decision Analysis (MCDA) in Healthcare Decision- Making. In: *Oxford Research Encyclopedia of Economics and Finance*. Oxford University Press

Haski-Leventhal, D., Grönlund, H., Holmes, K. et al. (2010). Service-Learning: Findings from a 14- Nation Study. *Journal of Nonprofit & Public Sector Marketing*, *22*, 161 – 179

A

Hauber, B. and Coulter, J. (2020). Using the Threshold Technique to Elicit Patient Preferences: An Introduction to the Method and an Overview of Existing Empirical Applications. *Applied Health Economics and Health Policy*, 18: 31 – 46

Hauber, A.B., González, J.M., Groothuis-Oudshoorn, C.G.M. et al. (2016). Statistical methods for the analysis of discrete choice experiments: A report of the ISPOR Conjoint Analysis Good Research Practices Task Force. *Value in Health*, *19*(4), 300 – 315

Hautnah. (2016). Hautkrebs: Jedes Jahr fünf Prozent mehr Neuerkrankungen. *Hautnah*, *15*, 100

Hayo, B. and Neumeier, F. (2019). Public Preferences for Government Spending Priorities: Survey Evidence from Germany. *German Economic Review*, *20*(4), 1 – 37

Head, M.L., Holman, L., Lanfear, R., Kahn, A.T. and Jennions, M.D. (2015). The Extent and Consequences of P-Hacking in Science. *PLoS Biology*, *13*(3), e1002106

Heidenreich, S., Phillips-Beyer, A., Flamion, B., Ross, M., Seo, J. and Marsh, K. (2021). Benefit-Risk or Risk-Benefit Trade-Offs? Another Look at Attribute Ordering Effects in a Pilot Choice Experiment. *The Patient*, *14*, 65 – 74

Heidenreich, S., Watson, V., Ryan, M. and Phimister, E. (2017). Decision heuristic or preference? Attribute non-attendance in discrete choice problems. *Health Economics*, *27*(1), 157 – 171

Heitmüller, A. and Inglis, K. (2007). The earnings of informal carers: Wage differentials and opportunity costs. *Journal of Health Economics*, *26*(4), 821 – 841

Hensher, D.A. (2006). How do respondents process stated choice experiments? Attribute consideration under varying information load. *Journal of Applied Econometrics*, *21*(6), 861 – 878

Hensher, D.A. and Collins, A.T. (2011). Interrogation of Responses to Stated Choice Experiments: Is there sense in what respondents tell us? *Journal of Choice Modelling*, *4*(1), 62 – 89

Hensher, D.A., Rose, J.M. and Greene, W.H. (2015). *Applied choice analysis*. Cambridge, UK: Cambridge University Press

Hernandez, J.I. (2023). *Portchoice 0.4.1*. Available at: https://pypi.org/project/portchoice/

Hess, S. and Palma, D. (2019). Apollo: a flexible, powerful and customisable freeware package for choice model estimation and application. *Journal of Choice Modelling*, *32*, 100170

Hess, S., Daly, A. and Batley, R. (2018). Revisiting consistency with random utility maximisation: theory and implications for practical work. *Theory and Decision*, *84*, 181 – 204

Hess, S., Hensher, D.A. and Daly, A. (2012). Not bored yet – Revisiting respondent fatigue in stated choice experiments. *Transportation Research Part A: Policy and Practice*, *46*, 626 – 644

Hildebrand, K., Chung, C., Boyer, T.A. and Palma, M. (2023). Does change in respondents' attention affect willingness to accept estimates from choice experiments? *Applied Economics*, *55*(28), 3279 – 3295

Himmler, S., Soekhai, V., Van Exel, J. and Brouwer, W. (2021). What works better for preference elicitation among older people? Cognitive burden of discrete choice experiment and case 2 best-worst scaling in an online setting. *Journal of Choice Modelling*, *38*, 100265

Hindriks, J. and Myles, G.D. (2006). *Intermediate Public Economics*. Cambridge, MA: MIT Press

Hinkema, M., Kompatscher, K., Egter van Wissekerke, N., Van Heumen, S. and Balkuv, F. (2020). *Prognose capaciteitsontwikkeling verpleeghuiszorg fase II.* TNO 2020 R11184. Delft, the Netherlands: TNO

Hirst, N.G., Gordon, L.G., Scuffham, P.A. and Green, A.C. (2012). Lifetime Cost-Effectiveness of Skin Cancer Prevention through Promotion of Daily Sunscreen Use. *Value in Health*, *15*, 261 – 268

Hoefman, R.J., Meulenkamp, T.M. and De Jong, J.D. (2017). Who is responsible for providing care? Investigating the role of care tasks and past experiences in a cross-sectional study in the Netherlands. *BMC Health Services Research*, *17*(1), 477

Hoefman, R.J., Van Exel, J. and Brouwer, W.B.F. (2019). The Monetary Value of Informal Care: Obtaining Pure Time Valuations Using a Discrete Choice Experiment. *PharmacoEconomics*, *37*, 531 – 540

Hole, A.R. (2004). Forecasting the demand for an employee Park and Ride service using commuters' stated choices. *Transport Policy*, *11*(4), 355 – 362

Hole, A.R., Kolstad, J.R. and Gyrd-Hansen, D. (2013). Inferred vs. stated attribute non-attendance in choice experiments: A study of doctors' prescription behaviour. *Journal of Economic Behavior & Organization*, *96*, 21 – 31

Hollin, I.L., Paskett, J., Schuster, A.L.R., Crossnohere, N.L. and Bridges, J.F.P. (2022). Best-Worst Scaling and the Prioritization of Objects in Health: A Systematic Review. *PharmacoEconomics*, *40*, 883 – 899

Hollingsworth, B., Ohinata, A., Picchio, M. and Walker, I. (2022). The Impacts of Free Universal Elderly Care on the Supply of Informal Care and Labour Supply. *Oxford Bulletin of Economics and Statistics*, *84*(4), 933 – 960

Holman, D.M., Kapelos, G.T., Shoemaker, M. and Watson, M. (2018). Shade as an Environmental Design Tool for Skin Cancer Prevention. *American Journal of Public Health*, *108*, 1607 – 1612

Hössinger, R., Peer, S. and Juschten, M. (2023). Give citizens a task: an innovative tool to compose policy bundles that reach the climate goal. *Transportation Research Part A: Policy and Practice*, *173*, 103694

Houston, N.A.M., Secrest, A.M., Harris, R.J. et al. (2016). Patient Preferences During Skin Cancer Screening Examination. *JAMA Dermatology*, *152*(9), 1052 – 1054

Howley, P., Hynes, S. and O'Donoghue, C. (2010). The citizen versus consumer distinction: An exploration of individuals' preferences in Contingent Valuation studies. *Ecological Economics*, *69*, 1524 – 1531

Hu, W., Fang, L., Ni, R., Zhang, H. and Pan, G. (2022). Changing trends in the disease burden of non-melanoma skin cancer globally from 1990 to 2019 and its predicted level in 25 years. *BMC Cancer*, *22*, 836

Huynh, E., Swait, J., Lancsar, E. and Ride, J. (2024). Menu choice stated preference tasks to capture demand complementarity in health. *Social Science & Medicine*, *344*, 116636

Jacobsen, J.B. and Thorsen, B.J. (2010). Preferences for site and environmental functions when selecting forthcoming national parks. *Ecological Economics*, *69*, 1532 – 1544

Janus, A.L. and Koslowski, A. (2020). Whose responsibility? Elder support norms regarding the provision and financing of assistance with daily activities across economically developed countries. *European Journal of Ageing*, *17*, 95 – 108

Jarvis, S.L. (2011). *Stated Preference Methods and Models: Analyzing Recreational Angling in New England Groundfisheries*. Dissertation defended at University of Maryland, College Park

Jiang, N., Ao, C., Xu, L., Wei, Y. and Long, Y. (2023). Will information interventions affect public preferences and willingness to pay for air quality improvement? An empirical study based on deliberative choice experiment. *Science of the Total Environment*, *868*, 161436

A

Jiang, Q., Penn, J. and Hu, W. (2022). *Learning and fatigue in real discrete choice experiments*. AAEA Conference Paper. Available at: https://ageconsearch.umn.edu/record/322516

Johnson, F.R., Lancsar, E., Marhsall, D. et al. (2013). Constructing experimental designs for discrete-choice experiments: Report of the ISPOR Conjoint Analysis Experimental Design Good Research Practices Task Force. *Value in Health*, *16*(1), 3 – 13

Johnston, R.J., Boyle, K.J., Adamowicz, W.L. et al. (2017). Contemporary guidance for stated preference studies. *Journal of the Association of Environmental and Resource Economists*, *4*(2), 319 – 405

Johnston, R.J., Rolfe, J. and Zawojska, E. (2018). Benefit transfer of environmental and resource values: Progress, prospects and challenges. *International Review of Environmental and Resource Economics*, *12*(2–3), 177 – 266

Jonker, M.F. (2024). Level overlap and level color coding revisited: Improved attribute attendance and higher choice consistency in discrete choice experiments. *Journal of Choice Modelling*, *52*, 100494

Jonker, M.F., Donkers, B., De Bekker-Grob, E.W. and Stolk, E.A. (2018a). Attribute level overlap (and color coding) can reduce task complexity, improve choice consistency, and decrease the dropout rate in discrete choice experiments. *Health Economics*, *28*, 350 – 363

Jonker, M.F., Donkers, B., De Bekker-Grob, E.W. and Stolk, E.A. (2018b). Effect of Level Overlap and Color Coding on Attribute Non-Attendance in Discrete Choice Experiments. *Value in Health*, *21*(7), 767 – 771

Jonker, M.F., Roudijk, B. and Maas, M. (2022). The Sensitivity and Specificity of Repeated and Dominant Choice Tasks in Discrete Choice Experiments. *Value in Health*, *25*(8), 1381 – 1389

Josten, E.J., Verbakel, E. and De Boer, A.H. (2024). A longitudinal study on the consequences of the take-up of informal care on work hours, labour market exit and workplace absenteeism due to illness. *Ageing & Society*, *44*(3), 495 – 518

Jung, M.H., Moon, A. and Nelson, L.D. (2020). Overestimating the valuations and preferences of others. *Journal of Experimental Psychology: General*, *149*(6), 1193 – 1214

Juschten, M. and Omann, I. (2023). Evaluating the relevance, credibility and legitimacy of a novel participatory online tool. *Environmental Science & Policy*, *146*, 90 – 100

Kaambwa, B., Chen, G., Khadka, J. et al. (2021). A preference for quality: Australian general public's willingness to pay for home and residential aged care. *Social Science & Medicine*, *289*, 114425

Kafle, A., Swallow, S.K. and Smith, E.C. (2015). Does Public Funding Affect Preferred Tradeoffs and Crowd-In or Crowd-Out Willingness to Pay? A Watershed Management Case. *Environmental & Resource Economics*, *60*, 471 – 495

Kasparian, N.A., McLoone, J.K. and Meiser, B. (2009). Skin cancer-related prevention and screening behaviors: a review of the literature. *Journal of Behavioral Medicine*, *32*, 406 – 428

Kattenberg, M. and Bakx, B. (2021). Substitute services: a barrier to controlling long-term care expenditures. *European Journal of Ageing*, *18*(1), 85 – 97

Keita Fakeye, M.B., Samuel, L.J., Drabo, E.F., Bandeen-Roche, K. and Wolff, J.L. (2023). Caregiving-Related Work Productivity Loss Among Employed Family and Other Unpaid Caregivers of Older Adults. *Value in Health*, *26*(5), 712 – 720

Keshavarzian, P. and Wu, C.-L. (2021). Exploring the Effect of Sequentially Receiving Airline and Destination Information on the Choice Behavior of Tourism Destinations. *Journal of Travel Research*, *60*(2), 251 – 266

Kim, J. and Morgül, K. (2017). Long-term consequences of youth volunteering: Voluntary versus involuntary service. *Social Science Research*, *67*, 160 – 175

Kjær, T., Bech, M., Gyrd-Hansen, D. and Hart-Hansen, K. (2006). Ordering effect and price sensitivity in discrete choice experiments: Need we worry? *Health Economics*, *15*(11), 1217 – 1228

Kling, C.L., Phaneuf, D.J. and Zhao, J. (2012). From Exxon to BP: Has some number become better than no number? *Journal of Economic Perspectives*, *26*(4), 3 – 26

Koç, H. and Van Kippersluis, H. (2017). Thought for Food: Nutritional Information and Educational Disparities in Diet. *Journal of Human Capital*, *11*(4), 508 – 552

Koetse, M.J. (2017). Effects of payment vehicle non-attendance in choice experiments on value estimates and the WTA–WTP disparity. *Journal of Environmental Economics and Policy*, *6*(3), 225 – 245

