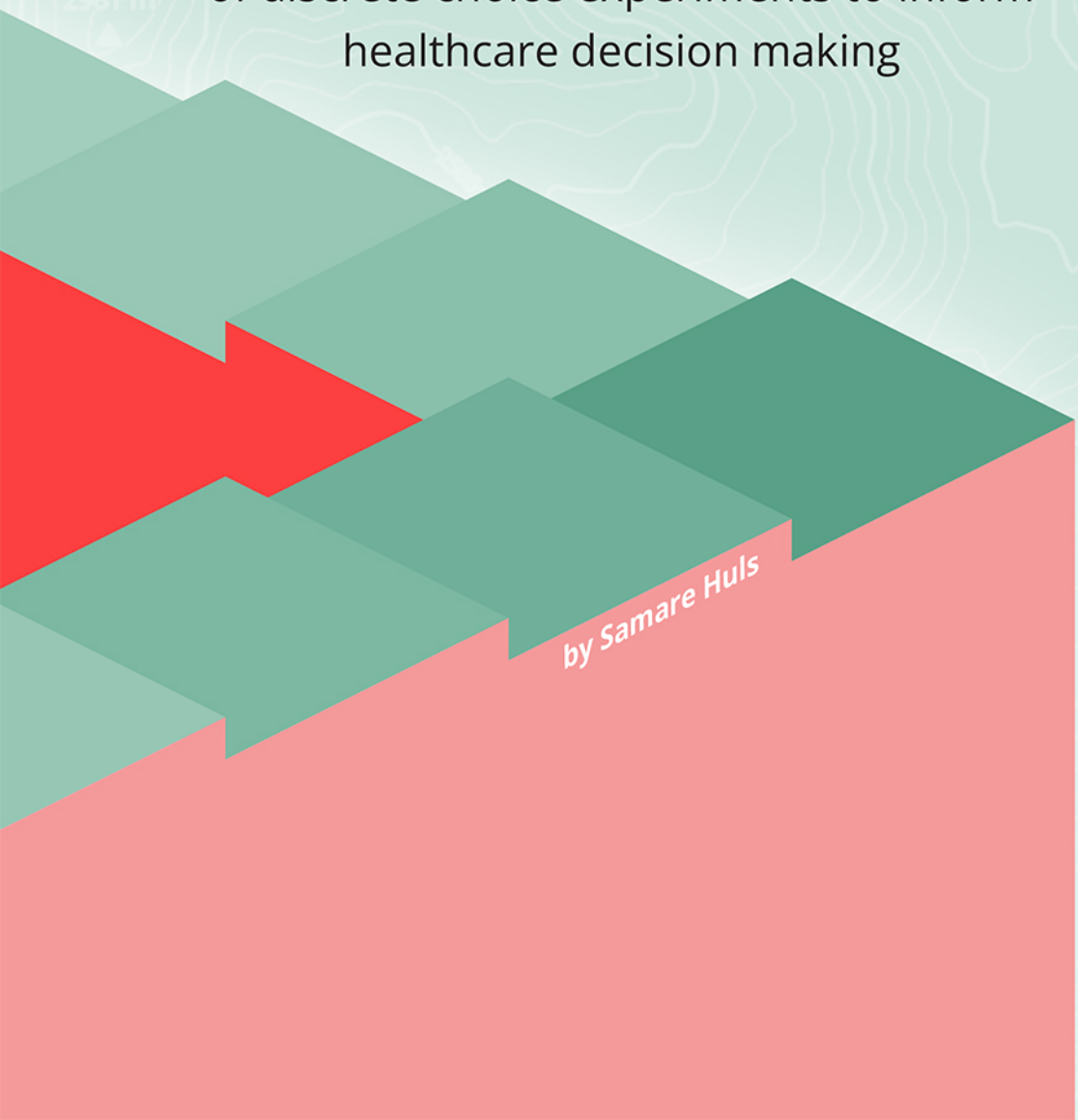
Saying yes, doing no?

Investigating the internal and external validity of discrete choice experiments to inform healthcare decision making



by Samare Huls

Saying yes, doing no?
Investigating the internal and external validity of discrete choice experiments to inform healthcare decision making

Ja zeggen, nee doen?
Onderzoek naar de interne en externe validiteit van discrete keuze-experimenten om de besluitvorming in de gezondheidszorg te informeren

Thesis

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

Prof. dr. A.L. Bredenoord

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

Thursday 1 December 2022 at 13.00 hours

Erasmus University Rotterdam

COLOPHON

Author: Samare Huls
Cover and chapter pages: Sebastiaan Huls & Rosa Stapel
Lay-out: Anke Muijsers - Persoonlijk Proefschrift
Printing: Ipskamp Printing

ISBN: 978-94-6421-862-6

This PhD trajectory was funded by the Erasmus Initiative - Smarter Choice for Better Health. Printing was partially funded by Erasmus School of Health Policy & Management.

Doctoral Committee

Promotors
Prof. dr. N.J.A. van Exel
Prof. dr. E.W. de Bekker-Grob
Prof. dr. C.A. Uyl-de Groot

Other members
Prof. dr. J.L.W. van Kippersluis
Prof. dr. M.P.M.H. Rutten-van Mölken
Prof. dr. G.A. de Wit

A PhD is like a topographic map: the presented information is subjective, unchartered territories are most interesting, and you can still get lost if you have one.

CONTENTS

CHAPTER 1

General introduction

Decisions about health always involve trade-offs between the pros and cons of alternative courses of action, both for individuals and for society. Individuals will generally prefer care that has many pros and few cons. However, preferences may differ between individuals. Pros can be improving length and quality of life, a preferred mode of administration, having a good connection with healthcare professionals, or any other type of expected benefits. Cons can be out-of-pocket costs, risk, and severity of side effects, waiting time, productivity losses, or burden of informal care. From a societal perspective, health care resources are scarce, and decisions need to be made about which care can be provided and which not. Consequently, not every individual can always get the care they prefer from their individual perspective. To achieve the most gain from the available budget, governments must trade-off the benefits and costs of health interventions. Aligning allocation decisions with public and patient preferences can ensure more support for difficult decisions at the societal level, and it can increase uptake, adherence, and satisfaction with treatment at the individual level (1–6). Public and patient preferences are increasingly used to inform decision making in healthcare, but how to elicit and incorporate preferences in a systematic and valid way, is still subject to debate.

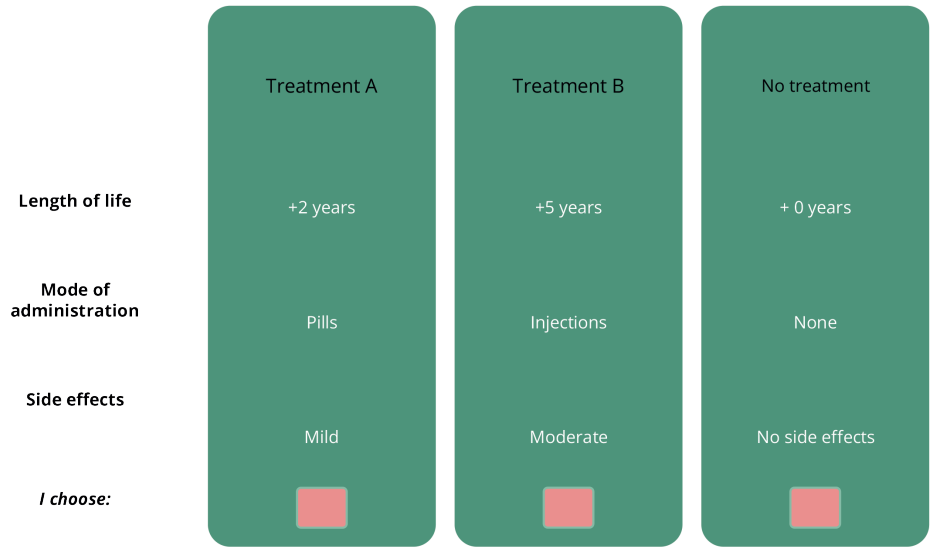
DISCRETE CHOICE EXPERIMENTS

One of the most used approaches to investigate public and patient preferences in the healthcare context is the discrete choice experiment (DCE) (7). This is a quantitative elicitation method with a strong theoretical foundation in random utility theory that elicits preferences from individuals in a systematic way by letting respondents repeatedly trade-off alternatives (8–12). To illustrate what a discrete choice experiment is, imagine the following - highly simplified - example in Figure 1. In this choice task there are three alternatives, namely “Treatment A”, “Treatment B” and “No treatment”. These alternatives are described by means of attributes and attribute levels. The three attributes in this example are “Length of life”, “Mode of administration” and “Side effects”. The levels of the first attribute are “+2 years” and “+5 years”, of the second attribute “Pills” and “Injections” and of the third “No”, “Mild” and “Moderate”. The last alternative, “No treatment”, is called the opt-out alternative, which is included so that the choice task more closely resembles real-life treatment choices. In a DCE, respondents complete several choice tasks, in which the attribute levels systematically change between alternatives and choice tasks. Based on the choices a group of respondents have made, the relative value of the attributes can be determined and the likelihood of choosing an alternative with certain attribute levels can be predicted. Common areas of application for DCEs in the healthcare context include eliciting preferences for treatments or services, and preferences for outcomes (8,13).

Ideally, DCEs are undertaken following available guidelines that consist of several stages (13–15). First, the relevant attributes and attribute levels are selected in accordance with the research question (16). Then, the experimental design – the specific combination of attribute levels to be used for the alternatives in each choice task – is generated (17). In this process, researchers aim to optimise the amount of information that can be elicited from each choice. The way in which the choice is presented, and the expected approach to analyse the data serve as input to this process. Thirdly, once the questionnaire is generated, data collection starts with a pilot study, after which the experimental design and the questionnaire can be updated if necessary. Lastly, econometric analyses are performed using the collected data (18–21). If these steps are carefully conducted, DCEs can provide valuable information for healthcare decision making.

DCEs are questionnaire-based and present respondents with hypothetical choices. Hence, DCEs elicit stated preferences (SP) and, therefore, primarily reflect the intention to choose something in the near or far future. The advantage of using this type of data is that preferences can be elicited in a controlled environment and that information can be obtained about many different choice situations not (yet) available in practice. However, the usefulness of DCEs for healthcare decision making depends on whether preferences elicited from respondents in a DCE are the same as preferences in reality, i.e., whether people actually do what they say they will do. Such information can be obtained by observing people’s choice behaviour in the past, their revealed preferences (RP). This type of data is free of hypothetical bias but may be influenced by many other factors outside the study context.

Figure 1: Example of a choice task in a discrete choice experiment



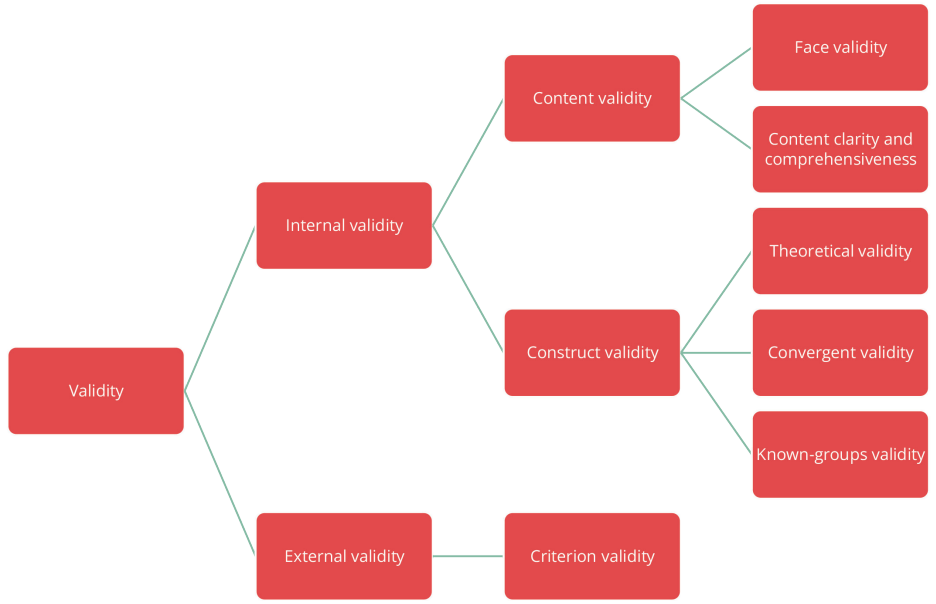
VALIDITY

The increasing application of DCEs to inform healthcare decision making stresses the importance of demonstrating the method's validity (7,22–24). In DCEs, validity generally is defined as "the extent to which quantitative measures of relative importance or trade-offs reflect the true preferences of patients" (25). There are different types and categorisations of validity. Figure 2 shows the categorisation that is used in this PhD thesis to distinguish between different types of validity of DCEs. This categorisation is based on existing categorisations of validity (26–32) and distinguishes three types of validity, namely content validity, construct validity and criterion validity (27). The first two refer to the internal validity of a DCE, while the last contributes to the external validity.

The internal validity of a DCE is defined as the extent to which the measured preferences represent actual preferences within the context of the particular study and is a reflection of "the rigour with which the study was conducted" (26). The first type of internal validity, content validity, is about face validity of the results and the clarity and comprehensiveness of the study. In DCEs, face validity can be assessed by examining whether the results match prior expectations and intuition (31,33–35). Clarity and comprehensiveness of the DCE is about whether respondents understand, and accept the study content and context, and whether all relevant attributes and levels are included in the DCE (30,31,36). The second type of internal validity is construct validity, which is about whether the results are consistent with hypotheses. The three subtypes of construct validity are theoretical validity, convergent validity, and known-groups validity. In DCEs, theoretical validity often reflects whether respondents' preferences are consistent with the axioms of random utility theory - the theoretical foundation on which the analyses of a DCE are based (30,31,37–39) - and whether the approach to analyse the data matches how respondents make decisions (28). Convergent validity is assessed by comparing the results of a DCE with those of other preference elicitation methods, ideally by letting the same respondents complete different methods (31). Known-groups validity refers to the degree to which differences in preferences between relevant groups of respondents can be distinguished, i.e., the degree to which heterogeneity in preferences can be attributed to respondents' characteristics (27,40).

The external validity of a DCE concerns the extent to which the measured preferences represent actual preferences outside the context of the particular study. In DCEs, external validity concerns the accuracy with which models based on stated preferences predict people's choices in the real world, that is, their revealed preferences (26,31). In this definition, external validity equals criterion validity, which takes revealed preferences as the gold standard. This specific type of validity is sometimes also referred to as predictive validity and reflects (32).

Figure 2: Types of validity



THESIS OBJECTIVES AND OUTLINE

The aim of this PhD thesis is to provide more insight into the internal and external validity of discrete choice experiments to inform healthcare decision making. The following three main objectives are distinguished:

1. Provide an overview of current challenges to integrating preferences in healthcare decision making;
2. Assess the internal validity of discrete choice experiments;
3. Assess the external validity of discrete choice experiments.

This thesis consists of seven chapters, including this general introduction to the thesis (Chapter 1) and a general discussion of the main findings, implications, strengths and limitations of this thesis, and recommendations for future research and policy (Chapter 7). The five substantive chapters 2 to 6 are structured into three parts, as presented schematically in Figure 3, following the abovementioned objectives.

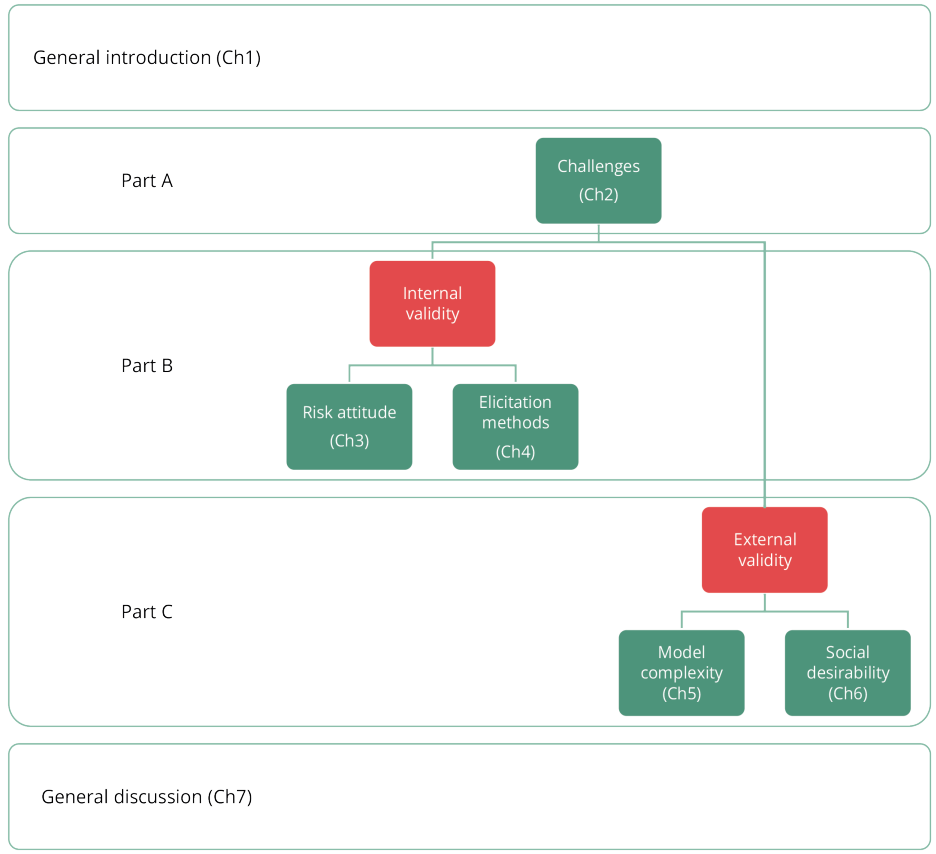
Part A consists of Chapter 2 which addresses the first objective of this thesis and sets the scene for the remainder of the thesis by reporting on a systematic review

of current challenges to integrating preferences in healthcare decision making, with a focus on patient preferences and health technology assessment.
Part B addresses the internal validity of discrete choice experiments and thus the second objective of this thesis. Chapter 3 addresses theoretical validity by studying how health-risk attitude affects individual decision making. The relationship between health-risk attitude and heterogeneity of preferences is studied by means of three case studies. Chapter 4 addresses theoretical and convergent validity by investigating how the presentation of choices affects stated preferences. In a head-to-head comparison of three different preference elicitation methods that move beyond traditional single-best DCE, theoretical validity is addressed by analysing the choices in each method and convergent validity by comparing the results between the three methods.

Part C of this PhD thesis extends to the external validity of discrete choice experiments and thus the third objective above, i.e., whether what people say they will do in a DCE matches what they do when facing the actual choice. In Chapter 5, stated and revealed preferences of respondents are compared in the context of colorectal cancer screening. Model complexity is gradually increased to assess its effect on prediction accuracy. Chapter 6 compares stated and revealed preferences in the context of food choice. It focusses particularly on socially desirable behaviour and its effect on the outcomes and prediction accuracy of DCEs.

Please note that Chapters 2 to 6 are based on papers submitted for publication in international peer-reviewed journals. Hence, these chapters can be read independently, and there may be some repetition between chapters.

Figure 3: Outline PhD thesis



☒A

☐B

☐C

PART A

SETTING THE SCENE

SETTING THE SCENE

To create the conditions in which something can happen

☒A

☐B

☐C

CHAPTER 2

Challenges

Based on
Huls, S.P.I., Whichello, C.L., van Exel, J., Uyl-de Groot, C.A., & de Bekker-Grob, E.W. (2019). What is next for patient preferences in health technology assessment? A systematic review of the challenges. Value in Health, 22(11), 1318-1328.

Published

HIGHLIGHTS

- The integration of patient preferences in health technology assessment is expected to increase uptake, adherence, and patient satisfaction.
- In this systematic review of the literature, we identified 37 unique research issues.
- Methodological and procedural research issues were most frequently mentioned.
- To advance the integration of patient preferences in HTA, a multi-stakeholder and holistic approach is needed.

ABSTRACT

Introduction

Integrating patient preferences in health technology assessment (HTA) is argued to improve uptake, adherence and patient satisfaction. However, how to elicit and incorporate these preferences in a systematic and scientifically valid manner is subject to debate. This article provides a systematic review of the challenges to integrating patient preferences in HTA raised in the literature about patient preferences in HTA.

Methods

A systematic review of articles published between 2013 and 2017 addressing challenges to the integration of patient preferences in HTA was conducted in seven databases. All issues with respect to the integration of patient preferences in HTA were extracted and divided into five categories: conceptual, normative, procedural, methodological and practical issues. The issues were ranked according to how often they were mentioned.

Results

Of 2,147 retrieved articles, 67 were included in the analysis. Thirty-seven unique research issues were identified. In the majority of the articles, methodological issues were posed (82%) followed by procedural (73%), normative (51%), practical (24%) and conceptual (9%) issues. Frequently posed methodological issues concerned preference heterogeneity and choice of method. Common procedural issues concerned how to evaluate the impact of preference studies and their degree of being evidence-based.

Discussion

This article provided an overview of issues with respect to the integration of patient preferences in HTA procedures. Most issues were of methodological and procedural nature; yet, the large number of different issues points to the overall importance of further researching the different aspects concerned with patient preferences in HTA. Through its ranking how many articles mention particular issues, this article proposes an implicit research agenda.

INTRODUCTION

Health technology assessment (HTA) informs reimbursement and coverage decisions on how to allocate health care resources to different health technologies by carefully assessing the costs and benefits of health interventions (41). With the increasing focus on patient preferences in clinical practice guidelines (42–44), academic research (45,46) and regulatory decision making (24,47,48), it is important that HTA not fall behind (49). The US Food and Drug Administration defines patient preferences as "qualitative or quantitative assessments of the relative desirability or acceptability to patients of specified alternatives or choices among outcomes or other attributes that differ among alternative health interventions" (50). In this context, qualitative assessments usually refer to exploring patient preferences and quantitative assessments refer to eliciting patient preferences. Not aligning the assessment of health intervention costs and benefits with patient preferences can cause adherence to be very different than expected, and it can explain why many health interventions that have developed throughout the medical product life cycle, end up not being used (5). Other arguments for integrating patient preferences are that it is considered ethical to listen to the patient voice (2,6), it will increase patient satisfaction (2,51) and that HTA decision making will be more informed and more transparent by including patient-relevant value judgements and experiential data (2–4,6,51).

Although the US Food and Drug Administration has provided guidance on how to use patient preference information in benefit-risk assessments (50), HTA is still lagging behind. Patients are increasingly being involved in the HTA decision-making process (52,53), but how to elicit and incorporate patient preferences in a systematic and scientifically valid manner is still subject to debate. For example, Weernink et al. (54) could not find a method that performed well from a statistical and patient burden point of view. Janssen et al. (55) suggested further researching validity and reliability tests for quantitative preference methods. Facey et al. (56) discussed whether and how qualitative, and quantitative, patient preference studies can be considered robust scientific evidence. Hansen and Lee (3) questioned the validity of qualitative research methods. In a recently published editorial, Mott (49) stated that HTA needs "substantive changes" to catch up with regulatory decision making in the incorporation of patient preferences. He mainly questions how to weigh patient preference information in current HTA procedures. Facey et al. (57) highlighted the "substantial challenges to realising the goal of informing evidence-based patient-centered policy." The variety of open questions concerning patient preferences in HTA raised by different researchers suggest the need for a comprehensive overview of all challenges in the field. Therefore, the objective of this article is to provide a systematic review of the challenges to integrate patient preferences in HTA raised by literature. By doing so, an implicit research agenda is proposed.

METHODS

Study design

To identify open questions concerning the use of patient preferences in HTA, the following study design was used. First, literature about patient preferences in HTA was identified. Second, the study characteristics of the included literature were elicited. Third, issues related to the integration of patient preferences in HTA, as raised in the literature, were derived. Last, the issues were categorised according to thematic differences and similarities. The results were analised on three different levels of categorisation, namely from broad to specific: "categories", "topics" and "issues". An illustration of the study design can be found in Figure 4.

Figure 4: Study design

* The three different levels of categorization, from broad to specific were: "categories", "topics" and "issues". Elicited issues were subdivided into topics; topics (and its issues) were also subdivided into categories.

Identification of literature

To identify literature about patient preferences in HTA, we conducted a systematic review using the databases Embase, Medline Ovid, Web of Science, Scopus, Cochrane CENTRAL, CINAHL EBSCOhost and Google Scholar. The search terms can be found in Appendix 1.

Articles were deemed eligible if they met the following six inclusion criteria. The studies had to concern patient preferences, had to concern health technology assessment, and had to discuss at least one issue concerning the integration of patient preferences in HTA. Further, the articles had to be English-language articles, the full text had to be available, and the articles had to be published between

2013 and 2017 were included because recent publication is inherent to providing a contemporary overview.

After excluding duplicates and articles outside the relevant publication years, two of the researchers (SH and CW) independently reviewed the remaining titles and abstracts for eligibility. If at least one of the researchers determined that an article met the eligibility criteria based on title and abstract screening, a full-text screening was done by the researchers. If no consensus could be reached about the eligibility of the full text, a third researcher (EBG) was consulted. The review was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) Statement (58).

Description of study characteristics

For all eligible articles, six study characteristics were extracted. Extracted data included the country in which the first author was employed, whether the study was a theoretical or applied study, the medical context (i.e., general or disease-specific) in which the study was conducted, and whether the article concerned qualitative (exploring) or quantitative (eliciting) patient preferences. In addition, we extracted which type of stakeholders raised the issue (e.g., respondents or the authors), and for which type of stakeholders the issue is relevant (e.g., patients, HTA bodies, or academics). Data were extracted by one researcher, after which two other researchers validated the findings.

Elicitation of issues

Issues concerning the integration of patient preferences in HTA were extracted from the literature in the broadest sense (i.e., questions, concerns, barriers, facilitators, and areas for further research). All study-specific elements were deleted from the extracted issues to allow for comparison of the issues across studies. Data were extracted by one researcher, followed by confirmation of two other researchers.

Three-level categorisation of issues

Two researchers (SH and CW) performed a three-level categorisation of the elicited issues, and a third researcher was consulted if no consensus could be reached. The three different levels, from broad to specific, were: “categories”, “topics” and “issues”. Issues were subdivided into topics; topics (and their issues) were also subdivided into categories. Again, an illustration of the study design is presented in Figure 4. To enhance consistency of the categorisation process, we defined, before analysis, whether a research issue would fall in only one category or topic, respectively. Consensus had to be reached on which category or topic best described the issue. As in Utens et al. (59), we used the following five categories as the broadest level of categorisation: conceptual, normative, procedural, methodological and practical issues. Conceptual issues relate to the definition and characterisation of patient preferences. Normative issues concern which members of society should have their

preferences elicited. Procedural issues relate to how to integrate patient preferences into the existing procedures of HTA. Methodological issues address establishing good and accurate research practice on the topic. Practical issues address all other concerns of practical nature such as time and money constraints. The topics were the second level of categorisation. Unlike the categories, topics were not predefined and were established using backwards induction. The issues were grouped according to thematic similarities and differences; the exact name of the topic was determined after the issues were grouped. Included in the third level of categorisation were the issues themselves. The categorised data were analised in two different ways. First, to give insight into the variety of issues in each category and topic, respectively, the number of issues in each category or topic (as % of the total number of issues) was established. Second, to measure frequency of occurrence of the issues, the number of articles that mentioned each issue (as % of total number of articles) was analised.

RESULTS

Identification of literature

The database search identified 2,147 articles, of which 375 unique articles published in 2013-2017 were screened. Sixty-seven articles met the inclusion criteria and were subject to data extraction and analysis (Figure 5).

Description of study characteristics

For most of the articles containing issues regarding patient preferences in HTA, the first authors worked on behalf of organisations/universities in Canada (n=13; 19%), the United Kingdom (n=13; 19%), and Germany (n=10; 15%) (Table 1). Other common countries of origin were the United States (n=8; 12%), Australia (n=7; 10%), and The Netherlands (n=6; 9%). Three-quarters of the articles (51 out of 67 articles) discussed the integration of patient preferences in HTA theoretically, rather than actually conducting a preference study. Almost two thirds of the articles (n=44) concerned a general medical context rather than a disease-specific context. Thirty-one articles concerned the qualitative elicitation of preferences (46%), whereas 17 concerned quantitative preference elicitation (25%), 9 concerned both (13%) and 10 did not specify (15%).

Many of the research issues were raised by the authors or authors cited in the articles (n=52; 78%). The vast majority of first authors (73%) worked in academia; the remainder worked for a variety of organisations (e.g., patient organisations, HTA agencies and private consultants). In the remainder of articles where study respondents were specifically asked about the advancement of patient preference integration in HTA (n=15, 22%), respondents were HTA professionals (n=5), patients (n=4), and a variety of other respondents (e.g., healthcare professionals, caregivers and policy makers, n=6). Most of the issues were relevant for HTA professionals and academics (n=34; 51%). Other issues were relevant for HTA professionals only (n=23;

34%), or for a variety of HTA professionals, clinical guideline developers, patients, patient organisations, or clinicians (n=10; 15%). Table 1 summarises these study characteristics; a more elaborate overview of the study characteristics per article is presented in Table A 1 in Appendix 2.

Figure 5: Study selection

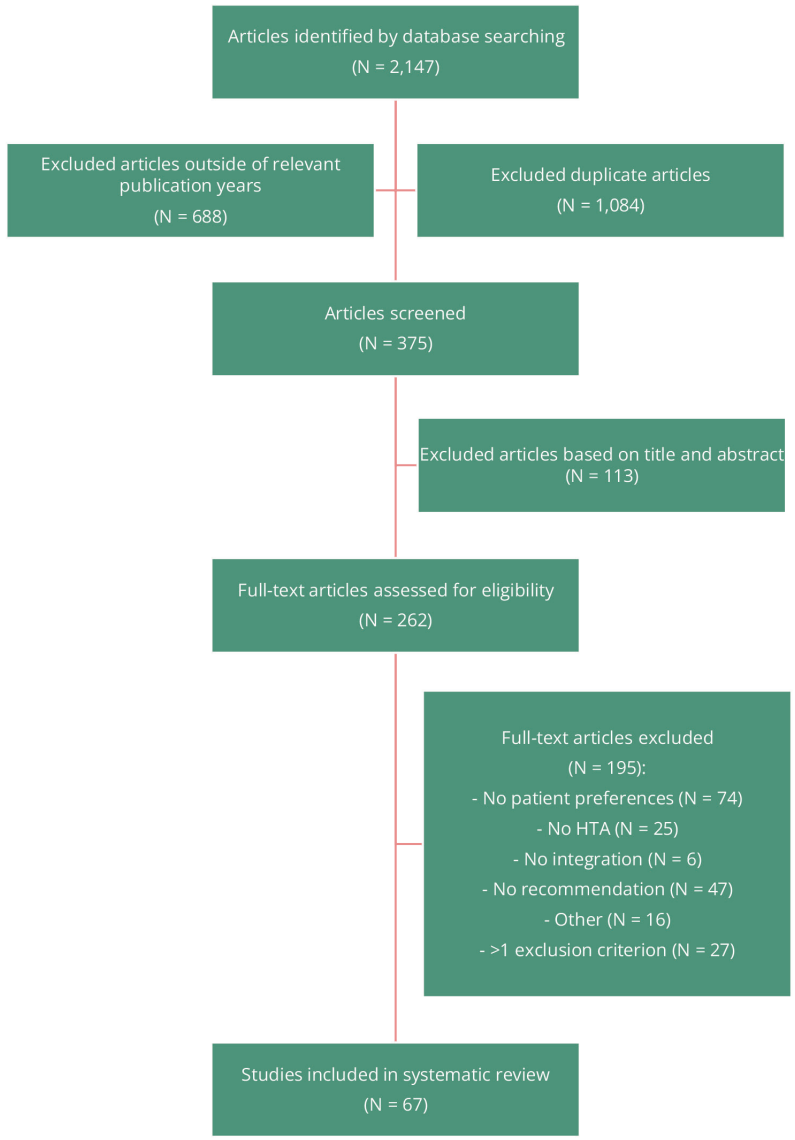


Table 1: Study characteristics - summary

Items		N=67[1]	%[2]
Country of origin	Canada	13	19%
	United Kingdom	13	19%
	Germany	10	15%
	United States	8	12%
	Australia	7	10%
	The Netherlands	6	9%
	Other	10	15%
Type of study	Theoretical	51	76%
	Application	15	22%
	Both	1	1%
Medical context	General	44	66%
	Disease-specific	23	34%
Preference elicitation	Qualitative	31	46%
	Quantitative	17	25%
	Both	9	13%
	Not defined	10	15%
Issue raised by stakeholder	Authors and cited authors	52	78%
	Respondents: HTA professionals	5	7%
	Respondents: Patients	4	6%
	Respondents: Other	6	9%
Issue relevant for stakeholder	HTA professionals and academics	34	51%
	HTA professionals	23	34%
	Other	10	15%

Categorisation of issues

Across the five categories of identified issues, 16 topics and 37 unique research issues were identified from the total selection of articles. The issues were the most specific level of categorisation. These were subdivided into topics. In turn, the topics were subdivided into categories. Table 2 presents a broad overview of the research categories and topics. Table 3 presents the most specific level, namely, the issues. The analysis of the three levels of categorisation is discussed ranging from broad to specific.