Kok, L., Berden, C. and Sadiraj, K. (2015). Costs and benefits of home care for the elderly versus residential care: a comparison using propensity scores. *European Journal of Health Economics*, *16*, 119 – 131

Kölln, A.-K. and Wlezien, C. (2024). Measuring public preferences for government spending under constraints: a conjoint-analytic approach. *Journal of Elections, Public Opinion and Parties*, *34*(2), 375 – 386

Koppelman, F.S. and Sethi, V. (2005). Incorporating variance and covariance heterogeneity in the Generalized Nested Logit model: an application to modeling long distance travel choice behavior. *Transportation Research Part B: Methodological*, *39*, 825 – 853

Køster, B., Meyer, M.K.H., Søgaard, J. and Dalum, P. (2020). Benefit-Cost Analysis of the Danish Sun Safety Campaign 2007 – 2015: Cost Savings from Sunburn and Sunbed Use Reduction and Derived Skin Cancer Reductions 2007 – 2040 in the Danish Population. *PharmacoEconomics - Open*, *4*(3), 419 – 425

Krabbe-Alkemade, Y., Makai, P., Shestalova, V. and Voesenek, T. (2020). Containing or shifting? Health expenditure decomposition for the aging Dutch population after a major reform. *Health Policy*, *124*, 268 – 274

Kragt, M.E. and Bennett, J.W. (2012). Attribute Framing in Choice Experiments: How Do Attribute Level Descriptions Affect Value Estimates? *Environmental and Resource Economics*, 51, 43 – 59

Krosnick, J.A. and Alwin, D.F. (1987). An evaluation of a cognitive theory of response-order effects in survey measurement. *Public Opinion Quarterly*, 51, 201 – 219

Krucien, N., Ryan, M. and Hermens, F. (2017a). Visual attention in multi-attributes choices: What can eye-tracking tell us? *Journal of Economic Behavior & Organization*, *135*, 251 – 267

Krucien, N., Watson, V. and Ryan, M. (2017b). Is best-worst scaling suitable for health state valuation? A comparison with discrete choice experiments. *Health Economics*, *26*(12), e1 – e16

Kruijshaar, M.E., Essink-Bot, M.L., Donkers, B. et al. (2009). A labelled discrete choice experiment adds realism to the choices presented: preferences for surveillance tests for Barrett esophagus. *BMC Medical Research Methodology*, *9*:31

Kumar, V. and Gaeth, G.J. (1991). Attribute order and product familiarity effects in decision tasks using conjoint analysis. *International Journal of Research in Marketing*, *8*(2), 113 – 124

A

Ladenburg, J. (2013). Does gender-specific starting point bias in choice experiments prevail among well-informed respondents: evidence from an empirical study. *Applied Economics Letters*, *20*(17), 1527 – 1530

Ladenburg, J. and Olsen, S.B. (2008). Gender-specific starting point bias in choice experiments: Evidence from an empirical study. *Journal of Environmental Economics and Management*, *56*(3), 275 – 285

Lades, L.K., Zawojska, E., Johnston, R.J., Hanley, N., Delaney, L. and Czajkowski, M. (2022). *Anomalies or Expected Behaviors? Understanding Stated Preferences and Welfare Implications in Light of Contemporary Behavioral Theory*. Working Paper Series Faculty of Economic Sciences, No. 20/2022. Warsaw, Poland: University of Warsaw.

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

Lancsar, E. and Louviere, J. (2008). Conducting Discrete Choice Experiments to Inform Healthcare Decision Making. *PharmacoEconomics*, *26*, 661 – 677

Lancsar, E. and Savage, E. (2004). Deriving welfare measures from discrete choice experiments: inconsistency between current methods and random utility and welfare theory. *Health Economics*, *13*(9), 901 – 907

Lancsar, E., Fiebig, D.G. and Hole, A.R. (2017). Discrete choice experiments: A guide to model specification, estimation and software. *PharmacoEconomics*, *35*, 697 – 716

Lancsar, E., Ride, J., Black, N., Burgess, L. and Peeters, A. (2022). Social acceptability of standard and behavioral economic inspired policies to reduce and prevent obesity. *Health Economics*, *31*(1), 197 – 214

Lanz, B. and Provins, A. (2013). Valuing Local Environmental Amenity with Discrete Choice Experiments: Spatial Scope Sensitivity and Heterogeneous Marginal Utility of Income. *Environmental and Resource Economics*, *56*, 105 – 130

Lanz, B. and Provins, A. (2015). Using discrete choice experiments to regulate the provision of water services: do status quo choices reflect preferences? *Journal of Regulatory Economics*, *47*, 300 – 324

Lattin, J.M. and Mcalister, L. (1985). Using a variety-seeking model to identify substitute and complementarity relationships among competing products. *Journal of Marketing Research*, *22*(3), 330 – 339

Lehnert, T., Günther, O.H., Hajek, A., Riedel-Heller, S.G. and König, H.-H. (2018). Preferences for home- and community-based long-term care services in Germany: a discrete choice experiment. *European Journal of Health Economics*, *19*, 1213 – 1223

Leiter, U., Keim, U. and Garbe, C. (2020). Epidemiology of Skin Cancer: Update 2020. In: J. Reichrath (ed). *Sunlight, Vitamin D and Skin Cancer.* Advances in Experimental Medicine and Biology. Springer. pp. 123 – 139

Lew, D.K. (2018). Discounting future payments in stated preference choice experiments. *Resource and Energy Economics*, *54*, 150 – 164

Lewis, K.E., Grebitus, C. and Nayga, R.M. (2016). U.S. consumers' preferences for imported and genetically modified sugar: Examining policy consequentiality in a choice experiment. *Journal of Behavioral and Experimental Economics*, *65*, 1 – 8

Li, Q., Huang, Z. and Christianson, K. (2016). Visual attention toward tourism photographs with text: An eye-tracking study. *Tourism Management*, *54*, 243 – 258

Liebe, U., Glenk, K., Oehlmann, M. and Meyerhoff, J. (2015). Does the use of mobile devices (tablets and smartphones) affect survey quality and choice behaviour in web surveys? *Journal of Choice Modelling*, *14*, 17 – 31

Liebe, U., Hundeshagen, C., Beyer, H. and Von Cramon-Taubadel, S. (2016). Context effects and the temporal stability of stated preferences. *Social Science Research*, *60*, 135 – 147

Lloyd-Smith, P. (2018). A new approach to calculating welfare measures in Kuhn-Tucker demand models. *Journal of Choice Modelling*, *26*, 19 – 27

Lloyd-Smith, P. (2020). Kuhn-Tucker and multiple discrete-continuous extreme value model estimation and simulation in R: The rmdcev package. *R Journal*, *12*(2), 266 – 292

Lloyd-Smith, P., Adamowicz, W. and Dupont, D. (2019). Incorporating Stated Consequentiality Questions in Stated Preference Research. *Land Economics*, *95*(3), 293 – 306

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

Logar, I., Brouwer, R. and Campbell, D. (2020). Does attribute order influence attribute-information processing in discrete choice experiments? *Resource and Energy Economics*, *60*, 101164

Løken, K.V., Lundberg, S. and Riise, J. (2017). Lifting the burden: Formal care of the elderly and labor supply of adult children. *Journal of Human Resources*, *52*(1), 247 – 271

Louviere, J.J., Islam, T., Wasi, N., Street, D. and Burgess, L. (2008). Designing Discrete Choice Experiments: Do Optimal Designs Come at a Price? *Journal of Consumer Research*, *35*(2), 360 - 375

Lundhede, T.H., Olsen, S.B., Jacobsen, J.B. and Thorsen, B.J. (2009). Handling respondent uncertainty in Choice Experiments: Evaluating recoding approaches against explicit modelling of uncertainty. *Journal of Choice Modelling*, *2*(2), 118 – 147

Luyten, J., Beutels, P., Vandermeulen, C. and Kessels, R. (2022). Social preferences for adopting new vaccines in the national immunization program: A discrete choice experiment. *Social Science & Medicine*, *303*, 114991

Maarse, J.A.M. and Jeurissen, P.P. (2016). The policy and politics of the 2015 long-term care reform in the Netherlands. *Health Policy*, *120*, 241 - 245

Maddala, T., Phillips, K.A. and Johnson, F.R. (2003). An experiment on simplifying conjoint analysis designs for measuring preferences. *Health Economics*, *12*(12), 1035 – 1047

Maguire, A., Persson, E. and Tinghög, G. (2023). Opportunity cost neglect: a meta-analysis. *Journal of the Economic Science Association*, *9*, 176 – 192

Mahieu, P.-A., Andersson, H., Beaumais, O., Crastes dit Sourd, R., Hess, S. and Wolff, F.-C. (2017). Stated preferences: a unique database composed of 1657 recent published articles in journals related to agriculture, environment, or health. *Review of Agricultural, Food and Environmental Studies*, *98*, 201 – 220

Makarov, A.A. and Simonova, G.I. (2018). Comparative analysis of the powers of the two-sample Kolmogorov-Smirnov and Anderson-Darling tests under various alternatives. *Journal of Mathematical Sciences*, *228*(5), 495 – 500

A

Manski, C.F. (1977). The structure of random utility models. *Theory and Decision*, *8*, 229 – 254

Mantonakis, A., Rodero, P., Lesschaeve, I. and Hastie, R. (2009). Order in Choice: Effects of Serial Position on Preferences. *Psychological Science*, *20*(11), 1309 - 1312

Mariel, P. and Meyerhoff, J. (2016). Hybrid discrete choice models: Gained insights versus increasing effort. *Science of the Total Environment*, *568*, 433 – 443

Mariel, P., Hoyos, D., Meyerhoff, J. et al. (2021). *Environmental valuation with discrete choice experiments: Guidance on design, implementation and data analysis*. Cham, Switzerland: SpringerNature

Marsh, D. and Phillips, Y. (2012). *Difficult Choices: What Influences the Error Variance in a Choice Experiment?* NZARES Conference Paper. Available at: https://ageconsearch.umn.edu/record/139651

Marsh, K., Gonzalez Sepulveda, J.M., Berlin, C. et al. (2025). What next for the science of patient preference? Interoperability, standardization, and transferability. *The Patient*, *18*, 101 – 108

Marsh, K., IJzerman, M., Thokala, P. et al. (2016). Multiple Criteria Decision Analysis for Health Care Decision Making – Emerging Good Practices: Report 2 of the ISPOR MCDA Emerging Good Practices Task Force. *Value in Health*, *19*(2), 125 – 137

Marsh, K., Van Til, J.A., Molsen-David, E. et al. (2020). Health Preference Research in Europe: A Review of Its Use in Marketing Authorization, Reimbursement, and Pricing Decisions – Report of the ISPOR Stated Preference Research Special Interest Group. *Value in Health*, *23*(7), 831 – 841

Martín-Martín, A., Orduna-Malea, E. and Delgado López-Cózar, E. (2018a). Coverage of highly-cited documents in Google Scholar, Web of Science, and Scopus: a multidisciplinary comparison. *Scientometrics*, *116*, 2175 – 2188

Martín-Martín, A., Orduna-Malea, E., Thelwall, M. and Delgado López-Cózar, E. (2018b). Google Scholar, Web of Science, and Scopus: A systematic comparison of citations in 252 subject categories. *Journal of Informetrics*, *12*, 1160 – 1177

Mattmann, M., Logar, I. and Brouwer, R. (2016). Wind power externalities: a meta-analysis. *Ecological Economics*, *127*, 23 – 36

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. In: P. Zarembka (ed.). *Frontiers in Econometrics*. New York: Academic Press

McFadden, D. (2001). Economic choices. *American Economic Review*, *91*(3), 351 – 378

McIntosh, E. (2006). Using discrete choice experiments within a cost-benefit analysis framework. *PharmacoEconomics*, *24*(9), 855 – 868

McNair, B.J., Bennett, J. and Hensher, D.A. (2011). A comparison of responses to single and repeated discrete choice questions. *Resource and Energy Economics*, *33*(3), 554 – 571

McNair, B.J., Hensher, D.A. and Bennett, J. (2012). Modelling Heterogeneity in Response Behaviour Towards a Sequence of Discrete Choice Questions: A Probabilistic Decision Process Model. *Environmental and Resource Economics*, *51*, 599 – 616

Meertens, A., Van Coile, L., Van Iseghem, T., Brochez, L., Verhaege, N. and Hoorens, I. (2024). Cost- of-Illness of Skin Cancer: A Systematic Review. *PharmacoEconomics*, *42*, 751 – 765

Meester, D.A.J., Hess, S., Buckell, J. and Hancock, T.O. (2023). Can decision field theory enhance our understanding of health-based choices? Evidence from risky health behaviors. *Health Economics*, *32*(8), 1710 – 1732