1 Absolute number of articles.
2 Relative number of articles (as % of total of 67 articles). Percentages might not add up to 100% because of rounding error.

Table 2: Relative occurrence of issues, per category of issues and per topic of issues

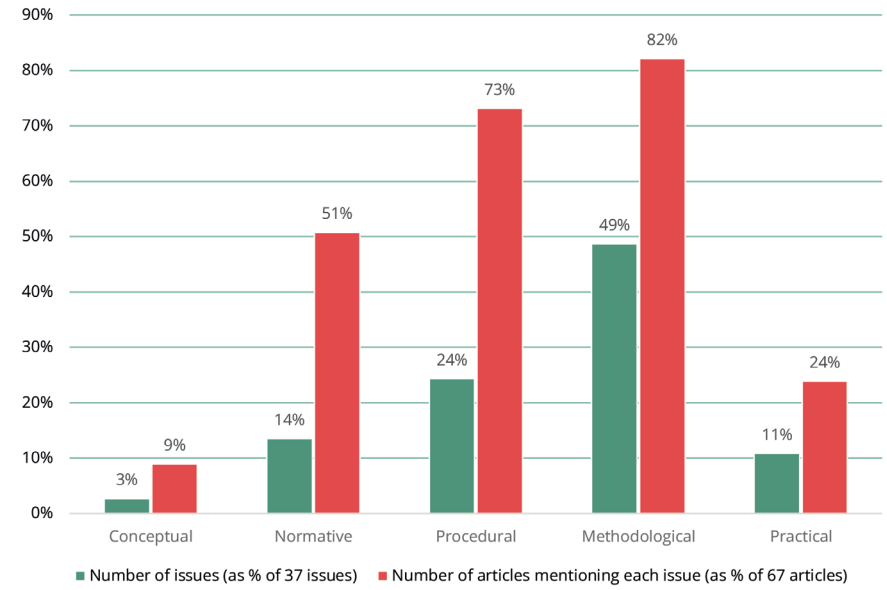
Category	Topic	# issues		# mentions	
		N=37[3]	%[4]	N=67[3]	%[4]
Conceptual		**1**	**3%**	**6**	**9%**
	Taxonomy	1	3%	6	9%
Normative		**5**	**14%**	**34**	**51%**
	Whose preferences	4	11%	26	39%
	Relevance of preferences	1	3%	11	16%
Procedural		**9**	**24%**	**49**	**73%**
	Weight	3	8%	21	31%
	Impact	1	3%	21	31%
	Patient education	3	8%	20	30%
	Evidence-based	1	3%	17	25%
	HTA stage	1	3%	16	24%
Methodological		**18**	**49%**	**55**	**82%**
	Choice of method	3	8%	30	45%
	Internal validity	3	8%	24	36%
	Generalisability	4	11%	19	28%
	Sample selection	1	3%	15	22%
	External validity	2	5%	9	13%
	Patient characteristics	2	5%	8	12%
	Reliability	3	8%	7	10%
Practical		**4**	**11%**	**16**	**24%**
	Resources	4	11%	16	24%

Categories of identified issues

Of the 37 issues, one was conceptual (3%), five were normative (13%), nine were procedural (24%), eighteen were methodological (49%), and four were of practical nature (11%) (Figure 6). In terms of how often the issues were mentioned, methodological issues arose relatively often from the literature, namely in 55 out of 67 articles (82%). Procedural issues were also mentioned frequently (n=49; 73%). Normative issues were raised relatively less (n=34; 51%) followed by practical issues (n=16; 24%) and conceptual issues (n=6; 9%).

3 Absolute number of issues identified and absolute number of articles mentioning each issue.
4 Relative number of issues (as % of 37 issues) and relative number of articles mentioning each issue (as % of 67 articles). Percentages might not add up to 100% because most studies mentioned multiple issues or because of rounding error.

Figure 6: Relative occurrence of issues, per category of issues



Topics of identified issues

The 16 research topics that were extracted can be found in Table 2. Each of these topics contains between one and four issues. The establishment of a taxonomy for patient preference studies was the only conceptual topic that was raised. It was raised in six out of 67 articles (9%, e.g., Utens et al. (59), Brooker et al. (60), and DeJean et al. (61)). Normative topics included whose preferences to elicit, and the relevance of preference studies to patients. Whose preferences to elicit was mentioned most, namely in 26 out of 67 articles (39%, e.g., Rashid et al. (62), Gagnon et al. (63), and Buck et al. (64)). Procedural topics concerned what weight to give patient preferences in current HTA procedures, how to evaluate impact, how to educate patients in preparation for preference studies, whether and how patient preferences are evidence-based, and in which HTA stage to incorporate patient preferences. The most often mentioned procedural topics were how to weight preference studies in comparison to or in addition to current ethical, clinical and cost-effectiveness (quality adjusted life year [QALY]) procedures (n=21; 31%, e.g., Dirksen (2), Mühlbacher and Kaczynski (65), and Mühlbacher and Sadler (66)), and how to evaluate the impact of preference studies on HTA decision making (n=21; 31%, e.g., Dipankui et al. (67), Kreis and Schmidt (68), and Abelson et al. (69)). Methodological topics concerned choice of method, internal and external validity, reliability, generalisability, and which patient characteristics affect preferences and how. The most prevailing methodological topic was choice of method (n=30; 45%, e.g., Utens et al. (70), Wortley et al. (71), and Brereton et al. (72)), followed by internal

validity (n=24; 36%, e.g., Brooker et al. (60), Wahlster et al. (73), and Danner et al. (74)). The only practical topic that was raised concerned resource constraints in conducting preference studies, which was mentioned in 16 out of 67 articles (24%, e.g., Utens et al. (59), Hailey et al. (75), and Single et al. (76)).

Issues

Table 3 gives an overview of the 37 unique research issues and their frequency of being mentioned. The most frequently posed normative issues concerned whether the representatives of patient organisations represent the preferences of a broader set of individuals (n=13; 19%, e.g., Rashid et al. (62), Gagnon et al. (63), and Buck et al. (64)) as well as whose preferences should be elicited (n=12; 18%, e.g., Kreis et al. (77), Mott and Najafzadeh (78), and Thokala et al. (79)), and whether patient relevant outcomes and processes should be accounted for in preference studies and how this should be done (n=11; 16%, e.g., Evers et al. (80), Mühlbacher et al. (81), and Berglas et al. (82)). The most frequently posed issue of a procedural nature was how to evaluate impact of preference studies (n=21; 31%, e.g., Dipankui et al. (67), Kreis and Schmidt (68), and Abelson et al. (69)), followed by whether preference studies can be considered robust scientific evidence (n=17; 25%, e.g., Iskrov and Stefanov (83), Moreira (84), and Tordrup et al. (85)), and in which HTA stage to incorporate them (n=16; 24%, e.g., Hämeen-Antilla et al. (86), Weeks et al. (87), and Husereau et al. (88)). The most frequently raised methodological issues were about which methods to use for preference elicitation (n=29; 43%, e.g., Utens et al. (70), Wortley et al. (71), and Brereton et al. (72)) and about heterogeneity in preferences (n=18; 27%, e.g., Wahlster et al. (73), Di Paolo et al. (89), and Doctor and MacEwan (90)). The most frequently mentioned practicial issues with conducting preference studies were cost constraints (n=13; 19%, e.g., Wortley et al. (71), Mossman et al. (91), and Kievit et al. (92)) and time constraints (n=11; 16%, e.g., Buck et al. (64), Brereton et al. (93), and Scott and Wale (94)).

Table 3: Three-level categorisation and relative occurrence of issues

Category	Topic	#	Issue	N=67[5]	%[6]	Article(s)
Conceptual	Taxonomy	1.	How should we define patient preferences and subsequently find and retrieve patient preference studies?	6	9%	(59–62,70,83)
Normative	Whose preferences	1.	Do preferences of representatives of patient organisations/advocacies represent preferences of a broader set of individuals?	13	19%	(55,63,64,77,86,91,93–99)
		2.	Whose preferences should be elicited (e.g., patients with or without treatment experience, carers, patient representatives)?	12	18%	(44,59,65,67,68,77-79,96,100,97,99, 101)
		3.	Are patient preferences influenced by external factors (e.g., media, family, or pharmaceutical companies)?	7	10%	(2,62,63,91,96,98,102)
		4.	How can preferences from various samples (e.g., clinicians, carers, and patients) be synthesised to be of value as a whole?	3	4%	(80,87,101)
	Relevance	1.	What are patient relevant outcomes (i.e., health versus wellbeing) and should preference studies also focus on process?	11	16%	(2,59,70,80–82,92,103–106)

5 Absolute number of articles mentioning each issue.
6 Relative number of articles mentioning each issue (as % of 67 articles). Percentages do not add up to 100% because most studies mentioned multiple issues or because of rounding error.

Table 3: *Continued*

Category	Topic	#	Issue	N=67[5]	%[6]	Article(s)
Procedural	Weight	1.	How should preference studies be evaluated in comparison/addition to clinical and economic evaluation studies?	15	22%	(2,59,61,63,65,70,72,73,94,96,97,99, 103,107,108)
		2.	How can preference studies add to or replace the QALY paradigm?	5	7%	(2,66,70,85,109)
		3.	How should ethical issues concerning patient preferences be weighed in HTA?	4	6%	(6,76,94,102)
	Impact	1.	How can we evaluate the impact of patient preferences studies on HTA decision making?	21	31%	(2,55,59,62,63,65,67-69,71,77,83,86, 87,91,94,95,97,100,110,111)
	Patient education	1.	How can patients be sufficiently trained to perform HTA studies?	14	21%	(55,63,64,77,86,89,93,94,96,97,100, 102,106,112)
		2.	How can communication between researchers and patients be aligned in preference studies?	6	9%	(55,63,64,94,97,106)
		3.	How should patients and caregivers be informed about HTA studies and the possibility to be involved?	12	19%	(71,80,86,87,94,96,97,100,103,104,112, 113)
	Evidence-based	1.	How is and should the quality and transparency of patient preference studies be assessed to be considered robust scientific evidence?	17	25%	(2,55,62,67,72,77,80,83–85,92,93,102, 96,109,99,112)
	HTA stage	1.	In which context and stage of HTA should preferences be used to inform decision making?	16	24%	(59,63,64,68,80,82,86-88,95,99,103, 105,107,111,112)

5 Absolute number of articles mentioning each issue.
6 Relative number of articles mentioning each issue (as % of 67 articles). Percentages do not add up to 100% because most studies mentioned multiple issues or because of rounding error.

Table 3: *Continued*

Category	Topic	#	Issue	N=67[5]	%[6]	Article(s)
Methodological	Choice of method	1.	Which methods are preferable for eliciting preferences?	29	43%	(2,59–61,63,65,66,68,70–74,79,85, 86,93,114,96,100,109,104,99,101,111, 112,115–117)
		2.	When should we use quantitative (eliciting) versus qualitative (exploring) research methods for patient preference studies?	3	4%	(2,59,60)
		3.	Which methods to elicit patient preferences are preferable in which stage of HTA?	2	3%	(63,95)
	Internal validity	1.	How do preferences on an individual level differ from those on a collective level (i.e., preference heterogeneity)?	18	27%	(55,66,70,72–74,78,81,89,90,92,94, 103,106,108,117–119)
		2.	How does framing affect preferences?	4	6%	(60,65,74,120)
		3.	How can validity of preference studies be tested?	1	1%	(81)
	Reliability	1.	How stable are preferences over time?	4	6%	(2,81,82,104)
		2.	How consistent are individuals in preference studies, how does this affect results, and how should inconsistent responses be handled?	3	4%	(55,74,81)
		3.	How should uncertainty in patient preferences be modelled?	1	1%	(65)

5 Absolute number of articles mentioning each issue.
6 Relative number of articles mentioning each issue (as % of 67 articles). Percentages do not add up to 100% because most studies mentioned multiple issues or because of rounding error.

Table 3: *Continued*

Category	Topic	#	Issue	N=67[5]	%[6]	Article(s)
	Generalisability	1.	How representative are preferences from the recruited sample for the entire population?	9	13%	(55,62,68,81,93,94,96,100,101)
		2.	Can preference studies be transferred across diseases and contexts (i.e., as a generic instrument)?	7	10%	(2,59,67,73,80,85,104)
		3.	Can preference studies be transferred across countries/sociocultural groups?	4	6%	(72,81,93,113)
		4.	How representative are preferences for a singular intervention compared to when it is administered alongside other interventions?	2	3%	(82,104)
	Sample selection	1.	How (i.e., via which channels and based on which characteristics) should the sample be selected?	15	22%	(62 – 64,75,77,80,86,87,93,96,97,100, 111,120,121)
	External validity	1.	What is the external validity of preference studies, and how can this be improved?	5	7%	(55,74,93,119,120)
		2.	How can we merge real-world data (e.g., adherence data) and stated preference studies?	4	6%	(2,82,83,90)

5 Absolute number of articles mentioning each issue.
6 Relative number of articles mentioning each issue (as % of 67 articles). Percentages do not add up to 100% because most studies mentioned multiple issues or because of rounding error.

Table 3: *Continued*

Category	Topic	#	Issue	N=67[5]	%[6]	Article(s)
	Patient characteristics	1.	Which sociodemographic patient characteristics (e.g., gender, age, education, income, family, risk attitude, and beliefs) affect preferences and how should we tailor these subgroups?	6	9%	(64,72,80,89,90,119)
		2.	Which disease-specific patient characteristics (e.g., stage and severity of illness) affect preferences, and how should we tailor these subgroups?	6	9%	(80,89,90,104,118,119)
Practical	Resources	1.	How can cost constraints of preference studies be overcome?	13	19%	(59,64,71,75–77,87,88,91–94,100)
		2.	How can time constraints of preference studies be overcome?	11	16%	(59,63,64,75–77,88,93,94,100,121)
		3.	How can staff/expertise constraints of preference studies be overcome (who should perform preference studies)?	6	9%	(59,63,76,94,100,122)
		4.	How can location constraints of preference studies be overcome?	1	1%	(64)

5 Absolute number of articles mentioning each issue.
6 Relative number of articles mentioning each issue (as % of 67 articles). Percentages do not add up to 100% because most studies mentioned multiple issues or because of rounding error.

DISCUSSION

In this study, from a selection of 67 articles, we identified 37 unique research issues that concern the integration of patient preferences in HTA. In most of the articles, methodological issues were raised (82%) followed by procedural (73%), normative (51%), practical (24%), and conceptual (9%) issues. Frequently posed methodological issues were about preference heterogeneity and choice of method. Common procedural issues concerned how to evaluate the impact of preference studies and their degree of being evidence-based.

The relatively large number of unique issues shows that patient preference integration is by and large a relevant topic to be researched. This review includes theoretical and applied studies and includes studies in numerous medical contexts from various countries. Furthermore, the identified issues relate to qualitative (exploring) and quantitative (eliciting) preference methods that might vary in rigorousness and addressability depending on the research question concerning patient preferences in HTA. Given the variety of study characteristics, this review provides a comprehensive research agenda that is relevant for multiple stakeholders. The issues in the articles were relevant for HTA professionals, academic researchers, clinical guideline developers, patients, patient organisations and/or clinicians. Nonetheless, the majority of the issues were raised by academic authors of the articles and the articles provide little guidance on how to address the issues. Hence, we believe that to reach consensus on the way forward, involvement, coordination and collaboration between the different stakeholders is warranted.

The issues identified in this review are very much in line with other non-HTA specific literature (5,24) in which experts generally argue for the use of patient preferences in health care research. According to Ostermann et al. (5) important issues are internal and external validity, reliability, and preference heterogeneity, as well as evidence-based prediction of uptake and adherence. In a research agenda concerning regulatory review of medical devices, Levitan et al. (24) stated that validity and reliability, the choice of method, sample selection, patient relevant outcomes and processes, framing, patient education, and correcting for patient characteristics were important issues to consider. Mott (49) prioritised issues about weighting patient preferences in current HTA procedures. He discussed whether patient preferences should be incorporated within the QALY or beyond the QALY, and proposed multiple-criteria decision analysis (MCDA) as a new methodological approach to HTA. The challenges mentioned by Facey et al. (57) include the impact of preference studies on HTA decisions, time and cost constraints, and how to weight preference studies alongside clinical and cost-effectiveness studies. The authors also strongly highlighted the need for patient preference studies to be evidence-based. Despite the fact that the previously mentioned authors differed in their prioritisation, all of the issues in their articles were also identified in our review, advocating its inclusiveness.

Some aspects of this review require discussion. A first limitation is that articles outside the scope of our definitions may have been overlooked for various reasons. As mentioned in some of the included articles (59–62,70,83), patient preferences are not clearly defined, and therefore studies concerning this topic are not easily retrievable. Furthermore, the integration of public preferences is sometimes discussed alongside the integration of patient preferences (62,64,68,69,78,87, 95,96,100,102,103,114,120,122). Although it is an important issue to address, it was not the explicit goal of this research to take a stance on or provide an overview of whether to use patient, public, or both preferences in HTA. For an overview of arguments for and against public and patient preferences in health valuation, other literature (49,123–128) can be consulted. Other reasons for potentially having overlooked important research questions that are inherent to reviewing literature are publication lag or bias and the inclusion of only English-language articles.

Second, the process of categorisation should be interpreted with caution. For pragmatic reasons, study characteristics and issues were extracted by one researcher, followed by confirmation by two other researchers. The categorisation of issues was performed by two researchers, followed by confirmation by a third researcher, yet the issues were subjectively categorised to put them into context. There could be discrepancies between what was originally meant by authors of the articles included in the analysis, how we interpreted the issues, and how other researchers would interpret them. In addition, we interpreted frequency of occurrence as a way to measure priority. The broader the issue, the more it is likely to occur; so it is possible that issues unintentionally became weighted according to their specificity in the extraction process. The current categorisation is by no means intended to be definitive. However, it is a systematically retrieved overview and, we believe, an informative descriptive basis for a more extensive prioritisation of issues to advance the integration of patient preferences in HTA. Other interesting research that goes beyond of the scope of this article could be in-depth analysis about how knowledge accumulated on a particular issue or topic as listed in this review. In addition, it might be interesting to research how particular issues or topics trend together.

Third, it should be noted that HTA studies vary in the degree to which patient preferences are meaningful. According to the articles included in this systematic review, integrating patient preferences in HTA is mostly relevant for in the following situations: when there is no one treatment that is considered superior (2,59,60), when the benefits of interventions are only marginal (2), when uncertainty of the treatment outcome is high (44), when there are multiple alternatives that vary largely in terms of risk-benefit trade-offs (44), when preferences of patients are expected to be very heterogeneous (44), and when the treatment concerns a rare disease that would benefit from early HTA (80).

Based on the literature and our interpretation of the data, we recommend two areas for further research that are fundamental to the advancement of

integrating patient preferences in HTA. First, as addressed in articles included in this review (2,66,70,85,109) and beyond (49,123,129), the discussion as to whether patient preferences should be incorporated within the QALY or beyond the QALY is essential to the integration of patient preferences in HTA. Second, in agreement with articles included in this review (66,79) and broader literature (47,49,53,130–132), we recommend exploration of the possibilities of using multiple-criteria decision analysis to integrate patient preferences. Both of these procedural matters relate to normative changes to current HTA procedures, hence warranting further research. Other normative issues such as whose preferences to incorporate in HTA concern a choice rather than further research. To address this entire spectrum of issues identified in this review, especially for normative issues, there needs to be better communication, collaboration, and consensus between the different stakeholders (104,110).

In line with the increasing use of patient preferences in various medical contexts, the integration of patient preferences in HTA is expected to contribute to better decision making, and to increase uptake, adherence, and patient satisfaction. So, what is next for patient preferences in HTA? Methodological and procedural issues were mentioned most; yet, the large number of different issues advocates the overall importance of a multi-stakeholder and holistic approach to the integration of patient preferences in HTA. By providing a contemporary overview of issues in the literature, this review is an important first step towards the integration of patient preferences in HTA in a systematic and scientifically valid manner. The next step requires coordination and collaboration between the different stakeholders to reach consensus on the way forward.

PART B

INTERNAL VALIDITY

☐ A

☒ B

☐ C

INTERNAL VALIDITY

The extent to which the observed results represent the truth within the context of a particular study

☐ A

☒ B

☐ C

CHAPTER 3

Risk attitude

Based on
Huls, S.P.I., Veldwijk, J., Swait, J.D., Viberg Johansson, J., Ancillotti, M., & de Bekker-Grob, E.W. (2022). Preference variation: where does health-risk attitude come into the equation? Value in Health.

In press

HIGHLIGHTS

- Decisions about health often involve risk, and risk preferences may vary between people.
- Interest in health preference heterogeneity and the role of risk is increasing.
- Modelling health-risk attitude as an individual characteristic underlying preference heterogeneity has the potential to improve model fit and model interpretations.
- Further research into the relationship between health-risk attitude and preference heterogeneity is warranted.

ABSTRACT

Introduction

Decisions about health often involve risk, and different decision-makers interpret and value risk information differently. Furthermore, an individual's attitude towards health-specific risks can contribute to variation in health preferences and behaviour. This study aimed to determine if and how health-risk attitude (HRA) and heterogeneity of health preferences are related.

Methods

To study the association between HRA and preference heterogeneity, we selected three Discrete Choice Experiment (DCE) case studies in the health domain that included risk attributes and accounted for preference heterogeneity. HRA was measured using the 13-item Health-Risk Attitude Scale (HRAS-13). We analysed two types of heterogeneity via panel latent class analyses, namely how HRA relates to 1) stochastic class allocation and 2) systematic preference heterogeneity.

Results

Our study did not find evidence that health-risk attitude as measured by the HRAS-13 distinguishes people between classes. However, we did find evidence that the HRAS-13 can distinguish people's preferences for risk attributes within classes. This phenomenon was more pronounced in the patient samples than in the general population sample. Moreover, we found that numeracy and health literacy did distinguish people between classes.

Discussion

Modelling health-risk attitude as an individual characteristic underlying preference heterogeneity has the potential to improve model fit and model interpretations. Nevertheless, the results of this study highlight the need for further research into the association between health-risk attitude and preference heterogeneity beyond class membership, a different measure of health-risk attitude, and the communication of risks.

INTRODUCTION

Decisions about health often involve risk, but at the same time it must be recognised that different decision-makers interpret and value risk information differently (133,134). For example, some people prefer a small chance of full recovery, over a more certain but moderately good outcome, and others vice versa. Examples such as these portray why patients' decisions sometimes conflict with physicians expectations or advice (135). The psychological construct that describes how people make decisions under uncertainty is called risk attitude (136), which has been shown to affect health policy decisions on an aggregate level when incorporated in analyses (137,138). Also on an individual level, attitude towards health-specific risks contributes to heterogeneity in health preferences and behaviour (139,140). Insights into the relationship between risk attitude and preference heterogeneity in health can improve the accuracy with which uptake and adherence to treatment are predicted (141,142). In addition, they can be informative when alternatives largely vary in terms of benefit-risk or when treatment outcomes are highly uncertain (44,143), for example, by providing information tailored to varying health-risk attitudes in clinical practice, or for patient subgroup considerations in benefit-risk assessments. Although researchers generally agree that risk attitude plays a role in healthcare decision making, there is no consensus on how to operationalise it (144). Risk attitude is often domain-specific (145,146); it covers both risk perception and risk-taking behaviour, although these can be conflicting concepts (146,147), and risk attitudes as measured in questionnaires do not always translate to the real world (148,149). As such, studying the relationship between health-risk attitude and health preferences is rather complex.

Health preferences are often elicited via Discrete Choice Experiments (DCEs), a method to quantify preferences for alternative health interventions by letting respondents repeatedly trade-off alternatives that are described using a variety of attributes and levels (22). They are increasingly used to incorporate patients' preferences in various medical contexts that concern benefit-risk decision making, such as clinical guidelines, regulatory decision making and health technology assessment of these health interventions (22,50,150). Many of the health interventions described in DCEs include one or more risk attributes. Harrison et al. systematically reviewed DCEs with a risk attribute in the field of health, and found that few of these DCEs incorporated individual characteristics underlying risk preferences (151). Only Tsuge et al. elicited subjective risk perception, and found that it influenced willingness to pay in their discrete choice experiment (152). None of the other identified studies elicited risk attitude; instead, they derived it from responses to the choice tasks and used this information in their statistical analysis. In line with the increasing use of and demand for analysing heterogeneity in DCEs (22,153,154), Russo et al. systematically reviewed individual characteristics underlying the decision-making process and their relation to preference heterogeneity (144). They

identified risk attitude as one important, yet marginally studied, factor relating to preference heterogeneity; experts agreed with this assessment (143). Furthermore, studies that assess risk in health-related DCEs focus on risk communication rather than risk attitude (155,156). While individual characteristics such as health numeracy, health literacy and decision-making style (also identified by Russo et al. (143)) are increasingly found to be related to preference heterogeneity (157,158), the complexities associated with the operationalisation of risk attitude hampers studying the relationship between risk attitude and preference heterogeneity (144).

Therefore, the purpose of this study therefore is to determine if and how health-risk attitude and heterogeneity of health preferences are related by means of three case studies, using the relatively new Health-Risk Attitude Scale (HRAS-13), which aims to overcome some of the operational complexities (147). To assess the relationship between the HRAS-13 and heterogeneity, we studied two types of heterogeneity, namely 1) stochastic class assignment and 2) systematic preference heterogeneity.

METHODS

Case studies

To study the association between health-risk attitude and heterogeneity of preferences, we selected three DCE case studies in the health domain that had at least one attribute that implicitly or explicitly concerned risks and for which we could gain the authors' consent to share the data for this purpose. The studies differed in terms of their topic, country, study population, number of respondents, and their DCE design leading to an increased generalisability of the results. An overview of the case studies and their DCE designs can be found in Table 4 (159–161). Attributes and levels were selected based on literature reviews, focus groups and interviews; these are presented in Table 5. The first case study concerned the treatment preferences of patients with Multiple Sclerosis (MS) in The Netherlands, France and the United Kingdom (159). Inclusion criteria were the following: aged 18 years or older, diagnosis of MS, and living in one of these three European countries. Respondents were recruited online via the commercial sampling company Survey Engine (N=753). Three out of four attributes were explicitly described as risks to questionnaire respondents. The second study analysed preferences regarding antibiotics usage in Sweden (160). An online sample of respondents between 18 and 65 years old was recruited from the Swedish general public (N=378). Respondents were recruited online via Dynata, a commercial sampling company. Three out of five attributes concerned risk, two of them in percentages and the third in words. The third study concerned care for hip-and knee osteoarthritis (HKOA) in The Netherlands (161). Respondents aged 45 years and over with or hip- or knee- osteoarthritis were recruited online, also via Dynata (N=648). In contrast to the other two studies, none of the attributes were explicitly described to respondents as being related to risks.

Nevertheless, based on the relationship between time preference and risk aversion, "waiting time in weeks" was classified as a risk attribute (162,163). The number and type of professionals involved was also classified as a risk attribute because health anxiety increases the belief that specialist referral is needed, and health anxiety was found to be driven by risk aversion (164,165). More details about the three studies are published elsewhere (159–161).

DCE design and questionnaire

In all studies, a Bayesian heterogeneous DCE design was created using Ngene ChoiceMetrics software (166) to maximise D-efficiency. Initial priors were based on literature, focus groups, and interviews with experts or members of the study population in the pre-piloting phase. Based on the results of a standard multinomial logit (MNL) model, the priors and the design were optimised once 10% of the required sample completed the questionnaire. In Study 1, the final experimental design contained 30 choice tasks that were divided into two blocks of 15 choice tasks. Each choice task had two generic alternatives ("Treatment 1" and "Treatment 2") that were characterised by a selection of attribute levels and the third alternative ("No treatment") allowed respondents to not choose any of the alternatives presented (opt-out). The design of the second study consisted of 48 unique choice tasks divided over three blocks of 16 choice tasks. Each choice task had two generic alternatives. In Study 3, the design consisted of 24 choice tasks and was divided into two blocks of 12 choice tasks. Again, each choice task had two generic alternatives. Examples of the choice tasks are given Figure A 1, Figure A 2, and Figure A 3 in Appendix 3.

To assess health-risk attitude, we used the 13-item Health-Risk Attitude Scale (HRAS-13) (147). The HRAS-13 is context-specific, and its items relate to medical treatment, the importance of health, general attitude towards risk-taking in health and care, and consideration of the future consequences of health behaviours. Advantages of using this scale are that 1) context-specific scales are found to better predict risk behaviour (148,167); 2) it overcomes the discussion about whether risk attitude is context dependent or whether a general risk attitude exists (146,147); and 3) it avoids the need to differentiate between risk-taking behaviour and risk perception (146,147). The items of the HRAS-13 were rated on a seven-point Likert scale from "completely disagree" to "completely agree". Total scores for the HRAS-13 were obtained by reverse-scoring seven of the items that are phrased negatively and then summing the scores to each item. Scores range between 13 and 91. Respondents with a lower HRAS-13 score are more health risk averse, while a higher HRAS-13 score indicates a more risk-prone attitude towards health risks. In addition, the questionnaires contained questions about health, age, gender, and education level. Health was measured using a visual analogue scale ranging from 0 to 100 in Study 1 and Study 3, while using a five-point Likert scale from "very poor" to "very good" in Study 2.

Table 4: Case study and DCE design characteristics

Characteristic	Study 1 Visser et al. (159)	Study 2 Ancillotti et al. (160)	Study 3 Arslan et al. (161)
Case studies			
Topic	Multiple sclerosis	Antibiotics	Hip-and knee osteoarthritis (HKOA)
Country	The Netherlands, France, United Kingdom	Sweden	The Netherlands
Study population	MS patients, ≥18 years old	General public, 18-65 years old	HKOA patients, ≥45 years old
Number of respondents	753	378	648
DCE design			
Number of attributes	4	5	6
Number of choice sets per block	15	16	12
Number of blocks	2	3	2
Number of alternatives	3 incl. opt-out	2	2
Risk attributes	Risk of relapse (%), Reducing disease progression (%), Risk of side effects (words and %)	Contribution to resistance (words), Risk of side effects (%), Treatment failure (%)	Waiting time in weeks (words), Professionals involved (words)
Number of latent classes in original study	2	3	4

Note: Attributes and levels were selected based on literature review, focus groups and interviews; they are presented in Table 2.

Age was measured on a continuous scale. Gender had three categories, namely "female", "male" or "other". Education level was measured according to the European Qualification Framework (168) and categorised as "low", "medium" or "high" in accordance with the Dutch Qualification Framework and Statistics Netherlands (169,170). In addition, Study 1 and Study 2 also contained questions about health literacy and numeracy. Health literacy was measured using the Communicative Health Literacy and Critical Health Literacy scale (171). The scale consists of five items on a four-point Likert scale that ranges from "never" to "always". Based on the average scoring system in De Bekker-Grob et al. and the three categories by Ancillotti et al., respondents with an average score of two or lower were categorised as "inadequate", those between two and three were categorised as "problematic", and those with an average score larger than three were deemed "sufficient" (160,172,173). The Dutch version of the scale (174) has more items than the original (Japanese) version (171) and the Swedish version (175). In addition, the Dutch version uses a four-point Likert scale rather than a five-point Likert scale. Study 1 was based on the Dutch version (and translated from Dutch to English and French); Study 2 was based on the Swedish version. For comparability between studies, health literacy was calculated using only the five items that were in the Swedish version, and responses in Study 2 were recoded so that they were also rated on a four-point Likert scale (divide each item score by 5, and multiply by 4/5). Numeracy was measured using the three-item version of the Subjective Numeracy Scale (SNS-3) (176). Based again on Ancillotti et al. and De Bekker-Grob et al. and (160,172,173), items were scored on a six-point Likert scale ranging from "not good at all/never" to "extremely good/very often" (160,172,173). Respondents with an average score below two were categorised as "inadequate", those with a score between three and four were categorised as "problematic" and those with an average score higher than five were deemed "sufficient".

Table 5: Attributes and levels

	Study 1 - MS		Study 2 - antibiotics		Study 3 - HKOA	
	Attributes	**Levels**	**Attributes**	**Levels**	**Attributes**	**Levels**
1.	Risk of relapse*	30% less risk 50% less risk 70% less risk	Contribution to resistance*	Low Medium High	Waiting time in weeks*	0 2 4
2.	Reducing disease progression*	20% less progression 40% less progression 60% less progression	Number of days treatment	3 days 7 days 14 days	Professionals involved*	GP Orthopaedist GP and orthopaedist
3.	Risk of side effects*	Very common mild (>10%) Common moderate (1-10%) Rare severe (0.1-1%)	Risk of side effects*	1% 5% 10% 20%	Price in Euros	0 45 90
4.	Mode of administration	Injections (1x per week) Injections (3x per week) Pills (1x per day) Pills (2x per day) Implant (1x per year) Implant (1x per 3 years)	Treatment failure*	80% 85% 90% 95%	Time per consult in minutes	10 15 30
5.			Costs	100 kr. 250 kr. 400 kr. 1000 kr.	Travel time in kilometres	1 7 20
6.					Equipment available	Direct Indirect

* Attributes with an asterisk implicitly or explicitly concerned risks.

Analysis of health-risk attitude and preference heterogeneity

Panel latent class models were used to analyse heterogeneity of preferences. These models account for the multiple choice sets each respondent completed (i.e., panel-structure), and they capture unobserved heterogeneity of preferences using a discrete number of classes (i.e., latent classes) (177–179). Following random utility theory, class allocation of respondent n in class c is based on choices for choice set s of each alternative j, and is given by $U_{nsj|c}$. The utility consists of an observable component v and a random component $\varepsilon_{nsj|c}$ which is formally written as follows:

$$U_{nsj|c} = V\left(X_{nsj}, \beta_c\right) + \varepsilon_{nsj|c} \quad (1)$$

Here β_c is a class-specific vector describing the preference weights of the attributes and levels X_{nsj} for respondent n for choice set s in alternative j. The exact model specification differed per study; the specification of the alternative specific constant(s), linearity of the attributes, and the number of classes were based on model fit and with consideration for class size and interpretability of the main-effects model.

To understand whether and how health-risk attitude and preference heterogeneity are related, we analysed two types of heterogeneity, namely 1) stochastic class assignment and 2) systematic preference heterogeneity. Both types of heterogeneity were included jointly to disentangle the different potential sources of preference heterogeneity. The impact of health-risk attitude on stochastic class assignment was included to analyse whether health-risk attitude could distinguish preferences between classes, i.e., whether it distinguished preferences for risk-related attributes as well as non-risk-related attributes. For matters of completeness, the class assignment model also included other variables based on their relationship with health-risk attitude and/or preference heterogeneity. The propensity of class membership φ_{nc} is specified as a linear-in-parameters function consisting of a constant term $\delta_{0|c}$ plus the variables health (139,147,157) (dichotomised based on median-split, good vs. rest), age (157,180) (continuous), gender (136,147,157,180) (male vs. female), education level (136,157) (high vs. rest), and if applicable numeracy (157,180) and health literacy (157,158) (sufficient vs. rest), thus:

$$\begin{aligned}\varphi_{nc} = \delta_{0|c} &+ \gamma_{1|c}\text{HRAS score}_n + \gamma_{2|c}\text{good health}_n + \gamma_{3|c}\text{age}_n \\ &+ \gamma_{4|c}\text{male}_n + \gamma_{5|c}\text{high education}_n + \gamma_{6|c}\text{sufficient literacy}_n \\ &+ \gamma_{7|c}\text{sufficient numeracy}_n + \omega_{nc}\end{aligned} \quad (2)$$

The stochastic term ω_{nc} is assumed to be EV Type 1 (Gumbel) independent and identically distributed across classes, yielding a polytomous MNL model for the probability of class membership (yielding a polytomous MNL model for the probability of class membership (note that the coefficient vector for one class must be set to zero):

$$\pi_{nc} = \frac{\exp\left(\overline{\varphi_{nc}}\right)}{\sum_{c\prime=1}^{C}\exp\left(\overline{\varphi_{nc\prime}}\right)} \quad (3)$$

Statistically significant γ coefficients (as indicated by $p<0.05$) indicate that a certain variable contributed to the class assignment model. For example, a positive and statistically significant γ coefficient of HRAS score in class 1, would mean that respondents with higher HRAS scores are more likely to be allocated to class 1 than the reference class. Nevertheless, a nonsignificant coefficient means that differences in HRAS scores do not explain differences in overall preference structures between the classes.

In parallel, we assessed the relationship between health-risk attitude and systematic preference heterogeneity by interacting the risk-related attributes with respondents' health-risk attitude. A statistically significant HRAS interaction term (again as indicated by $p<0.05$) with a risk-related attribute, for example, in class 1, is interpreted as health-risk attitude explaining preference heterogeneity of that attribute within that class.

To assess the impact of including health-risk attitude, in each study we compared log-likelihood of the model that included health-risk attitude in the class allocation model and used interactions with a model that did not do either but was equal in all other aspects. Log-likelihood statistics were compared using likelihood ratio tests, as the number of classes is equal between models. All analyses were performed in Nlogit 6.

RESULTS

Respondents

Given the varying study contexts, inclusion criteria, and study designs, the three studies had different types of respondents (see Table 4 for an overview of the case studies). The studies consisted of 753, 378 and 648 respondents, respectively. In Study 2, the general public sample, HRAS-13 scores were generally higher (more positive attitude towards health risks) and more dispersed than in the MS sample (Study 1), and the hip-and knee osteoarthritis sample (Study 3). In Study 1, respondents were less healthy (mean=60.6) than in Study 3 (mean=68.8); they were younger (mean=42.0), mostly female (67.9%) and highly educated (47.3%).

Furthermore, the sample of Study 1 was less literate and slightly less numerate than in Study 2. In the second study, most people had a good (43.1%) or very good (15.3%) health. The sample of Study 2 was slightly older than in the first study, but younger than in the third. As in Study 1, most respondents were highly educated (51.3%). In Study 3, respondents were oldest (mean=61.7), 55.4% were female, and fewer (25.3%) were highly educated than in the other studies. No data was collected on health literacy and numeracy. An overview of these respondent characteristics can be found in Table 6, while more information about the relationship between HRAS-13 scores and other background variables is presented in Table A 2 in Appendix 4.

Table 6: Respondent characteristics per study

Characteristic	Category	Study 1 - MS N (%)	Study 2 - antibiotics N (%)	Study 3 - HKOA N (%)
N		753 (100)	378 (100)	648 (100)
HRAS-13 score, mean (SD)		44.5 (9.2)	60.2 (9.8)	49.0 (5.4)
HRAS-13 score, median		46	60	50
HRAS-13 score, range		18-70	19-86	29-65
Health, mean (SD)		60.6 (20.3)	-	68.8 (19.6)
Health, median		65	Good	73
Health, categories	Very poor	-	6 (1.6)	-
	Poor	-	38 (10.1)	-
	Neutral	-	113 (29.9)	-
	Good	-	163 (43.1)	-
	Very good	-	58 (15.3)	-
Health, categories median split	High: >median	386 (51.3)	58 (15.3)	323 (49.8)
	Low: ≤median	367 (48.7)	320 (84.7)	325 (50.2)
Age, mean (SD)		42.0 (12.1)	43.3 (13.6)	61.7 (8.9)
Gender	Female	512 (67.9)	208 (55.0)	359 (55.4)
	Male	241 (32.1)	169 (44.7)	289 (44.6)
	Other	0 (0.0)	1 (0.3)	0 (0.0)
Education level	Low	188 (25.0)	70 (18.5)	207 (31.9)
	Medium	201 (26.7)	108 (28.6)	275 (42.4)
	High	356 (47.3)	194 (51.3)	164 (25.3)
	Other	8 (1.1)	6 (1.6)	2 (0.3)

Table 6: *Continued.*

Characteristic	Category	Study 1 - MS N (%)	Study 2 - antibiotics N (%)	Study 3 - HKOA N (%)
Health literacy	Inadequate	96 (12.7)	8 (2.1)	-
	Problematic	497 (66.0)	117 (31.0)	-
	Sufficient	160 (21.2)	253 (66.9)	-
Numeracy	Inadequate	51 (6.8)	23 (6.1)	-
	Problematic	331 (44.0)	154 (40.7)	-
	Sufficient	371 (49.3)	201 (53.2)	-

Health-risk attitude and preference heterogeneity

An overview of the results per study are presented in Table 7 and Table 8 described below; for further information, the full results are presented in Appendix 5 (see Table A 3, Table A 4, and Table A 5). In none of the studies were HRAS-13 scores statistically significantly related to stochastic classification of preferences. This indicates that parameters in the utility function were not jointly dependent on health-risk attitude for any of the classes in any of the studies. Nevertheless, numeracy was related to class allocation (p=0.020) in Study 1. In Study 2, age (p=0.004) and health literacy contributed to class allocation (p=0.012) in class 1 and 2 respectively. In Study 3, age explained class allocation in two classes (p=0.004 and p=0.040).

In contrast, systematic heterogeneity as measured by interactions between health-risk attitude and risk attributes was present in some risk attributes of the studies. In Study 1, the MS patient sample with three risk attributes phrased using percentages, we found systematic preference heterogeneity for all risk attributes in the first and largest class. In this class, health-risk attitude significantly moderated the effect of reducing the risk of relapse and reducing disease progression (p<0.001 for both) and the risk of rare severe side effects (p=0.020). In the second class, only the interaction between health-risk attitude and reducing risk of relapse (p=0.003) was significant. In addition, the second study, concerning the antibiotics context with a general public sample, had three classes and three risk attributes. Health-risk attitude explained part of the heterogeneity for treatment failure rate (p=0.019) in one of the classes but not in the other two. The interaction effects with the other risk attributes, however, were not significant in any of the classes. In the third study about patients' preferences for hip-and knee osteoarthritis treatment, two attributes implicitly concerned risk. In two of the four classes, health-risk attitude explained heterogeneity of preferences for waiting time (p= 0.001 for both) but not for professionals involved. As shown in Table 8, inclusion of HRAS-13 scores significantly improved the model fit only in Study 3 ($\chi 2$=37.9, df=15, p<0.001). In the other studies, the improvement was not statistically significant.

Table 7: Overview results per study

Type of heterogeneity	Class	Study 1 - MS			Study 2 - antibiotics			Study 3 - HKOA		
			Coeff.	P-value		Coeff.	P-value		Coeff.	P-value
Stochastic class allocation	1	HRAS	-0.002	0.839	HRAS	0.018	0.316	HRAS	0.007	0.794
	2	HRAS	0.000	-	HRAS	-0.002	0.924	HRAS	-0.001	0.980
	3	-	-	-	HRAS	0.000	-	HRAS	0.036	0.301
	4	-	-	-	-	-	-	HRAS	0.000	-
Systematic heterogeneity	1	Risk relapse (%)	-0.009	<0.001	Resistance (med)	0.012	0.262	Waiting time	-0.023	0.001
		Progression (%)	-0.017	<0.001	Resistance (high)	0.004	0.794	Orthopaedist	0.026	0.327
		Side effects (mod.)	-0.002	0.391	Side effects (5%)	0.001	0.949	GP & ortho.	-0.002	0.953
		Side effects (sev.)	-0.002	0.002	Side effects (10%)	0.000	0.989			
					Side effects (20%)	-0.001	0.934			
					Treatment fail (%)	-0.004	0.577			
	2	Risk relapse (%)	0.012	0.003	Resistance (med)	-0.010	0.320	Waiting time	-0.004	0.441
		Progression (%)	-0.007	0.026	Resistance (high)	-0.005	0.578	Orthopaedist	0.031	0.167
		Side effects (mod.)	0.002	0.772	Side effects (5%)	0.002	0.834	GP & ortho.	0.026	0.198
		Side effects (sev.)	0.003	0.318	Side effects (10%)	0.012	0.285			
					Side effects (20%)	0.014	0.197			
					Treatment fail (%)	-0.016	0.019			

Table 7: *Continued.*

Type of heterogeneity	Class	Study 1 - MS			Study 2 - antibiotics			Study 3 - HKOA		
			Coeff.	P-value		Coeff.	P-value		Coeff.	P-value
	3	-	-	-	Resistance (med)	0.007	0.466	Waiting time	0.036	0.001
					Resistance (high)	-0.004	0.733	Orthopaedist	-0.041	0.317
					Side effects (5%)	-0.001	0.949	GP & ortho.	0.023	0.616
					Side effects (10%)	-0.005	0.655			
					Side effects (20%)	-0.023	0.057			
					Treatment fail (%)	-0.001	0.847			
	4	-	-	-	-	-		Waiting time	0.008	0.053
								Orthopaedist	0.025	0.913
								GP & ortho.	0.027	0.056

Table 8: Model fit per study

Statistic		Study 1 - MS	Study 2 - antibiotics	Study 3 - HKOA
Log-likelihood (-)	Excl. HRAS	9389.98	3009.38	4215.38
	Incl. HRAS	9383.08	2999.73	4196.44
Number of parameters	Excl. HRAS	29	41	47
	Incl. HRAS	38	61	62
Likelihood ratio test	χ2	13.8	19.3	37.9
	Df	9	20	15
	p-value	0.130	0.502	<0.001

DISCUSSION

Hence, where does health-risk attitude come into the equation when researching preference variation? Our study did not find evidence that health-risk attitude as measured by the HRAS-13 distinguishes people between classes. Nevertheless, we did find evidence that the HRAS-13 can distinguish people's preferences for some risk attributes within classes. This association between health-risk attitude and preference heterogeneity was stronger in the case studies where respondents were sampled from a patient population than in the case study that used a general public sample. Respondents in the patient samples were also more health-risk averse than members of the general public. In the first case study, which used a patient sample, health risk attitude explained the heterogeneity of preferences for most attributes in both classes, but it did not significantly improve the model fit. In the third study, which also used a patient sample, health-risk attitude was related to heterogeneity of preferences for one attribute in two out of four health preference classes. Although the two risk attributes of this study only implicitly concerned risk, it was the only study in which the model fit statistically significantly improved by incorporating health-risk attitude.

Furthermore, we found that numeracy, health literacy, and age impacted stochastic class allocation, meaning that these characteristics could distinguish preferences between classes for risk-related attributes as well as non-risk-related attributes. In the study where numeracy impacted class allocation, all risk attributes were phrased using percentages. In the study where health literacy impacted class allocation, one of the risk attributes was described in words. Moreover, numeracy and literacy were among the characteristics that improved external validity when accounted for in preference heterogeneity in De Bekker-Grob et al. (157), and among the psychological constructs with the strongest consensus to be included in preference studies in the review of Russo et al. (143). Our results suggest that

risks are in some way related to preference heterogeneity, either directly when health-risk attitude distinguishes people's preferences within classes or indirectly when people have varying levels of numeracy and literacy.

The relevance of these results is threefold. Firstly, the impact of health-risk attitude on preferences should be explored beyond class membership by interacting the health-risk attitude with the risk-related attributes. This is expected to be mostly relevant in contexts where alternatives largely vary in terms of benefit-risk, when treatment outcomes are highly uncertain, or when patients are risk-averse. In those contexts, accounting for health-risk attitude has the potential to improve model fit and model interpretations. Secondly, the impact of health-risk attitude on preferences should be explored using a different measure than the HRAS-13. Given that we did not find strong evidence for this using the HRAS-13, which is a health-specific instrument of which the items cover a broad range of health domains (147), an option would be to use a more targeted measure of health-risk attitude in DCEs. In addition, one could research the relationship between health preference heterogeneity and measures that use a narrower definition of risk attitude (e.g., the standard gamble method (181–183) or the Balloon Analogue Risk Task (BART) (184)). Such studies can confirm if indeed health-risk attitude is not linked to preferences as strongly as anticipated (143,144), or whether it could be explained by the relatively low levels of variance in the HARS-13 scores in the case studies. As outlined in the methods section, we do recommend sticking to a health-specific measure of risk attitude. Thirdly, as numeracy and health literacy were found to impact stochastic class allocation, our results add to existing literature that stresses the importance of the communication of risks (i.e., presentation, framing, training materials and analysis) in DCEs (e.g., Harrison et al. (151), Veldwijk et al. (158) and Peters et al. (185)). In this study, we analysed a wide range of risk attributes. Although we did not observe clear differences in the relationship between health-risk attitude and preference heterogeneity based on the type or phrasing of the risk attribute, we find that numeracy explained heterogeneity in the study in which risks were presented using percentages, while literacy explained heterogeneity in the study where some risk attributes were phrased using percentages and some using words.

A strength of this study is that it is among the first to research health-risk attitude as an individual characteristic underlying heterogeneity in health preferences, and thereby responds to the call for this type of research (143,144,151). The case studies provide a cross-European comparison in three different health contexts with varying degrees of risk and study population leading to an increased generalisability of the results. Nevertheless, the differences in samples also make it harder to identify the source of similarities and differences in results between the studies. Given that secondary data was used for the current study, comparability across the studies is limited. In future research, it would be interesting to set up studies with the specific aim to compare the impact of health-risk attitude across different populations and risk attributes. It should also be noted that it is unclear whether respondents' level

of perceived riskiness of the attributes is in line with what was determined by the researchers. As risk perception and risk behaviour are not always aligned (146), we recommend future research in this area to also elicit respondents' risk perception at an early stage of DCE development. Furthermore, this research focussed on improving model fit and model interpretations from the perspective of internal validity. Given the mixed evidence regarding the predictive ability of questionnaire-based measures of risk-attitude (148,149,167), it would be interesting to also study if and how health-risk attitude and heterogeneity of health preferences are related from the perspective of external validity and individual-level prediction accuracy, for example as in De Bekker-Grob et al. (172).

In conclusion, our study did not find evidence that health-risk attitude as measured by the HRAS-13 distinguishes people between classes. Nevertheless, we did find evidence that the HRAS-13 can distinguish people's preferences for risk attributes within classes. This phenomenon was more pronounced in the patient samples than in the general population sample. Furthermore, we found that preference heterogeneity is impacted by numeracy and health literacy. These results warrant the relevance of further research into preference heterogeneity beyond class membership, a different measure of health-risk attitude, and the communication of risks.

INTERNAL VALIDITY

The extent to which the observed results represent the truth within the context of a particular study

☐ A

☒ B

☐ C

CHAPTER 4

Elicitation methods

Based on
Huls, S.P.I., Donkers, A.C.D., Ride, J., & Lancsar, E. (2022). Two for the price of one: If moving beyond traditional single-best discrete choice experiments, should we use best-worst, best-best or ranking for preference elicitation? Health Economics.

In press

HIGHLIGHTS

- We compare preferences elicited using best-worst, best-best or ranking DCE using six criteria.
- Preferences differ depending on the elicitation method used.
- Preferences elicited from first choices differ from second choices, especially in best-worst.
- Our results suggest using single-best DCE rather than best-worst, best-best or ranking.
- If moving beyond a single-best DCE, best-best and ranking are preferred.

ABSTRACT

Introduction
This study undertook a head-to-head comparison of best-worst, best-best and ranking discrete choice experiments (DCEs) to help decide which method to use if moving beyond traditional single-best DCEs.

Methods
Respondents were randomised to one of three preference elicitation methods. Rank-ordered (exploded) mixed logit models and respondent-reported data were used to compare methods and first and second choices.

Results
First choices differed from second choices and preferences differed between elicitation methods, even beyond scale and scale dynamics. First choices of best-worst had good choice consistency, scale dynamics and statistical efficiency, but this method's second choices performed worst. Ranking performed best on respondent-reported difficulty and preference; best-best's second choices on statistical efficiency.

Discussion
All three preference elicitation methods improve efficiency of data collection relative to using first choices only. However, differences in preferences between first and second choices challenge moving beyond single-best DCE. If nevertheless doing so, best-best and ranking are preferred over best-worst DCE.

INTRODUCTION

Discrete Choice Experiments (DCEs) are widely used to answer a range of research questions relating to health care preferences (22,186). More recently academic interest in the way in which respondents process information and in which DCE questions are asked has been increasing (187–190). In DCEs it is standard practice to elicit preferences by repeatedly presenting respondents with a range of alternatives presented in choice sets. In each choice set, respondents evaluate alternatives that differ in their attribute levels. These traditional DCEs only ask respondents to choose the best alternative per set. In addition to asking respondents their single-best alternative, there is an increasing interest in obtaining more preference information per set by also asking respondents to choose among the remaining alternatives, which can lead to elicitation of a full preference order. These additional observations per choice set offer three potential benefits. Firstly, a smaller sample size is required to obtain similar efficiency, i.e., similar standard errors, which might be useful if there are budget constraints, or if the research population is small (191–194). Otherwise, one can choose to work with the same sample size, but present fewer choice sets to each respondent to decrease respondent burden and still have similar efficiency. Lastly, it facilitates more advanced analyses like the estimation of individual-level models (195–197). Especially in health, these potential benefits have encouraged researchers to move beyond single-best DCE.

Full preference orders that go beyond single-best DCE can be elicited by 1) ranking all alternatives simultaneously, 2) repeatedly choosing the best and the worst alternatives, or 3) repeatedly choosing the best alternatives. Ranking has been used in discrete choice experiments for many years (198–200). It provides a full preference order, and the sequence in which alternatives are ranked can be specified by the researcher or left up to the respondent. However, it has been argued to be burdensome and the stability of the results have been found to differ per rank (191,201–203). Best-worst DCE is another method to elicit partial or full preference orders that has been proposed to overcome cognitive burden (204). In best-worst DCE, respondents are asked to choose the best and worst alternative from a set of at least three alternatives[7]. This can be sequential (i.e., first choosing the best alternative, then worst and repeat this for the remaining alternatives until all alternatives are eliminated) or simultaneous (i.e., choosing the best and worst alternative at the same time and repeat this until a full preference order is reached). Regardless of the way in which best and worst choices are elicited, there are studies that question whether positive (i.e.,best) choices and negative (i.e.,worst) choices have the same underlying mental processes (207–210). As outlined in Dyachenko, in addition to the decision sequence affecting preferences if best and worst choices

7 Best-worst scaling was first introduced by Finn and Louviere (205). There are three types of best-worst scaling that differ in the respect that respondents either evaluate objects (case 1), attributes (case 2) or alternatives (also called multi-profile or BWDCE, case 3) (192,206).

are elicited sequentially, the mere framing of best and worst choices may lead to two different mental processes (208). To overcome this, the method best-best DCE (BBDCE) was introduced relatively recently to avoid the need to swap between positive and negative mental processes (211). This third method to move beyond traditional single-best DCE is characterised by eliciting single-best choice, second-best choice from the remaining alternatives and so on until a full preference order is obtained.

If the objective is to obtain more preference information per choice set than from a traditional single-best DCE, key, to date unanswered, questions are 1) which of best-worst DCE, best-best DCE and ranking is most appropriate to do so and 2) what are potential downsides if using one of these elicitation methods?

We identified fifteen studies that empirically compared elicitation methods that move beyond single-best DCE. These can be arranged into three categories, namely studies that explored differences in choices (a) within a preference elicitation method (e.g., differences between best and worst choices in BWDCE); (b) between preference elicitation methods (e.g., differences between BWDCE and ranking); or (c) both within and between methods. None of these studies provide a head-to-head comparison of choices within and between best-worst (BWDCE), best-best (BBDCE) and ranking DCE, i.e., (c) above with all three methods. Instead, they were restricted to a comparison of (usually subsets of) best-worst, ranking, and traditional single-best DCE and only one included best-best DCE. Most were performed outside health. We provide an overview of the identified studies below and their characteristics in Table A 6 in Appendix 6 with their results discussed in more detail in the Online Supplementary Material of the published article.

Six studies were of type (a), meaning preferences were elicited using the same method but first, second and/or follow-up choices were analysed separately (191,192,201,207,210,212). These studies compared choices within a preference elicitation method, but they did not compare between different elicitation methods. Only Lancsar et al. (192) studied choices in the health domain. They generally found differences in preferences, scale, and/or efficiency between choices. Another six studies experimentally compared a subset of the methods that move beyond DCE or compared one or two of these methods to traditional single-best DCE, i.e., (b) above (213–218). These studies compared different elicitation methods but did not compare choices within a preference elicitation method. They found differences in preferences, respondent burden, predictive ability, and/or scale. Only Xie et al. studied choices in the health domain (213).

Hawkins et al., Krucien et al., and Giergiczny et al. studied differences both within and between methods, i.e., (c) above, but found conflicting results regarding similarity of preferences and none compared ranking, BBDCE and BWDCE (209,219,220). None compared all three of ranking, BBDCE and BWDCE. Krucien et al. was the only study that compared choices within and between methods in the health context (219). Hawkins et al. compared three types of elicitation methods:

best choice, worst choice, and best-worst choices (209). They found that the first choices were similar, regardless of the presence of a second choice. Also, best and worst choices were inversely related, based on which the authors conclude that they are based on the same underlying information. Krucien et al. compared the same methods as Hawkins et al. but found that preferences differed between choices, even beyond scale (210,219). Choice consistency was higher in best choices. Giergiczny et al. compared best-best and best-worst elicitation methods and single-best, best-best and best-worst model specifications (220). They found in both elicitation methods that willingness to pay estimates from models using all choices were significantly different from estimates in models using only first choices, even if correcting for scale. However, preferences and scale parameters were similar between elicitation methods if the data were modelled using the same (recoded) exploded logit specification.

While all these studies provide interesting and varied insights into the various elicitation methods and their first, second and/or follow-up choices, none provide a comparison within and between methods in best-worst, best-best and ranking DCEs. As such, the primary objective of this study was to undertake a head-to-head comparison of choices within and between methods in best-worst, best-best, and ranking DCEs to inform the decision on which preference method to use if wanting to move beyond traditional single-best DCE. We focus on the health domain, namely by eliciting preferences for Australian obesity reduction policies. To provide a practical overview of the advantages and disadvantages of all three methods, we compare them using six criteria. Four criteria are defined using choice modelling estimates, namely: 1) trade-off consistency; 2) choice consistency; 3) scale dynamics, interpreted as learning or fatigue effects; and 4) statistical efficiency. The remaining two criteria are based on respondent-reported data, namely 5) difficulty, and 6) respondent stated preference between the three preference elicitation approaches. Each criterion is defined in the methods section, along with an explanation of the study format and the details of the analyses. Results are presented per assessment criterion, followed by a discussion of the findings.

METHODS

Study format

To enable a direct comparison between best-worst DCE, best-best DCE and ranking DCE, the study consisted of three treatment arms that each represented one of the methods. Importantly, the experiments per arm were identical (same choice sets, versions etc.) except that the elicitation task, i.e., the choice questions asked, differed. Respondents were randomised to one of three arms. The questionnaire involved four parts. Firstly, respondents answered screening questions that enabled quota sampling. Secondly, respondents were presented with 16 DCE choice sets (see “Attributes, levels and data collection”). In the BWDCE and BBDCE

arms respondents faced two choices per choice set. In the first choice they were asked to choose their most preferred alternative. This chosen alternative was then removed from the choice set. Respondents were subsequently asked to choose their worst or best alternative from the remaining alternatives, respectively in best-worst and best-best. The resulting choices implied a complete preference order over the three alternatives per choice set in each of these elicitation methods. In the ranking arm, respondents were asked to provide an explicit preference order by placing a 1 under the alternative they most preferred, a 2 under the alternative they next most preferred and a 3 under the alternative they preferred least. All alternatives were visible until the respondent provided the full preference order and moved to the next choice set. Examples of the choice tasks can be found in the Online Supplementary Material of the published article. Thirdly, respondents answered questions about the DCE. In particular, they stated the perceived difficulty of the experienced elicitation method and their preference between each of the three elicitation methods: one of which they had experienced, the other two were described to them. Lastly, respondents answered socio-demographic, health-related and attitudinal questions relating to obesity.

Attributes, levels and data collection

The study concerned taxpayer preferences for healthcare policies to reduce obesity in Australia. For more information about the attributes, levels, experimental design, and the online questionnaire design, please see Lancsar et al. which explores preferences based on single-best data (221). Attributes in the study were policy type, effectiveness in reducing obesity rates, and cost in terms of higher taxes. The attributes and their levels are described in Table 9.

A total of 256 choice sets were blocked into 16 versions of 16 choice sets. As mentioned earlier, respondents were randomised to one of three arms. Within each arm respondents were randomised to one of the 16 versions of 16 choice sets. Each choice set consisted of three alternatives. Two alternatives were generic and designed to vary per choice set (Policy A and Policy B), the other alternative was a constant status quo that reflected the current situation. This current situation was described as no additional policy interventions, no change to the projected obesity rate (i.e., 32% would be obese in 2020) and no additional cost. The questionnaire was administered to a sample of taxpayers from an online panel, representative of the Australian population of taxpayers in age and gender. The sample size was chosen to allow estimation of reliable models in each arm, while considering the number of attributes, levels, respondents per version and the parameters to be estimated (196). Respondents were asked to suppose the Australian government was considering introducing new policies to reduce obesity and is interested in taxpayer preferences.

Table 9: Attributes and levels

Attribute	Levels
Policy Type	Nutritional information labelling using traffic light symbols National mass media campaign to encourage healthy lifestyle choices Ban unhealthy food and drink advertising to children Improve nutritional quality of foods sold in public institutions Funding for physical activity infrastructure and outdoor spaces Tax sugar-sweetened beverages Payment incentive for the obese to increase physical activity Pre-paid cards for healthy foods in supermarkets (reference level)
Impact on obesity rates in 2020 (effectiveness)	32% will be obese in 2020 (no change to the projected obesity rate) 31% will be obese in 2020 (moderate reduction in the projected obesity rate) 29% will be obese in 2020 (large reduction in the projected obesity rate) 28% will be obese in 2020 (very large reduction in the projected obesity rate)
Additional cost to you per year, paid as an increase in income taxes by (cost)	$12 per year ($1 per month) $120 per year ($10 per month) $240 per year ($20 per month) $480 per year ($40 per month)

Statistical analysis

The discrete choice data in this study are modelled using random utility theory. To capture all information contained in a full preference order, we used rank-ordered (also called exploded) mixed logit models (222). We first elaborate on this modelling approach, followed by a description of the six assessment criteria used to compare the preference elicitation methods. Lastly, we describe the four different stages of analysis.

Following random utility theory, each respondent i makes choices concerning the alternatives in a choice set t based on the latent utility of each alternative j, given by U_{ijt}. The utility is comprised of a systematic component $X'_{ijt}\beta_i$ and a random component ε_{ijt}:

$$U_{ijt} = X'_{ijt}\beta_i + \varepsilon_{ijt} \tag{4}$$

Here β_i is an individual-specific vector describing the preference weights of the attributes an d X'_{ijt} represents the attribute levels for individual i, alternative j in choice set t. Following Swait & Louviere (40), one can extend (1) to allow for variation in choice consistency by including a scale parameter λ. This scale parameter can differ per elicitation method and/or choice and scales the systematic component of utility[8]:

$$U_{ijt} = \lambda_s \cdot \left(X'_{ijt}\beta_i\right) + \varepsilon_{ijt} \tag{5}$$

To identify the scale effects, we impose $\lambda_1 = 1$ for the reference scenario, where a scenario refers to the combination of the elicitation method and the choice within that method. If mean preference weights over all individuals in the sample (β) are stable across elicitation methods and/or choices, beyond scale, then one would obtain mean preference weights specific to a scenario s: $\beta_s = \lambda_s\beta_1$ (223). However, when not only the relative importance of the error term is different, but also the trade-offs as characterized by the mean preference weights β, then this approach can no longer be used to capture scale differences. Instead, we build on the ideas of the McKelvey-Zavoina pseudo R-squared (224), given by:

$$R^2 = \frac{Var(X'\beta)}{Var(X'\beta) + \sigma_\varepsilon^2} \tag{6}$$

This captures the relative importance of the systematic part in choice models using the variance of $X'\beta$. Using the variation in the systematic utility component within choice sets, we generalise the definition of the utility scale λ as follows:

$$\lambda_s = \sqrt{\frac{\sum_{t=1}^{S}\sum_{j\neq k}\left(\left(X_{t,j} - X_{t,k}\right)\beta_s\right)^2}{\sum_{t=1}^{S}\sum_{j\neq k}\left(\left(X_{t,j} - X_{t,k}\right)\beta_1\right)^2}} \tag{7}$$

Here β_s and β_1 represent the mean preference weights in the elicitation method and/or choice under evaluation and the reference scenario, respectively.

8 As outlined in Bradley & Daly (202), one can also set the scale parameter to be related to the random component of utility: $U_{ijt} = X_{ijt}\beta_i + \varepsilon_{ijt} \cdot \lambda_s$. The position of the scale parameter does not change the results, only the interpretation.

Furthermore, $X_{t,j}$ and $X_{t,k}$ represent the vectors of attributes of alternatives j and k in choice set t, respectively. This definition relies on the notion that choices are more consistent when the differences between the systematic utility components of the alternatives in the choice set become larger. Note that this indeed generalises the usual definition of the utility scale when $\beta_s = \lambda_s \beta_1$.

To also capture variation in choice consistency over the course of the choice sets, the scale parameter λ_s is defined using a combination of λ_s, the choice consistency of the first choice set (i.e., $t = 1$), and δ_{st}, the linearly specified[9] scale dynamics over the course of the choice sets, relative to $t = 1$ (i.e.,from task $t = 2$ to $t = 16$). To ensure a positive value for λ_s, a log-linear specification is used:

$$\log(\lambda_{st}) = \gamma_s + (t-1) \cdot \delta_{st} \tag{8}$$

Here, for identification purposes, γ_s is set to zero if β can differ between elicitation methods and/or choices.