Meginnis, K., Hanley, N., Mujumbusi, L. and Lamberton, P.H.L. (2020). Non-monetary numeraires: Varying the payment vehicle in a choice experiment for health interventions in Uganda. *Ecological Economics*, *170*, 106569

Meißner, M., Musalem, A. and Huber, J. (2016). Eye tracking reveals processes that enable conjoint choices to become increasingly efficient with practice. *Journal of Marketing Research*, *53*(1), 1 – 17

Meldrum, J.R., Champ, P., Bond, C. and Schoettle, A. (2020). Paired Stated Preference Methods for Valuing Management of White Pine Blister Rust: Order Effects and Outcome Uncertainty. *Journal of Forest Economics*,

Mendelsohn, R. (2019). An Examination of Recent Revealed Preference Valuation Methods and Results. *Review of Environmental Economics and Policy*, *13*(2), 267 – 282

Mentzakis, E., Ryan, M. and McNamee, P. (2011). Using discrete choice experiments to value informal care tasks: exploring preference heterogeneity. *Health Economics*, *20*(8), 930 – 944

Metcalfe, P.J., Baker, W., Andrews, K. et al. (2012). An assessment of the nonmarket benefits of the Water Framework Directive for households in England and Wales. *Water Resources Research*, *48*, W03526

Meyerhoff, J. and Glenk, K. (2015). Learning how to choose – effects of instructional choice sets in discrete choice experiments. *Resource and Energy Economics*, *41*, 122 – 142

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

Milte, R., Ratcliffe, J., Chen, G. and Crotty, M. (2018). What Characteristics of Nursing Homes Are Most Valued by Consumers? A Discrete Choice Experiment with Residents and Family Members. *Value in Health*, *21*(7), 843 – 849

Milte, R., Ratcliffe, J., Kumaran, S. et al. (2024). Public Attitudes for Quality and Funding of Long- Term Care: Findings from an Australian Survey. *Health & Social Care in the Community*, *2024*(1), 5798242

Ministry of Education, Culture and Science of the Netherlands. (2023). *MDT brengt samen en verbindt: Voortgang Maatschappelijke Diensttijd (MDT), Mei 2024*. The Hague, the Netherlands: Rijksoverheid

Miyawaki, A., Kobayashi, Y., Noguchi, H., Watanabe, T., Takahashi, T. and Tamiya, N. (2020). Effect of reduced formal care availability on formal/informal care patterns and caregiver health: a quasi-experimental study using the Japanese long-term care insurance reform. *BMC Geriatrics*, *20*, 207

Mokas, I., Lizin, S., Brijs, T., Witters, N. and Malina, R. (2021). Can immersive virtual reality increase respondents' certainty in discrete choice experiments? A comparison with traditional presentation formats. *Journal of Environmental Economics and Management*, *109*, 102509

Molin, E., Mokhtarian, P. and Kroesen, M. (2016). Multimodal travel groups and attitudes: A latent class cluster analysis of Dutch travelers. *Transportation Research Part A: Policy and Practice*, *83*, 14 – 29

Mommaerts, C. (2025). Long-Term Care Insurance and the Family. *Journal of Political Economy*, *133*(1), 1 – 52

Morais, P. (2022). Artificial tanning devices (Sunbeds): where do we stand? *Cutaneous and Ocular Toxicology*, *41*(2), 123 – 128

A

Morrison, M. and MacDonald, D.H. (20 11). A comparison of compensating surplus and budget reallocation with opportunity costs specified. *Applied Economics*, *43*, 4677 – 4688

Mosca, I., Van der Wees, P.J., Mot, E.S., Wammes, J.J.G. and Jeurissen, P.P.T. (2017). Sustainability of Long-term Care: Puzzling Tasks Ahead for Policy-Makers. *International Journal of Health Policy and Management*, *6*(4), 195 – 205

Moura, A. (2022). Do subsidized nursing homes and home care teams reduce hospital bed-blocking? Evidence from Portugal. *Journal of Health Economics*, 84, 102640

Mouter, N. (2021). Willingness to allocate public budget and Participatory Value Evaluation. In: N. Mouter (ed.). *New Methods, Reflections and Application Domains in Transport Appraisal*. Advances in Transport Policy and Planning, Vol. 7. Cambridge: Academic Press; pp. 83 – 102

Mouter, N., Boxebeld, S., Kessels, R. et al. (2022a). Public Preferences for Policies to Promote COVID-19 Vaccination Uptake: A Discrete Choice Experiment in The Netherlands. *Value in Health*, *25*(8), 1290 – 1297

Mouter, N., Hernández, J.I. and Itten, A.V. (2021a). Public participation in crisis policymaking. How 30,000 Dutch citizens advised their government on relaxing COVID-19 lockdown measures. *PLoS ONE*, *16*(5), e0250614

Mouter, N., Koster, P. and Dekker, T. (2021b). Contrasting the recommendations of participatory value evaluation and cost-benefit analysis in the context of urban mobility investments. *Transportation Research Part A: Policy and Practice*, 144, 54 – 73

Mouter, N., Koster, P. and Dekker, T. (2021c). Participatory value evaluation for the evaluation of flood protection schemes. *Water Resources and Economics*, *36*, 100188

Mouter, N., Shortall, R., Spruit, S. and Itten, A. (2021d). Including young people, cutting time and protecting useful outcomes: Participatory Value Evaluation as a new practice of public participation in the Dutch energy transition. *Energy Research & Social Science*, *75*, 101965

Mouter, N, Trujillo Jara, K., Hernandez, J.I. et al. (2022b). Stepping into the shoes of the policy maker: Results of a Participatory Value Evaluation for the Dutch long term COVID-19 strategy. *Social Science & Medicine*, *314*, 115430

Mouter, N., Van Cranenburgh, S. and Van Wee, B. (2017). Do individuals have different preferences as consumer and citizen? The trade-off between travel time and safety. *Transportation Research Part A: Policy and Practice*, *106*, 333 – 349

Mouter, N., Van Cranenburgh, S. and Van Wee, B. (2018). The consumer-citizen duality: Ten reasons why citizens prefer safety and drivers desire speed. *Accident Analysis and Prevention*, *121*, 53 – 63

Mühlbacher, A. and Johnson, F.R. (2016). Choice Experiments to Quantify Preferences for Health and Healthcare: State of the Practice. *Applied Health Economics and Health Policy*, *14*, 253 – 266

Mühlbacher, A.C., Kaczynski, A., Zweifel, P. and Johnson, F.R. (2016). Experimental measurement of preferences in health and healthcare using best-worst scaling: an overview. *Health Economics Review*, *6*, 5

Mulderij, L.S., Hernández, J.I., Mouter, N., Verkooijen, K.T. and Wagemakers, A. (2021). Citizen preferences regarding the public funding of projects promoting a healthy body weight among people with a low income. *Social Science & Medicine*, *280*, 114015

Mulhern, B., Norman, R., De Abreu Lourenco, R. et al. (2019). Investigating the relative value of health and social care related quality of life using a discrete choice experiment. *Social Science & Medicine*, *233*, 28 – 37

Mulhern, B., Norman, R., Lorgelly, P. et al. (2017). Is Dimension Order Important when Valuing Health States Using Discrete Choice Experiments Including Duration? *PharmacoEconomics*, *35*, 439 – 451

Mulhern, B., Shah, K., Janssen, M.F., Longworth, L. and Ibbotson, R. (2016). Valuing Health Using Time Trade-Off and Discrete Choice Experiment Methods: Does Dimension Order Impact on Health State Values? *Value in Health*, *19*(2), 210 – 217

Nabatchi, T. (2012). Putting the "public" back in public values research: designing participation to identify and respond to values. *Public Administration Review*, *72*(5), 699 – 708

Needham, K., Czajkowski, M., Hanley, N. and LaRiviere, J. (2018). What is the causal impact of information and knowledge in stated preference studies? *Resource and Energy Economics*, *54*, 69 – 89

Neill, C.L. and Lahne, J. (2022). Matching reality: A basket and expenditure based choice experiment with sensory preferences. *Journal of Choice Modelling*, *44*, 100369

Nguyen, L., Jokimäki, H., Linnosmaa, I. et al. (2022). Valuing informal carers' quality of life using best-worst scaling – Finnish preference weights for the Adult Social Care Outcomes Toolkit for carers (ASCOT-Carer). *European Journal of Health Economics*, *23*, 357 – 374

Nguyen, T.C., Le, H.T., Nguyen, H.D., Ngo, M.T. and Nguyen, H.Q. (2021). Examining ordering effects and strategic behaviour in a discrete choice experiment. *Economic Analysis and Policy*, *70*, 394 – 413

Nguyen, T.C., Robinson, J., Kaneko, S. and Chinh, N.T. (2015). Examining ordering effects in discrete choice experiments: A case study in Vietnam. *Economic Analysis and Policy*, *45*, 39 – 57

Nichols, D. (n.d.). *Coloring for Colorblindness*. Retrieved in 2025 from: https://davidmathlogic.com/colorblind/#%23D81B60-%231E88E5-%23FFC107-%23004D40

Nieboer, A.P., Koolman, X. and Stolk, E. (2010). Preferences for long-term care services: Willingness to pay estimates derived from a discrete choice experiment. *Social Science & Medicine*, *70*, 1317 – 1325

Noels, E., Hollestein, L., Luijx, K. et al. (2020). Increasing Costs of Skin Cancer due to Increasing Incidence and Introduction of Pharmaceuticals, 2007 – 2017. *Acta Dermato-Venereologica*, *100*(10), adv00147

Nord, E. (1995). The Person-trade-off Approach to Valuing Health Care Programs. *Medical Decision Making*, *15*(3), 201 – 208

Norman, R., Kemmler, G., Viney, R. et al. (2016a). Order of Presentation of Dimensions Does Not Systematically Bias Utility Weights from a Discrete Choice Experiment. *Value in Health*, *19*(8), 1033 – 1038

Norman, R., Viney, R., Aaronson, N.K. et al. (2016b). Using a discrete choice experiment to value the QLU-C10D: feasibility and sensitivity to presentation format. *Quality of Life Research*, *25*, 637 – 649

Nova, G. and Guevara, C.A. (2023). *In-depth, breadth-first, or both? Characterising the information search process in a public transport SP experiment*. Research Square working paper. Available at: https://doi.org/10.21203/rs.3.rs-2772318/v1

Nunes, P.A.L.D. and Travisi, C.M. (2009). Comparing Tax and Tax Reallocation Payments in Financing Rail Noise Abatement Programmes: Results from a Stated Choice Valuation Study in Italy. *Environmental & Resource Economics*, *43*, 503 – 517

A

Nyborg, K. (2000). Homo Economicus and Homo Politicus: interpretation and aggregation of environmental values. *Journal of Economic Behavior & Organization*, *42*, 305 – 322

OECD. (2023). *Beyond Applause? Improving Working Conditions in Long-Term Care*. Paris, France: OECD Publishing

OECD. (2023). *Monthly comparative price levels (November 2023).* Retrieved from: https://stats.oecd.org/Index.aspx?QueryId=24057#

OECD. (n.d.). *Health expenditure and financing*. Available at: https://stats.oecd.org/

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

Ohdoko, T. and Yoshida, K. (2012). Public preferences for forest ecosystem management in Japan with emphasis on species diversity. *Environmental Economics & Policy Studies*, *14*, 147 – 169

Olsen, S.B., Lundhede, T.H., Jacobsen, J.B. and Thorsen, B.J. (2011). Tough and Easy Choices: Testing the Influence of Utility Difference on Stated Certainty-in-Choice in Choice Experiments. *Environmental and Resource Economics*, *49*, 491 – 510

Oppewal, H., Huybers, T. and Crouch, G.I. (2015). Tourist destination and experience choice: A choice experimental analysis of decision sequence effects. *Tourism Management*, *48*, 467 – 476

Oppewal, H., Morrison, M., Wang, P. and Waller, D. (2010). Preference Stability: Modeling how Consumer Preferences Shift after Receiving New Product Information. In: *Choice Modelling: The State-of-the-art and The State-of-practice*. Emerald.