By moving beyond single-best DCE, more information is elicited from one respondent or one choice set. To account for this multitude of choices per choice set, we used a rank-ordered logit (also called exploded logit) model as first introduced by Punj and Staelin (200). Furthermore, to also capture heterogeneity between individuals, and hence to estimate the individual-specific preference weights β_1 as specified in equation (4), rank-ordered mixed logit models were used (222). Note that as explained in study format, in BWDCE and BBDCE respondents faced two sequential choices per choice set. In ranking, respondents were asked to provide an explicit one-off preference order. As each choice set had three alternatives, all three methods resulted in a complete preference ordering. This allows analysis using the same exploded logit version of the likelihood. By keeping everything constant across elicitation methods, including the analyses, observed differences between elicitation methods and/or choices can be directly attributed to the way in which the preferences were elicited.

Criteria

Six assessment criteria are used to compare the methods best-worst, best-best and ranking. The first four criteria are based on the choice modelling approach described above; they are 1) trade-off consistency, 2) choice consistency, 3) scale dynamics and 4) and statistical efficiency. The last two criteria are based on respondent-reported variables: 5) difficulty, and 6) preference. Now we provide further explanation of the

9 This linear specification of scale dynamics is chosen to reduce the number of parameters as no clear pattern could be observed from the data using a fully flexible specification. A quadratic specification led to a slight improvement of log-likelihood of a model with all betas and scale pooled but given the size of the improvement a linear specification was preferred for ease of interpretation.

criteria; a summary of the criteria and their operationalisation is provided in Table 10. Choice modelling analyses were performed in Julia version 1.6.1. Other statistical analyses were performed in R version 4.1.2.

Trade-off consistency

We assess trade-off consistency by comparing the marginal rates of substitution (MRS) of attributes between elicitation methods and/or choices. This MRS between two attributes X_k and X_l can be computed as:

$$MRS = \frac{\partial U / \partial X_k}{\partial U / \partial X_l} \tag{9}$$

Note that the MRS simplifies to the ratio of the preference weights β_k and β_l if the utility function is specified to be linear in parameters. In this study, the ratio of the mean preference weights is studied and standard errors of the MRS are approximated using the parametric bootstrap (225,226). This is done by randomly sampling for each MRS from the corresponding multivariate normal distribution with mean preference weights and covariance matrices of the random parameters β_k and β_l and calculating the ratio using 10,000 replications. We focus on the MRS between cost and effectiveness for its ease of interpretation, other MRS are also reported. From a methodological perspective, there is no preferred value of the MRS. However, similar values between elicitation methods and/or choices are preferred, as these indicate that preferences do not vary between elicitation methods or choices.

Furthermore, in line with Swait and Louviere consistency in preferences across elicitation methods and/or choices is assessed by testing whether parameters in the various elicitation methods and/or choices are similar (i.e., can be pooled) while allowing for differences in scale per elicitation method, per choice or both (40). Poolability of parameters across elicitation methods and/or choices is formally tested using likelihood-ratio tests. Again, similarity of estimates and hence poolablity of preference weights is preferred.

Choice consistency

The second criterion, choice consistency, assesses whether differences between the systematic utility components of the alternatives in the choice set are larger, relative to the scale of the error term, resulting in higher choice consistency. Higher choice consistency means lower variability in choices. Firstly, if β_s is estimated separately for the elicitation methods and/or choices, we capture choice consistency using λ_s, the scale of the systematic utility component defined in equation (7). A value below one indicates that choices are less consistent than the reference scenario, a

value above one indicates choices are more consistent than the reference scenario. Higher choice consistency is preferred. Secondly, if β_s is restricted to be the same (up to scale) for the elicitation methods and/or choices (i.e., aggregate-level analysis), we capture choice consistency using γ_s, the choice consistency of the first choice set as defined in the log-linear specification in equation (7). Again, higher choice consistency is preferred. Values of γ_s below zero, i.e., log(1), indicate lower choice consistency, values above zero indicate higher choice consistency. It should be noted that this measure only indicates how (in)consistent choices are for a given number of alternatives, and hence cannot be used to make a comparison between choice sets with different numbers of alternatives.

Scale dynamics

Thirdly, scale dynamics indicates choice consistency over the course of the choice sets. Increasing unexplained variance and hence lower choice consistency over the choice sets could indicate fatigue (202). Vice versa, if choice consistency is increasing, there could be learning effects. In this study, scale dynamics is quantified with δ_{st}, defined in equation (8), which is related to the scale of the systematic component of utility[8]. A negative value of δ_{st} indicates that scale is decreasing, and that unexplained variance is increasing over the choice sets, and hence we interpret it as fatigue. Similarly, we interpret positive values as evidence of learning. We consider values close to zero to be preferable. Learning and fatigue both indicate that choices might be too difficult at the beginning or end of the series of choice sets, respectively.

Statistical efficiency

The last choice modelling criterion is statistical efficiency of the parameter estimates. As mentioned, if more information is elicited from one respondent or one choice set, a smaller sample size or fewer choice sets per respondent is required to obtain similar efficiency (i.e., similar precision of estimates). Our measure of efficiency uses the standard error of the MRS as in equation (8). We use as measure of efficiency of a method the obtained standard error of the MRS, corrected for the impact of sample size, as this is informative about the information content and hence efficiency of a single respondent. Specifically, our efficiency measure is defined as:

$$Efficiency = \frac{1}{\sqrt{N_s} \cdot \sigma_s} \tag{10}$$

Here N_s is the sample size used in the elicitation method and/or choice s under evaluation and σ_s is the standard error of the estimated MRS on the sample of N respondents in that particular elicitation method and/or choice. This measure is

tightly connected to the efficiency analysis and sample size calculations in Rose & Bliemer, with the required sample size for a certain precision of the MRS being proportional to this efficiency measure (227) .

Difficulty

The first criterion based on respondent-reported data is difficulty. All respondents were asked to report the difficulty of the (sub) choices they experienced by picking which description best reflects their experience: easy, difficult, or neither. These results are compared between elicitation methods and first and second choices using chi-squared tests. The exact phrasing of the question is reported in Table A 7 in Appendix 7.

Preference

The last criterion, preference, is also measured using respondent-reported data. To assess this, respondents in each arm (i.e., in each method) were asked whether they would have preferred another (described) method than the one they experienced. Results are compared descriptively between elicitation methods. The phrasing of the questions is reported in Table A 7 in Appendix 7.

Four-stage analysis

To evaluate the preference elicitation methods, the criteria are compared in four stages. Firstly, trade-off consistency, choice consistency, scale dynamics and efficiency are evaluated on a disaggregate level, meaning that parameters are separately estimated for first and second choices in all three elicitation methods. In the second stage of analysis, first and second choices are aggregated within the elicitation method. These aggregate-level analyses allow a direct comparison between the three methods best-worst, best-best and ranking in terms of trade-off consistency, choice consistency, scale dynamics and efficiency. Thirdly, to assess first and second choices regardless of the elicitation method, parameters are estimated by aggregating per choice across the three methods. This allows comparison between first and second choices in terms of trade-off consistency, choice consistency, scale dynamics and efficiency. Lastly, descriptive analyses are performed to assess the criteria difficulty and preference. A summary of the assessment criteria can be found in Table 10.

Table 10: Summary assessment criteria

Criterion	Operationalisation	Values and interpretation	Preferred direction
Trade-off consistency	Marginal rate of substitution (MRS); poolability	0 to infinity	No preferred direction for MRS value; similarity and poolability are preferred
Choice consistency	Relative scale differences in systematic utility (λ)	Compared to ref. scenario: 0-1: choices less consistent 1: choices equally consistent >1: choices more consistent	Higher values are preferred
	Log of the scale of the first choice set (γ)	Compared to ref. scenario: <0: choices less consistent 0: choices equally consistent >0: choices more consistent	Higher values (less negative) are preferred
Scale dynamics	Change in error scales over choice sets (δ)	<0: fatigue 0: neutral >0: learning	Close to zero is preferred over positive or negative values
Statistical efficiency	Precision of MRS $\left(\frac{1}{\sqrt{N}\cdot\sigma}\right)$	0 to infinity	Higher values are preferred
Difficulty	Respondent-reported (easy, neutral, difficult)	0-100% easy 0-100% neutral 0-100% difficult	Higher values are preferred if easy, lower values are preferred if difficult
Preference	Respondent-reported (yes, no)	0-50%: experienced method is preferred 50-100%: described method is preferred	Lower values are preferred if 0-50%, higher values are preferred if 50-100%

RESULTS

No statistically significant differences between arms in age, gender, employment status, general health, education and questionnaire response time were identified using Bonferroni adjusted tests. Respondent characteristics per arm can be found in Table A 8 in Appendix 8.

Table 11 provides a concise overview of the results across methods and across criteria. The tick marks indicate which elicitation method or choice performed best on a particular criterion, multiple tick marks per criterion implies that methods or choices performed equally well, a hyphen indicates that we could not draw a conclusion based on the available data. We further elaborate on these findings below and in the discussion.

Before continuing to the results, we note the following. In the disaggregate analyses, the parameters were estimated for every elicitation method and choice separately. Preferences were not the same across the six choices, not even up to scale implying strictly speaking one should not aggregate preferences within or between methods. However, following the aim of this study to help researchers decide which method to use if moving beyond traditional single-best DCEs, we do present how well the methods performed on every criterion if one would pool choices within a method (i.e., aggregate per elicitation method). Additionally, we also want to inform the decision on whether to even move beyond single-best DCE by presenting how well first choices performed compared to second choices (i.e., aggregate per choice). The remainder of this section describes the results per stage of analysis and per criterion. For brevity, only a subset of the estimation results is presented; the full set of estimation results are found in Table OSM 1 to Table OSM 6 in the Online Supplementary Material of the published article.

Disaggregate analyses per method and choice

Table 12 presents the results of the model on a disaggregate level, meaning that the model parameters were estimated for every elicitation method and choice separately.

Trade-off consistency

In the first choices of the three methods, the MRS between cost and effectiveness was lower ($5.44-$5.77 per percent increased effectiveness) than in the second choices ($6.40/%-$7.71/%). A similar pattern was found in the MRS for the majority of the policies, as shown in Table OSM 2 in the Online Supplementary Material of the published article. Furthermore, based on the likelihood ratio (LR) tests for poolability, preferences were not the same across all six sub choices (test-statistic= 2262.8, df=175, p<0.001), not even up to scale (test-statistic= 1827.8, df= 170, p<0.001.

Criterion	Disaggr. per method and choice						Aggr. per elicitation method			Aggr. per choice	
	BW1	BW2	BB1	BB2	R1	R2	BW	BB	Rank	1st	2nd
Trade-off consistency	-	-	-	-	-	-	-	-	-	-	-
Choice consistency	√						√	√		√	
Scale dynamics	√		√	√		√			√	√	
Statistical efficiency	√			√	√			√		√	
Difficulty					√				√	√	
Preference					√	√			√	-	-

Table 11: Summary results per criterion and per stage of analysis

Choice consistency

Choice consistency, quantified with λ as specified in equation (6), was below one in all choices in all elicitation methods, indicating lower choice consistency than the reference scenario (the first choice of best-worst). In best-best and ranking, first and second choices within a method were relatively similar. Overall, responses were most consistent in the first choices of best-worst and least consistent in the second choices of best-worst.

Scale dynamics

Besides the second choices of ranking, scale dynamics (i.e., scale over the course of the choice sets) were positive in all choices and elicitation methods, but statistically significantly different to zero only for the second best-worst choice and first ranking choice. This indicates that in these cases, unexplained variance was decreasing over the choice sets, which could be interpreted as learning.

Statistical efficiency

Statistical efficiency as specified in equation (10), connected with the inverse of the standard errors of the MRS between cost and effectiveness, was higher in the first choices as compared to the second choices. Since the first choice is a choice out of all three, and the second choice is a choice out of two statistical efficiency is expected to be higher in the first than the second choices (228). In the first choices efficiency was highest in best-worst, in the second choices it was highest in ranking. However, most efficiency parameters for the different policies, as shown in Table OSM 2, show that efficiency of the first choices was highest in ranking. For the second choices, it was generally highest in best-best and lowest in best-worst.

Aggregate analyses per elicitation method

A summary of the results in which preferences were estimated for every elicitation method separately, but preferences of the choices were aggregated, i.e., assumed identical, within the method can be found in Table 13. Choice consistency and scale dynamics were allowed to be different for first and second choices, as pooling tests showed that these parameters could not be aggregated[10].

10 Results not reported, available on request.

Table 12: Disaggregate results

	BW1		BW2		BB1		BB2		Rank1		Rank2	
Variables	**Est.**	**SE**	**Est.**	**SE**	**Est.**	**SE**	**Est.**	**SE**	**Est.**	**SE**	**Est.**	**SE**
MRS $(\frac{\partial U/\partial eff}{\partial U/\partial X^{cost}})$	\$5.77/%	0.469	\$7.71/%	0.874	\$5.72/%	0.593	\$6.69/%	0.743	\$5.44/%	0.484	\$6.40/%	0.579
Choice consistency (λ)	REF		0.393		0.730		0.731		0.691		0.610	
Scale dynamics (δ)	0.001	0.005	0.033	0.006	0.008	0.005	0.004	0.006	0.010	0.004	-0.001	0.006
Efficiency $(\frac{1}{\sqrt{N}\cdot\sigma})$	0.067		0.036		0.055		0.044		0.065		0.054	
N parameters	2*36				2*36				2*36			
N respondents	997				926				1012			
Log-likelihood	-20552.3				-17928.9				-19778.7			

Trade-off consistency

The MRS between cost and effectiveness was \$6.10 per percent increase in effectiveness in best-worst, \$6.64/% in best-best and \$6.43/% in ranking. For the policies, the MRS was generally lowest for best-best and highest for best-worst, as shown in Table OSM 4 the Online Supplementary Material of the published article. However, for all methods the model that pooled choices within the method was rejected in favour of the model with separate first and second choices (likelihood ratio test statistics 860.4, 508.0, and 187.0 for BW, BB and ranking, respectively, all df=34, $p<0.001$). These results of the pooling test show that even if allowing differences in choice consistency and scale dynamics between choices, preferences of the first and second choices could not be pooled in any of the methods, and hence first and second choices should be analysed separately. The results below should therefore be interpreted with caution. They show how the elicitation methods would perform if disregarding the fact that preferences from first and second choices could not be pooled for any of the methods.

Choice consistency

Like preferences, choice consistency could not be pooled between the first and second choices. As such, choice consistency was measured using two different estimates for choice consistency in these analyses. Firstly, choice consistency (λ) as specified in equation (6) was below one in best-best and ranking, indicating less consistent responses in the first choices of those elicitation methods than in the reference scenario best-worst. Secondly, a separate estimate of choice consistency (γ) as specified in equation (8) was estimated for the second choices in each elicitation method, as compared to the choice consistency of that method's first choices. Here, the negative parameters indicate that second choices were less consistent than first choices. In best-best, second choices were closest to the first choices in terms of consistency.

Scale dynamics

Most scale dynamics parameters were positive, suggesting that in most elicitation methods respondents experienced learning. However, it was only significant for the second choices of best-worst and best-best. In ranking, none of the parameters were statistically different from zero suggesting the least learning over the course of the choice sets in this method.

Statistical efficiency

Statistical efficiency of the preference estimates as computed using the standard errors of the MRS between cost and effectiveness was highest for best-best, then ranking and least for best-worst. A similar pattern was observed in all policy parameters as shown in Table OSM 4 in the Online Supplementary Material of the published article.

Table 13: Aggregate results per elicitation method

Variables	BW Est.	BW SE	BB Est.	BB SE	Rank Est.	Rank SE
MRS $(\frac{\partial U/\partial eff}{\partial U/\partial X^{cost}})$	$6.10/%	0.687	$6.64/%	0.494	$6.43/%	0.511
Choice consistency (λ) 1st	REF		0.850		0.828	
Choice consistency (γ) 2nd	-1.024	0.076	-0.172	0.000	-0.247	0.060
Scale dynamics (δ) 1st	-0.004	0.005	0.006	0.005	0.009	0.005
Scale dynamics (δ) 2nd	0.033	0.006	0.014	0.007	0.005	0.006
Efficiency $(\frac{1}{\sqrt{N}\cdot\sigma})$	0.046		0.066		0.061	
N parameters	38		38		38	
N respondents	997		926		1012	
Log-likelihood	-20982.5		-18182.9		-19872.2	
Pooling test for combining two choices within method	860.4, df=34, p<0.001		508.0, df=34, p<0.001		187.0, df=34, p<0.001	

Aggregate analyses per choice

Table 14 presents a summary of the results of the model in which preferences were estimated for first and second choice separately but pooled between elicitation methods. Like before, choice consistency and scale dynamics were allowed to differ between methods, as pooling tests showed that these parameters could not be aggregated (results not reported).

Trade-off consistency

The MRS between cost and effectiveness was $5.61/% effectiveness in the first choices and $7.59/% in the second choices. As shown in Table OSM 6 in the Online Supplementary Material of the published article, the MRS was generally lower in first than in second choices. However, neither the first nor second choices could be pooled between elicitation methods (likelihood ratio test statistics 132.2, and 224.0 for first and second choices, respectively, both df=68, p<0.001). This shows that even after allowing for differences in choice consistency and scale dynamics between methods, preferences could not be pooled between methods, and hence first and second choices should be analysed separately. Based on this finding, again, the results below should be interpreted with caution.

Choice consistency

As in the aggregate analysis, we report two different types of choice consistency: choice consistency (λ) as specified in equation (6) if preferences are estimated separately per choice across methods, and choice consistency (γ) as specified in equation (7) if preferences are assessed in aggregate. Firstly, in the second choices, choice consistency (λ) was below one, indicating less consistent choices than in the reference scenario (i.e., first choices of best-worst). Secondly, separate estimates of choice consistency (γ) were estimated for the choices of the other elicitation methods, as compared to the choice consistency of best-worst. Within first choices, the positive parameter for ranking indicates that first choices in ranking were more consistent than first choices of best-worst. Within the second choices, responses in best-best and ranking were more consistent than the second choices in best-worst. Furthermore, the estimates of choice consistency (γ) in the first choices were closer to zero than in the second choices, indicating more similar choice consistency of responses in first than second choices.

Scale dynamics

Like observed when aggregating preferences over the elicitation methods, most scale dynamics parameters were positive, suggesting that most respondents experienced learning. Besides in ranking, scale parameters of first choices were closer to zero than second choices meaning that first choices are preferred. Scale dynamics were especially large in the second choice of best-worst.

Statistical efficiency

The MRS between cost and effectiveness showed highest statistical efficiency for the first choices. This also holds for all policy parameters, as shown in Table OSM 6 in the Online Supplementary Material of the published article.

Descriptive analysis

Descriptive results provided additional insights into respondent stated perceived difficulty and preference of respondents regarding which of best-worst, best-best and ranking DCEs they thought were 'best' and are presented in Table 15.

Table 14: Aggregate results per choice

Variables	1st		2nd	
	Est.	SE	Est.	SE
MRS $(\frac{\partial U/\partial eff}{\partial U/\partial X^{cost}})$	$5.61/%	0.370	$7.59/%	0.450
Choice consistency (λ) BW	REF	NA	0.510	NA
Choice consistency (γ) BB	-0.066	0.065	0.340	0.089
Choice consistency (γ) Rank	0.011	0.067	0.295	0.086
Scale dynamics (δ) BW	0.011	0.005	0.037	0.007
Scale dynamics (δ) BB	0.010	0.005	0.018	0.007
Scale dynamics (δ) Rank	0.009	0.005	0.007	0.006
Efficiency $(\frac{1}{\sqrt{N}\cdot\sigma})$	0.086		0.073	
N parameters	40		40	
N respondents	2935		2935	
Log-likelihood	-33499.4		-25113.8	
Pooling test for combining three choices between methods	132.2, df=68, p<0.001		224.0, df=68, p<0.001	

Difficulty

Difficulty was found to vary between choices and elicitation methods. The majority of respondents who were randomised to the ranking task perceived the ranking task to be easy (69%), higher than in best-worst (59% to 63% depending on first or second choice) and best-best (49% to 62%). Respondents who were randomised to the best-worst and best-best tasks both more often perceived the first choice to be easy (63% and 62% respectively) than the second choice (59% and 49% respectively). The second choice in best-best was most often reported to be difficult (14%). The differences in perceived difficulty were statistically significant for nearly all choices ($p<0.001$), no statistically significant differences were found between the first choice of best-worst and best-best ($p=0.77$).

Preference

At the end of the questionnaire, the two elicitation methods respondents did not experience were described to them and they were asked whether they would prefer each over the one they just completed (see Table A 7 in Appendix 7 for phrasing of questions). Some respondents would have preferred another method than the

one they experienced (27% to 41% depending on the experienced and described method). In the ranking arm, the lowest number of respondents would have preferred another method over the one they experienced. In the best-best arm, 38% would have preferred best-worst and 38% would have preferred ranking. In the best-worst arm, the highest number of respondents would have preferred another method than best-worst. These results suggest ranking is the most preferred method, followed by best-best, and then best-worst.

Table 15: Difficulty and preference per elicitation method – descriptive results

	Best-worst N=997		Best-best N=926		Ranking N=1012
How did you find the task of... [a]	Best	Worst	Best	2nd best	Ranking
Easy	63%	59%	62%	49%	69%
Neither	32%	32%	33%	37%	26%
Difficult	5%	9%	5%	14%	4%
% who would have preferred [b]					
Best-best	38%				36%
Best-worst			38%		27%
Ranking	41%		38%		

a: Statistically significant differences between all choices ($p<0.01$) besides difference between first choice in best-worst and best-best ($p=0.77$). See Table A 7 in Appendix 7 for exact phrasing of questions. b: Percentages don't add up to 100% because questions were not mutually exclusive. See Table A 7 in Appendix 7 for phrasing of questions.

DISCUSSION

This study undertook a head-to-head comparison of best-worst (BWDCE), best-best (BBDCE) and ranking DCEs between and within methods, to inform the decision on which preference method to use when considering moving beyond traditional single-best DCE. Eliciting additional observations per choice set can translate into needing a smaller sample size, reducing respondent burden with fewer choice sets for each respondent and/or enabling more advanced analyses. We compared the preference elicitation methods using six criteria, namely: 1) trade-off consistency, 2) choice consistency, 3) scale dynamics, also known as learning or fatigue effects, 4) statistical efficiency, 5) stated difficulty, and 6) stated preference. These were tested in a health-based study that elicited preferences for obesity reduction policies in Australia.

Most importantly, we found that preferences were not the same across the methods and their sub-choices, not even up to scale, based on poolability testing. This implies that in this study, the first to compare all three methods, we should not

aggregate preferences within or between methods. However, for some researchers the benefits of moving beyond single-best DCE weigh strongly. To inform their decision on which preference method to use if moving beyond traditional single-best DCEs, we analysed the methods' performance on every criterion by means of aggregate analyses per elicitation method. Ranking performed best in terms of scale dynamics, respondent-reported difficulty, and respondent-reported preference. Best-worst performed best in terms of choice consistency, while statistical efficiency of the preference estimates was highest for best-best.

In the disaggregate analyses, where we separately analyse first and second choices, the first choices of best-worst performed well in terms of choice consistency and scale dynamics and statistical efficiency, but this method's second choices performed worst on these criteria. First choices from ranking and best-best performed less well than those of best-worst on these criteria but their second choices performed better than those of best-worst. Of these two methods, ranking performed better on efficiency and best-best on choice consistency. The low efficiency of each individual choice in best -best is in contrast with its high performance when pooled. This might be due to the two choices being more similar – although not poolable – for best-best. The results of this study can also be used to inform the decision on whether to even move beyond single-best DCE by comparing first choices to second choices (i.e., aggregate per choice). We found that first choices outperformed second choices on all criteria, but that neither the first nor second choices could be pooled across elicitation methods.

With preferences from second choices and especially worst choices reflecting different trade-offs, combining them with first choices will make the overall results less accurate and potentially invalid. Furthermore, with preferences from first choices differing between elicitation methods, we suggest using a traditional single-best DCE would generally be most appropriate. However, if one considers moving beyond first choice, for example due to a tight budget or if the research population is small, based on our empirical results we recommend best-best or ranking. Our results provide an empirical basis for the decision to move beyond single-best DCE and choosing an elicitation method to do so, however, the actual decision will always depend on the criteria at hand.

These differences in elicitation methods and choices are in line with the empirical work that found that preferences were different, depending on method or choice (191,201,207). In particular, the large differences between best and worst choices are in line with literature that states that attributes and alternatives are weighted differently depending on whether the choice is to accept or reject (229–232) as this likely shifts a person's mindset from a promotion to a prevention focus (233,234). Furthermore, like Giergiczny et al. (220) and Krucien et al. (219), we found that first choices are different from second choices, which raises doubts regarding whether to move beyond traditional single-best preference elicitation methods. If the aim is

to measure what people want most, then this is measured well with the first choice (157,172).

This study is the first to provide a head-to-head comparison of the three elicitation methods, including the relatively new method best-best DCE that had been hypothesised to overcome some of these difficulties by using the same mindset for each decision (211). This extensive comparison between various elicitation methods has not yet been made in health nor in the literature more broadly. There are, however, some aspects of the study worth noting to allow careful interpretation of the findings. Firstly, from a methodological point of view, we note that respondents had to choose between three alternatives with two sequential choices in BWDCE and BBDCE and an explicit one-off preference order in the ranking task. As such, there were only three ranks to be studied and there was a fixed order in best-best and best-worst but not in ranking. Due to only having three alternatives in the choice set, this study design allowed us to impose the same econometric model specification on a full preference order for all elicitation methods. As such, observed differences between elicitation methods can be directly attributed to the way in which the preferences were elicited. If wanting to use more than three alternatives, the econometric model specification no longer matches the data generation process of all elicitation methods and such a comparison is no longer possible. Alternative model specifications that match the data generation process such as the sequential best worst DCE model proposed in Lancsar et al. (192) can then be used. Secondly, the findings from this sample may not generalise to other groups or respondents that complete DCEs. We collected data in a sample of taxpayers from an online panel in which the level of engagement in the choice tasks may differ from a sample of patients or clinicians. Lastly, although Hawkins et al. (209) showed that the mere existence of a second choice did not change preferences of the first, it should be noted that we did not elicit single-best only. The direct comparison between single-best DCE and best-best DCE would be a useful contribution to the health economics literature.

Other avenues for future research would be to further study the external validity of these elicitation methods, and to compare if any of these elicitation methods is more suitable for a particular type of research question or context. In some contexts, choices are framed in a positive manner and focus on the promotion of the best alternative, for example preference for a treatment. In other contexts, choices are framed negatively and focus on prevention of the worst alternative, for example when choosing to take medication to prevent a particular health state. Following this reasoning, it would also be interesting to elicit worst and second-worst preferences. Furthermore, it would be interesting to study the implications of our results for dual-response choice designs that are also argued to increase efficiency of data collection (235).

PART C

EXTERNAL VALIDITY

☐A

☐B

☒C

CHAPTER 7

General discussion

SAYING YES, DOING NO? – MAIN FINDINGS AND IMPLICATIONS

In this PhD thesis, I examined the internal and external validity of discrete choice experiments to inform healthcare decision making. The three main objectives were: 1) to provide an overview of current challenges to integrating preferences in healthcare decision making; 2) to assess the internal validity of discrete choice experiments; and 3) to assess the external validity of discrete choice experiments. In this chapter, I discuss the main findings per objective and their implications, followed by the strengths and limitations of the work presented in this thesis, several recommendations for research and policy, and the conclusions. Note that the work performed in this PhD thesis highly appreciates, builds upon, and hopefully contributes to the body of knowledge on preference elicitation methods to inform healthcare decision making.

Current challenges to integrating preferences in healthcare decision making

Integrating public and patient preferences in healthcare decision making has the potential to ensure more support for difficult decisions at the societal level, and it can increase uptake, adherence, and satisfaction with treatment at the individual level (1–6). However, how to elicit and incorporate these preferences in a systematic and scientifically valid manner is subject to debate. In Part A of this thesis (Chapter 2), challenges to integrating preferences in healthcare decision making were identified.

This chapter listed 37 challenges that were identified in a systematic review of the literature on the integration of preferences in healthcare decision making, with a focus on patient preferences and health technology assessment. Methodological challenges were most frequently mentioned in the included literature, closely followed by procedural, normative, practical, and conceptual challenges. Important methodological challenges related to internal validity, external validity, heterogeneity, and the choice of method; all of which were addressed in the following chapters of this thesis. Other methodological issues were reliability, generalisability, and sample selection. Common procedural challenges concerned how to evaluate the impact of preference studies, their degree of being evidence-based, and what weight preference studies should get in healthcare decision making, e.g., within or alongside clinical and cost-effectiveness studies.

Since the publication of this chapter, many researchers echo the important yet unclear role of preferences in healthcare decision making, not only for patient preferences in health technology assessment. Throughout the entire medical product life cycle, there are similar challenges regarding the integration of patient, public and other preferences (329,330). The main questions are when to incorporate preferences (in which stage of the medical product life cycle, and under which circumstances) (331–334,330,335–337,329,154,338–341), who to incorporate (public, patients, or others) (154,334,338,340,342), and how to incorporate preferences into healthcare decision making (using which preference elicitation methods, and

with what weight) (7,154,330,333,334,336–341,343–346)? In line with Chapter 2, many publications stressed the importance of the quality of preference studies (154,331–333,335,338,340,341,344–346), the need for further guidance on the integration of preferences in healthcare decision making (329,333,335–341,346,347), and the importance of a multi-stakeholder and holistic approach to the integration of preferences in healthcare decision making (329,330,334,339,341,347).

In conclusion, the findings of this chapter and the literature that followed imply that methodological challenges are most pressing, and that guidance and a multi-stakeholder approach are needed to integrate preferences in healthcare decision making.

The internal validity of discrete choice experiments

The validity of preference elicitation methods like DCEs is an important methodological challenge to integrating preferences in healthcare decision making. Regardless of available guidelines, there are numerous ways in which researchers can conduct a DCE, and even more ways in which respondents in a DCE can interpret, value, and use information to make decisions. This makes improving the internal validity of DCEs an ongoing quest. Part B of this thesis described the assessment of internal validity of DCEs by means of their theoretical and convergent validity.

In Chapter 3, the relationship between health-risk attitude and heterogeneity of preferences was studied through a cross-European comparison in three different health contexts. Health-risk attitude as measured in this study could not distinguish respondents between latent classes. It did explain preference heterogeneity for some risk attributes within the classes. Only in one of three case studies, model fit was statistically significantly improved by incorporating health-risk attitude. Numeracy and health literacy distinguished people between classes. In Chapter 4, theoretical and convergent validity were assessed in a head-to-head comparison of three different preference elicitation methods that move beyond traditional single-best DCE. In all three methods – best-worst DCE, best-best DCE and ranking DCE – respondents were asked to choose the alternative they preferred most (single-best) and to choose among the remaining alternatives. First choices differed from second choices in all three preference elicitation methods, and the results also differed between the three methods.

The improvement in model fit in only one of three case studies in Chapter 3 implies that theoretical contribution of modelling individual characteristics and preference heterogeneity depends on the study context. In general, incorporating individual characteristics underlying preferences can contribute to theoretical validity of DCEs in contexts where respondents or the described health intervention are very heterogenous, or when policy decisions are likely to vary for different subgroups (25,143). Given the usage of secondary data in Chapter 3, it is hard to identify the source of similarities and differences in results between the case

studies. It would be interesting to experimentally study the health-risk attitude, the presentation of risks and the relation to heterogeneity (348,156,185,151).

In Chapter 4, the differences in preferences within and between elicitation methods illustrate that respondents make different decisions depending on how the choice task is presented. DCE research has focussed on how attribute framing and the presentation of the opt-out alternative affect results (349,350,187,351–354). However, research on how the presentation of the choice tasks affects decision making is rather limited (355). A direct comparison between single-best DCE and best-best DCE can shed light on whether the mere presence of an extra choice changes preferences of the first choice. Incorporating cognitive burden of the methods in this comparison would be valuable (154,340,341).

Overarchingly, these findings imply that the internal validity of DCEs can be improved by identifying the multitude of factors that influence respondents when making decisions in particular contexts, and by trying to accommodate for these factors in the presentation of choices and data analysis of DCEs.

The external validity of discrete choice experiments

The importance of external validity of DCEs is emphasized by many researchers (22,31,238,239,241). However, the number of studies in which external validity is assessed is rather limited, especially in health. Moreover, there is no consensus on the underlying causes of the discrepancy between stated and revealed preferences. In Part C of this thesis, external validity was assessed by comparing stated and revealed preferences in two different contexts, with a focus on model complexity and social desirability bias.

The study described in Chapter 5 was conducted in the context of colorectal cancer screening and considered the effect of model complexity on internal and external validity. Choices were accurately predicted for 95% of the respondents in the holdout task of the DCE and for 90% in the actual screening choice. Most respondents chose to participate in screening in the DCE and in the actual screening choice. Model fit improved when heterogeneity was included; individual-level prediction accuracy was only marginally better. Differences between stated and revealed preferences could be attributed to respondents' health and the support of their general practitioner.

In Chapter 6, the effect of social desirability bias and cheap talk mitigation on internal and external validity was studied in the context of food choice by randomly allocating respondents to one of four questionnaire versions: default without textual manipulation, priming socially desirable behaviour, cheap talk mitigation, or both. There were little to no differences between respondent-reported characteristics, DCE results and prediction accuracy for the holdout task and actual food choices between questionnaire versions. Although differences were marginal, prediction accuracy for the holdout choice task was slightly higher in the default version where there were no textual manipulations. Contrarily, prediction accuracy of the

actual choices was slightly higher when respondents were subjected to cheap talk mitigation. However, revealed preferences were only available for a limited number of respondents.

A first implication of these results relates to the large number of respondents that opt-in when choices were hypothetical. The opt-in rate was also high for revealed preferences in Chapter 5, but not in Chapter 6. Reasons for respondents to overstate the desirability of hypothetical goods or services include the lack of consequentiality of choices, incentive compatibility, choice context, the amount of available information, and focussing illusion (240,324–327). A low-key solution to better align stated and revealed preferences for opting-in would be to make small changes to the questionnaire, for example with a budget reminder or by increasing the saliency of the opt-out alternative (298,318,356). Alternatively, results based on stated preference data can be statistically calibrated using revealed preference data (13,357). A more fundamental solution would be to pay closer attention to the sampling procedure. Collaborating with the National Institute for Public Health and the Environment in Chapter 5 allowed eliciting revealed preferences from all respondents in the sample, but sending a letter with an invitation to fill in the questionnaire induced substantial scope for selection bias. Sampling respondents using a sampling company in Chapter 6 ensured a large sample size for stated preferences, but at the cost of sample size for revealed preferences. These findings highlight the importance of sampling from a population that is representative of the target population, i.e., respondents who normally face the choice under investigation (355). Additionally, sampling respondents shortly before they face the actual choice might also aid the alignment of stated and revealed preferences.

A second implication of the findings regarding external validity in this thesis relates to model complexity. In Chapter 5, accounting for heterogeneity improved model fit. However, simultaneously accounting for scale and preference heterogeneity proved too demanding for the data. Accounting for heterogeneity generally decreases bias, provides a richer interpretation of the data and might prevent suboptimal policy advice, but increasing model complexity comes at the cost of estimation time, interpretation problems and potentially overfitting (358,359). More importantly, increased model complexity may have little effect on model performance and individual-level prediction accuracy, as also found in Chapter 3 and Chapter 5 (359). In other examples, increasing model complexity did positively impact model performance and prediction accuracy (157,172,173,260). Guidance on accounting for heterogeneity needs to be based on an informed discussion about the identification and interpretation of heterogeneity (358). At all times, model complexity should be in line with the research question, the richness of the data, and consider the effect it has on external validity.

Thirdly, the findings of Chapter 5 and 6 highlight that many factors can contribute to the discrepancy between stated and revealed preferences and thereby to the external validity of DCEs. This versatility of factors causing hypothetical

bias complicates reaching consensus about the determinants of hypothetical bias and ways to overcome it (240,241,279,321–323). By going beyond merely comparing stated and revealed preferences in particular contexts, qualitative and experimental studies into the underlying causes of hypothetical bias might ultimately advance the external validity of DCEs. Qualitative research can be used to help interpret unexpected results, in addition to its common usage to inform the selection of attributes and attribute levels, and to provide insights into clarity and comprehensiveness of the DCEs (337,338,340). Additionally, experimental research can be used to test hypotheses that are formed based on observed choices.

Overall, implications from the findings in these chapters are to carefully consider the sampling procedure, model complexity and the versatility of factors causing hypothetical bias.

DOING YES, DOING NO? – STRENGTHS AND LIMITATIONS

The main findings and implications I just discussed largely rely on the choices made in this PhD thesis. Some of these choices have strengthened the work (do's), others have been a limitation (don'ts). To put the findings and implications into perspective, I now discuss these strengths and limitations.

Firstly, I consider the variety of methodological approaches used to study the internal and external validity of one single method a strength of the research. This thesis comprised one systematic review and four quantitative studies. In two of the quantitative chapters, effects were studied within the samples, one of them using secondary data of three case studies. In the other two, the topic was experimentally studied using a between-subjects design in which respondents were randomly allocated to different questionnaire versions. Secondly, the diversity of study contexts and samples increases the generalisability of the findings. The conducted studies were about different types of health interventions, namely treatment, health screening, food choice and public policy. Respondents were recruited from patient and public samples and differed in whether they were the intended end-users of the good or service under evaluation. Thirdly, all chapters in this thesis were based on an extensive review of the literature. Chapter 2 is a systematic review. In Chapter 3, two systematic reviews by other researchers highlighted the need for researching the topic. In Chapter 4 and 5, I provide a table that summarises the relevant literature. Chapter 6 builds upon the review of the literature in Chapter 5, and two recently conducted systematic reviews by other researchers.

The work presented in this PhD thesis also comes with some limitations that need noting. Firstly, the validity of DCEs can be assessed in many more ways than the subset used in this PhD thesis. Moreover, measures of reliability such as test-retest, and version consistency and choice reliability can also provide valuable insights to improve the quality of the study beyond validity (31). Secondly, while this thesis intentionally focussed on DCEs, it led to the exclusion of other preference elicitation

methods. There are multiple other promising methods available to explore or elicit preferences that can be used to advance the integration of preferences in healthcare decision making (7,345). The choice to focus on DCEs implies limited generalisability to preference elicitation methods. Thirdly, given the quantitative nature of the method, there was limited researcher-respondent interaction and hence limited story behind the data. As addressed earlier in this discussion, the usage of qualitative data could enrich the findings from quantitative studies.
Fourthly, although study contexts and samples were diverse, respondents were mostly sampled from online panels and from the Dutch population and used cross-sectional data, again leading to limited generalisability of the results. Lastly, Chapter 6 was only distantly connected to health. As addressed in Part C of this thesis, there are few studies on the external validity of DCEs in the health domain, given the perceived complexity of obtaining revealed preference data (172,238). To ensure a large study population, Chapter 6 was conducted in the context of food choice. However, revealed preferences were only available for a limited number of respondents, potentially due to hypothetical bias. The study context beyond health implies that the results cannot be generalised to the health context. However, in my opinion, the recommendations about the sampling procedure, using different or additional mitigation methods, and considering the versatility of factors causing hypothetical bias formulated in and based on this chapter can be generalised to any context including health.

SAYING YES, DOING YES? - RECOMMENDATIONS

What do the findings, implications, and limitations that I outlined above mean for how to say yes to integrating public and patient preferences in healthcare decision making, and how to do yes?

For researchers

By conducting a DCE, researchers aim to determine the relative importance of characteristics of a health intervention, and – more importantly – how these characteristics influence decision making once people actually face the choice. Hence, in my opinion, the goal of methodological research of DCEs should predominantly be to enhance external validity of the method, either directly or indirectly via internal validity. After all, *"What animals do in labs is nothing like what they do in the wild—so what are you actually learning when you do these experiments? You're learning how animals behave in labs"* (360).
As addressed in this general discussion, external validity of DCEs can benefit from carefully considering the sampling procedure, model complexity and the versatility of factors causing hypothetical bias. Internal validity can be enhanced via the identification of factors that influence respondents when making decisions, and by accommodating for these factors in the presentation of choices and analyses of

DCEs. Specific focal points to address both internal and external validity would be to advance research into cognitive burden and heuristics that respondents use to simplify decision making. Addressing cognitive burden, also for respondents with cognitive impairments, assures validity of the data and increases the value of DCE results to stakeholders (154,340,341,361,362). Furthermore, researchers should try to understand which factors influence decision making, rather than delete respondents who make different decisions than they expected them to (363–367). Hence, when choosing how to present information and collect the data, I recommend researchers to take into account the existence of cognitive biases and heuristics and the way in which they influence decision making (240,368). Similarly, the approach to analysing the data can be adapted to respondents who use heuristics such as attribute screening and choice set formation, either based on random utility or beyond (369–371). Other interesting lines of research to increase the validity of DCEs would be the stability of preferences over time such as with time-to-think (327,372), and the role of social networks in decision making (see Chapter 5).

Additionally, I believe researchers have a facilitating role towards the public, patients, and policymakers in the integration of preferences in healthcare decision making. There are many benefits to incorporating preferences in healthcare decision making, but to truly achieve societal impact in this domain, the beneficiaries – i.e., the public, patients, and policymakers – should be involved as research partners, for example by using a patient-oriented approach in all stages of designing a DCE (373).

For policymakers

Public and patient involvement has gained attention in research, but to date there is limited evidence on the actual use of these types of input and its implications in healthcare decision making (341,374,375). As identified in Chapter 2, most pressing challenges in the procedural, normative, and practical domains are whose preferences to incorporate, how to weight preferences studies in comparison to or in addition to current ethical, clinical and cost-effectiveness studies, and how to evaluate the impact of preference studies. Despite that there is no direct and structural role for preference data in healthcare decision making yet, quantitative preference studies can be valuable supportive evidence alongside other types of evidence in many ways, for example in a broader health technology assessment framework (330,332,335,344).

In parallel with the increased interest in the integration of public and patient preferences in healthcare decision making, there is a call to involve patients and members of the public as research partners in a collaborative manner rather than via a merely tokenistic consultation of research subjects (336,337,347,376–378). However, resource challenges and uncertainty about the impact of the input complicate this type of integration (375,379,380). To overcome the challenges, there is much to be learned from previous experiences and guidelines (376,381–384).

FINAL REMARKS

Public and patient preferences can greatly contribute to healthcare decision making by ensuring more support for difficult decisions at the societal level, and increasing uptake, adherence, and satisfaction with treatment at the individual level. Although research into the validity of preference elicitation methods has come a long way, there are still many challenges remaining. I have argued that the road ahead should start from the perspective of external validity, trying to understand which factors influence decision making, and tailoring the presentation of information and the approach to analyse the data to these insights. In doing so, the sampling procedure, model complexity and the versatility of factors causing hypothetical bias should be considered. To eventually incorporate preferences in healthcare decision making in a more systematic and scientifically valid way, a multi-stakeholder approach is needed whereby researchers, policymakers, the public, patients, and other stakeholders collaborate as research partners.

OTHER

☒ OTHER, NAMELY...

Summary

ENGLISH SUMMARY

Decisions about health always involve trade-offs, both for individuals and for society. In a society with scarce healthcare resources, difficult decisions have to be made on a societal level. Consequently, not every individual can always get the care they prefer. Integrating public and patient preferences in healthcare decision making has the potential to ensure more support for difficult decisions at the societal level, and it can increase uptake, adherence, and treatment satisfaction at the individual level. One of the most used approaches to inform healthcare decision making about public or patient preferences is the discrete choice experiment (DCE). DCE is a quantitative method that elicits preferences from individuals by repeatedly letting them trade-off alternatives containing different characteristics. However, to be useful for healthcare decision-making, it is of great importance that the method's validity is sufficiently established, which currently is not the case. This doctoral thesis provides insights into the internal and external validity of DCEs to inform healthcare decision making.

Chapter 1 introduces the relevant concepts of this thesis by providing background information on DCEs and validity. As DCEs are survey-based and present respondents with hypothetical choices, elicited preferences are called stated preferences. The extent to which the measured preferences represent actual preferences within the context of a study is referred to as the internal validity of a DCE, the rigour with which a study was conducted. However, the usefulness of DCEs for healthcare decision making also depends on their external validity, that is, whether preferences elicited from respondents in a DCE are the same as their preferences in reality. Such information can be obtained by observing people's choice behaviour, their revealed preferences.

Chapter 2 provides an overview of current challenges to integrating preferences in healthcare decision making, with a focus on patient preferences and health technology assessment. Using a systematic review of the literature, 37 challenges were identified. Methodological challenges were most frequently mentioned in the included literature, closely followed by procedural, normative, practical, and conceptual challenges. Important methodological challenges were related to internal validity, external validity, preference heterogeneity, and the choice of method; these challenges were addressed in the following chapters of this thesis.

In Chapter 3, the relationship between health-risk attitude and heterogeneity of preferences is studied through a cross-European comparison in three different health contexts. Health-risk attitude as measured using the 13-item Health-Risk Attitude Scale could not distinguish respondents between distinct preference patterns, but there was a relationship between health-risk attitude and heterogeneity for some risk attributes within the classes with similar preference patterns. Only in one of three health contexts, model fit statistically significantly improved by incorporating health-risk attitude.

Chapter 4 investigates how the presentation of choices affects stated preferences by comparing three different DCE formats. In all three DCE formats – best-worst DCE, best-best DCE and ranking DCE – respondents were asked to choose the alternative they preferred most (single-best) and to choose among the remaining alternatives. First choices differed from second choices in all three DCE formats. The three DCE formats also led to different results.

In Chapter 5, internal and external validity of DCEs are assessed by comparing stated and revealed preferences of respondents in the context of colorectal cancer screening with a focus on the role of model complexity. Revealed preferences were accurately predicted on an individual level for 90% of respondents. Incorporating heterogeneity improved model fit and prediction accuracy for stated preferences, but not revealed preferences.

In Chapter 6, internal and external validity are again assessed by comparing stated and revealed preferences, now in the context of food choice and with a focus on social desirability bias. Socially desirable behaviour, cheap talk mitigation and their effect on the outcomes and prediction accuracy of DCEs were studied by using four different questionnaire versions. Revealed preferences were accurately predicted on an individual level for 9-20% of respondents, depending on the type of vegetable and questionnaire version. Cheap talk mitigation slightly improved prediction accuracy for revealed preferences, but not for stated preferences.

Chapter 7 discusses the main findings, strengths and limitations, and implications of this thesis as well as recommendations for future research and policy. As addressed in the earlier chapters, the integration of public and patient preferences in healthcare decision making can benefit both individuals and society. Although research into the validity of preference elicitation methods has come a long way, many challenges remain. Internal validity of DCEs can be improved by identifying the multitude of factors that influence respondents when making decisions in particular contexts, and by trying to accommodate for these factors in the presentation of choices and data analysis of DCEs. External validity can be improved by paying close attention to the sampling procedure, model complexity and the versatility of factors causing hypothetical bias. To eventually incorporate preferences in healthcare decision making in a more systematic and scientifically valid way, a multi-stakeholder approach is needed whereby researchers, policymakers, the public, patients, and other stakeholders collaborate as research partners.

NEDERLANDSE SAMENVATTING

Beslissingen over gezondheid en zorg gaan altijd gepaard met afwegingen, zowel voor individuen als voor de samenleving. In een samenleving met schaarse middelen in de gezondheidszorg moeten op maatschappelijk niveau moeilijke beslissingen worden genomen. Daardoor kan niet iedereen altijd de zorg krijgen die hij of zij het liefste wil. Het integreren van publieke- en patiëntvoorkeuren in de besluitvorming in de gezondheidszorg zorgt potentieel voor meer steun voor moeilijke beslissingen op maatschappelijk niveau. Het kan daarnaast de acceptatie, therapietrouw en behandelingstevredenheid op individueel niveau vergroten. Een van de meest gebruikte methodes om de besluitvorming in de gezondheidszorg te informeren over publieke- en patiëntvoorkeuren, is het discrete keuze-experiment (discrete choice experiment; DCE). DCE is een kwantitatieve methode die voorkeuren meet door mensen herhaaldelijk alternatieven met verschillende kenmerken te laten afwegen. Om deze gemeten voorkeuren te gebruiken voor de besluitvorming in de zorg, is het echter van groot belang dat de validiteit van de methode voldoende vaststaat, wat nu niet het geval is. Dit proefschrift biedt inzicht in de interne en externe validiteit van DCE's om de besluitvorming in de gezondheidszorg te informeren.

Hoofdstuk 1 introduceert de relevante concepten van dit proefschrift door achtergrondinformatie te geven over DCE's en validiteit. Omdat DCE's afgenomen worden door middel van vragenlijsten en omdat de keuzes hypothetisch zijn, worden de gemeten voorkeuren "stated preferences" genoemd. De mate waarin de gemeten voorkeuren daadwerkelijke voorkeuren vertegenwoordigen binnen de context van een onderzoek, wordt de interne validiteit van een DCE genoemd, ofwel de nauwkeurigheid waarmee een onderzoek is uitgevoerd. Het nut van DCE's voor besluitvorming in de gezondheidszorg hangt echter ook af van hun externe validiteit, dat wil zeggen of de gemeten voorkeuren van respondenten in een DCE, hetzelfde zijn als hun voorkeuren in werkelijkheid. Dergelijke informatie kan worden verkregen door het keuzegedrag van mensen in de praktijk, hun "revealed preferences", te observeren.

Hoofdstuk 2 geeft een overzicht van de huidige uitdagingen bij het integreren van voorkeuren in de besluitvorming in de gezondheidszorg, met een focus op patiëntvoorkeuren en "Health Technology Assessment". Met behulp van een systematische review van de literatuur werden 37 uitdagingen geïdentificeerd. Methodologische uitdagingen werden het vaakst genoemd in de opgenomen literatuur, op de voet gevolgd door procedurele, normatieve, praktische en conceptuele uitdagingen. Belangrijke methodologische uitdagingen waren gerelateerd aan interne validiteit, externe validiteit, variabiliteit van voorkeuren en de keuze van de methode. Deze uitdagingen zijn behandeld in de volgende hoofdstukken van dit proefschrift.

In Hoofdstuk 3 wordt de relatie tussen gezondheidsrisicohouding, de mate waarin iemand bereid is om risico te nemen rondom zijn of haar gezondheid, en variabiliteit van voorkeuren bestudeerd door middel van een cross-Europese vergelijking in drie verschillende gezondheidscontexten. De gezondheidsrisicohouding is gemeten met de 13-item "Health-Risk Attitude Scale". Er kon geen onderscheid gemaakt worden tussen verschillende groepen respondenten aan de hand van hun gezondheidsrisicohouding, maar er was wel een verband tussen gezondheidsrisicohouding en variabiliteit van voorkeuren voor sommige risicoattributen binnen de verschillende groepen. Slechts in een van de drie gezondheidscontexten was er een statistisch significante verbetering in de passendheid van het model door het opnemen van gezondheidsrisicohouding.

Hoofdstuk 4 onderzoekt of keuzes van respondenten verschillen afhankelijk van hoe ze gepresenteerd zijn door drie verschillende DCE-vormen te vergelijken. In alle drie de vormen van DCE – "best-worst DCE", "best-best DCE" en "ranking DCE" - werden respondenten gevraagd het alternatief te kiezen dat zij het beste vonden ("single-best") en te kiezen uit de overige alternatieven. In alle drie de DCE-vormen was het keuzeproces van eerste keuzes anders dan de tweede keuzes. Ook leidden ze tot verschillende resultaten.

In Hoofdstuk 5 worden de interne en externe validiteit van DCE's beoordeeld door de "stated preferences" en "revealed preferences" van respondenten met elkaar te vergelijken. Dit is gedaan in de context van het bevolkingsonderzoek naar darmkankerscreening en met nadruk op de rol van modelcomplexiteit. "Revealed preferences" werden nauwkeurig voorspeld op individueel niveau voor 90% van de respondenten. Het opnemen van heterogeniteit verbeterde de passendheid van het model en voorspellingsnauwkeurigheid voor "stated preferences", maar niet voor "revealed preferences.

Hoofdstuk 6 beschrijft de beoordeling van de interne en externe validiteit door "stated preferences" en "revealed preferences" te vergelijken, nu in de context van voedselkeuze en met nadruk op de rol van sociaal wenselijk gedrag. Sociaal wenselijk gedrag, het verminderen ervan met de "cheap talk mitigation method" en het effect op de uitkomsten en voorspellingsnauwkeurigheid van DCE's werden bestudeerd door vier verschillende versies van een vragenlijst met elkaar te vergelijken. "Revealed preferences" werden nauwkeurig voorspeld op individueel niveau voor 9-20% van de respondenten, afhankelijk van het onderwerp en de versie van de vragenlijst. Het gebruik van de "cheap talk mitigation method" verbeterde enigszins de voorspellingsnauwkeurigheid voor "stated preferences", maar niet voor "revealed preferences".

Hoofdstuk 7 bespreekt de belangrijkste bevindingen, sterke punten en beperkingen, en implicaties van dit proefschrift, evenals aanbevelingen voor toekomstig onderzoek en beleid. Zoals in de eerdere hoofdstukken is besproken, kan de integratie van publieke- en patiëntvoorkeuren in de besluitvorming in de gezondheidszorg zowel individuen als de samenleving ten goede komen. Hoewel het

onderzoek naar de validiteit van methoden om voorkeuren te meten al een lange weg heeft afgelegd, blijven er nog veel uitdagingen. Interne validiteit van DCE's kan verbeterd worden door factoren te identificeren die respondenten beïnvloeden bij het nemen van beslissingen in bepaalde contexten, en door rekening te houden met deze factoren in de presentatie van keuzes en data-analyse van DCE's. Externe validiteit kan worden verbeterd door goed te letten op de steekproefprocedure, de complexiteit van het model en de veelzijdigheid van factoren die "hypothetical bias" veroorzaken. Om uiteindelijk voorkeuren in de besluitvorming in de gezondheidszorg op een meer systematische en wetenschappelijk verantwoorde manier te integreren, is een multi-stakeholderbenadering nodig waarbij onderzoekers, beleidsmakers, het algemene publiek, patiënten en andere belanghebbenden samenwerken als onderzoekspartners.

Portfolio

PORTFOLIO

About the author

Samare Huls (1994) has a Bachelor of Science degree in Economics and Business Economics from the Erasmus University Rotterdam (2016), and a cum laude Master of Science degree in Behavioural Economics from the same university (2017). During this period, she worked as a student assistant and marketing analytics intern, and spent four months studying at Pompeu Fabra University in Barcelona. Motivated by understanding, analysing, and communicating the implications of human behaviour in a data-driven way, Samare started a PhD trajectory at Erasmus School of Health Policy & Management that resulted in this PhD thesis (2017-2022). During this trajectory, she was committed to work on the societal impact of the multidisciplinary initiative Smarter Choices for Better Health, was a member of the Erasmus Choice Modelling Centre, founded the ISPOR student chapter Rotterdam, was a board member of youngESHPM, and organised the EuHEA PhD conference in Rotterdam. Samare also completed a postacademic master in Data Science & Business Analytics (2020).

After her PhD research, Samare continued her academic journey as a researcher and lecturer by trying to understand human behaviour, sharing knowledge, and working in teams and projects with societal relevance at Erasmus School of Health Policy & Management.

Publications

2019 Huls, S. P., Whichello, C. L., van Exel, J., Uyl-de Groot, C. A., & de Bekker-Grob, E. W. (2019). What is next for patient preferences in health technology assessment? A systematic review of the challenges. Value in Health, 22(11), 1318-1328.

2020 Arslan, I. G., Huls, S. P. I., de Bekker-Grob, E. W., Rozendaal, R., Persoons, M. C. T., Spruijt-van Hell, M. E., ... & Schiphof, D. (2020). Patients', healthcare providers', and insurance company employees' preferences for knee and hip osteoarthritis care: a discrete choice experiment. Osteoarthritis and Cartilage, 28(10), 1316-1324.

2021 Visser, L. A., Huls, S. P. I., Uyl-de Groot, C. A., de Bekker-Grob, E. W., & Redekop, W. K. (2021). An implantable device to treat multiple sclerosis: A discrete choice experiment on patient preferences in three European countries. Journal of the Neurological Sciences, 428, 117587.

2021 Huls, S., Sajjad, A., Kanters, T., Hakkaart-van Roijen, L., Brouwer, W., & van Exel, J. (2021). Tijdens lockdown maakten werkenden minder uren en kregen ze minder gedaan. Economisch-Statistische Berichten, 106(4793), 11-13.

2022 Huls, S. P., van Osch, S. M., Brouwer, W. B., van Exel, J., & Stiggelbout, A. M. (2022). Psychometric evaluation of the Health-Risk Attitude Scale (HRAS-13): Assessing the reliability, dimensionality and validity in the general population and a patient population. Psychology & Health, 37(1), 34-50.

2022 Brouwer, W., Huls, S., Sajjad, A., Kanters, T., Roijen, L. H. V., & van Exel, J. (2022). In Absence of Absenteeism: Some Thoughts on Productivity Costs in Economic Evaluations in a Post-corona Era. PharmacoEconomics, 40(1), 7-11.

2022 Huls, S. P., Sajjad, A., Kanters, T. A., Hakkaart-van Roijen, L., Brouwer, W. B., & van Exel, J. (2022). Productivity of working at home and time allocation between paid work, unpaid work and leisure activities during a pandemic. Pharmacoeconomics, 40(1), 77-90.

2022 Huls, S.P.I., Veldwijk, J., Swait, J.D., Viberg Johansson, J., Ancillotti, M., & de Bekker-Grob, E.W. (2022). Preference variation: where does health-risk attitude come into the equation? Value in Health, in press

2022 Huls, S.P.I., Donkers, A.C.D., Ride, J., & Lancsar, E. (2022). Two for the price of one: If moving beyond traditional single-best discrete choice experiments, should we use best-worst, best-best or ranking for preference elicitation? Health Economics, in press

Courses

2017	Health Technology Assessment, Erasmus School of Health Policy & Management
2018	Making an academic poster that stands out, Erasmus Graduate School of Social Sciences and the Humanities
2018	How to survive your PhD, Erasmus Graduate School of Social Sciences and the Humanities
2018	Systematic literature search in PubMed, Erasmus Medical Centre
2018	Systematic literature search in other databases, Erasmus Medical Centre
2018	Measurement of Patient Preferences using Discrete Choice Experiments, Erasmus School of Health Policy & Management
2018	Network Meta Analysis, Erasmus School of Health Policy & Management
2018	Choice modelling and survey design, University of Leeds
2019	Choice modelling and survey design: advanced, University of Leeds
2019	Ethical dilemmas in academia, Young ESHPM
2020	Data Science and Business Analytics, Erasmus Quantitative Intelligence
2021	Causal Inference, EuHEA PhD conference
2022	University Teaching Qualification, RISBO, Erasmus University

Conferences

2017	Meeting of International Acadamy of Health Prefence Research
2017	ISPOR Annual European Conference
2017	NVTAG Symposium HTA methodology
2018	Lowlands Health Economics Study Group Conference
2018	Society of Medical Decision Making Biannual Conference
2018	European Health Economics Association Biannual Conference
2018	ISPOR Annual European Conference
2019	Lowlands Health Economics Study Group Conference
2019	Cancer Drug Development Forum - “Involving Patients in Oncology Drug Development”
2019	Meeting of International Acadamy of Health Prefence Research
2019	European Health Economics Association PhD Conference
2020	Meeting of International Acadamy of Health Prefence Research
2020	European Health Economics Association Biannual Conference

2020	MultiJuSe II
2020	Wittgenstein Centre Conference 2020 “Demographic Aspects of COVID-19 Pandemic and its Consequences”
2021	EuHEA PhD Conference
2022	Lowlands Health Economics Study Group Conference

Teaching

2017	Introduction methods & techniques
2018	Methods & techniques 2
2018	Introduction methods & techniques
2019	Measuring patient preferences using discrete choice experiment
2019	Bachelor honours class
2019	Master thesis
2019	Introduction methods & techniques
2019	Understanding health behaviour
2019	Advanced research methods
2020	Measuring patient preferences using discrete choice experiment
2020	Bachelor honours class
2020	Master thesis
2020	Understanding health behaviour
2020	Introduction methods & techniques
2020	Advanced research methods
2021	Measuring patient preferences using discrete choice experiment
2021	Bachelor honours class
2021	Master thesis
2021	Introduction methods & techniques
2022	Master thesis

Managerial

2018-2020	ISPOR student chapter Rotterdam: co-founder and board member
2019-2020	YoungESHPM: board member
2019-2020	EuHEA PhD Rotterdam: organising committee
2019-2021	Smarter Choices for Better Health: junior committee member
2020-2021	EuHEA PhD Rotterdam: organising committee

Grants & awards

2019	Mentioned in Mentioned in Academic Health Economics Blog by Rachel Houten, 11th November 2019
2020	Best poster award, Wittgenstein Centre Conference 2020 “Demographic Aspects of COVID-19 Pandemic and its Consequences”

Miscellaneous

2017-2022	Member Erasmus Choice Modelling Centre

Appendices

APPENDIX 1: DATABASE SEARCH TERMS

Embase

('patient preference'/de OR 'patient participation'/de OR 'patient engagement'/de OR 'patient attitude'/de OR 'attitude to health'/de OR 'consumer attitude'/de OR 'patient decision making'/de OR (((patient* OR client* OR consumer* OR user*) NEAR/3 (preference* OR participat* OR involvement* OR engage* OR choice* OR perspective* OR utilit* OR acceptab* OR desirab* OR tradeoff* OR trade-off* OR decision* OR attitude* OR voice*))):ab,ti) AND ('biomedical technology assessment'/de OR (((biomedical OR health OR healthcare OR medical) NEAR/3 technolog* NEAR/3 (assessment* OR decision*)) OR hta):ab,ti) NOT [Conference Abstract]/lim AND [english]/lim

Medline Ovid

(Patient Preference/ OR Patient Participation/ OR Patient Acceptance of Health Care/ OR Attitude to Health/ OR (((patient* OR client* OR consumer* OR user*) ADJ3 (preference* OR participat* OR involvement* OR engage* OR choice* OR perspective* OR utilit* OR acceptab* OR desirab* OR tradeoff* OR trade-off* OR decision* OR attitude* OR voice*))).ab,ti.) AND (Technology Assessment, Biomedical/ OR (((biomedical OR health OR healthcare OR medical) ADJ3 technolog* ADJ3 (assessment* OR decision*)) OR hta).ab,ti.) AND english.la.