Orquin, J.L., Bagger, M.P. and Loose, S.M. (2013). Learning affects top down and bottom up modulation of eye movements in decision making. *Judgment and Decision Making*, *8*(6), 700 – 716

Ovaskainen, V. and Kniivilä, M. (2005). Consumer versus citizen preferences in contingent valuation: evidence on the role of question framing. *Australian Journal of Agricultural and Resource Economics*, *49*(4), 379 – 394

Özdemir, S., Johnson, F.R. and Whittington, D. (2016). Ideology, public goods and welfare valuation: An experiment on allocating government budgets. *Journal of Choice Modelling*, *20*, 61 – 72

Özdemir, S., Ng, S., Huynh, V.A., Mühlbacher, A., Tan, H.K. and Finkelstein, E.A. (2023). Trade-Offs between Vaccine Effectiveness and Vaccine Safety: Personal versus Policy Decisions. *PharmacoEconomics – Open*, *7*, 915 – 926

Palma, D. and Hess, S. (2022). Extending the Multiple Discrete Continuous (MDC) modelling framework to consider complementarity, substitution, and an unobserved budget. *Transportation Research Part B: Methodological*, *161*, 13 – 35

Palma, D., Calastri, C. and Pawlak., J. (2023). The role of time budgets in models of multi-tasking while travelling: A comparison between the MDCEV and eMDC approach. *Transportation Research Part A: Policy and Practice*, *176*, 103796

Patterson, S.E. and Reyes, A.M. (2024). Changes in Americans' Views on Who Should Provide and Pay for Assistance to Older Adults with Activity Limitations, 2012 to 2022. *Socius*, *10*

Pearce, A.M., Mulhern, B.J., Watson, V. and Viney, R.C. (2020). How are debriefing questions used in health discrete choice experiments? An online survey. *Value in Health*, *23*(3), 289 – 293

Pechey, R., Burge, P., Mentzakis, E., Suhrcke, M. and Marteau, T.M. (2014). Public acceptability of population-level interventions to reduce alcohol consumption: A discrete choice experiment. *Social Science & Medicine*, *113*, 104 – 109

Pedersen, L.B., Kjær, T., Kragstrup, J. and Gyrd-Hansen, D. (2011). Does the inclusion of a cost attribute in forced and unforced choices matter?: Results from a web survey applying the discrete choice experiment. *Journal of Choice Modelling*, *4*(3), 88 – 109

Pellegrini, A., Rose, J. and Scarpa, R. (2022). Multiple Herbicide Use in Cropland: A Discrete- Continuous Model for Stated Choice Data. *Land Economics*, *98*(2), 355 – 375

Penn, J. and Hu, W. (2023). Adjusting and celebrating elicited values based on follow-up certainty questions: a meta-analysis. *Environmental and Resource Economics*, *84*, 919 – 946

Persson, E. and Tinghög, G. (2020). Opportunity cost neglect in public policy. *Journal of Economic Behavior & Organization*, *170*, 301 – 312

Petrolia, D.R., Interis, M.G. and Hwang, J. (2018). Single-Choice, Repeated-Choice, and Best-Worst Scaling Elicitation Formats: Do Results Differ and by How Much? *Environmental and Resource Economics*, *69*, 365 – 393

Pinjari, A.R. and Bhat, C. (2021). Computationally efficient forecasting procedures for Kuhn-Tucker consumer demand model systems: Application to residential energy consumption analysis. *Journal of Choice Modelling*, *39*, 100283

Pinjari, A.R. and Bhat, C.R. (2010). A multiple discrete-continuous nested extreme value (MDCNEV) model: Formulation and application to non-worker activity time-use and timing behavior on weekdays. *Transportation Research Part B: Methodological*, *44*, 562 – 583

Pinto-Prades, J.L., McHugh, N., Donaldson, C. and Manoukian, S. (2019). Sequence effects in time trade-off valuation of hypothetical health states. *Health Economics*, *28*(11), 1308 – 1319

Populytics. (n.d.). *Populytics – Home*. Available at: https://populytics.nl/en/

Powe, N.A. and Bateman, I.J. (2003). Ordering effects in nested 'top-down' and 'bottom-up' contingent valuation designs. *Ecological Economics*, *45*(2), 255 – 270

Quaife, M., Terris-Prestholt, F., Di Tanna, G.L. and Vickerman, P. (2018). How well do discrete choice experiments predict health choices? A systematic review and meta-analysis of external validity. *European Journal of Health Economics*, *19*, 1053 – 1066

R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria

Read, S., Erens, B., Wittenberg, R. et al. (2021). Public preferences for paying for social care in later life in England: A latent class analysis. *Social Science & Medicine*, *274*, 113803

Reckers-Droog, V., Jansen, M., Bijlmakers, L., Baltussen, R., Brouwer, W. and Van Exel, J. (2020). How does participating in a deliberative citizens panel on healthcare priority setting influence the views of participants? *Health Policy*, *124*(2), 143 – 151

REDECAN. (2023). *Estimaciones de la incidencia del cáncer en España, 2023*. Red Española de Registros de Cáncer

A

Remoundou, K., Adaman, F., Koundouri, P. and Nunes, P.A.L.D. (2014). Is the value of environmental goods sensitive to the public funding scheme? Evidence from a marine restoration programme in the Black Sea. *Empirical Economics*, *47*, 1173 – 1192

Ride, J., Goranitis, I., Meng, Y., LaBond, C. and Lancsar, E. (2024). A reporting checklist for discrete choice experiments in health: The DIRECT checklist. *PharmacoEconomics*, *42*, 1161 – 1175

Rijksoverheid. (2023). *Niets doen is geen optie: IBO Ouderenzorg*. The Hague, the Netherlands: Rijksoverheid

RIVM. (2022). *Raming zorguitgaven ouderenzorg 2021-2040*. Retrieved from: https://www.vzinfo.nl/documenten/raming-zorguitgaven-ouderenzorg-2021-2040-verslag

Rogers, A.A., Burton, M.P., Cleland, J.A., Rolfe, J.C., Meeuwig, J.J. and Pannell, D.J. (2020). Expert judgements and community values: preference heterogeneity for protecting river ecology in Western Australia. *Australian Journal of Agricultural and Resource Economics*, *64*, 266 – 293

Roky, A.H., Islam, M.M., Ahasan, A.M.F. et al. (2025). Overview of skin cancer types and prevalence rates across continents. *Cancer Pathogenesis and Therapy*, *3*, 89 – 100

Rolfe, J. and Brouwer, R. (2012). Design effects in a meta-analysis of river health choice experiments in Australia. *Journal of Choice Modelling*,

Rolfe, J. and Windle, J. (2005). Valuing options for reserve water in the Fitzroy Basin. *Australian Journal of Agricultural and Resource Economics*, *49*(1), 91 – 114

Rolfe, J., Brouwer, R. and Johnston, R.J. (2015). Meta-analysis: Rationale, issues and applications. In: R.J. Johnston, J. Rolfe, R.S. Rosenberger and R. Brouwer (eds.). *Benefit transfer of environmental and resource values: a guide for researchers and practitioners*. The Economics of Non-Market Goods and Resources series, vol. 14. Springer

Rollins, C. (2023). Investigating cost non-attendance as a driver of inflated welfare estimates in mixed-logit models. *Journal of Agricultural Economics*, *74*, 921 – 934

Rose, J.M. (2015). *Incorporating a budget constraint into discrete choice experiments*. Paper presented at the 4th International Choice Modelling Conference. Austin, TX

Rose, J.M., Beck, M.J. and Hensher, D.A. (2015). The joint estimation of respondent-reported certainty and acceptability with choice. *Transportation Research Part A: Policy and Practice*, *71*, 141 – 152

Rotteveel, A.H., Lambooij, M.S., Over, E.A.B. et al. (2022). If you were a policymaker, which treatment would you disinvest? A participatory value evaluation on public preferences for active disinvestment of healthcare interventions in the Netherlands. *Health Economics, Policy and Law*, *17*(4), 428 – 443

Rudolph, L., Freitag, M. and Thurner, P.W. (2024). Ordering effects versus cognitive burden: How should we structure attributes in conjoint experiments? *Public Opinion Quarterly*, *88*(3), 991 – 1016

Russell, C.S., Bjørner, T.B. and Clark, C.D. (2003). Searching for evidence of alternative preferences, public as opposed to private. *Journal of Economic Behavior & Organization*, *51*, 1 – 27

Ryan, M. and Bate, A. (2001). Testing the assumptions of rationality, continuity and symmetry when applying discrete choice experiments in health care. *Applied Economics Letters*, *8*(1), 59 – 63

Ryan, M., Krucien, N. and Hermens, F. (2018). The eyes have it: Using eye tracking to inform information processing strategies in multi-attributes choices. *Health Economics*, *27*(4), 709 – 721

Ryan, M., McIntosh, E. and Shackley, P. (1998). Methodological issues in the application of conjoint analysis in health care. *Health Economics*, *7*, 373 – 378

Ryan, M., Scott, D.A., Reeves, C. et al. (2001). Eliciting public preferences for healthcare: a systematic review of techniques. *Health Technology Assessment*, *5*(5)

Ryan, M., Watson, V. and Entwistle, V. (2009). Rationalising the 'irrational': a think aloud study of discrete choice experiment responses. *Health Economics*, *18*(3), 321 – 336

Sælensminde, K. (2001). Inconsistent choices in Stated Choice data. *Transportation*, *28*, 269 – 296

Salloum, R.G., Shenkman, E.A., Louviere, J.J. and Chambers, D.A. (2017). Application of discrete choice experiments to enhance stakeholder engagement as a strategy for advancing implementation: A systematic review. *Implementation Science*, *12*(1), 140

Sandorf, E.D., Campbell, D. and Hanley, N. (2017). Disentangling the influence of knowledge on attribute non-attendance. *Journal of Choice Modelling*, *24*, 36 – 50

Sandorf, E.D., Crastes dit Sourd, R. and Mahieu, P.-A. (2018). The effect of attribute-alternative matrix displays on preferences and processing strategies. *Journal of Choice Modelling*, *29*, 113 – 132

Sangers, T.E., Wakkee, M., Kramer-Noels, E.C., Nijsten, T. and Lugtenberg, M. (2021). Views on mobile health apps for skin cancer screening in the general population: an in-depth qualitative exploration of perceived barriers and facilitators. *British Journal of Dermatology*, *185*(5), 961 – 969

Santos-Eggimann, B. and Meylan, L. (2017). Older Citizens' Opinions on Long-Term Care Options: A Vignette Survey. *Journal of the American Medical Directors Association*, *18*(4), 326 – 334

Savage, S.J. and Waldman, D.M. (2008). Learning and fatigue during choice experiments: a comparison of online and mail survey modes. *Journal of Applied Econometrics*, *23*(3), 351 – 371

Sawtooth Software. (n.d.). *Lighthouse Studio v.9.14.2*. Sequim, WA: Sawtooth Software

Scarpa, R., Zanoli, R., Bruschi, V. and Naspetti, S. (2013). Inferred and stated attribute non-attendance in food choice experiments. *American Journal of Agricultural Economics*, *95*(1), 165 – 180

Scheufele, G. and Bennett, J. (2012). Response Strategies and Learning in Discrete Choice Experiments. *Environmental and Resource Economics*, *52*, 435 – 453

Schmitz, H. and Westphal, M. (2017). Informal care and long-term labor market outcomes. *Journal of Health Economics*, *56*, 1 – 18

Schreuder, K., De Groot, J., Hollestein, L. and Louwman, M. (2019). *Huidkanker in Nederland: Cijfers uit 30 jaar Nederlandse Kankerregistratie*. Integraal Kankercentrum Nederland

Schreuder, K., Hollestein, L., Nijsten, T.E.C., Wakkee, M. and Louwman, M.W.J. (2022). A nationwide study of the incidence and trends of first and multiple basal cell carcinomas in the Netherlands and prediction of future incidence. *British Journal of Dermatology*, *186*, 476 – 484

Scott, A. and Vick, S. (1999). Patients, Doctors and Contracts: An Application of Principal-Agent Theory to the Doctor-Patient Relationship. *Scottish Journal of Political Economy*, *46*(2), 111 – 134

Segovia, M.S. and Palma, M.A. (2021). Testing the consistency of preferences in discrete choice experiments: an eye tracking study. *European Review of Agricultural Economics*, *48*(3), 624 – 664

Selivanova, A. and Krabbe, P.F.M. (2018). Eye tracking to explore attendance in health-state descriptions. *PLoS ONE*, *13*(1), e0190111

Sendi, P., Al, M.J., Gafni, A. and Birch, S. (2003). Optimizing a portfolio of health care programs in the presence of uncertainty and constrained resources. *Social Science & Medicine*, *57*(11), 2207 – 2215

Sendín-Martin, M., Hernández-Rodríguez, J.C., Durán-Romero, A.J., Ortiz-Álvarez, J., Conejo-Mir, J. and Pereyra-Rodríguez, J.-J. (2021). Non-Melanoma Skin Cancer Mortality in Spain: A Predictive Model up to 2044. *Journal of Clinical Medicine*, *10*, 5750

Sever, I., Verbič, M. and Sever, E.K. (2019a). Cost attribute in health care DCEs: Just adding another attribute or a trigger of change in the stated preferences? *Journal of Choice Modelling*, *32*, 100135