Cochrane CENTRAL

((((patient* OR client* OR consumer* OR user*) NEAR/3 (preference* OR participat* OR involvement* OR engage* OR choice* OR perspective* OR utilit* OR acceptab* OR desirab* OR tradeoff* OR trade-off* OR decision* OR attitude* OR voice*))):ab,ti) AND ((((biomedical OR health OR healthcare OR medical) NEAR/3 technolog* NEAR/3 (assessment* OR decision*)) OR hta):ab,ti)

Web of Science

TS=(((((patient* OR client* OR consumer* OR user*) NEAR/2 (preference* OR participat* OR involvement* OR engage* OR choice* OR perspective* OR utilit* OR acceptab* OR desirab* OR tradeoff* OR trade-off* OR decision* OR attitude* OR voice*)))) AND ((((biomedical OR health OR healthcare OR medical) NEAR/2 technolog* NEAR/2 (assessment* OR decision*)) OR hta))) AND DT=(article) AND LA=(english)

Scopus

TITLE-ABS-KEY(((((patient* OR client* OR consumer* OR user*) W/2 (preference* OR participat* OR involvement* OR engage* OR choice* OR perspective* OR utilit* OR acceptab* OR desirab* OR tradeoff* OR trade-off* OR decision* OR attitude* OR voice*)))) AND ((((biomedical OR health OR healthcare OR medical) W/2 technolog* W/2 (assessment* OR decision*)) OR hta))) AND DocType(ar) AND Language(english)

CINAHL EBSCOhost

(MH Consumer Participation+ OR MH Attitude to Health OR TI (((patient* OR client* OR consumer* OR user*) N2 (preference* OR participat* OR involvement* OR engage* OR choice* OR perspective* OR utilit* OR acceptab* OR desirab* OR tradeoff* OR trade-off* OR decision* OR attitude* OR voice*))) OR AB (((patient* OR client* OR consumer* OR user*) N2 (preference* OR participat* OR involvement* OR engage* OR choice* OR perspective* OR utilit* OR acceptab* OR desirab* OR tradeoff* OR trade-off* OR decision* OR attitude* OR voice*)))) AND (TI (((biomedical OR health OR healthcare OR medical) N2 technolog* N2 (assessment* OR decision*)) OR hta) OR AB (((biomedical OR health OR healthcare OR medical) N2 technolog* N2 (assessment* OR decision*)) OR hta)) AND LA (english)

Google Scholar

"patient|client|consumer|user|preferences|participation|involvement|engagement|choice|perspective|utility|acceptability|desirability|tradeoff*|decision|attitude|voice*" "biomedical|health|medical technology assessment"

APPENDIX 2: STUDY CHARACTERISTICS – PER ARTICLE

Table A 1: Study characteristics - per article

Authors (year)	Country	Type of study	Medical context	Preference elicitation	Issue raised by	Issue relevant for
Abelson et al. (69)	Canada	Theory	General	Qualitative	Authors	HTA professionals
Beresniak et al. (109)	France	Theory	General	Quantitative	Authors	HTA professionals and academics
Berglas et al. (82)	Canada	Application	General	Qualitative	Authors	HTA professionals
Brereton et al. (72)	UK	Theory	Palliative care & general	Qualitative	Authors	HTA professionals and academics
Brereton et al. (93)	UK	Application	Palliative care & general	Qualitative	Authors	HTA professionals and academics
Brooker et al. (60)	Canada	Application	COPD ventilation	Quantitative	Authors and cited authors	HTA professionals and academics
Buck et al. (64)	UK	Theory	General	Qualitative	Respondents (patients) and authors	HTA professionals
Burke et al. (107)	US	Theory	General	N/D	Authors	HTA professionals, clinicians and patients
Cassels et al. (98)	Canada	Theory	General	Qualitative	Authors	HTA professionals
Chen et al. (121)	US	Theory	Oncology	N/D	Authors	HTA professionals and academics
Dalle Fratte et al. (114)	Italy	Application	HPV and general	Qualitative	Authors	HTA professionals
Danner et al. (74)	Germany	Application	Age-Related Macular Degeneration	Quantitative	Authors	Academics
DeJean et al. (61)	Canada	Theory	General	Qualitative	Authors	HTA professionals and academics
Di Paolo et al. (89)	Italy	Theory	General	Qualitative	Authors	HTA professionals
Dipankui et al. (67)	Canada	Application	General	Qualitative	Respondents (various) and authors	HTA professionals and academics

Table A 1: *Continued.*

Authors (year)	Country	Type of study	Medical context	Preference elicitation	Issue raised by	Issue relevant for
Dirksen (2)	NL	Theory	General	Qualitative	Authors and cited authors	HTA professionals and academics
Doctor and MacEwan (90)	US	Theory	General	N/D	Authors	HTA professionals and academics
Douglas et al. (95)	Canada	Theory	Orphan drugs	Qualitative	Authors	HTA professionals
Drummond et al. (112)	US	Theory	General	Mixed	Authors	HTA professionals
Ducey et al. (102)	Canada	Theory	General	Qualitative	Respondents (HTA profs) and authors	HTA professionals and academics
Evers et al. (80)	NL	Theory	Rare diseases	Qualitative	Authors	HTA professionals and academics
Facey et al. (118)	Scotland	Theory	Rare diseases	Qualitative	Authors and cited authors	HTA professionals
Gagnon et al. (110)	Canada	Applications	Early HTA	Qualitative	Authors	HTA professionals
Gagnon et al. (63)	Canada	Theory	General	Qualitative	Respondents (various) and authors	HTA professionals
Gagnon et al. (97)	Canada	Theory	General	Qualitative	Respondents (various) and authors	HTA professionals
Hailey et al. (75)	Australia	Theory	General	Mixed	Respondents (HTA profs) and authors	HTA professionals and academics
Hämeen-Anttila et al. (86)	Finland	Theory	General	Qualitative	Respondents (patient reps & health care profs) and authors	HTA professionals and academics

Table A 1: *Continued.*

Authors (year)	Country	Type of study	Medical context	Preference elicitation	Issue raised by	Issue relevant for
Husereau et al. (88)	Canada	Theory	General	Qualitative	Authors	HTA professionals
Iskrov and Stefanov (83)	Bulgaria	Theory	General	Quantitative	Authors	HTA professionals and academics
Janssen et al. (55)	US	Theory	General	Authors	Authors and cited authors	Academics
Janssen et al. (101)	Germany	Application	Hemodialysis	Mixed	Respondents (patients) and authors	HTA professionals and academics
Kennedy-Martin et al. (115)	UK	Theory	Diabetic kidney disease	Quantitative	Authors	HTA professionals and academics
Kievit et al. (92)	NL	Theory	General	Qualitative	Authors	HTA professionals and academics
Kleme et al. (111)	Finland	Theory	General	Qualitative	Authors	HTA professionals
Kreis et al. (77)	Germany	Theory	General	Qualitative	Respondents (HTA profs) and authors	HTA professionals and academics
Kreis and Schmidt (68)	Germany	Theory	General	Qualitative	Authors	HTA professionals
Li and Ngorsuraches (113)	US	Theory	General	N/D	Authors	HTA professionals
Lopes et al. (96)	Australia	Theory	General	Qualitative	Respondents (various) and authors	HTA professionals
Low (105)	UK	Theory	General	Qualitative	Authors	Patients and patient organisations
MacLeod et al. (103)	Australia	Theory	General	Qualitative	Authors	HTA professionals and academics
Moreira (84)	UK	Theory	Alzheimer	Qualitative	Authors	HTA professionals and academics
Morel and Cano (104)	Belgium	Theory	Rare diseases	Mixed	Authors	HTA professionals and academics

Table A 1: *Continued.*

Authors (year)	Country	Type of study	Medical context	Preference elicitation	Issue raised by	Issue relevant for
Morgan et al. (120)	UK	Application	Breastfeeding and smoking cessation in pregnancy	Mixed	Respondents (patients) and authors	HTA professionals and academics
Mossman et al. (91)	Belgium	Theory	General	N/D	Authors and cited authors	HTA professionals and academics
Mott and Najafzadeh (78)	UK	Application	Anticoagulant therapy and general	Quantitative	Authors	HTA professionals and academics
Mühlbacher (108)	Germany	Theory	General	Quantitative	Authors	HTA professionals and academics
Mühlbacher et al. (119)	Germany	Application	Acute Coronary Syndrome	Quantitative	Authors	HTA professionals and academics
Mühlbacher et al. (81)	Germany	Application	Chronic hepatitis C	Quantitative	Respondents (patients), authors and cited authors	HTA professionals and academics
Mühlbacher et al. (99)	Germany	Theory	General	Quantitative	Authors	HTA professionals and academics
Mühlbacher and Kaczynski (65)	Germany	Theory	General	Quantitative	Authors	Academics
Mühlbacher and Sadler (66)	Germany	Application	Hepatitis C	Quantitative	Authors	HTA professionals and academics
Narbutas et al. (106)	Lithuania	Theory	Cancer care	N/D	Authors	HTA professionals
Payakachat et al. (116)	US	Theory	General	Quantitative	Authors	Academics
Rashid et al. (62)	UK	Theory	General	Mixed	Authors	HTA professionals and clinical guideline developers
Regier et al. (122)	US	Theory	Oncology	N/D	Authors	HTA professionals

Table A 1: *Continued.*

Authors (year)	Country	Type of study	Medical context	Preference elicitation	Issue raised by	Issue relevant for
Scott and Wale (94)	Australia	Theory	General	Qualitative	Authors	HTA professionals and patient organisations
Single et al. (76)	Australia	Theory	General	Qualitative	Authors	HTA professionals
Thokala et al. (79)	UK	Theory	General	Quantitative	Authors	HTA professionals and academics
Tordrup et al. (85)	UK	Application	General	Quantitative	Authors	HTA professionals and academics
Utens et al. (59)	NL	Theory	General	Both	Respondents (various) and authors	HTA professionals, clinical guideline developers and academics
Utens et al. (70)	NL	Theory	General	N/D	Authors	HTA professionals, clinical guideline developers and academics
Wahlster et al. (73)	UK	Theory and application	Palliative care & general	Both	Authors	HTA professionals
Wale et al. (6)	UK	Theory	General	Both	Authors	HTA professionals
Weeks et al. (87)	Canada	Theory	General	Qualitative	Respondents (HTA profs)	HTA professionals and academics
Weernink et al. (117)	NL	Application	Parkinson's Disease	Quantitative	Authors	HTA professionals and academics
Whitty (100)	Australia	Theory	General	N/D	Respondents (HTA profs) and authors	HTA professionals and academics
Wortley et al. (71)	Australia	Theory	General	N/D	Authors	HTA professionals

APPENDIX 3: EXAMPLES OF CHOICE TASKS

Figure A 1: Example of a choice task for multiple sclerosis treatment - Study 1

Imagine you have to choose between two MS treatments or no treatment, which of the following MS treatments would you prefer? Treatment 1, Treatment 2, or No treatment?

	Treatment 1	Treatment 2	No treatment
Risk of relapse	30% less risk	70% less risk	Unknown
Reducing disease progression	20% less progression	60% less progression	No reduction in disease progression
Risk of side effects	Very common mild side effects (more than 10% risk)	Common moderate side effects (1 to 10% risk)	No side effects
Mode of administration	1 pill per day	Replacing the implant 1 time per year	None
I choose:	○	○	○

Figure A 2: Example of a choice task for antibiotics usage - Study 2

Antibiotics are used for infections caused by bacteria. Imagine you have a bacterial infection. Your life is not in danger, but your doctor recommends that you take antibiotics. In the choice situations that follow, you will be asked what you would do if you had the opportunity to choose between different antibiotic treatments.

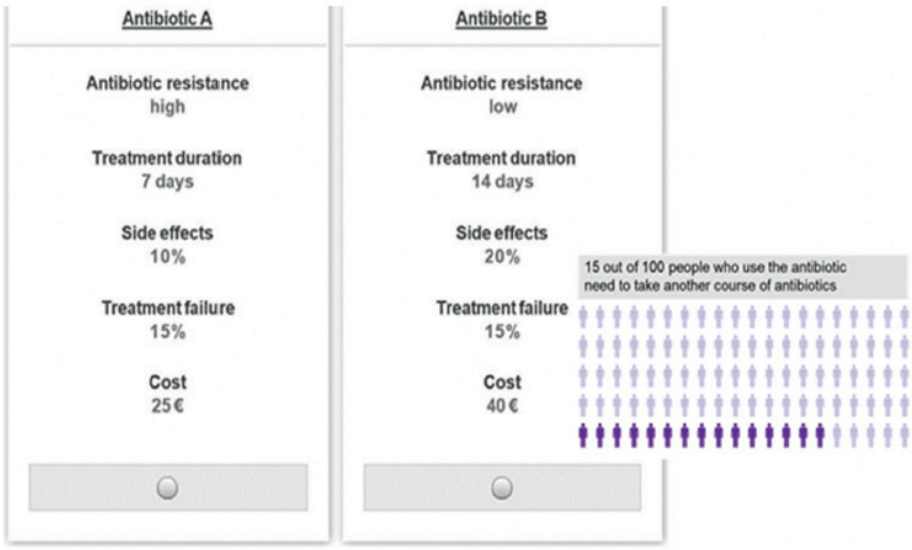


Figure A 3: Example of a choice task for hip-and knee osteoarthritis - Study 3

Imagine that you can choose which healthcare you receive for your complaints, which of the following would you choose, Scenario 1 or Scenario 2? Please select the Scenario that you prefer most by checking the box below.

	Scenario 1	Scenario 2
Waiting time to visit	1 week	No waiting time
Healthcare providers during consultation	Orthopaedist	General practitioner
Out of pocket costs	€90,-	€0,-
Duration of consultation	10 minutes	10 minutes
Travel distance	7 kilometres	1 kilometre
Access to specialists equipment	Same location and same day as the consultation. (direct)	Another location and another day than the consultation. (indirect)

I would choose: ☐ ☐

APPENDIX 4: RELATIONSHIP HRAS AND BACKGROUND VARIABLES

Table A 2: Relationship HRAS and background variables

		HRAS, mean (SD)		
Characteristic	**Category**	**Study 1**	**Study 2**	**Study 3**
Health	Very poor	-	51.8 (13.2)	-
	Poor	-	59.5 (10.5)	-
	Neutral	-	57.6 (9.7)	-
	Good	-	61.1 (8.7)	-
	Very good	-	64.1 (10.0)	-
Gender	Female	44.2 (9.0)	61.3 (10.0)	48.4 (5.4)
	Male	47.7 (8.0)	58.8 (9.3)	49.7 (5.3)
	Other	0 (0.0)	-	0 (0.0)
Education level	Low	45.4 (8.9)	58.7 (11.7)	49.3 (5.3)
	Medium	45.8 (8.0)	59.1 (9.3)	49.3 (5.4)
	High	45.1 (9.2)	61.2 (9.2)	48.2 (5.5)
	Other	43.6 (11.4)	63.8 (7.7)	45.5 (0.7)
Health literacy	Inadequate	44.79 (9.05)	53.6 (9.1)	-
	Problematic	45.60 (8.84)	57.6 (8.2)	-
	Sufficient	45.00 (8.84)	61.6 (10.1)	-
Numeracy	Inadequate	44.23 (8.71)	58.8 (10.7)	-
	Problematic	46.77 (8.66)	59.6 (9.3)	-
	Sufficient	44.37 (8.91)	60.8 (10.0)	-

APPENDIX 5: LATENT CLASS RESULTS

Table A 3: Latent class results - Study 1 (Multiple sclerosis)

	Class 1			Class 2		
	Coeff.	**Std.err.**	**P-value**	**Coeff.**	**Std.err.**	**P-value**
Constant (no treatment)	-0.642	0.085	<0.001	1.747	0.133	<0.001
Constant (Treatment 1)	-0.116	0.028	<0.001	-0.201	0.061	0.001
Reducing risk of relapse	0.775	0.076	<0.001	-0.403	0.191	0.035
Reducing disease progression	1.347	0.069	<0.001	0.561	0.156	<0.001
Risk of side effects						
Very common mild side effects (Ref)	0	-	-	0	-	-
Common moderate side effects	0.097	0.108	0.369	-0.117	0.271	0.665
Rare severe side effects	-0.144	0.047	0.002	-0.209	0.151	0.165
Mode of administration						
Injections 3x per week (Ref)	0	-	-	0	-	-
Injections 1x per week	0.162	0.034	<0.001	-0.032	0.072	0.654
Implant 1x per year	0.159	0.018	<0.001	0.166	0.040	<0.001
Implant 1x per 3 years	0.121	0.015	<0.001	0.118	0.030	<0.001
Pills 2x per day	0.072	0.012	<0.001	0.105	0.024	<0.001
Pills 1x per day	0.111	0.010	<0.001	0.125	0.018	<0.001

Table A 3: *Continued.*

	Class 1			Class 2		
	Coeff.	**Std.err.**	**P-value**	**Coeff.**	**Std.err.**	**P-value**
Interactions HRAS						
*Reducing risk of relapse (%)	-0.009	0.002	<0.001	0.012	0.004	0.003
*Reducing disease progression (%)	-0.017	0.001	<0.001	-0.007	0.003	0.026
*Risk of side effects (common moderate)	-0.002	0.002	0.391	0.002	0.006	0.772
*Risk of side effects (rare severe)	0.002	0.001	0.020	0.003	0.003	0.318
Class allocation model						
Constant	1.131	0.702	0.107	0	-	-
HRAS	-0.002	0.012	0.839	0	-	-
Health (good)	-0.138	0.198	0.486	0	-	-
Age	-0.001	0.008	0.939	0	-	-
Gender (male)	-0.059	0.212	0.781	0	-	-
Education (high)	0.123	0.196	0.528	0	-	-
Literacy (sufficient)	-0.287	0.235	0.222	0	-	-
Numeracy (sufficient)	0.634	0.201	0.002	0	-	-
Average class probability			76.6%			23.4%
Log-likelihood (-)						9383.08
Akaike Information Criterion (AIC)						18842.20

Table A 4: Latent class results - Study 2 (Antibiotics)

	Class 1			Class 2			Class 3		
	Coeff.	**Std.err.**	**P-value**	**Coeff.**	**Std.err.**	**P-value**	**Coeff.**	**Std.err.**	**P-value**
Contribution to resistance									
Low (Ref)	0	-	-	0	-	-	0	-	-
Medium	-2.253	0.662	0.001	0.143	0.588	0.809	-0.175	0.582	0.763
High	-3.756	0.997	<0.001	-0.280	0.587	0.633	0.724	0.622	0.244
Number of days treatments									
3 days (Ref)	0	-	-	0	-	-	0	-	-
7 days	0.094	0.113	0.406	0.166	0.098	0.089	0.158	0.087	0.071
14 days	-0.160	0.109	0.141	-0.354	0.099	<0.001	-0.075	0.084	0.374
Risk of side effects									
1% (Ref)	0	-	-	0	-	-	0	-	-
5%	-0.404	0.820	0.623	-0.329	0.695	0.637	-0.316	0.648	0.626
20%	0.023	0.791	0.977	-0.701	0.705	0.320	-0.137	0.623	0.826
20%	-0.258	0.860	0.764	-0.974	0.670	0.146	0.413	0.711	0.561
Treatment failure	0.214	0.456	0.639	0.949	0.407	0.020	0.117	0.325	0.718
Costs	-0.064	0.015	<0.001	-0.403	0.023	<0.001	0.073	0.013	<0.001
Interactions HRAS									
*Resistance (med)	0.012	0.011	0.262	-0.010	0.010	0.320	0.007	0.010	0.466
*Resistance (high)	0.004	0.016	0.794	-0.005	0.010	0.578	-0.004	0.010	0.733
*Risk of side effects (5%)	0.001	0.013	0.949	0.002	0.012	0.834	-0.001	0.011	0.949

Table A 4: *Continued.*

	Class 1			Class 2			Class 3		
	Coeff.	**Std.err.**	**P-value**	**Coeff.**	**Std.err.**	**P-value**	**Coeff.**	**Std.err.**	**P-value**
*Risk of side effects (10%)	0.000	0.013	0.989	0.012	0.012	0.285	-0.005	0.010	0.655
*Risk of side effects (20%)	-0.001	0.014	0.934	0.014	0.011	0.197	-0.023	0.012	0.057
*Treatment failure (%)	-0.004	0.007	0.577	-0.016	0.007	0.019	-0.001	0.005	0.847
Class allocation model									
Constant	0.362	1.053	0.731	1.206	1.025	0.239	0	-	-
HRAS	0.018	0.018	0.316	-0.002	0.017	0.924	0	-	-
Health (good)	-0.260	0.435	0.551	-0.528	0.467	0.258	0	-	-
Age	-0.035	0.012	0.004	-0.016	0.012	0.192	0	-	-
Gender (male)	-0.199	0.321	0.536	-0.494	0.334	0.139	0	-	-
Education (high)	0.271	0.314	0.388	-0.483	0.318	0.129	0	-	-
Literacy (sufficient)	0.413	0.343	0.228	0.890	0.355	0.012	0	-	-
Numeracy (sufficient)	0.206	0.305	0.499	-0.560	0.318	0.079	0	-	-
Average class probability			38.5%			33.6%			28.0%
Log-likelihood (-)									2999.73
Akaike Information Criterion (AIC)									621.46

Table A 5: Latent class results - Study 3 (Hip-and knee osteoarthritis)

	Class 1			Class 2			Class 3			Class 4		
	Coeff.	**Std.err.**	**P-value**	**Coeff.**	**Std.err.**	**P-value**	**Coeff.**	**Std.err.**	**P-value**	**Coeff.**	**Std.err.**	**P-value**
Constant (Treatment 1)	0.097	0.105	0.355	-0.198	0.061	0.001	0.460	0.154	0.003	-0.207	0.122	0.090
Waiting time (per week)	0.989	0.329	0.003	0.061	0.236	0.796	-2.252	0.581	0.000	-1.096	0.391	0.005
Professionals involved												
GP (Ref)	0	-	-	0	-	-	0	-	-	0	-	-
Orthopaedist	-0.760	1.297	0.558	0.555	1.075	0.606	2.196	2.044	0.283	0.366	1.196	0.760
GP and orthopaedist	0.846	1.436	0.556	0.947	0.983	0.335	-0.933	2.244	0.678	3.827	1.331	0.004
Price in euros	-0.023	0.003	<0.001	-0.008	0.001	<0.001	-0.012	0.002	<0.001	-0.081	0.009	<0.001
Time per consult in minutes	0.003	0.007	0.727	0.008	0.004	0.037	0.016	0.008	0.041	0.004	0.007	0.531
Travel time in kilometers	-0.106	0.014	<0.001	-0.017	0.005	0.002	-0.009	0.009	0.336	-0.036	0.009	<0.001
Specialist equipment												
Indirect access (Ref)	0	-	-	0	-	-	0	-	-	0	-	-
Direct access	0.430	0.122	<0.001	0.876	0.071	<0.001	0.249	0.124	0.045	0.530	0.114	<0.001
Interactions HRAS												
*Waiting time	-0.023	0.007	0.001	-0.004	0.005	0.441	0.036	0.011	0.001	0.015	0.008	0.053
*Professionals (Orthopaedist)	0.026	0.027	0.327	0.031	0.023	0.167	-0.041	0.041	0.317	0.003	0.025	0.913
*Professionals (GP and ortho.)	-0.002	0.030	0.953	0.026	0.020	0.198	0.023	0.045	0.616	-0.051	0.027	0.056

Table A 5: *Continued.*

	Class 1			Class 2			Class 3			Class 4		
	Coeff.	**Std.err.**	**P-value**	**Coeff.**	**Std.err.**	**P-value**	**Coeff.**	**Std.err.**	**P-value**	**Coeff.**	**Std.err.**	**P-value**
Class allocation model												
Constant	-3.340	1.795	0.063	-1.189	1.366	0.384	-2.419	1.995	0.226	0	-	-
HRAS	0.007	0.027	0.794	-0.001	0.022	0.980	0.036	0.035	0.301	0	-	-
Health (good)	0.226	0.300	0.451	-0.295	0.228	0.196	-0.471	0.347	0.175	0	-	-
Age	0.048	0.017	0.004	0.027	0.013	0.040	0.003	0.020	0.890	0	-	-
Gender (male)	-0.422	0.308	0.170	-0.224	0.235	0.341	0.056	0.353	0.873	0	-	-
Education (high)	-0.520	0.362	0.151	0.101	0.254	0.691	-0.188	0.395	0.635	0	-	-
Average class probability			22.9%			34.9%			14.0%			28.3%
Log-likelihood (-)												4196.44
Akaike Information Criterion (AIC)												8496.88

APPENDIX 6: LITERATURE OVERVIEW

Table A 6: Literature overview

	Method				Criteria								
Authors (year)	**DCE**	**BB**	**BW**	**Rank**	**Coef./ MRS**	**Scale**	**Scale dyn.**	**Efficiency**	**Self-rep.**	**Ext. validity**	**Other**	**Context**	**Different preferences**
Single method – within method comparison													
Ben-Akiva et al. (201)				√	√	√						Transport	Yes
Delle Site et al. (207)			√		√			√		√	Probit models	Transport	Yes
Hausman & Ruud (191)				√	√	√						Transport	Yes
Hawkins et al. (210)			√		√	√					Out-of-sample prediction	Marketing	No
Scarpa et al. (212)			√			√	√					Environment	No
Lancsar et al. (192)			√		√			√				Health	No
Multimethod – between method comparison													
Akaichi et al. (214)	√			√	√					√	Proportion chosen, in-sample prediction	Agriculture	Yes
Caparrós et al. (215)	√			√	√	√			√			Environment	No
Chang et al. (216)	√			√	√					√	Market shares	Marketing	Yes
Petrolia et al. (217)	√		√		√	√		√			Status-quo, attribute non-attendance	Environment	No
Xie et al. (213)	√		√		√			√	√		Intraclass Corr. Coeff.	Health	No
Yangui et al. (218)	√		√	√	√					√	Intra-rater reliability	Agriculture	No

Table A 6: *Continued.*

Authors (year)	Method				Criteria								
	DCE	BB	BW	Rank	Coef./ MRS	Scale	Scale dyn.	Efficiency	Self-rep.	Ext. validity	Other	Context	Different preferences
Multimethod – within and between method comparison													
Giergiczny et al. (220)	√	√	√		√	√			√			Transport	Yes
Hawkins et al. (209)	√		√		√						Proportion chosen, in-sample prediction	Marketing	No
Krucien et al. (219)	√		√		√	√						Health	Yes
This study		√	√	√	√	√	√	√	√			Health	Yes

APPENDIX 7: QUESTIONS RESPONDENT-REPORTED CRITERIA, PER CRITERION AND PER METHOD

Table A 7: Questions respondent-reported criteria, per criterion and per method

Question [response options]	BW	BB	Rank
Difficulty [easy, neither, difficult]			
Thinking about the choices you just made, how did you find the task of choosing the best option in each scenario?	√	√	
Thinking about the choices you just made, how did you find the task of choosing the worst option in each scenario?	√		
Thinking about the choices you just made, how did you find the task of choosing the second best option in each scenario?		√	
Thinking about the choices you just made, how did you find the task of ranking the 3 options in each scenario?			√
Preference [yes, no]			
Instead of choosing best and then worst in each scenario, would you have preferred to choose the first best (like you did) and then instead of choosing worst of the remaining two options, choose best of the remaining two options?	√		
Instead of choosing best and then worst in each scenario, would you have preferred to rank the options from best to worst?	√		
Instead of choosing best and then second best in each scenario, would you have preferred to choose the first best (like you did) and then instead of choosing second best of the remaining two options, choose worst of the remaining two options?		√	
Instead of choosing best and then second best in each scenario, would you have preferred to rank the options from best to worst?		√	
Instead of ranking the options in each scenario from the best to worst, would you have preferred to choose the first best and then choose second best of the remaining two options?			√
Instead of ranking the options in each scenario from the best to worst, would you have preferred to choose the first best and then choose worst of the remaining two options?			√

APPENDIX 8: CHARACTERISTICS OF RESPONDENTS PER ARM

Table A 8: Characteristics of respondents per arm

Variables	Best-worst N=997	Best-best N=926	Ranking N=1012
Age, years, mean (SD)	44.0 (14.3)	43.6 (14.4)	44.1 (14.1)
Gender, female (%)	447 (44.8)	408 (44.1)	449 (44.4)
Employment status, employed (%)	843 (84.6)	798 (86.2)	855 (84.5)
Education, n (%)			
Never completed (high) school	88 (8.8)	89 (9.6)	92 (9.1)
High school	133 (13.3)	125 (13.5)	135 (13.3)
TAFE	237 (23.8)	221 (23.9)	210 (20.8)
Associate diploma	83 (8.3)	76 (8.2)	81 (8.0)
University	443 (44.4)	414 (44.7)	493 (48.7)
Other	13 (1.3)	1 (0.1)	1 (0.1)
Self-assessed health, n (%)	106 (10.6)	96 (10.4)	102 (10.1)
Excellent	340 (34.1)	334 (36.4)	354 (35.0)
Very good	376 (37.7)	341 (36.8)	378 (37.4)
Good	126 (12.6)	127 (13.7)	147 (14.5)
Fair	49 (4.9)	28 (3.0)	31 (3.1)
Poor			
Duration in minutes, mean (SD)	40.7 (56.1)	37.4 (44.9)	36.9 (33.4)
Duration in minutes, median	28.4	27.5	29.1

APPENDIX 9: OVERVIEW OF LITERATURE COMPARING STATED PREFERENCES (SP) AND REVEALED PREFERENCES (RP)

Table A 9: Overview of literature comparing stated preferences (SP) and revealed preferences (RP)

Study	Joint estimation	Aggregate or individual level comparison	Type of revealed choice	Context	Accuracy of SP predictions
Within sample					
De Bekker-Grob et al., 2018 (260)	No	Aggregate	Holdout task	Health	No self-contained measure
De Bekker-Grob et al., 2019 (157)	No	Both	Holdout task	Health	93%
De Bekker-Grob et al., 2020 (172)	No	Both	Actual choice & holdout task	Health	91%
De Bekker-Grob et al., 2021 (173)	No	Both	Holdout task	Health	97%
Fifer et al., 2014 (316)	No	Individual	Actual choice	Transport	No self-contained measure
Krucien et al., 2015 (29)	No	Both	Choice intention	Health	16.7-48.5%
Lambooij et al., 2015 (385)	No	Both	Actual choice	Health	80%
Meghani et al., 2013 (386)	No	Individual	Holdout task	Health	MAE: 3-10%
Mohammadi et al., 2017 (387)	No	Both	Actual choice	Health	83%
Natter and Feurstein, 2002 (388)	No	Aggregate	Actual choice & holdout task	Marketing	$r2$: 50%, average bias: 5%
Ryan & Watson, 2009 (389)	No	Aggregate	Actual choice	Health	80%
Salampessy et al., 2015 (390)	No	Both	Actual choice	Health	74%
Telser & Zweifel, 2007 (26)	No	Aggregate	Choice intention	Health	No self-contained measure
Zipursky et al., 2017 (391)	No	Individual	Holdout task	Health	MAE: 1-2%

Table A 9: *Continued.*

Study	Joint estimation	Aggregate or individual level comparison	Type of revealed choice	Context	Accuracy of SP predictions
Adamowicz et al., 1994 (392)	Yes	Aggregate	Actual choice	Environment	Prefs differ up to scale
Cameron et al., 2002 (393)	Yes	Aggregate	Actual choice	Environment	Prefs differ up to scale
Carlsson and Martinsson, 2001 (394)	Yes	Both	Non-hypothetical choice task	Environment	Prefs differ up to scale
Kesternich et al., 2013 (395)	Yes	Both	Actual choice	Health	Prefs differ up to scale
Whitehead, 2005 (396)	Yes	Individual	Actual choice	Environment	Forecast error: 3-8%
Between samples					
Araña & León, 2013	No	Aggregate	Real-world uptake	Environment	Significant differences in market share on long but not short term
Cornelsen et al., 2020 (187)	No	Aggregate	Real-world uptake	Health	No self-contained measure
Linley & Hughes, 2013 (397)	No	Both	Real-world uptake	Health	39-90%
Regier et al., 2020 (328)	No	Aggregate	Real-world uptake	Health	RMSE: 0.11, correlation >90%
Vossler and Watson, 2003 (398)	No	Aggregate	Actual choice	Public policy	Average bias: 0-10%
De Corte et al., 2021 (399)	Yes	Aggregate	Actual choice	Health	No self-contained measure
Mark & Swait, 2003 (400)	Yes	Aggregate	Estimated real-world uptake	Health	Prefs differ beyond scale

Table A 9: *Continued.*

Study	Joint estimation	Aggregate or individual level comparison	Type of revealed choice	Context	Accuracy of SP predictions
Within and between samples					
Chang et al., 2009 (216)	No	Aggregate	Non-hypothetical choice task & Real-world uptake	Marketing	MSE: 0.047
Mueller et al., 2010 (401)	No	Aggregate	Actual choice & real-world uptake	Marketing	Pearson's r: -0.293 and 0.088
Kruk et al., 2009 (402)	No	Aggregate	Actual choice & real-world uptake	Health	4.1 p.p. difference within, 10 p.p. between samples
Buckell & Hess, 2019 (357)	Yes	Aggregate	Actual choice & real-world uptake	Health	No self-contained measure

MAE = Mean Absolute Error
RMSE = Root Mean Square Error
MSE = Mean Squared Error

APPENDIX 10: INVITEE CHARACTERISTICS

Table A 10: Invitee characteristics

	All invitees N=14288 (%)	Respondents N=568 (%)	Non-respondents N=13,720 (%)	P-val.
Gender, female	7095 (49.7)	269 (47.4)	6826 (49.8)	0.282
Region of residence				
East	2523 (17.7)	96 (16.9)	2427 (17.7)	0.011
South	4028 (28.2)	194 (34.2)	3834 (27.9)	
South-West	7680 (53.8)	277 (48.8)	7403 (54.0)	
Other	57 (0.4)	1 (0.2)	56 (0.4)	
Participate CRC screening, yes	8900 (62.3)	520 (91.5)	8380 (61.1)	<0.001

APPENDIX 11: RESULTS DISCRETE CHOICE EXPERIMENT PER MODEL

Table A 11: Results discrete choice experiment per model

	MNL		MNL+		HMNL		HMNL+		HMNL++
	Est.	P-value	Est.	P-value	Est.	P-value	Est.	P-value	
Parameter estimates									
Alternative-specific constant (no screening)	-5.000	<0.001	-2.353	0.014	-0.668	0.024	-4.312	0.646	-
Effectiveness (%)	0.040	<0.001	0.008	0.553	0.006	0.025	0.035	0.784	-
Risk of false negative (%)	-0.110	<0.001	-0.028	0.252	-0.014	0.025	-0.088	0.683	-
Waiting time FIT result (1 week)									
Waiting time FIT result (2 weeks)	0.207	<0.001	0.082	0.812	0.023	0.046	0.045	0.954	-
Waiting time FIT result (3 weeks)	0.213	<0.001	-0.060	0.863	0.020	0.060	-0.027	0.982	-
Waiting time follow-up test (2 weeks)									
Waiting time follow-up test (4 weeks)	-0.094	0.014	-0.568	0.093	-0.012	0.088	-0.265	0.904	-
Waiting time follow-up test (8 weeks)	-0.539	<0.001	-0.726	0.027	-0.071	0.026	-0.451	0.689	-
Frequency screening (every year)									
Frequency screening (every 2 years)	-0.086	0.033	0.430	0.215	-0.012	0.091	0.036	0.993	-
Frequency screening (every 3 years)	-0.554	<0.001	0.823	0.034	-0.072	0.026	-0.213	0.973	-

Table A 11: *Continued.*

	MNL		MNL+		HMNL		HMNL+		HMNL++
	Est.	P-value	Est.	P-value	Est.	P-value	Est.	P-value	
Scale heterogeneity									
Scale * attitude for					2.117	<0.001	1.168	0.175	-
Scale * literate					-0.060	0.187	0.108	0.590	-
Scale * numerate					0.238	<0.001	-0.283	0.052	-
Scale * deliberative					0.143	0.002	-0.016	0.967	-
Scale * high education					0.234	<0.001	-0.150	0.817	-
Scale * healthy					-0.091	0.165	1.738	0.899	-
Scale * male					-0.133	0.005	-0.188	0.105	-
Scale * GP visit					0.121	0.024	-0.017	0.988	-
Scale * hospital visit					-0.191	0.002	-1.219	0.732	-
Scale * no cancer					-0.212	0.005	-0.438	0.277	-
Scale * CRC in family					0.116	0.020	0.083	0.869	-
Preference heterogeneity									
Attitude for * effectiveness			0.046	<0.001			0.007	0.944	-
Attitude for * false negative			-0.095	<0.001			-0.008	0.947	-
Attitude for * wait FIT (2 weeks)			-0.142	0.633			-0.091	0.948	-
Attitude for * wait FIT (4 weeks)			-0.099	0.741			-0.129	0.943	-
Attitude for * wait follow-up (4 weeks)			0.137	0.635			0.129	0.943	-
Attitude for * wait follow-up (8 weeks)			-0.408	0.145			0.060	0.944	-
Attitude for * frequency (every 2 years)			-0.596	0.049			-0.269	0.947	-

Table A 11: *Continued.*

	MNL		MNL+		HMNL		HMNL+		HMNL++
	Est.	P-value	Est.	P-value	Est.	P-value	Est.	P-value	
Attitude for * frequency (every 3 years)			-1.113	0.001			-0.369	0.947	-
Attitude for * opt-out			-4.556	<0.001			-0.526	0.945	-
Literate * effectiveness			-0.001	0.729			-0.001	0.942	-
Literate * false negative			-0.005	0.520			0.001	0.947	-
Literate * wait FIT (2 weeks)			0.095	0.265			0.009	0.948	-
Literate * wait FIT (4 weeks)			0.129	0.170			0.018	0.947	-
Literate * wait follow-up (4 weeks)			0.156	0.054			0.020	0.945	-
Literate * wait follow-up (8 weeks)			0.497	<0.001			0.073	0.944	-
Literate * frequency (every 2 years)			0.024	0.778			0.003	0.950	-
Literate * frequency (every 3 years)			-0.071	0.504			<0.001	0.992	-
Literate * opt-out			0.459	0.138			0.189	0.946	-
Numerate * effectiveness			0.020	<0.001			0.004	0.944	-
Numerate * false negative			-0.033	<0.001			-0.008	0.943	-
Numerate * wait FIT (2 weeks)			0.146	0.107			0.012	0.948	-
Numerate * wait FIT (4 weeks)			0.198	0.045			0.018	0.945	-
Numerate * wait follow-up (4 weeks)			0.134	0.124			0.006	0.952	-
Numerate * wait follow-up (8 weeks)			0.017	0.852			-0.023	0.943	-
Numerate * frequency (every 2 years)			-0.017	0.851			-0.001	0.956	-
Numerate * frequency (every 3 years)			-0.004	0.969			-0.020	0.939	-
Numerate * opt-out			-0.088	0.781			-0.208	0.942	-
Deliberative * effectiveness			0.008	0.060			0.001	0.951	-
Deliberative * false negative			-0.012	0.127			-0.002	0.955	-

Table A 11: *Continued.*

	MNL		MNL+		HMNL		HMNL+		HMNL++
	Est.	P-value	Est.	P-value	Est.	P-value	Est.	P-value	
Deliberative * wait FIT (2 weeks)			0.095	0.277			0.010	0.945	-
Deliberative * wait FIT (4 weeks)			0.221	0.021			0.022	0.946	-
Deliberative * wait follow-up (4 weeks)			-0.008	0.928			-0.008	0.944	-
Deliberative * wait follow-up (8 weeks)			-0.080	0.354			-0.017	0.949	-
Deliberative * frequency (every 2 years)			-0.009	0.917			-0.002	0.952	-
Deliberative * frequency (every 3 years)			-0.045	0.679			-0.011	0.945	-
Deliberative * opt-out			-0.298	0.355			-0.059	0.956	-
High education * effectiveness			0.019	<0.001			0.003	0.948	-
High education * false negative			-0.035	<0.001			-0.007	0.950	-
High education * wait FIT (2 weeks)			-0.057	0.525			-0.007	0.934	-
High education * wait FIT (4 weeks)			-0.169	0.088			-0.020	0.944	-
High education * wait follow-up (4 weeks)			0.089	0.301			0.003	0.910	-
High education * wait follow-up (8 weeks)			-0.108	0.223			-0.029	0.952	-
High education * frequency (every 2 years)			0.022	0.808			0.010	0.941	-
High education * frequency (every 3 years)			0.020	0.862			0.001	0.990	-
High education * opt-out			-1.116	<0.001			-0.261	0.951	-
Healthy * effectiveness			-0.024	<0.001			-0.040	0.531	-
Healthy * false negative			0.033	0.007			0.090	0.624	-
Healthy * wait FIT (2 weeks)			0.058	0.667			0.035	0.938	-
Healthy * wait FIT (4 weeks)			0.180	0.226			0.135	0.697	-
Healthy * wait follow-up (4 weeks)			-0.129	0.331			0.103	0.382	-
Healthy * wait follow-up (8 weeks)			-0.285	0.042			0.286	0.820	-

Table A 11: *Continued.*

	MNL		MNL+		HMNL		HMNL+		HMNL++
	Est.	P-value	Est.	P-value	Est.	P-value	Est.	P-value	
Healthy * frequency (every 2 years)			0.252	0.068			0.234	0.343	-
Healthy * frequency (every 3 years)			0.081	0.646			0.585	0.446	-
Healthy * opt-out			0.290	0.566			4.427	0.576	-
Male * effectiveness			-0.001	0.886			<0.001	0.951	-
Male * false negative			0.007	0.402			-0.001	0.945	-
Male * wait FIT (2 weeks)			-0.003	0.975			0.004	0.951	-
Male * wait FIT (4 weeks)			0.063	0.511			0.017	0.944	-
Male * wait follow-up (4 weeks)			0.248	0.003			0.025	0.941	-
Male * wait follow-up (8 weeks)			0.573	<0.001			0.061	0.943	-
Male * frequency (every 2 years)			0.181	0.039			0.019	0.945	-
Male * frequency (every 3 years)			0.457	<0.001			0.045	0.945	-
Male * opt-out			0.947	0.003			0.056	0.945	-
GP visit * effectiveness			0.003	0.652			-0.001	0.963	-
GP visit * false negative			-0.006	0.545			0.001	0.971	-
GP visit * wait FIT (2 weeks)			0.066	0.554			0.017	0.922	-
GP visit * wait FIT (4 weeks)			-0.144	0.253			0.006	0.854	-
GP visit * wait follow-up (4 weeks)			-0.028	0.797			-0.001	0.986	-
GP visit * wait follow-up (8 weeks)			0.076	0.487			0.016	0.949	-
GP visit * frequency (every 2 years)			-0.138	0.228			-0.016	0.934	-
GP visit * frequency (every 3 years)			-0.387	0.007			-0.031	0.932	-
GP visit * opt-out			-1.639	0.001			-0.111	0.895	-
Hospital visit * effectiveness			-0.013	0.047			0.007	0.960	-

Table A 11: *Continued.*

	MNL		MNL+		HMNL		HMNL+		HMNL++
	Est.	P-value	Est.	P-value	Est.	P-value	Est.	P-value	
Hospital visit * false negative			0.017	0.152			-0.023	0.959	-
Hospital visit * wait FIT (2 weeks)			-0.065	0.620			0.020	0.967	-
Hospital visit * wait FIT (4 weeks)			0.007	0.960			0.022	0.957	-
Hospital visit * wait follow-up (4 weeks)			0.093	0.461			-0.058	0.951	-
Hospital visit * wait follow-up (8 weeks)			-0.171	0.188			-0.297	0.955	-
Hospital visit * frequency (every 2 years)			-0.394	0.005			-0.152	0.954	-
Hospital visit * frequency (every 3 years)			-0.498	0.003			-0.378	0.955	-
Hospital visit * opt-out			0.229	0.648			-2.394	0.946	-
No cancer * effectiveness			-0.013	0.108			<0.001	0.966	-
No cancer * false negative			0.021	0.167			<0.001	0.934	-
No cancer * wait FIT (2 weeks)			0.028	0.852			0.014	0.945	-
No cancer * wait FIT (4 weeks)			-0.030	0.855			0.012	0.950	-
No cancer * wait follow-up (4 weeks)			0.109	0.448			-0.004	0.906	-
No cancer * wait follow-up (8 weeks)			0.357	0.025			-0.001	0.974	-
No cancer * frequency (every 2 years)			-0.192	0.204			-0.013	0.946	-
No cancer * frequency (every 3 years)			-0.491	0.008			-0.067	0.943	-
No cancer * opt-out			1.611	0.019			-0.061	0.910	-
CRC in family * effectiveness			0.009	0.066			<0.001	0.797	-
CRC in family * false negative			-0.025	0.005			-0.001	0.873	-
CRC in family * wait FIT (2 weeks)			0.022	0.810			<0.001	0.996	-
CRC in family * wait FIT (4 weeks)			-0.032	0.756			0.002	0.889	-
CRC in family * wait follow-up (4 weeks)			-0.054	0.541			-0.005	0.951	-

Table A 11: *Continued.*

	MNL		MNL+		HMNL		HMNL+		HMNL++
	Est.	P-value	Est.	P-value	Est.	P-value	Est.	P-value	
CRC in family * wait follow-up (8 weeks)			-0.110	0.227			0.004	0.953	-
CRC in family * frequency (every 2 years)			-0.061	0.510			-0.008	0.940	-
CRC in family * frequency (every 3 years)			-0.040	0.737			-0.001	0.932	-
CRC in family * opt-out			-0.876	0.020			-0.016	0.810	-
Model statistics									
N parameters	9		108		20		119		-
Log-likelihood	-6184		-5860		-6029		-5803		-
AIC	12385		11935		12098		11845		-
BIC	12449		12704		12240		12692		-
Holdout task prediction									
Percentage accurately predicted	95.07%		95.28%		95.32%		95.42%		-
Positive predictive value (PPV)	97.23		97.28		97.30		97.34		-
Negative predicted value (NPV)	2.77		10.22		10.79		13.61		-
Real choice prediction									
Percentage accurately predicted	89.82%		89.81%		89.65%		89.87%		-
Positive predictive value (PPV)	97.92		97.84		97.64		97.83		-
Negative predictive value (NPV)	2.08		2.80		3.12		3.62		-

APPENDIX 12: RESPONDENT CHARACTERISTICS, SOCIAL SUPPORT

Table A 12: Respondent characteristics, social support

		All respondents N=568 (%)	SP=RP N=513 (%)	SP≠RP N=55 (%)	P-val.
Partner	Very important	287 (50.5)	259 (50.5)	28 (50.9)	0.626
	Important	107 (18.8)	94 (18.3)	13 (23.6)	
	Neutral	54 (9.5)	51 (9.9)	3 (5.5)	
	Unimportant	35 (6.2)	32 (6.2)	3 (5.5)	
	Very unimportant	28 (4.9)	27 (5.3)	1 (1.8)	
	Not applicable	57 (10.0)	50 (9.7)	7 (12.7)	
Sibling	Very important	118 (20.8)	104 (20.3)	14 (25.5)	0.100
	Important	140 (24.6)	126 (24.6)	14 (25.5)	
	Neutral	125 (22.0)	116 (22.6)	9 (16.4)	
	Unimportant	77 (13.6)	70 (13.6)	7 (12.7)	
	Very unimportant	61 (10.7)	59 (11.5)	2 (3.6)	
	Not applicable	47 (8.3)	38 (7.4)	9 (16.4)	
Children	Very important	236 (41.5)	211 (41.1)	25 (45.5)	0.566
	Important	110 (19.4)	98 (19.1)	12 (21.8)	
	Neutral	50 (8.8)	47 (9.2)	3 (5.5)	
	Unimportant	38 (6.7)	36 (7.0)	2 (3.6)	
	Very unimportant	43 (7.6)	41 (8.0)	2 (3.6)	
	Not applicable	91 (16.0)	80 (15.6)	11 (20.0)	

Table A 12: *Continued.*

		All respondents N=568 (%)	SP=RP N=513 (%)	SP≠RP N=55 (%)	P-val.
Friend(s), same age	Very important	82 (14.4)	71 (13.8)	11 (20.0)	0.246
	Important	143 (25.2)	130 (25.3)	13 (23.6)	
	Neutral	128 (22.5)	115 (22.4)	13 (23.6)	
	Unimportant	83 (14.6)	78 (15.2)	5 (9.1)	
	Very unimportant	65 (11.4)	62 (12.1)	3 (5.5)	
	Not applicable	67 (11.8)	57 (11.1)	10 (18.2)	
Friend(s), other age	Very important	72 (12.7)	64 (12.5)	8 (14.5)	0.186
	Important	135 (23.8)	124 (24.2)	11 (20.0)	
	Neutral	139 (24.5)	125 (24.4)	14 (25.5)	
	Unimportant	87 (15.3)	80 (15.6)	7 (12.7)	
	Very unimportant	66 (11.6)	63 (12.3)	3 (5.5)	
	Not applicable	69 (12.1)	57 (11.1)	12 (21.8)	
General practitioner	Very important	99 (17.4)	87 (17.0)	12 (21.8)	0.008
	Important	169 (29.8)	160 (31.2)	9 (16.4)	
	Neutral	124 (21.8)	110 (21.4)	14 (25.5)	
	Unimportant	75 (13.2)	70 (13.6)	5 (9.1)	
	Very unimportant	60 (10.6)	55 (10.7)	5 (9.1)	
	Not applicable	41 (7.2)	31 (6.0)	10 (18.2)	

Table A 12: *Continued.*

		All respondents N=568 (%)	SP=RP N=513 (%)	SP≠RP N=55 (%)	P-val.
Other doctor	Very important	67 (11.8)	61 (11.9)	6 (10.9)	0.822
	Important	131 (23.1)	120 (23.4)	11 (20.0)	
	Neutral	137 (24.1)	122 (23.8)	15 (27.3)	
	Unimportant	79 (13.9)	73 (14.2)	6 (10.9)	
	Very unimportant	61 (10.7)	56 (10.9)	5 (9.1)	
	Not applicable	93 (16.4)	81 (15.8)	12 (21.8)	
Pharmacist	Very important	41 (7.2)	36 (7.0)	5 (9.1)	0.414
	Important	78 (13.7)	73 (14.2)	5 (9.1)	
	Neutral	155 (27.3)	139 (27.1)	16 (29.1)	
	Unimportant	111 (19.5)	102 (19.9)	9 (16.4)	
	Very unimportant	96 (16.9)	89 (17.3)	7 (12.7)	
	Not applicable	87 (15.3)	74 (14.4)	13 (23.6)	
Religious leader	Very important	14 (2.5)	14 (2.7)	0 (0.0)	0.138
	Important	23 (4.0)	23 (4.5)	0 (0.0)	
	Neutral	97 (17.1)	86 (16.8)	11 (20.0)	
	Unimportant	92 (16.2)	86 (16.8)	6 (10.9)	
	Very unimportant	106 (18.7)	98 (19.1)	8 (14.5)	
	Not applicable	236 (41.5)	206 (40.2)	30 (54.5)	
Other	Very important	32 (5.6)	29 (5.7)	3 (5.5)	0.885
	Important	26 (4.6)	22 (4.3)	4 (7.3)	
	Neutral	44 (7.7)	39 (7.6)	5 (9.1)	
	Unimportant	28 (4.9)	26 (5.1)	2 (3.6)	
	Very unimportant	44 (7.7)	41 (8.0)	3 (5.5)	
	Not applicable	394 (69.4)	356 (69.4)	38 (69.1)	

APPENDIX 13: SUMMARY OF STUDIES ON CHEAP TALK MITIGATION (CT) IN DISCRETE CHOICE EXPERIMENTS, ADAPTED FROM HAGHANI ET AL. (241)

Table A 13: Summary of studies on cheap talk mitigation (CT) in discrete choice experiments, adapted from Haghani et al. (241)

Authors (year)	Method	RP data	Choice context	Socially desirable attribute	Effect CT	Highlights
List et al. (2006) (315)	Cheap talk	Yes	Public non-market goods (monetary donation and sports cards)	Monetary donation	Yes	Responses were not statistically different between the real and hypothetical with cheap talk treatments; Subjects in the hypothetical with cheap talk treatment were more likely to make inconsistent decisions
Chowdhury et al. (2011) (286)	Cheap talk	Yes	Food choice (sweet potato)	Type of potato: nutritional value	Yes	Results confirmed the presence of significant hypothetical bias; Cheap talk reduced the magnitude of bias but did not fully eliminate it
Bosworth and Taylor (2012) (403)	Cheap talk	Yes	Purchase a tree to be planted in public space	None: measured via opt-in or out	Mixed	A dramatically larger number of subjects opted-into the market in the hypothetical questionnaire compared to the real payment condition; Cheap talk induced respondents to opt-out of the market; Respondents in the hypothetical treatment with cheap talk were more price sensitive compared to the real payment treatment (cheap-talk overcorrection)

Table A 13: *Continued.*

Authors (year)	Method	RP data	Choice context	Socially desirable attribute	Effect CT	Highlights
Silva et al. (2012) (292)	Cheap talk	Yes	Food choice (fruit)	Presence of preservatives	Mixed	Perceived task complexity had a significant impact on cheap talk's effectiveness in reducing HB; The cheap talk script was effective only when subjects considered the task to be easy
Moser et al. (2014) (293)	Cheap talk	Yes	Food choice (apples)	Use of pesticides: effect on environment and human health	Mixed	Results confirmed the presence of hypothetical bias; Results confirmed the mixed effectiveness of a cheap talk script
Broadbent (2014) (404)	Multiple	Yes	Recreation site expansion	Monetary donation	No	Hypothetical bias was not present in the MWTP valuation for the quasi-public good; Cheap-talk and follow-up certainty were found to reduce MWTP estimates to be less than actual estimates
Fifer et al. (2014) (316)	Multiple	Yes	Driving behaviour	Speeding	Yes	SC model estimates were prone to HB; Cheap talk and certainty scales when combined have the potential to compensate for HB
De-Magistris et al. (2014) (294)	Multiple	Yes	Food products (sushi)	Omega 3 content: nutritional value	No	MWTPs in the honesty priming treatment were significantly lower than those in baseline hypothetical CE; Values from hypothetical CE with honesty priming were not significantly different from non-hypothetical CE; Cheap talk script was not able to mitigate the HB in hypothetical CE

Table A 13: *Continued.*

Authors (year)	Method	RP data	Choice context	Socially desirable attribute	Effect CT	Highlights
Carlsson et al. (2010) (317)	Multiple	Yes	Donation to environmental projects	Monetary donation	Yes	Both hypothetical treatments (own and third-person preference) showed large differences with the real-money treatment; HB effect was smaller when using a third-person preference viewpoint
Carlsson et al. (2005) (313)	Cheap talk	No	Food choice (chicken and ground beef)	None	Yes	Estimated MWTP for food was lower in the questionnaire version with cheap talk
Özdemir et al. (2009) (326)	Cheap talk	No	Medical treatments (rheumatoid arthritis)	None	Yes	Cheap talk not only affected the coefficient of the cost attribute, but also preferences for other attributes; WTP estimates were generally lower in the cheap talk sample
Tonsor and Shupp (2011) (314)	Cheap talk	No	Food demand (apples)	None	Yes	Cheap talk scripts not only influenced the level of WTP, but also may produce more reliable estimates; The magnitude of the impact on WTP depended on respondent familiarity
Bello and Abdulai (2016) (295)	Multiple	No	Organic food product	Use of pesticides, organic certification, vitamins, soil erosion	Yes	Honesty priming resulted in lower WTP values by nearly a factor of two relative to cheap talk for three of the four attributes
Bello and Abdulai (2016) (296)	Multiple	No	Organic food product	Use of pesticides, organic certification, vitamins, soil erosion	Yes	The level of questionnaire engagement was higher under honesty priming effect compared cheap talk

Table A 13: *Continued.*

Authors (year)	Method	RP data	Choice context	Socially desirable attribute	Effect CT	Highlights
Howard et al. (2017) (320)	Multiple	No	Environmental policy	Agricultural nutrient pollution and harmful algal blooms	No	The cheap talk effect faded with repeated choices; Online implementation of an honesty priming intervention yielded no significant change in price sensitivity compared to a control
Lin et al. (2019) (297)	Multiple	No	Food choice (pork loin)	Country of origin: local or not	No	No significant differences in WTP values for between the various mitigation methods and a control group; HB effect was likely not significant
Ladenburg and Olsen (2014) (318)	Multiple	No	Public good (re-establishment of a stream)	None	Yes	Opt-out reminder significantly reduced total WTP and to some extent also MWTP beyond the capability of the cheap talk alone; Introducing opt-out reminders as a supplement to a short CT script reduced welfare measures at the decision-to-opt-in level but not at the MWTP level
Varela et al. (2014) (319)	Multiple	No	Forest fire prevention program	Yearly burned area	Mixed	The inclusion of a single opt-out reminder did not sufficiently improve the cheap talk effect
Gschwandtner and Burton (2020) (298)	Multiple	No	Organic food product (vegetables and meat)	Organic, chemical usage, environmentally friendly, animal friendly, best before	Yes	A budget reminder combined with cheap talk script appeared to have reduced hypothetical bias more successfully than honesty priming

APPENDIX 14: EXAMPLES CHOICE TASKS

Figure A 4: Example choice task bell pepper

Imagine a normal day when you would do the groceries. You can choose between the options below. Which would you choose? The first option, the second option, or none of these options?

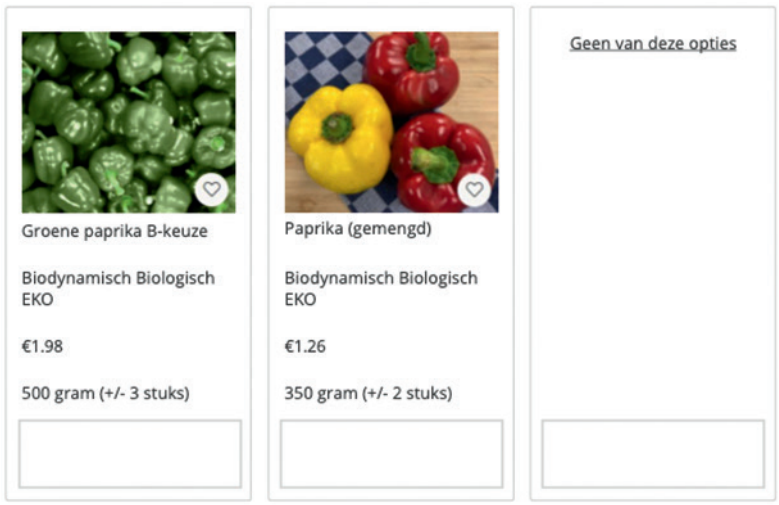


Figure A 5: Example choice task eggplant

Imagine a normal day when you would do the groceries. You can choose between the options below. Which would you choose? The first option, the second option, or none of these options?

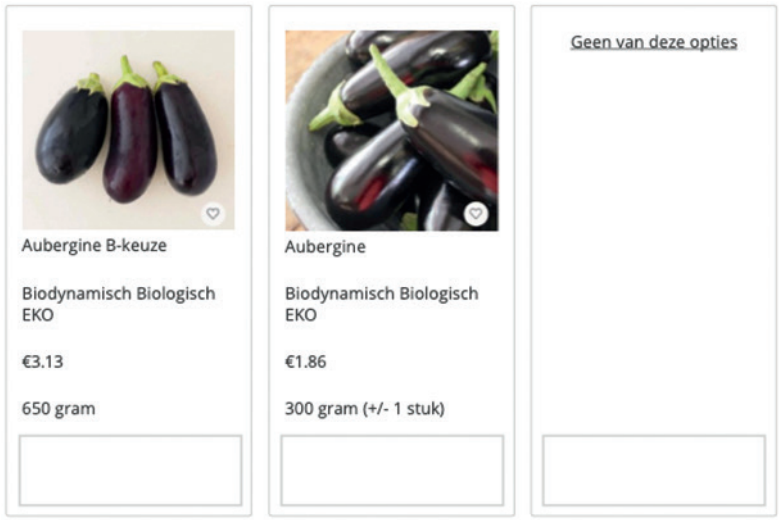


Figure A 6: Example choice task cucumber

Imagine a normal day when you would do the groceries. You can choose between the options below. Which would you choose? The first option, the second option, or none of these options?

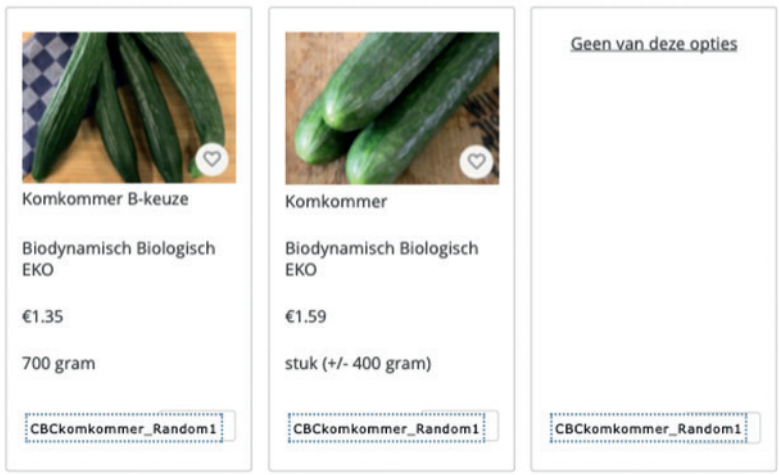


Figure A 7: Fixed holdout choice task bell pepper

Imagine a normal day when you would do the groceries. You can choose between the options below. Which would you choose? The first, second, third, or fourth option, or none the of these options?

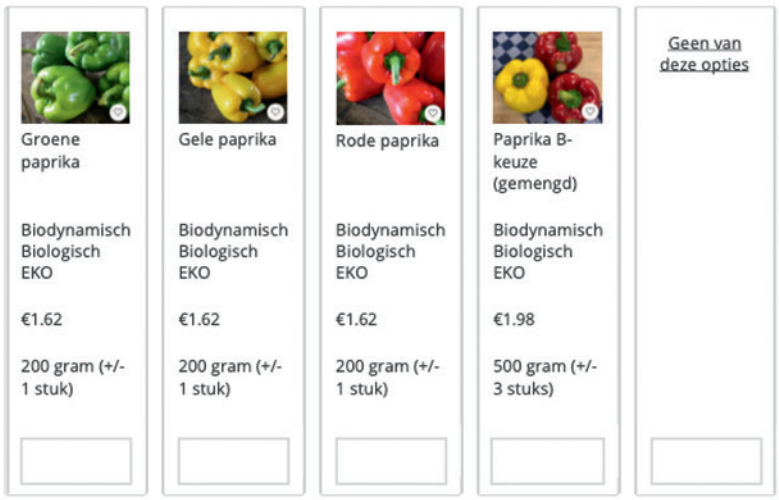


Figure A 8: Fixed holdout choice task eggplant

Imagine a normal day when you would do the groceries. You can choose between the options below. Which would you choose? The first option, the second option, or none of these options?

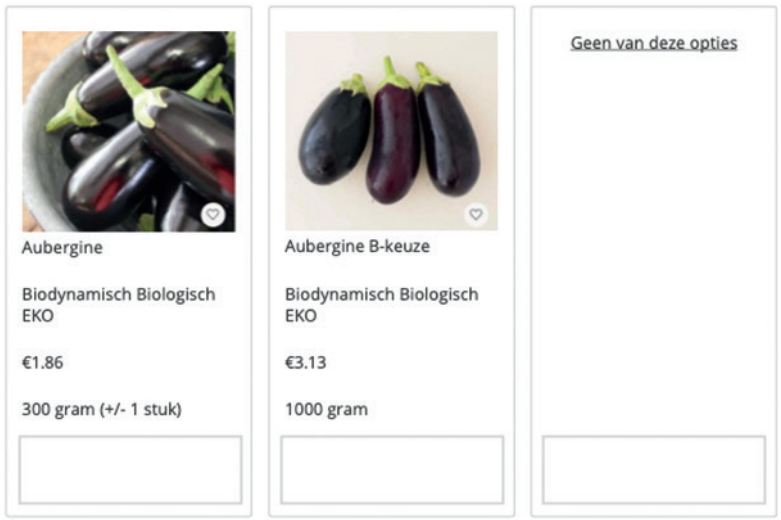
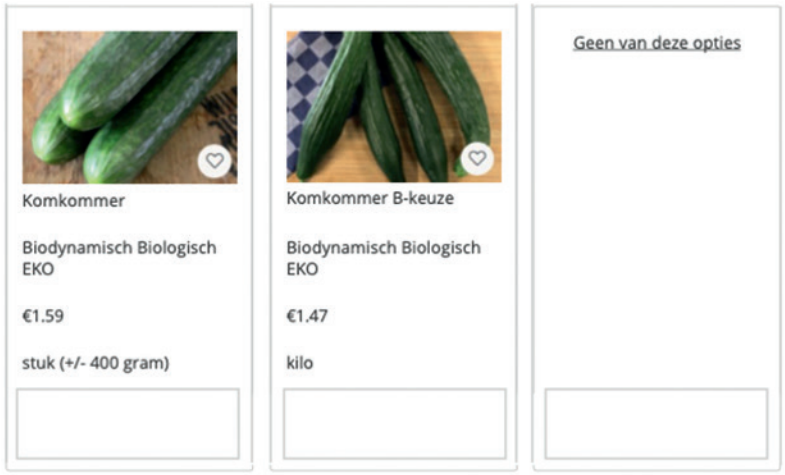


Figure A 9: Fixed holdout choice task cucumber

Imagine a normal day when you would do the groceries. You can choose between the options below. Which would you choose? The first option, the second option, or none of these options?