Sever, I., Verbič, M. and Sever, E.K. (2019b). Estimating willingness-to-pay for health care: A discrete choice experiment accounting for non-attendance to the cost attribute. *Journal of Evaluation in Clinical Practice*, *25*(5), 843 – 849

Shih, S.T.F., Carter, R., Heward, S. and Sinclair, C. (2017). Economic evaluation of future skin cancer prevention in Australia. *Preventive Medicine*, *99*, 7 – 12

Shih, S.T.-F., Carter, R., Sinclair, R., Mihalopoulos, C. and Vos, T. (2009). Economic evaluation of skin cancer prevention in Australia. *Preventive Medicine*, *49*, 449 – 453

Shr, Y.-H., Ready, R., Orland, B. and Echols, S. (2019). How Do Visual Representations Influence Survey Responses? Evidence from a Choice Experiment on Landscape Attributes of Green Infrastructure. *Ecological Economics*, *156*, 375 – 386

Simmons, C., Ilinca, S., Kadi, S. and Comas-Herrera, A. (2024). Financing and Provision of Long-Term Care in Europe: Reflections on Intergenerational Solidarity in Care for Older People. *Social Policy & Administration* [article in press]

Simonsen, N.F., Kjær, T. and Gyrd-Hansen, D. (2021). Pure altruism and misjudgement: A bad combination? *Journal of Health Economics*, *80*, 102550

Sinclair, C. and Foley, P. (2009). Skin cancer prevention in Australia. *British Journal of Dermatology*, *161*(s3), 116 – 123

Sjöstrand, H. (2001). *The ordering effect in Stated Preference studies – a study of public transport passengers' valuation of standard*. Paper presented at the 9th World Conference on Transport Research. Available at: https://lup.lub.lu.se/search/files/6365306/744719.pdf

Skedgel, C. and Regier, D.A. (2015). Constant-Sum Paired Comparisons for Eliciting Stated Preferences: A Tutorial. *The Patient*, *8*, 155 – 163

Skedgel, C., Wailoo, A.J. and Akehurst, R.L. (2015). Choosing vs. allocating: discrete choice experiments and constant-sum paired comparisons for the elicitation of societal preferences. *Health Expectations*, *18*(5), 1227 – 1240

Smith, A.K. and Elias, L.J. (2013). Native Reading Direction and Corresponding Preferences for Left- or Right-Lit Images. *Perceptual and Motor Skills*, *116*(2), 355 – 367

Smith, R.D. and Sach, T.H. (2010). Contingent valuation: what needs to be done? *Health Economics, Policy and Law*, *5*(1), 91 – 111

Soekhai, V., De Bekker-Grob, E.W., Ellis, A.R. and Vass, C.M. (2019a). Discrete choice experiments in health economics: past, present and future. *PharmacoEconomics*, *37*(2), 201 – 226

Soekhai, V., Donkers, B., Johansson, J.V. et al. (2023). Comparing outcomes of a discrete choice experiment and case 2 best-worst scaling: an application to neuromuscular disease treatment. *The Patient*, *16*, 239 – 253

Soekhai, V., Whichello, C., Levitan, B. et al. (2019b). Methods for exploring and eliciting patient preferences in the medical product lifecycle: a literature review. *Drug Discovery Today*, *24*(7), 1324 – 1331

Soliño, M., Farizo, B.A. and Campos, P. (2017). Internal validity in discrete choice experiments: varying the position of the price attribute. *Applied Economics Letters*, *24*(13), 940 – 944

Solky, B.A., Phillips, P.K., Christenson, L.J., Weaver, A.L., Roenigk, R.K. and Otley, C.C. (2007). Patient preferences for facial sunscreens: A split-face, randomized, blinded trial. *Journal of the American Academy of Dermatology*, *57*(1), 67 – 72

Spalek, T.M. and Hammad, S. (2005). The Left-to-Right Bias in Inhibition of Return Is Due to the Direction of Reading. *Psychological Science*, *16*(1), 15 – 18

Spinks, J. and Mortimer, D. (2016). Lost in the crowd? Using eye-tracking to investigate the effect of complexity on attribute non-attendance in discrete choice experiments. *BMC Medical Informatics and Decision Making*, 16, 14

Srivastava, J., Connolly, T. and Beach, L.R. (1995). Do Ranks Suffice? A Comparison of Alternative Weighting Approaches in Value Elicitation. *Organizational Behavior and Human Decision Processes*, *63*(1), 112 – 116

Stanley, T.D. (2005). Beyond Publication Bias. *Journal of Economic Surveys*, *19*(3), 295 – 533

Statistics Netherlands. (2023a). *Mantelzorg en vrijwilligerswerk in de zorg*. Retrieved from; https://www.cbs.nl/nl-nl/longread/aanvullende-statistische-diensten/2023/mantelzorg-en- vrijwilligerswerk-in-de-zorg/3-resultaten

Statistics Netherlands. (2023b). *Overledenen; doodsoorzaak (uitgebreide lijst), leeftijd, geslacht*. Retrieved from: https://opendata.cbs.nl/statline/#/CBS/nl/dataset/7233/table?ts=1690468161897

Statistics Netherlands. (n.d.$_a$). *Bevolking; hoogst behaald onderwijsniveau en herkomst*. Retrieved on August 21, 2024 from: https://opendata.cbs.nl/statline/#/CBS/nl/dataset/85453NED/table

Statistics Netherlands. (n.d.$_b$). *Bevolking op de eerste van de maand; geslacht, leeftijd, herkomstland*. Retrieved on August 21, 2024 from: https://opendata.cbs.nl/statline/#/CBS/nl/dataset/85721NED/table?ts=1724249926688

Statistics Spain. (2023). *Death statistics according to cause of death. Results*. Available at: https://ine.es/dyngs/INEbase/en/operacion.htm?c=Estadistica_C&cid=1254736176780&menu=resultados&idp=1254735573175#!tabs-1254736194710

Stöckel, J. and Bom, J. (2022). Revisiting Longer-term Health Effects of Informal Caregiving: Evidence from the UK. *Journal of the Economics of Ageing*, *21*, 100343

Stoll, C.R.T., Izadi, S., Fowler, S., Green, P., Suls, J. and Colditz, G.A. (2019). The value of a second reviewer for study selection in systematic reviews. *Research Synthesis Methods*, *10*(4), 539 – 545

Swait, J. and Adamowicz, W. (2001). The Influence of Task Complexity on Consumer Choice: A Latent Class Model of Decision Strategy Switching. *Journal of Consumer Research*, *28*(1), 135 – 148

A

Swallow, S.K. and McGonagle, M.P. (2006). Public Funding of Environmental Amenities: Contingent Choices Using New Taxes or Existing Revenues for Coastal Land Conservation. *Land Economics*, *82*(1), 56 – 67

Swartz, K. (2013). Searching for a Balance of Responsibilities: OECD Countries' Changing Elderly Assistance Policies. *Annual Review of Public Health*, *34*, 397 – 412

Swartz, K., Miake, N. and Farag, N. (2012). Long-Term Care: Common Issues and Unknowns. *Journal of Policy Analysis and Management*, *31*(1), 139 – 152

Tanasache, O., Balcombe, K. and Ariane, K. (2023). *Uncovering decision-making processes in nutritional labelling use: An empirical investigation of mouse-tracking in a discrete choice experiment*. SSRN Working paper. Available at: https://dx.doi.org/10.2139/ssrn.4601691

Tejera-Vaquerizo, A., Descalzo-Gallego, M.A., Otero-Rivas, M.M. et al. (2016). Skin Cancer Incidence and Mortality in Spain: A Sytematic Review and Meta-Analysis. *Acta Dermo- Sifiliográficas*, *107*(4), 318 – 328

Tenand, M., Bakx, P. and Wouterse, B. (2023). *The impact of co-payments for nursing home care on care use and welfare*. EsCHER Working Paper, No. 2023009. Rotterdam, the Netherlands: Erasmus Centre for Health Economics Rotterdam

Teng, Y., Yu, Y., Li, S. et al. (2021). Ultraviolet Radiation and Basal Cell Carcinoma: An Environmental Perspective. *Frontiers in Public Health*, *9*, 666528

Tervonen, T., Gelhorn, H., Sri Bhashyam, S. et al. (2017). MCDA swing weighting and discrete choice experiments for elicitation of patient benefit-risk preferences: a critical assessment. *Pharmacoepidemiology & Drug Safety*, *26*(12), 1483 – 1491

Thau, M., Mikkelsen, M.F., Hjortskov, M. and Pedersen, M.J. (2021). Question order bias revisited: A split-ballot experiment on satisfaction with public services among experienced and professional users. *Public Administration*, *99*(1), 189 – 204

The PREFER consortium. (2022). PREFER Recommendations: Why, when and how to assess and use patient preferences in medical product decision-making. Zenodo. https://doi.org/10.5281/zenodo.6470922

Thokala, P., Devlin, N., Marsh, K. et al. (2016). Multiple Criteria Decision Analysis for Health Care Decision Making – An Introduction: Report 1 of the ISPOR MCDA Emerging Good Practices Task Force. *Value in Health*, *19*(1), 1 – 13

Thurstone, L.L. (1927). A law of comparative judgment. *Psychological Review*, *34*, 273 – 286

Tian, Y.J., Duong, V., Buhr, E., Felber, N.A., Schwab, D.R. and Wangmo, T. (2025). Monitored and Cared for at Home? Privacy Concerns When Using Smart Home Health Technologies to Care for Older Persons. *AJOB Empirical Bioethics*, *16*(2), 61 – 76

Tienhaara, A., Ahtiainen, H. and Pouta, E. (2015). Consumer and citizen roles and motives in the valuation of agricultural genetic resources in Finland. *Ecological Economics*, *114*, 1 – 10

Tocock, M., Borriello, A., Tinch, D., Rose, J.M. and Hatton MacDonald, D. (2023). How environmental beliefs influence the acceptance of reallocating government budgets to improving coastal water quality: a hybrid choice model. *Australasian Journal of Environmental Management*, *30*(3-4), 348 – 366

Train, K. and Sonnier, G. (2005). Mixed logit with bounded distributions of correlated partworths. In: R. Scarpa and A. Alberini (eds.). *Applications of Simulation Methods in Environmental and Resource Economics*. The Economics of Non-Market Goods and Resources series, vol. 6. Springer

Trujilo Jara, K., Hernandez, J.I., Mouter, N., Brouwer, W. and Van Exel, J. (2025). Preferences of citizens in Peru for school opening during a public-health crisis: A participatory value evaluation study. *Social Science & Medicine*, *365*, 117581

Tseng, L.P.D. and Lii, Y. (2006). The Role of Attribute Order and Number Effects in Consumers' Multiattribute Preferential Decisions. In: R.W. Belk (ed.). *Research in Consumer Behavior*. Vol. 10. Emerald

Tsuchiya, A., Bansback, N., Hole, A.R. and Mulhern, B. (2019). Manipulating the 5 Dimensions of the EuroQol Instrument: The Effects on Self-Reporting Actual Health and Valuing Hypothetical Health States. *Medical Decision Making*, *39*(4), 380 – 392

Tuhkanen, H., Piirsalu, E., Nõmmann, T. et al. (2016). Valuing the benefits of improved marine environmental quality under multiple stressors. *Science of the Total Environment*, *551-552*, 367 – 375

Uggeldahl, K., Jacobsen, C., Lundhede, T.H. and Olsen, S.B. (2016). Choice certainty in Discrete Choice Experiments: Will eye tracking provide useful measures? *Journal of Choice Modelling*, *20*, 35 – 48

Urban, K., Mehrmal, S., Uppal, P., Giesey, R.L. and Delost, G.R. (2020). The global burden of skin cancer: A longitudinal analysis from the Global Burden of Disease Study, 1990 – 2017. *JAAD International*, *2*, 98 – 108

Van Beek, L., Mouter, N., Pelzer, P., Hajer, M. and Van Vuuren D. (2022). *The role of experts and expertise in practices of citizen engagement in climate policy: a comparative analysis of two contrasting cases*. Research Square Working Paper. https://doi.org/10.21203/rs.3.rs- 2226349/v1

Van de Walle, S. and Van Ryzin, G.G. (2011). The order of questions in a survey on citizen satisfaction with public services: lessons from a split-ballot experiment. *Public Administration*, *89*(4), 1436 – 1450

Van der Waerden, P., Borgers, A., Timmermans, H. and Bérénos, M. (2006). Order Effects in Stated-Choice Experiments: Study of Transport Mode Choice Decisions. *Transportation Research Record*, *1985*(1), 12 – 18

Van Niekerk, C.C., Groenewoud, H.M.M. and Verbeek, A.L.M. (2021). Trends and projections in cutaneous melanoma death in the Netherlands from 1950 to 2045. *Medicine*, *100*(48), e27784

Van Ooijen, R., De Bresser, J. and Knoef, M. (2017). *Meer eigen verantwoordelijkheid in ouderenzorg: wensen en mogelijkheden*. Netspar Brief, No. 9. Tilburg, the Netherlands: Netspar

Van Til, J.A. and IJzerman, M.J. (2014). Why Should Regulators Consider Using Patient Preferences in Benefit-risk Assessment? *PharmacoEconomics*, *32*, 1 – 4

Vanermen, I., Kessels, R., Verheyen, K., Muys, B. and Vranken, L. (2021). The effect of information transfer related to soil biodiversity on Flemish citizens' preferences for forest management. *Science of the Total Environment*, *776*, 145791

Vass, C.M. and Boeri, M. (2021). Mobilising the next generation of stated-preference studies: the association of access device with choice behaviour and data quality. *The Patient*, *14*, 55 – 63

Veldwijk, J., Determann, D., Lambooij, M.S. et al. (2016a). Exploring how individuals complete the choice tasks in a discrete choice experiment: an interview study. *BMC Medical Research Methodology*, *16*, 45

A

Veldwijk, J., DiSantostefano, R.L., Janssen, E. et al. (2023). Maximum Acceptable Risk Estimation Based on a Discrete Choice Experiment and a Probabilistic Threshold Technique. *The Patient*, *16*, 641 – 653

Veldwijk, J., Essers, B.A.B., Lambooij, M.S., Dirksen, C.D., Smit, H.A. and De Wit, G.A. (2016b). Survival or Mortality: Does Risk Attribute Framing Influence Decision-Making Behavior in a Discrete Choice Experiment? *Value in Health*, *19*(2), 202 – 209

Veldwijk, J., Groothuis-Oudshoorn, C.G.M., Kihlbom, U. et al. (2019). How psychological distance of a study sample in discrete choice experiments affects preference measurement: a colorectal cancer screening case study. *Patient Preference and Adherence*, *13*, 273 – 282

Veldwijk, J., Lambooij, M.S., Van Til, J.A., Groothuis-Oudshoorn, C.G.M., Smit, H.A. and De Wit, G.A. (2015). Words or graphics to present a Discrete Choice Experiment: Does it matter? *Patient Education and Counseling*, *98*(11), 1376 - 1384

Veldwijk, J., Marceta, S.M., Swait, J.D., Lipman, S.A. and De Bekker-Grob, E.W. (2023). Taking the shortcut: Simplifying heuristics in discrete choice experiments. *The Patient*, *16*, 301 – 315

Veldwijk, J., Ozdemir, S., Bui, M. et al. (2025). Transferability of preferences; for better or …? *The Patient*, *18*, 97 – 100

Veldwijk, J., Smith, I.P., Oliveri, S. et al. (2024). Comparing discrete choice experiment with swing weighting to estimate attribute relative importance: a case study in lung cancer patient preferences. *Medical Decision Making*, *44*(2), 203 – 216

Vermunt, J.K. (2010). Latent Class Modeling with Covariates: Two Improved Three-Step Approaches. *Political Analysis*, *18*(4), 450 – 469

Vermunt, J.K. and Magidson, J. (2002). Latent Class Cluster Analysis. In: J.A. Hagenaars, A.L. McCutcheon. (eds.). *Applied Latent Class Analysis*. Cambridge, UK: Cambridge University Press

Vermunt, J.K. and Magidson, J. (2016). *Technical Guide for Latent GOLD 5.1: Basic, Advanced, and Syntax*. Belmont, MA: Statistical Innovations Inc.

Versteegh, M., Knies, S. and Brouwer, W. (2016). From Good to Better: New Dutch Guidelines for Economic Evaluations in Healthcare. *PharmacoEconomics*, *34*, 1071 – 1074

Vondolia, G.K. and Navrud, S. (2019). Are non-monetary payment modes more uncertain for stated preference elicitation in developing countries? *Journal of Choice Modelling*, *30*, 73 – 87

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

Vossler, C.A., Doyon, M. and Rondeau, D. (2012). Truth in Consequentiality: Theory and Field Evidence on Discrete Choice Experiments. *American Economic Journal: Microeconomics*, *4*(4), 145 – 171

Warburton, J. and Smith, J. (2003). Out of the Generosity of Your Heart: Are We Creating Active Citizens through Compulsory Volunteer Programmes for Young People in Australia? *Social Policy & Administration*, *37*(7), 772 – 786

Watson, T.P.G., Tong, M., Bailie, J., Ekanayake, K. and Bailie, R.S. (2024). Relationship between climate change and skin cancer and implications for prevention and management: a scoping review. *Public Health*, *227*, 243 – 249

Weng, W., Morrison, M.D., Boyle, K.J., Boxall, P.C. and Rose, J. (2021). Effects of the number of alternatives in public good discrete choice experiments. *Ecological Economics*, *182*, 106904

Whichello, C., Bywall, K.S., Mauer, J. et al. (2020a). An overview of critical decision-points in the medical product lifecycle: Where to include patient preference information in the decision- making process? *Health Policy*, *124*(12), 1325 – 1332

Whichello, C., Levitan, B., Juhaeri, J. et al. (2020b). Appraising patient preference methods for decision-making in the medical product lifecycle: an empirical comparison. *BMC Medical Informatics and Decision-Making*, *20*, 114

Whichello, C., Smith, I., Veldwijk, J., De Wit, G.A., Rutten-Van Mölken, M.P.M.H. and De Bekker- Grob, E.W. (2023). Discrete choice experiment versus swing-weighting: A head-to-head comparison of diabetic patient preferences for glucose-monitoring devices. *PLoS ONE*, *18*(7), e0283926

Whitty, J.A. and Gonçalves, A.S.O. (2018). A systematic review comparing the acceptability, validity and concordance of discrete choice experiments and best-worst scaling for eliciting preferences in healthcare. *The Patient*, *11*, 301 – 317

Whitty, J.A., Lancsar, E., Rixon, K., Golenko, X. and Ratcliffe, J. (2014a). A Systematic Review of Stated Preference Studies Reporting Public Preferences for Healthcare Priority Setting. *The Patient*, *7*, 365 – 386

Whitty, J.A., Walker, R., Golenko, X. and Ratcliffe, J. (2014b). A think aloud study comparing the validity and acceptability of discrete choice and best worst scaling methods. *PLoS ONE*, *9*(4), e90635

WHO. (2024). *Ageing and health*. Retrieved on January 7, 2025 from: https://www.who.int/news- room/fact-sheets/detail/ageing-and- health#:~:text=By%202050%2C%2080%25%20of%20all,was%20older%20than%2060%20y ears

Wittenberg, R., Read, S., Erens, B. et al. (2024). Who should pay for social care for older people in England? Results from surveys of public attitudes to the funding of adult social care. *Journal of Social Policy*, *53*(1), 1 – 25

Xie, F., Pullenayegum, E., Gaebel, K., Oppe, M. and Krabbe, P.F.M. (2014). Eliciting preferences to the EQ-5D-5L health states: discrete choice experiment or multiprofile case of best-worst scaling? *European Journal of Health Economics*, *15*, 281 – 288

Yang, K. (2016). Creating public value and institutional innovations across boundaries: an integrative process of participation, legitimation, and implementation. *Public Administration Review*, *76*(6), 873 – 885

Yew, G.C.K. (2021). Trust in and Ethical Design of Carebots: The Case for Ethics of Care. *International Journal of Social Robotics*, *13*, 629 – 645

Yu, Z., Zheng, M., Fan, H., Liang, X. and Tang, Y. (2024). Ultraviolet (UV) radiation: a double-edged sword in cancer development and therapy. *Molecular Biomedicine*, *5*, 49

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

Zawojska, E., Welling, M. and Sagebiel, J. (2024). Effects of Varying the Location of Perceived Consequentiality Elicitation in a Discrete Choice Experiment Survey. *Land Economics*, *100*(2), 223 – 244

A

Zhang, J. and Adamowicz, W.L. (2011). Unraveling the Choice Format Effect: A Context-Dependent Random Utility Model. *Land Economics*, *87*(4), 730 – 743

Zhang, Y., Ho, T.Q.A., Ferris-Prestholt, F. et al. (2025). Prediction accuracy of discrete choice experiments in health-related research: a systematic review and meta-analysis. *eClinicalMedicine*, *79*, 102965

Zhao, Y., Van den Berg, P.E.W., Ossokina, I.V. and Arentze, T.A. (2022). Comparing self-navigation and video mode in a choice experiment to measure public space preferences. *Computers, Environment and Urban Systems*, *95*, 101828

Summary | Samenvatting

Summary

Scarcity of resources prompts us to consider the opportunity costs of our allocation decisions, both individually and societally. Therefore, to maximize welfare, it is important to obtain insights into the relative value of alternative allocations. In the context of public preferences for health policies, which is the focus of this dissertation, resource allocations are typically valued using stated preferences. An established preference-elicitation method is the Discrete Choice Experiment (DCE), which comes with a strong theoretical foundation and is widely applied in the health domain. DCEs present respondents with a sequence of choice tasks, each composed of two or more alternatives described by a number of attributes, of which the levels are varied between alternatives and choice tasks. In each choice task, respondents are asked to select the alternative they prefer, which resembles the real-life choice environment for many types of individual healthcare choices. In the context of policy decisions, however, policymakers may have to select multiple policy measures at the same time as well as decide on how much to invest in each of them.

In recent years, several preference-elicitation methods have been developed that allow respondents to choose combinations of alternatives. Participatory Value Evaluation (PVE) is one of these methods. In a single choice task, respondents are presented with a set of policy alternatives addressing a particular policy problem, and they are asked to compose the portfolio of policy alternatives they would prefer for addressing this problem, considering a resource constraint. Like in a DCE, each of the policy alternatives is described by a number of attributes, of which the levels are varied, but in a PVE only between respondents. PVE has been introduced in transportation and environmental economics but may also be valuable for health policy questions.

The aim of this dissertation is to advance the literature on the elicitation of public preferences for health policies and, to this end, contains three more specific objectives. The first objective is to position PVE relative to other more commonly used multi-attribute preference-elicitation methods used in the health domain, like DCEs. **Chapter 2** introduces PVE in the health domain and, by comparing it conceptually to four established methods, aims to contribute to a well-informed selection of method for preference-elicitation by researchers and policymakers. Advantages of PVE include that the method better resembles the reality of policymakers, allows respondents to evaluate synergies between alternatives, and enables researchers to elicit respondents' preferences for policy alternatives as well as how much to invest in the policy problem simultaneously. Disadvantages of PVE include that it may result in a higher cognitive

A

burden on respondents and is less efficient than other methods, requiring a larger sample of respondents. The Chapter ends with recommendations for further development of PVE, in particular regarding the feasibility and validity of the method.

The second objective of this dissertation is to examine the influence of design characteristics of a DCE or PVE on the preferences elicited with the choice experiment, which is addressed in Chapters 3 and 6. **Chapter 3** discusses the findings of a review of the literature across different domains on the impact of the presentation order of alternatives, attributes and choice sets on respondents' choices in a DCE. It shows that the majority of the 85 included studies found statistically significant ordering effects, and discusses the main mechanisms for these effects, methods to mitigate or reduce these effects in future studies, and several recommendations for further research. **Chapter 6** examines whether expenditure preferences and consequentiality perceptions of respondents in a PVE are sensitive to the payment vehicle used in the experiment and to the priming of opportunity costs. Consequentiality is the extent to which a choice-experiment affects real-life outcomes that are important for the respondent, such as policy decisions. It shows that expenditure distributions and consequentiality perceptions were very similar across three versions of the survey. These results warrant further investigation of respondents' perceptions of the substitution mechanisms implied by the different payment vehicles when eliciting preferences for public policies.

The third objective of this dissertation is to explore public preferences for health policy alternatives from a citizen perspective using DCE and PVE, which is addressed in Chapters 4 and 5. **Chapter 4** uses a DCE to elicit public preferences for collective skin cancer prevention policies in Austria, the Netherlands, and Spain. It shows that the tax attribute was the most important and most disliked attribute in each country. Information campaigns and a reduction in the price of sunscreen were the most preferred policy measures, and the prohibition of solar bed sales and of solaria the least preferred. Overall, the preference structure was very similar across countries. Most respondents would recommend their government to take policy action regarding skin cancer prevention. **Chapter 5** uses PVE to elicit public preferences for policy action regarding long-term care (LTC) for older people in the Netherlands in 2040. Respondents derived positive utility from all seven policy alternatives, and from increases in the fulfilment of nursing care needs and decreases in need for informal caregiving. Differences in preferences between respondents were found, particularly for increasing the use of supportive care technologies and introducing compulsory social service for young adults. Overall, the results suggest a preference for a substantial increase in LTC expenditure and distributing the additional resources over a variety of policy alternatives, rather than investing heavily in one or two particular policies.

Finally, **Chapter 7** summarizes the main findings of Chapters 2 to 6 and discusses some strengths and limitations as well as several implications for future research and policy. In particular, the Chapter considers the validity of eliciting preferences for public policy from a citizen perspective and our current understanding of respondents' choice processes and choice behaviours in PVE choice tasks. It puts forward several manners for researchers to improve this understanding and incorporate the newly acquired insights in their design and analysis of PVEs. This final Chapter also includes some methodological reflections highlighting that many questions remain regarding preference-elicitation for health policies. It is, therefore, important to continue working on methodological improvements that further raise the feasibility, internal and external validity and reliability of preference-elicitation methods, resulting in preferences that more accurately inform policymakers about the societal welfare implications of their decisions.

Samenvatting

Schaarste aan middelen dwingt ons om na te denken over de opportuniteitskosten van onze bestedingen, op zowel individueel als maatschappelijk niveau. Om de maatschappelijke welvaart te maximaliseren, is het belangrijk inzicht te krijgen in de relatieve waarde van alternatieve bestedingen. In de context van publieke voorkeuren voor gezondheidsbeleid, het onderwerp van dit proefschrift, worden bestedingen meestal gewaardeerd via het meten van voorkeuren. Een gevestigde methode hiervoor is het discrete keuze-experiment (Discrete Choice Experiment, DCE), dat een sterke theoretische basis heeft en veel wordt toegepast binnen de gezondheidszorg. DCE's bestaan uit een reeks keuzetaken, elk bestaande uit twee of meer alternatieven die worden gekenmerkt door verschillende attributen, waarvan de niveaus worden gevarieerd tussen alternatieven en keuzetaken. In elke taak kiezen respondenten het alternatief dat hun voorkeur heeft, wat de aard van veel daadwerkelijke individuele keuzes rondom gezondheid en zorg goed benadert. Voor beleidsbeslissingen geldt echter vaak dat beleidsmakers meerdere maatregelen tegelijkertijd kiezen en op hetzelfde moment moeten besluiten hoeveel in elke maatregel geïnvesteerd wordt.

In de afgelopen jaren zijn meerdere methoden ontwikkeld voor het meten van voorkeuren die het mogelijk maken om combinaties van alternatieven te kiezen. Participatieve Waarde Evaluatie (Participatory Value Evaluation, PVE) is één van deze methoden. In één enkele keuzetaak krijgen respondenten een reeks beleidsalternatieven gepresenteerd die een specifiek beleidsprobleem adresseren. De respondenten worden gevraagd om een portfolio van hun voorkeur samen te stellen binnen beperkte middelen. Net als bij DCE's worden de beleidsalternatieven gekenmerkt door attributen waarvan de niveaus worden gevarieerd, maar bij een PVE alleen tussen respondenten. PVE is geïntroduceerd in transport- en milieueconomie, maar kan ook waardevol zijn voor beleidsvraagstukken rondom gezondheid en zorg.

Het doel van dit proefschrift is om bij te dragen aan de literatuur omtrent het meten van publieke voorkeuren voor gezondheidsbeleid en omvat hiertoe drie specifiekere doelstellingen. De eerste doelstelling is het positioneren van PVE ten opzichte van andere veelgebruikte methoden, zoals DCE's. **Hoofdstuk 2** introduceert PVE in de gezondheidszorg en vergelijkt PVE conceptueel met vier gevestigde methoden. Voordelen van PVE zijn onder andere dat het de realiteit van beleidsmakers beter kan benaderen, respondenten in staat stelt om synergiën tussen beleidsopties mee te nemen in hun voorkeuren, en het voorkeuren kan meten voor zowel beleidsalternatieven als hoeveel geld besteed wordt aan het beleidsprobleem. Nadelen zijn onder andere de mogelijke grotere cognitieve belasting voor respondenten en de lagere efficiëntie ten opzichte van andere methoden, waardoor een grotere steekproef van respondenten vereist is. Het hoofdstuk sluit af met aanbevelingen voor verdere ontwikkeling van PVE, met name op het gebied van de uitvoerbaarheid en validiteit van de methode.

De tweede doelstelling van dit proefschrift is het onderzoeken van de invloed van ontwerpkenmerken van een DCE of PVE op de voorkeuren gemeten met het keuze-experiment, en wordt behandeld in hoofdstukken 3 en 6. **Hoofdstuk 3** bespreekt de bevindingen van een literatuurstudie in verschillende domeinen naar het effect van de volgorde van presenteren van alternatieven, attributen en keuzesets op de keuzes van respondenten in DCE's. Het laat zien dat de meeste van de 85 studies significante volgorde-effecten vinden en bespreekt de belangrijkste mechanismen voor dit effect, methoden om deze effecten te voorkomen of beperken, en diverse aanbevelingen voor vervolgonderzoek. **Hoofdstuk 6** onderzoekt of bestedingsvoorkeuren en percepties van consequentialiteit binnen een PVE beïnvloed worden door het soort betalingsmechanisme dat gebruikt wordt en door het benadrukken van opportuniteitskosten. Consequentialiteit is de mate waarin een keuze-experiment effect heeft op daadwerkelijke uitkomsten die voor de respondent belangrijk zijn, zoals beleidskeuzes. De verdelingen van uitgaven en de percepties van consequentialiteit waren zeer vergelijkbaar tussen drie versies van de vragenlijst. Deze resultaten vragen om vervolgonderzoek naar de percepties van respondenten omtrent de substitutiemechanismen bij verschillende betalingsmechanismen bij het meten van voorkeuren voor beleidsmaatregelen.

A

De derde doelstelling van dit proefschrift is het verkennen van publieke voorkeuren voor gezondheids- en zorgbeleid vanuit een burgerperspectief door middel van DCE en PVE, wat wordt behandeld in hoofdstukken 4 en 5. **Hoofdstuk 4** gebruikt een DCE om voorkeuren te meten voor collectieve maatregelen ter preventie van huidkanker in Oostenrijk, Nederland en Spanje. Het belastingattribuut was het belangrijkste en minst geliefde attribuut in elk land. Informatiecampagnes en een reductie in de prijs van zonnebrand waren de meest geliefde beleidsmaatregelen, terwijl verboden op zonnebanken en zonnestudio's het minst geliefd waren. De voorkeuren waren in het algemeen vergelijkbaar tussen landen. De meeste respondenten zouden hun overheid aanraden om actie te ondernemen tegen huidkanker. **Hoofdstuk 5** gebruikt PVE om voorkeuren te meten voor ouderenzorgbeleid in Nederland in 2040. Respondenten ontleenden positief nut uit alle zeven beleidsalternatieven, een toename van de vervulling van de behoefte aan verpleegzorg en een afname van de behoefte aan mantelzorg. Ook vond deze studie verschillen in voorkeuren tussen respondenten, met name wat betreft de verhoogde inzet van ondersteunende zorgtechnologieën en de introductie van een maatschappelijke dienstplicht voor jongeren. Over het geheel genomen suggereren de resultaten een voorkeur voor substantiële verhoging van de uitgaven aan ouderenzorg en het verdelen van de extra middelen over meerdere beleidsmaatregelen, in plaats van veel te investeren in één of twee specifieke maatregelen.

Tot slot vat **Hoofdstuk 7** de belangrijkste bevindingen van Hoofdstukken 2 tot en met 6 samen, bespreekt het de sterke en zwakke punten en ook de implicaties voor toekomstig onderzoek en beleid. Het hoofdstuk gaat met name in op de validiteit van het meten van voorkeuren voor beleidsmaatregelen vanuit een burgerperspectief en ons huidige begrip van de keuzeprocessen en het keuzegedrag van respondenten in PVE-keuzetaken. Het brengt verschillende manieren naar voren waarop onderzoekers dit begrip kunnen verbeteren en de nieuw verworven inzichten kunnen meenemen in het ontwerpen en analyseren van PVE's. Dit laatste hoofdstuk bevat ook enkele methodologische reflecties die benadrukken dat veel vragen blijven bestaan over het meten van voorkeuren voor gezondheids- en zorgbeleid. Het is dan ook belangrijk om te blijven werken aan methodologische verbeteringen die de uitvoerbaarheid, interne en externe validiteit en betrouwbaarheid van het meten van voorkeuren verbeteren, resulterende in voorkeuren die beleidsmakers accurater informeren over de implicaties van hun besluiten voor de maatschappelijke welvaart.

Portfolio

Research

Peer-reviewed articles in scientific journals

- Boxebeld, S., Mouter, N. and Van Exel, J. (2025). Trade-offs in long-term care for older people in an ageing society: a constrained portfolio choice experiment. *Journal of the Economics of Ageing* [article in press]
- Boxebeld, S., Mouter, N. and Van Exel, J. (2025). Public preferences for skin cancer prevention policies: a discrete choice experiment in three European countries. *Social Science & Medicine*, *378*, 118155
- Boxebeld, S. (2024). Ordering effects in discrete choice experiments: a systematic literature review across domains. *Journal of Choice Modelling*, *51*, 100489
- Boxebeld, S., Geijsen, T., Tuit, C., Van Exel, J., Makady, A., Maes, L., Van Agthoven, M. and Mouter, N. (2024). Public preferences for the allocation of societal resources over different healthcare purposes. *Social Science & Medicine*, *341*, 116536
- Boxebeld, S., Mouter, N. and Van Exel, J. (2024). Participatory Value Evaluation (PVE): a new preference elicitation method for decision-making in healthcare. *Applied Health Economics and Health Policy*, *22*, 145 – 154
- Mouter, N., Boxebeld, S., Kessels, R., Van Wijhe, M., De Wit, A., Lambooij, M. and Van Exel, J. (2022). Public Preferences for Policies to Promote COVID-19 Vaccination Uptake: A Discrete Choice Experiment in The Netherlands. *Value in Health*, *25*(8), 1290 – 1297

A

Under review

- Boxebeld, S., Wouterse, B. and Mierau, J. (2025). Quantifying the broader welfare benefits of prevention: an illustration for overweight and obesity. *Submitted*
- Boxebeld, S. (2025). The use of the mixed logit model to analyze discrete choice experiment data in health economics: towards an improved model specification, estimation, and reporting. *Submitted*

Articles in professional journal

- Wouterse, B., Boxebeld, S. and Mierau, J. (2024). Preventie is goed voor de gezondheid en voor de portemonnee. *ESB*, *109*(4838), 440 – 442
- Mouter, N., Boxebeld, S. and Van Exel, J. (2020). Analyses vormen startpunt van inhoudelijke discussie over coronabeleid. *ESB*, *105*(4791), 516 – 518

Policy report

- Mouter, N., Boxebeld, S., Kessels, R., Van Wijhe, M., De Wit, A., Lambooij, M. and Van Exel, J. (2020). *Groot draagvlak onder Nederlanders voor een 'vaccinatiebewijs light'*. Delft, the Netherlands: Delft University of Technology

Teaching

- Advanced Research Methods: course coordination, lecturing, workgroup teaching, exam development & grading
- Economics of Well-being: lecturing, course redevelopment, course co-coordination, exam development & grading
- Personal and Professional Development (in Dutch: Persoonlijke en Professionele Ontwikkeling): course coordination, grading
- Master Thesis supervision
- Quantitative Research Project (in Dutch: Kwantitatief Leeronderzoek): workgroup teacher/project tutor
- Bachelor graduation project (in Dutch: Afstudeerproject): supervision and coaching
- Choices and Dilemmas in Healthcare: workgroup teaching, exam grading, essay supervision

Funding acquisition

- Erasmus Centre for Health Economics Rotterdam (EsCHER), 2025, workshop methodological advances in choice experiments and choice models (€1,500)
- Royal Netherlands Academy of Arts and Sciences (KNAW) Van der Gaag Grant, 2023, research visit University of Oxford (€2,100)
- Erasmus Trustfund, 2023, research visit University of Oxford (€1,000)
- Erasmus Trustfund, 2023, research project long-term care policy preferences (€10,000)
- Erasmus Trustfund, 2023, research project skin cancer prevention policy preferences (€10,000)
- COST network, 2021, research visit Royal Holloway (€1,250)

Research visits

- February/March 2024: University of Oxford, Health Economics Research Centre (HERC)
- August/September 2021: Royal Holloway University of London, Department of Economics

Training during PhD

Research-related

- Structural Equation Modelling, Erasmus Graduate School of Social Sciences and the Humanities (2025)
- Advanced Choice Modelling, University of Leeds (2023)
- Choice Modelling and Stated Choice Survey Design, University of Leeds (2022)
- MATLAB Data Skills and Tools for the Social Sciences and Humanities, Erasmus Graduate School of Social Sciences and Humanities (2022)
- Measurement of Patient Preferences Using Discrete Choice Experiments, Erasmus School of Health Policy and Management (2022)
- How to Get Your Article Published, Erasmus Graduate School of Social Sciences and the Humanities (2021)
- Experimental Economics, Erasmus School of Economics (2021)

A

Teaching-related

- University Teaching Qualification (completed, diploma to be received), Risbo (2025)
- Partial University Teaching Qualification (delivery component), Risbo (2023)
- Lesson observation, Risbo (2023)
- Microlab How to supervise students, Risbo (2023)
- Basic didactics, Risbo (2023)
- Coaching and intervision skills, Boertien Vergouwen Overduin (2021)

External seminars

- University of Warsaw, Faculty of Economic Sciences – 5 December 2024 (Warsaw, PL)
- University of Sheffield, Sheffield Centre for Health and Related Research (SCHARR), Health Economics and Decision Science (Heds) group – 4 December 2024 (Sheffield, UK)
- Université Paris Cité, Interdisciplinary Laboratory for Applied Research in Economics/Management and Health (LIRAES) – 27 November 2024 (Paris, FR)
- University of Bristol, Health Economics and Health Policy team (HEHP) – 2 October 2024 (Bristol, UK)
- University of Aberdeen, Health Economics Research Unit (HERU) – 20 March 2024 (Aberdeen, UK)
- University of Birmingham, Health Economics Unit (HEU) – 13 March 2024 (Birmingham, UK)
- University of Oxford, Health Economics Research Centre (HERC) – 27 February 2024 (Oxford, UK)
- University of Oxford, Nuffield Department of Primary Care Health Sciences, Health Economics & Policy Evaluation group – 16 February 2024 (Oxford, UK)

Conference/workshop presentations and discussions

p=presenting author, d=discussant

- International Academy of Health Preference Research (IAHPR) Annual Meeting 2025 (Enschede, NL)(p)
- United Kingdom Health Economics Study Group (HESG) Summer 2025 (Brighton, UK) (p)
- Workshop on Non-Market Valuation (WONV) 2025 (Nancy, FR)(p)
- Lowlands Health Economics Study Group (lolaHESG) 2025 (Ermelo, NL)(d)
- European Health Policy Group (EHPG) meeting at LSE 2024 (London, UK)(p)
- European Health Economics Association (EuHEA) bi-annual conference 2024 (Vienna, AT)(p)
- Lowlands Health Economics Study Group (lolaHESG) 2024 (Leusden, NL)(p)
- Smarter Choices for Better Health Conference 2023 (Rotterdam, NL)(p)

- European Health Economics Association (EuHEA) PhD-Supervisor Conference 2023 (Bologna, IT) (p x 2)
- Lowlands Health Economics Study Group (lolaHESG) 2023 (Egmond aan Zee, NL) (p, d x 2)
- International Academy of Health Preference Research (IAHPR) Annual Meeting 2022 (Berlin, DE) (p)
- European Health Economics Association (EuHEA) bi-annual conference 2022 (Oslo, NO) (p)
- Lowlands Health Economics Study Group (lolaHESG) 2022 (Maastricht, NL) (p)

Professional service

Organization & Committee membership

- Organizer Multidisciplinary Junior Senior (Multi-JuSe) seminar series, Erasmus School of Health Policy & Management
- Member organizing committee, European Health Economics Association (EuHEA) bi-annual conference 2026
- Member scientific committee, Lowlands Health Economics Study Group (lolaHESG) 2023
- Ad hoc member assessment committee MHBA, Erasmus School of Health Policy & Management
- Member activity committee, Erasmus School of Health Policy & Management

Reviewing

- Articles for scientific journals:
• Social Science & Medicine (4x)
• Value in Health (3x)
• Applied Stochastic Models in Business and Industry (1x)
• International Journal of Health Policy and Management (1x)
• Journal of Choice Modelling (1x)
• SSM – Health Systems (1x)
- Conference abstracts:
• Lowlands Health Economics Study Group (lolaHESG) 2023
• International Academy of Health Preference Research (IAHPR) 2024, 2025

A

About the author

Acknowledgements

About the author

Sander Boxebeld grew up in Heino, the Netherlands. He completed programs in European Public Administration (BSc) and European Studies (MSc) at the University of Twente and Economics (MSc) at the VU University Amsterdam.

In November 2020, he started his PhD within the Health Economics department of Erasmus School of Health Policy & Management (ESHPM), Erasmus University Rotterdam. His doctoral research focused on the use and further advancement of discrete and portfolio choice experiments to elicit citizens' preferences for public policies in the health domain. Next to his doctoral research, he was also involved in studies on public preferences for COVID-19 vaccination policies, healthcare resource allocation, and the broader welfare benefits of prevention. This cumulated in multiple presentations at international conferences and various peer-reviewed publications in scientific journals. During his PhD, he undertook a research visit to the University of Oxford (UK), for which he was awarded the Van der Gaag grant by the Royal Netherlands Academy of Arts and Sciences (KNAW).

During his PhD, he was involved in various teaching activities, amongst others as a course coordinator of Advanced Research Methods and of Personal and Professional Development, lecturer within the minor program Economics of Well-being, and supervisor of bachelor and master theses. Also, he was and is part of multiple committees and boards. Within the academic environment, this included the activity committee of ESHPM, the scientific committee of the lowlands Health Economics Study Group (lolaHESG) 2023, and the organizing committee of the European Health Economics Association (EuHEA) conference 2026. Outside academia, this included various committees of student athletics association D.A.V. Kronos, the board and multiple committees of Dutch student athletics federation N.S.A.F. ZeuS, and the board of a foundation that financially supports students in the Philippines.

After completing his PhD, he will continue his research and teaching activities within the same department as an assistant professor.

Acknowledgements

September 2025, Rotterdam

About five years ago, I applied for the PhD position leading me to this dissertation. While I was happily surprised to be selected after two rounds of assignments and interviews, I did not accept the offer immediately. Instead, like I make most choices in life, I postponed the decision, doubted a bit more, changed my mind a few times, and then ultimately accepted the offer. Even though I had to get into the topic at first, I've never regretted my decision and truly enjoyed the past five years.

The people that contributed most directly are, of course, my supervisors, Job and Niek. Thank you both for giving me all the freedom to pursue whatever I found interesting, staying patient when I wanted to redesign the setup of the choice tasks or respecify the choice models for the 100th time, and trusting me and supporting me throughout the entire trajectory. Niek, from the two of you, I was fewer in your surroundings. Yet, even from a distance, your enthusiasm is truly contagious. I admire your ability to engage with policymakers, experts in your own and other fields, journalists, and the general public alike, all seemingly without much effort.

Job, I still don't have a clue how you manage to combine all your tasks with being so accessible and flexibly available. Not only do you care about the people you are working with, but also about everyone in the department and the School. I appreciate your dedication to creating and maintaining an environment where everyone can flourish. I can only hope to, one day, become a supervisor as good as you are. Thank you for providing me with the opportunity to stay in the department and further build on my academic career, it's a chance I didn't dare to wish for.

Next, I'd like to thank the people I collaborated with, be it on papers that did not end up in my PhD: Mattijs, Ardine, Roselinde and Maarten for the DCE on COVID-19 vaccination policies. Tom, Charlotte, Amr, Laurence and Michel for the PVE on resource allocation over different healthcare sectors. Bram and Jochen, thank you for working together on the project about the broader welfare benefits of prevention. I admire the ability of both of you to disseminate your research via various channels and your commitment to contribute to societal welfare with your research. I also want to thank Tom and Koen for answering all my questions about the Wevaluate software, Charlotte and Karen for introducing me into LatentGold, and Ignacio for generously helping me with any questions related to optimal portfolio analyses.

An absolute highlight of my PhD was the research visit to Oxford. John, thank you so much for hosting me. Your way of combining an in-depth technical expertise with the ability to conduct research of high societal relevance is inspirational and something I'll continue to pursue. Your relaxed style of working suited me well, and I hope to continue working together in the years to come. Also thanks to everyone else in Oxford who made my stay so wonderful, both within the university walls (special mentions for Joaquim, Apostolos, Stavros, Shuye, and Nam) and outside of them (especially Max, Andy, Chris, Sandor, Wolf and Jovan).

After Rotterdam, Delft, and Oxford, Leeds has been an important place to me during my PhD. Much of what I know about choice modelling I've learned there. I'd like to thank Stephane, Michiel, and Thijs for their courses, which were essential in my development as a researcher. Thijs, there's much to thank you for; apart from your lectures also for the collaboration with me and John, the meetings in Leeds, Oxford, Delft and online, your endless patience when I tried to follow your explanations (but actually managed to do so only some of the times), and being part of my defence committee and symposium. Romain, thanks for the interesting and entertaining chats we've had in Leeds and Nancy and your advice on some of my modelling choices. Let's meet at a next WONV meeting, hopefully.

Thanks to all the many great people I've met at conferences, workshops, seminars, courses and research visits in Rotterdam, Groningen, Maastricht, Ermelo, Egmond aan Zee, Leusden, Enschede, Berlin, Oslo, Vienna, Bologna, Nancy, Paris, London, Liverpool, Leeds, Oxford, Aberdeen and Brighton. These meetings gave me inspiration for my daily work, taught me a lot, and were definitely among the highlights of working in academia. Naming you all would be unfeasible, so I'm only thanking a few people in particular. Verity, thank you being such a wonderful host during my visit to Aberdeen and for being part of my dissertation and defence committees. Hans and Marije, also thank you for being part of the dissertation committee, defence committee, and/or my symposium. Anaïs, thank you for inviting me to the workshop in Paris and visiting me in Rotterdam. I'm looking forward to a great collaboration with you, Jonathan, and Thomas. Iris, our experience of being stuck in Oslo for an entire weekend is definitely one for the books. We both seem to excel in combining things that are seemingly uncombinable and (whether or not along

A

those lines), I'm looking forward to continue meeting in the years to come and maybe even do a study together one day.

Teaching was an important element of my PhD, and I'd like to thank those with whom I worked together and who gave me opportunities for taking the next step. Especially Job for trusting me with taking over part of his course from my first year onwards, Elly and Martijn for the happy-spirited collaboration in recent years, and above all Vivian. Thank you for your support and the great atmosphere you're creating, I look forward to many more years of sharing (research) interests, stories, and music: let's keep on dancing at the pink pony club!

Then, I'd like to thank everyone from the Erasmus Choice Modelling Centre for the monthly gatherings and support along the way, in particular a few people: Marcel, thank you for the many conversations in the hallway and teaching me the basics of conducting a DCE. Jorien, apparently primary school De Springplank in Heino is a fruitful basis for becoming a choice modelling researcher ☺ Thank you for your help, advice, and your pragmatism and humour. Esther, thank you for building the thriving health preference research community in Rotterdam and well beyond, and for being part of my dissertation and defence committees.

Moving on to those who contributed (much) more indirectly to my PhD trajectory, thanks to all colleagues who made these years a lot of fun. I'd like to thank everyone by name, but that would be impossible. To throw in another cliché (that's nevertheless true): you know who you are! Even though I couldn't join as much as I wished, especially in the first years of the PhD, I truly enjoyed the many, many events. From random theme drinks to a bonsai workshop, from overpriced cocktails with dubious names in foreign cities to sports activities, from organizing events together to searching for conference afterparties, from landscape painting to karaoke – there are always so many things going on that it's sometimes hard to even keep track of it all.

Outside university, the past few years have been great too, with so many events, new adventures, and milestones. Thanks to everyone who's been part of it in some way and whose company I got to enjoy, whether it was partying, having dinner together, being on holiday, or organizing things in many different committees and boards. Looking forward to much more of that!

Naturally, my parents and sister have always played a large role, and that's no different these days, despite us living further apart and seeing each other less frequently. You might not have much affinity with what I'm doing or where I'm living, sometimes still wondering when I'm going to have an actual job or why I'm living in that overcrowded grim city, in the end you're always very supportive. I feel blessed having a warm 'home' to still visit on a regular basis and, while I know you'd want me to join you even much more often, I'm looking forward to many more celebrations, shopping trips, dinners, Christmas markets, and holidays.

Steef, from everyone who contributed indirectly, you've arguably contributed the most. The longer we're together, the more I start appreciating our balance in complementing and contrasting each other. Even though I'm not always fully aware of it in the moment, deep down I know how much I appreciate your stability, inquisitiveness, and positivity. I can't wait for many more years of exploring new places, being indecisive, creating organized chaos, and just being together.