APPENDIX 15: RESPONDENT CHARACTERISTICS PER VERSION

Table A 14: Respondent characteristics per version

	Version 1	Version 2	Version 3	Version 4	P-value
N	236	269	282	240	
Consent share data = no (%)	79 (33.5)	102 (37.9)	101 (35.8)	79 (32.9)	0.620
Age (%)					0.168
18-29	39 (16.5)	44 (16.4)	56 (19.9)	49 (20.4)	
30-39	29 (12.3)	56 (20.8)	50 (17.7)	40 (16.7)	
40-49	37 (15.7)	35 (13.0)	50 (17.7)	30 (12.5)	
50-59	33 (14.0)	42 (15.6)	49 (17.4)	35 (14.6)	
60-69	51 (21.6)	54 (20.1)	42 (14.9)	53 (22.1)	
70-79	42 (17.8)	36 (13.4)	32 (11.3)	28 (11.7)	
80plus	5 (2.1)	2 (0.7)	3 (1.1)	5 (2.1)	
Sex (%)					0.923
male	119 (50.4)	130 (48.3)	137 (48.6)	117 (48.8)	
female	117 (49.6)	138 (51.3)	144 (51.1)	123 (51.2)	
rather not say	0 (0.0)	1 (0.4)	1 (0.4)	0 (0.0)	

Table A 14: *Continued.*

	Version 1	Version 2	Version 3	Version 4	P-value
Education (%)					0.882
low1	8 (3.4)	10 (3.7)	8 (2.8)	3 (1.2)	
low2	18 (7.6)	21 (7.8)	20 (7.1)	17 (7.1)	
middle1	44 (18.6)	37 (13.8)	47 (16.7)	37 (15.4)	
middle2	59 (25.0)	67 (24.9)	60 (21.3)	50 (20.8)	
middle3	24 (10.2)	30 (11.2)	33 (11.7)	30 (12.5)	
high1	43 (18.2)	58 (21.6)	68 (24.1)	63 (26.2)	
high2	40 (16.9)	45 (16.7)	44 (15.6)	38 (15.8)	
else	0 (0.0)	1 (0.4)	2 (0.7)	2 (0.8)	
City (%)					0.125
Amsterdam	19 (8.1)	15 (5.6)	22 (7.8)	9 (3.8)	
Rotterdam	121 (51.3)	151 (56.1)	136 (48.2)	120 (50.0)	
The Hague	91 (38.6)	99 (36.8)	110 (39.0)	104 (43.3)	
Utrecht	2 (0.8)	2 (0.7)	3 (1.1)	0 (0.0)	
rather not say	0 (0.0)	0 (0.0)	2 (0.7)	0 (0.0)	
other, namely	3 (1.3)	2 (0.7)	9 (3.2)	7 (2.9)	
Buy bell pepper (%)					0.231
weekly	107 (45.3)	108 (40.1)	134 (47.5)	107 (44.6)	
twice a month	71 (30.1)	70 (26.0)	63 (22.3)	73 (30.4)	
monthly	35 (14.8)	50 (18.6)	41 (14.5)	37 (15.4)	
twice a year	14 (5.9)	24 (8.9)	23 (8.2)	14 (5.8)	
never	9 (3.8)	17 (6.3)	21 (7.4)	9 (3.8)	

Table A 14: *Continued.*

	Version 1	Version 2	Version 3	Version 4	P-value
Buy eggplant (%)					0.393
weekly	29 (12.3)	42 (15.6)	49 (17.4)	26 (10.8)	
twice a month	45 (19.1)	56 (20.8)	56 (19.9)	50 (20.8)	
monthly	57 (24.2)	50 (18.6)	58 (20.6)	66 (27.5)	
twice a year	44 (18.6)	43 (16.0)	42 (14.9)	34 (14.2)	
never	61 (25.8)	78 (29.0)	77 (27.3)	64 (26.7)	
Buy cucumber (%)					0.359
weekly	128 (54.2)	135 (50.2)	156 (55.3)	141 (58.8)	
twice a month	44 (18.6)	66 (24.5)	52 (18.4)	38 (15.8)	
monthly	30 (12.7)	37 (13.8)	44 (15.6)	35 (14.6)	
twice a year	27 (11.4)	22 (8.2)	27 (9.6)	20 (8.3)	
never	7 (3.0)	9 (3.3)	3 (1.1)	6 (2.5)	
Household size (mean (SD))	2.00 (1.20)	2.44 (2.70)	2.29 (1.34)	2.36 (1.31)	0.034
Household with kids 0-3 years = yes (%)	16 (12.4)	32 (17.1)	26 (13.6)	30 (17.5)	0.496
Household with kids 4-12 years = yes (%)	24 (18.6)	38 (20.3)	41 (21.5)	37 (21.6)	0.914
Household with kids 13-17 years = yes (%)	11 (8.5)	19 (10.2)	31 (16.2)	22 (12.9)	0.151
Household with adults = yes (%)	103 (79.8)	139 (74.3)	144 (75.4)	129 (75.4)	0.706

Table A 14: *Continued.*

	Version 1	Version 2	Version 3	Version 4	P-value
Groceries (%)					0.077
household	173 (73.3)	221 (82.2)	228 (80.9)	201 (83.8)	
self	58 (24.6)	43 (16.0)	47 (16.7)	37 (15.4)	
never	5 (2.1)	5 (1.9)	7 (2.5)	2 (0.8)	
Groceries store (%)					0.953
daily	88 (38.1)	93 (35.2)	95 (34.5)	86 (36.1)	
weekly	128 (55.4)	153 (58.0)	161 (58.5)	138 (58.0)	
bimonthly	7 (3.0)	10 (3.8)	12 (4.4)	6 (2.5)	
monthly	4 (1.7)	7 (2.7)	3 (1.1)	5 (2.1)	
biyearly	1 (0.4)	0 (0.0)	2 (0.7)	1 (0.4)	
never	3 (1.3)	1 (0.4)	2 (0.7)	2 (0.8)	
Groceries online (%)					0.011
daily	19 (8.2)	18 (6.8)	24 (8.7)	9 (3.8)	
weekly	37 (16.0)	39 (14.8)	51 (18.5)	35 (14.7)	
bimonthly	24 (10.4)	17 (6.4)	23 (8.4)	30 (12.6)	
monthly	19 (8.2)	17 (6.4)	36 (13.1)	21 (8.8)	
biyearly	34 (14.7)	30 (11.4)	24 (8.7)	36 (15.1)	
never	98 (42.4)	143 (54.2)	117 (42.5)	107 (45.0)	
Vegetarian = no (%)	188 (81.4)	209 (79.2)	226 (82.2)	204 (85.7)	0.292
Flexitarian = yes (%)	24 (55.8)	27 (49.1)	25 (51.0)	13 (38.2)	0.484
Pescetarian = yes (%)	17 (39.5)	21 (38.2)	23 (46.9)	19 (55.9)	0.359
Eggs = yes (%)	15 (34.9)	21 (38.2)	19 (38.8)	17 (50.0)	0.574

Table A 14: *Continued.*

	Version 1	Version 2	Version 3	Version 4	P-value
Dairy = yes (%)	12 (27.9)	18 (32.7)	18 (36.7)	12 (35.3)	0.827
Rennet-free cheese = yes (%)	4 (9.3)	9 (16.4)	5 (10.2)	7 (20.6)	0.410
Vegan = yes (%)	2 (4.7)	3 (5.5)	4 (8.2)	1 (2.9)	0.765
Fixed holdout – bell pepper (%)					0.225
green	24 (10.6)	27 (10.7)	23 (8.8)	15 (6.5)	
yellow	12 (5.3)	21 (8.3)	21 (8.0)	22 (9.5)	
red	66 (29.1)	73 (29.0)	64 (24.5)	47 (20.3)	
mixed B	110 (48.5)	109 (43.3)	128 (49.0)	127 (55.0)	
no	15 (6.6)	22 (8.7)	25 (9.6)	20 (8.7)	
Fixed holdout – eggplant (%)					0.123
regular	81 (46.3)	78 (40.8)	71 (34.6)	70 (39.8)	
B	70 (40.0)	73 (38.2)	100 (48.8)	70 (39.8)	
no	24 (13.7)	40 (20.9)	34 (16.6)	36 (20.5)	
Fixed holdout – cucumber (%)					0.258
regular	59 (25.8)	63 (24.2)	64 (22.9)	57 (24.4)	
B	155 (67.7)	165 (63.5)	188 (67.4)	162 (69.2)	
no	15 (6.6)	32 (12.3)	27 (9.7)	15 (6.4)	
Making ends meet (%)					0.149
very hard	32 (13.6)	34 (12.6)	47 (16.7)	21 (8.8)	
hard	64 (27.1)	91 (33.8)	87 (30.9)	66 (27.5)	
easy	98 (41.5)	95 (35.3)	98 (34.8)	105 (43.8)	
very easy	42 (17.8)	49 (18.2)	50 (17.7)	48 (20.0)	

Table A 14: *Continued.*

	Version 1	Version 2	Version 3	Version 4	P-value
Smoking (%)					0.079
no	85 (36.0)	122 (45.4)	118 (41.8)	103 (42.9)	
used to	76 (32.2)	80 (29.7)	76 (27.0)	88 (36.7)	
occasional	39 (16.5)	36 (13.4)	43 (15.2)	25 (10.4)	
daily	36 (15.3)	31 (11.5)	45 (16.0)	24 (10.0)	
Alcohol - days per week (%)					0.301
never	61 (25.8)	89 (33.1)	92 (32.6)	72 (30.0)	
less than once	48 (20.3)	50 (18.6)	68 (24.1)	49 (20.4)	
one	32 (13.6)	33 (12.3)	36 (12.8)	35 (14.6)	
two	22 (9.3)	36 (13.4)	34 (12.1)	32 (13.3)	
three	30 (12.7)	23 (8.6)	21 (7.4)	23 (9.6)	
four	11 (4.7)	15 (5.6)	5 (1.8)	6 (2.5)	
five	8 (3.4)	5 (1.9)	6 (2.1)	8 (3.3)	
six	6 (2.5)	4 (1.5)	1 (0.4)	3 (1.2)	
seven	18 (7.6)	14 (5.2)	19 (6.7)	12 (5.0)	
Alcohol - glasses per time (mean (SD))	2.53 (1.64)	2.78 (2.98)	2.46 (4.26)	2.42 (2.05)	0.651

Table A 14: *Continued.*

	Version 1	Version 2	Version 3	Version 4	P-value
Physical activity - days per week (%)					0.023
never	8 (3.4)	15 (5.6)	19 (6.7)	10 (4.2)	
less than once	20 (8.5)	32 (11.9)	15 (5.3)	11 (4.6)	
one	36 (15.3)	29 (10.8)	40 (14.2)	16 (6.7)	
two	27 (11.4)	28 (10.4)	37 (13.1)	26 (10.8)	
three	29 (12.3)	31 (11.5)	33 (11.7)	47 (19.6)	
four	23 (9.7)	24 (8.9)	34 (12.1)	30 (12.5)	
five	29 (12.3)	34 (12.6)	36 (12.8)	29 (12.1)	
six	18 (7.6)	25 (9.3)	15 (5.3)	18 (7.5)	
seven	46 (19.5)	51 (19.0)	53 (18.8)	53 (22.1)	
Nutrition - days per week (%)					0.499
never	1 (0.4)	3 (1.1)	8 (2.8)	2 (0.8)	
less than once	5 (2.1)	10 (3.7)	7 (2.5)	6 (2.5)	
one	21 (8.9)	17 (6.3)	24 (8.5)	13 (5.4)	
two	10 (4.2)	18 (6.7)	26 (9.2)	11 (4.6)	
three	26 (11.0)	33 (12.3)	34 (12.1)	25 (10.4)	
four	41 (17.4)	38 (14.1)	41 (14.5)	40 (16.7)	
five	55 (23.3)	62 (23.0)	67 (23.8)	59 (24.6)	
six	40 (16.9)	45 (16.7)	43 (15.2)	44 (18.3)	
seven	37 (15.7)	43 (16.0)	32 (11.3)	40 (16.7)	
Health (mean (SD))	6.58 (2.20)	6.69 (2.05)	6.78 (2.11)	7.16 (1.80)	0.014

Table A 14: *Continued.*

	Version 1	Version 2	Version 3	Version 4	P-value
Attitude - environment (%)					0.510
1	34 (14.4)	45 (16.7)	44 (15.6)	28 (11.7)	
2	27 (11.4)	46 (17.1)	40 (14.2)	33 (13.8)	
3	88 (37.3)	78 (29.0)	89 (31.6)	77 (32.1)	
4	48 (20.3)	51 (19.0)	54 (19.1)	59 (24.6)	
5	39 (16.5)	49 (18.2)	55 (19.5)	43 (17.9)	
Attitude - fairtrade (%)					0.537
1	22 (9.3)	34 (12.6)	26 (9.2)	20 (8.3)	
2	45 (19.1)	44 (16.4)	52 (18.4)	33 (13.8)	
3	74 (31.4)	86 (32.0)	105 (37.2)	96 (40.0)	
4	57 (24.2)	63 (23.4)	65 (23.0)	56 (23.3)	
5	38 (16.1)	42 (15.6)	34 (12.1)	35 (14.6)	
Attitude - health 1 (%)					0.591
1	7 (3.0)	15 (5.6)	15 (5.3)	8 (3.3)	
2	28 (11.9)	22 (8.2)	26 (9.2)	26 (10.8)	
3	75 (31.8)	89 (33.1)	103 (36.5)	84 (35.0)	
4	74 (31.4)	76 (28.3)	77 (27.3)	78 (32.5)	
5	52 (22.0)	67 (24.9)	61 (21.6)	44 (18.3)	

Table A 14: *Continued.*

	Version 1	Version 2	Version 3	Version 4	P-value
Attitude - health 2 (%)					0.794
1	10 (4.2)	17 (6.3)	9 (3.2)	8 (3.3)	
2	22 (9.3)	29 (10.8)	32 (11.3)	27 (11.2)	
3	59 (25.0)	67 (24.9)	77 (27.3)	54 (22.5)	
4	81 (34.3)	84 (31.2)	91 (32.3)	91 (37.9)	
5	64 (27.1)	72 (26.8)	73 (25.9)	60 (25.0)	
Attitude - health 3 (%)					0.258
1	18 (7.6)	24 (8.9)	24 (8.5)	16 (6.7)	
2	41 (17.4)	38 (14.1)	43 (15.2)	34 (14.2)	
3	63 (26.7)	102 (37.9)	86 (30.5)	85 (35.4)	
4	59 (25.0)	65 (24.2)	73 (25.9)	67 (27.9)	
5	55 (23.3)	40 (14.9)	56 (19.9)	38 (15.8)	
Attitude - health 4 (%)					0.154
1	4 (1.7)	14 (5.2)	7 (2.5)	3 (1.2)	
2	16 (6.8)	19 (7.1)	26 (9.2)	12 (5.0)	
3	68 (28.8)	71 (26.4)	78 (27.7)	73 (30.4)	
4	92 (39.0)	111 (41.3)	116 (41.1)	110 (45.8)	
5	56 (23.7)	54 (20.1)	55 (19.5)	42 (17.5)	

Table A 14: *Continued.*

	Version 1	Version 2	Version 3	Version 4	P-value
Attitude - local (%)					0.662
1	5 (2.1)	10 (3.7)	6 (2.1)	7 (2.9)	
2	10 (4.2)	18 (6.7)	23 (8.2)	15 (6.2)	
3	87 (36.9)	87 (32.3)	109 (38.7)	80 (33.3)	
4	84 (35.6)	104 (38.7)	89 (31.6)	90 (37.5)	
5	50 (21.2)	50 (18.6)	55 (19.5)	48 (20.0)	
Attitude - taste (%)					0.348
1	1 (0.4)	4 (1.5)	6 (2.1)	7 (2.9)	
2	19 (8.1)	13 (4.8)	20 (7.1)	11 (4.6)	
3	46 (19.5)	64 (23.8)	60 (21.3)	45 (18.8)	
4	106 (44.9)	111 (41.3)	119 (42.2)	119 (49.6)	
5	64 (27.1)	77 (28.6)	77 (27.3)	58 (24.2)	
Attitude - money (%)					0.709
1	15 (6.4)	13 (4.8)	12 (4.3)	6 (2.5)	
2	12 (5.1)	15 (5.6)	13 (4.6)	9 (3.8)	
3	45 (19.1)	53 (19.7)	56 (19.9)	46 (19.2)	
4	71 (30.1)	81 (30.1)	93 (33.0)	92 (38.3)	
5	93 (39.4)	107 (39.8)	108 (38.3)	87 (36.2)	

Table A 14: *Continued.*

	Version 1	Version 2	Version 3	Version 4	P-value
Attitude - time (%)					0.326
1	53 (22.5)	70 (26.0)	55 (19.5)	67 (27.9)	
2	62 (26.3)	72 (26.8)	83 (29.4)	78 (32.5)	
3	67 (28.4)	76 (28.3)	80 (28.4)	53 (22.1)	
4	34 (14.4)	36 (13.4)	46 (16.3)	27 (11.2)	
5	20 (8.5)	15 (5.6)	18 (6.4)	15 (6.2)	
Attitude - personal (%)					0.666
1	17 (7.2)	25 (9.3)	25 (8.9)	14 (5.8)	
2	23 (9.7)	27 (10.0)	30 (10.6)	26 (10.8)	
3	107 (45.3)	117 (43.5)	115 (40.8)	118 (49.2)	
4	55 (23.3)	73 (27.1)	78 (27.7)	60 (25.0)	
5	34 (14.4)	27 (10.0)	34 (12.1)	22 (9.2)	
Attitude - label (%)					0.977
1	14 (5.9)	18 (6.7)	16 (5.7)	16 (6.7)	
2	40 (16.9)	34 (12.6)	44 (15.6)	37 (15.4)	
3	106 (44.9)	129 (48.0)	124 (44.0)	112 (46.7)	
4	50 (21.2)	54 (20.1)	62 (22.0)	52 (21.7)	
5	26 (11.0)	34 (12.6)	36 (12.8)	23 (9.6)	

Table A 14: *Continued.*

	Version 1	Version 2	Version 3	Version 4	P-value
Priority - environment (%)					0.746
1-2	33 (14.0)	39 (14.5)	44 (15.6)	24 (10.0)	
3-4	66 (28.0)	88 (32.7)	90 (31.9)	78 (32.5)	
5-6	107 (45.3)	108 (40.1)	114 (40.4)	108 (45.0)	
7-8	30 (12.7)	34 (12.6)	34 (12.1)	30 (12.5)	
Priority - fairtrade (%)					0.328
1-2	22 (9.3)	27 (10.0)	28 (9.9)	13 (5.4)	
3-4	57 (24.2)	61 (22.7)	83 (29.4)	59 (24.6)	
5-6	112 (47.5)	131 (48.7)	125 (44.3)	129 (53.8)	
7-8	45 (19.1)	50 (18.6)	46 (16.3)	39 (16.2)	
Priority - animal welfare (%)					0.767
1-2	8 (3.4)	13 (4.8)	15 (5.3)	8 (3.3)	
3-4	66 (28.0)	64 (23.8)	78 (27.7)	72 (30.0)	
5-6	103 (43.6)	121 (45.0)	129 (45.7)	106 (44.2)	
7-8	59 (25.0)	71 (26.4)	60 (21.3)	54 (22.5)	
Priority - waste (%)					0.321
1-2	4 (1.7)	13 (4.8)	13 (4.6)	4 (1.7)	
3-4	49 (20.8)	52 (19.3)	60 (21.3)	45 (18.8)	
5-6	123 (52.1)	130 (48.3)	146 (51.8)	123 (51.2)	
7-8	60 (25.4)	74 (27.5)	63 (22.3)	68 (28.3)	

Table A 14: *Continued.*

	Version 1	Version 2	Version 3	Version 4	P-value
Priority - local (%)					0.978
1-2	21 (8.9)	27 (10.0)	25 (8.9)	22 (9.2)	
3-4	80 (33.9)	97 (36.1)	100 (35.5)	87 (36.2)	
5-6	100 (42.4)	116 (43.1)	124 (44.0)	104 (43.3)	
7-8	35 (14.8)	29 (10.8)	33 (11.7)	27 (11.2)	
Priority - price (%)					0.462
1-2	9 (3.8)	5 (1.9)	7 (2.5)	4 (1.7)	
3-4	24 (10.2)	38 (14.1)	39 (13.8)	43 (17.9)	
5-6	111 (47.0)	129 (48.0)	129 (45.7)	106 (44.2)	
7-8	92 (39.0)	97 (36.1)	107 (37.9)	87 (36.2)	
Priority - taste (%)					0.242
1-2	1 (0.4)	7 (2.6)	7 (2.5)	1 (0.4)	
3-4	12 (5.1)	22 (8.2)	23 (8.2)	19 (7.9)	
5-6	111 (47.0)	112 (41.6)	131 (46.5)	110 (45.8)	
7-8	112 (47.5)	128 (47.6)	121 (42.9)	110 (45.8)	
Priority - ease (%)					0.492
1-2	8 (3.4)	8 (3.0)	17 (6.0)	10 (4.2)	
3-4	47 (19.9)	67 (24.9)	69 (24.5)	63 (26.2)	
5-6	131 (55.5)	141 (52.4)	149 (52.8)	128 (53.3)	
7-8	50 (21.2)	53 (19.7)	47 (16.7)	39 (16.2)	

Table A 14: *Continued.*

	Version 1	Version 2	Version 3	Version 4	P-value
Priority - health (%)					0.028
1-2	3 (1.3)	4 (1.5)	8 (2.8)	2 (0.8)	
3-4	21 (8.9)	46 (17.1)	50 (17.7)	30 (12.5)	
5-6	136 (57.6)	136 (50.6)	148 (52.5)	147 (61.3)	
7-8	76 (32.2)	83 (30.9)	76 (27.0)	61 (25.4)	
Consideration of future consequences (mean (SD))	24.58 (3.00)	24.57 (3.18)	24.64 (3.15)	24.98 (3.02)	0.418
Consideration of others (mean (SD))	11.20 (2.46)	11.05 (2.29)	11.11 (2.35)	11.21 (2.12)	0.840
Self deception (mean (SD))	34.57 (6.27)	34.75 (6.13)	34.58 (6.36)	34.48 (6.21)	0.968
Impression management (mean (SD))	37.36 (7.12)	37.08 (7.66)	37.56 (8.04)	37.65 (8.18)	0.843
Questionnaire complexity (%)					0.492
very easy	69 (29.2)	86 (32.0)	83 (29.4)	61 (25.4)	
easy	112 (47.5)	108 (40.1)	124 (44.0)	111 (46.2)	
indifferent	50 (21.2)	68 (25.3)	72 (25.5)	59 (24.6)	
hard	5 (2.1)	7 (2.6)	2 (0.7)	8 (3.3)	
very hard	0 (0.0)	0 (0.0)	1 (0.4)	1 (0.4)	
Questionnaire length (%)					0.535
very short	23 (9.7)	20 (7.4)	24 (8.5)	9 (3.8)	
short	24 (10.2)	26 (9.7)	27 (9.6)	18 (7.5)	
indifferent	156 (66.1)	187 (69.5)	192 (68.1)	180 (75.0)	
long	29 (12.3)	34 (12.6)	35 (12.4)	28 (11.7)	
very long	4 (1.7)	2 (0.7)	4 (1.4)	5 (2.1)	

Table A 14: *Continued.*

	Version 1	Version 2	Version 3	Version 4	P-value
Honest answers (%)					0.203
always	173 (73.3)	203 (75.5)	223 (79.1)	189 (78.8)	
mostly	54 (22.9)	47 (17.5)	42 (14.9)	40 (16.7)	
sometimes	6 (2.5)	13 (4.8)	15 (5.3)	7 (2.9)	
hardly	3 (1.3)	6 (2.2)	1 (0.4)	4 (1.7)	
never	0 (0.0)	0 (0.0)	1 (0.4)	0 (0.0)	
Familiarity collaborating supermarket (%)					0.359
frequent buyer	29 (12.3)	28 (10.4)	30 (10.6)	15 (6.2)	
occasional buyer	11 (4.7)	16 (5.9)	9 (3.2)	16 (6.7)	
heard of	26 (11.0)	32 (11.9)	37 (13.1)	25 (10.4)	
never heard of	170 (72.0)	193 (71.7)	206 (73.0)	184 (76.7)	
Intention to use coupon = no (%)	121 (51.3)	144 (53.5)	140 (49.6)	123 (51.2)	0.840
Batch (%)					0.553
1	18 (7.6)	32 (11.9)	30 (10.6)	23 (9.6)	
2	104 (44.1)	103 (38.3)	111 (39.4)	107 (44.6)	
3	114 (48.3)	134 (49.8)	141 (50.0)	110 (45.8)	
Use coupon = yes (%)	0 (0.0)	7 (2.6)	5 (1.8)	5 (2.1)	0.241
Ordered veggies = yes (%)	0 (0.0)	2 (0.7)	2 (0.7)	4 (1.7)	0.281
Ordered B-choice = yes (%)	0 (0.0)	1 (0.4)	0 (0.0)	3 (1.2)	0.088

Note: N doesn't always add up to 1027. Not all respondents answered every question. Attitudes = 1: totally disagree – 5: totally agree. Priorities = 1: not important at all – 8: very important

APPENDIX 16: RESPONDENT CHARACTERISTICS PER CONSENT TO SHARE DATA

Table A 15: Respondent characteristics per consent to share data

	Consent	No consent	P-value
N	666	361	
Age (%)			<0.001
18-29	129 (19.4)	59 (16.3)	
30-39	135 (20.3)	40 (11.1)	
40-49	119 (17.9)	33 (9.1)	
50-59	98 (14.7)	61 (16.9)	
60-69	105 (15.8)	95 (26.3)	
70-79	73 (11.0)	65 (18.0)	
80plus	7 (1.1)	8 (2.2)	
Sex (%)			0.866
male	324 (48.6)	179 (49.6)	
female	341 (51.2)	181 (50.1)	
rather not say	1 (0.2)	1 (0.3)	
Education (%)			0.565
low1	22 (3.3)	7 (1.9)	
low2	51 (7.7)	25 (6.9)	
middle1	106 (15.9)	59 (16.3)	
middle2	140 (21.0)	96 (26.6)	
middle3	78 (11.7)	39 (10.8)	
high1	156 (23.4)	76 (21.1)	
high2	110 (16.5)	57 (15.8)	
else	3 (0.5)	2 (0.6)	

Table A 15: *Continued.*

	Consent	No consent	P-value
City (%)			<0.001
Amsterdam	62 (9.3)	3 (0.8)	
Rotterdam	333 (50.0)	195 (54.0)	
The Hague	252 (37.8)	152 (42.1)	
Utrecht	5 (0.8)	2 (0.6)	
rather not say	2 (0.3)	0 (0.0)	
other, namely	12 (1.8)	9 (2.5)	
Buy bell pepper (%)			0.009
weekly	296 (44.4)	160 (44.3)	
twice a month	199 (29.9)	78 (21.6)	
monthly	97 (14.6)	66 (18.3)	
twice a year	39 (5.9)	36 (10.0)	
never	35 (5.3)	21 (5.8)	
Buy eggplant (%)			0.003
weekly	95 (14.3)	51 (14.1)	
twice a month	153 (23.0)	54 (15.0)	
monthly	157 (23.6)	74 (20.5)	
twice a year	91 (13.7)	72 (19.9)	
never	170 (25.5)	110 (30.5)	

Table A 15: *Continued.*

	Consent	No consent	P-value
Buy cucumber (%)			0.249
weekly	348 (52.3)	212 (58.7)	
twice a month	138 (20.7)	62 (17.2)	
monthly	95 (14.3)	51 (14.1)	
twice a year	69 (10.4)	27 (7.5)	
never	16 (2.4)	9 (2.5)	
Household size (mean (SD))	2.36 (2.05)	2.12 (1.11)	0.040
Household with kids 0-3 years = yes (%)	83 (19.3)	21 (8.5)	<0.001
Household with kids 4-12 years = yes (%)	110 (25.6)	30 (12.1)	<0.001
Household with kids 13-17 years = yes (%)	55 (12.8)	28 (11.3)	0.651
Household with adults = yes (%)	299 (69.5)	216 (87.1)	<0.001
Groceries (%)			0.361
household	525 (78.8)	298 (82.5)	
self	128 (19.2)	57 (15.8)	
never	13 (2.0)	6 (1.7)	
Groceries store (%)			0.453
daily	248 (38.0)	114 (32.1)	
weekly	362 (55.4)	218 (61.4)	
bimonthly	23 (3.5)	12 (3.4)	
monthly	13 (2.0)	6 (1.7)	
biyearly	3 (0.5)	1 (0.3)	
never	4 (0.6)	4 (1.1)	

Table A 15: *Continued.*

	Consent	No consent	P-value
Groceries online (%)			<0.001
daily	66 (10.1)	4 (1.1)	
weekly	115 (17.6)	47 (13.2)	
bimonthly	71 (10.9)	23 (6.5)	
monthly	60 (9.2)	33 (9.3)	
biyearly	77 (11.8)	47 (13.2)	
never	264 (40.4)	201 (56.6)	
Vegetarian = no (%)	501 (76.7)	326 (91.8)	<0.001
Flexitarian = yes (%)	82 (53.9)	7 (24.1)	0.006
Pescetarian = yes (%)	67 (44.1)	13 (44.8)	1.000
Eggs = yes (%)	60 (39.5)	12 (41.4)	1.000
Dairy = yes (%)	49 (32.2)	11 (37.9)	0.703
Rennet-free cheese = yes (%)	21 (13.8)	4 (13.8)	1.000
Vegan = yes (%)	5 (3.3)	5 (17.2)	0.010
Fixed holdout – bell pepper (%)			<0.001
green	67 (10.6)	22 (6.5)	
yellow	67 (10.6)	9 (2.6)	
red	167 (26.5)	83 (24.4)	
mixed B	294 (46.6)	180 (52.9)	
no	36 (5.7)	46 (13.5)	

Table A 15: *Continued.*

	Consent	No consent	P-value
Fixed holdout – eggplant (%)			<0.001
regular	205 (41.3)	95 (37.8)	
B	223 (45.0)	90 (35.9)	
no	68 (13.7)	66 (26.3)	
Fixed holdout – cucumber (%)			<0.001
regular	170 (26.2)	73 (20.7)	
B	440 (67.7)	230 (65.3)	
no	40 (6.2)	49 (13.9)	
Making ends meet (%)			<0.001
very hard	106 (15.9)	28 (7.8)	
hard	212 (31.8)	96 (26.6)	
easy	233 (35.0)	163 (45.2)	
very easy	115 (17.3)	74 (20.5)	
Smoking (%)			<0.001
no	277 (41.6)	151 (41.8)	
used to	186 (27.9)	134 (37.1)	
occasional	119 (17.9)	24 (6.6)	
daily	84 (12.6)	52 (14.4)	

Table A 15: *Continued.*

	Consent	No consent	P-value
Alcohol - days per week (%)			0.073
never	190 (28.5)	124 (34.3)	
less than once	140 (21.0)	75 (20.8)	
one	97 (14.6)	39 (10.8)	
two	85 (12.8)	39 (10.8)	
three	66 (9.9)	31 (8.6)	
four	25 (3.8)	12 (3.3)	
five	22 (3.3)	5 (1.4)	
six	8 (1.2)	6 (1.7)	
seven	33 (5.0)	30 (8.3)	
Alcohol - glasses per time (mean (SD))	2.57 (3.34)	2.51 (1.96)	0.781
Physical activity - days per week (%)			0.001
never	29 (4.4)	23 (6.4)	
less than once	52 (7.8)	26 (7.2)	
one	99 (14.9)	22 (6.1)	
two	78 (11.7)	40 (11.1)	
three	94 (14.1)	46 (12.7)	
four	69 (10.4)	42 (11.6)	
five	83 (12.5)	45 (12.5)	
six	51 (7.7)	25 (6.9)	
seven	111 (16.7)	92 (25.5)	

Table A 15: *Continued.*

	Consent	No consent	P-value
Nutrition - days per week (%)			0.001
never	10 (1.5)	4 (1.1)	
less than once	21 (3.2)	7 (1.9)	
one	66 (9.9)	9 (2.5)	
two	42 (6.3)	23 (6.4)	
three	81 (12.2)	37 (10.2)	
four	100 (15.0)	60 (16.6)	
five	158 (23.7)	85 (23.5)	
six	103 (15.5)	69 (19.1)	
seven	85 (12.8)	67 (18.6)	
Health (mean (SD))	6.65 (2.27)	7.08 (1.55)	0.001
Attitude - environment (%)			<0.001
1	103 (15.5)	48 (13.3)	
2	86 (12.9)	60 (16.6)	
3	192 (28.8)	140 (38.8)	
4	137 (20.6)	75 (20.8)	
5	148 (22.2)	38 (10.5)	

Table A 15: *Continued.*

	Consent	No consent	P-value
Attitude - fairtrade (%)			0.099
1	65 (9.8)	37 (10.2)	
2	119 (17.9)	55 (15.2)	
3	222 (33.3)	139 (38.5)	
4	151 (22.7)	90 (24.9)	
5	109 (16.4)	40 (11.1)	
Attitude - health 1 (%)			0.103
1	29 (4.4)	16 (4.4)	
2	72 (10.8)	30 (8.3)	
3	218 (32.7)	133 (36.8)	
4	188 (28.2)	117 (32.4)	
5	159 (23.9)	65 (18.0)	
Attitude - health 2 (%)			0.211
1	27 (4.1)	17 (4.7)	
2	78 (11.7)	32 (8.9)	
3	153 (23.0)	104 (28.8)	
4	232 (34.8)	115 (31.9)	
5	176 (26.4)	93 (25.8)	

Table A 15: *Continued.*

	Consent	No consent	P-value
Attitude - health 3 (%)			<0.001
1	50 (7.5)	32 (8.9)	
2	100 (15.0)	56 (15.5)	
3	194 (29.1)	142 (39.3)	
4	173 (26.0)	91 (25.2)	
5	149 (22.4)	40 (11.1)	
Attitude - health 4 (%)			0.003
1	17 (2.6)	11 (3.0)	
2	48 (7.2)	25 (6.9)	
3	169 (25.4)	121 (33.5)	
4	276 (41.4)	153 (42.4)	
5	156 (23.4)	51 (14.1)	
Attitude - local (%)			<0.001
1	15 (2.3)	13 (3.6)	
2	47 (7.1)	19 (5.3)	
3	208 (31.2)	155 (42.9)	
4	234 (35.1)	133 (36.8)	
5	162 (24.3)	41 (11.4)	

Table A 15: *Continued.*

	Consent	No consent	P-value
Attitude - taste (%)			0.074
1	13 (2.0)	5 (1.4)	
2	51 (7.7)	12 (3.3)	
3	137 (20.6)	78 (21.6)	
4	293 (44.0)	162 (44.9)	
5	172 (25.8)	104 (28.8)	
Attitude - money (%)			0.003
1	38 (5.7)	8 (2.2)	
2	39 (5.9)	10 (2.8)	
3	119 (17.9)	81 (22.4)	
4	226 (33.9)	111 (30.7)	
5	244 (36.6)	151 (41.8)	
Attitude - time (%)			0.019
1	153 (23.0)	92 (25.5)	
2	190 (28.5)	105 (29.1)	
3	166 (24.9)	110 (30.5)	
4	107 (16.1)	36 (10.0)	
5	50 (7.5)	18 (5.0)	

Table A 15: *Continued.*

	Consent	No consent	P-value
Attitude - personal (%)			0.002
1	52 (7.8)	29 (8.0)	
2	64 (9.6)	42 (11.6)	
3	274 (41.1)	183 (50.7)	
4	185 (27.8)	81 (22.4)	
5	91 (13.7)	26 (7.2)	
Attitude - label (%)			<0.001
1	52 (7.8)	12 (3.3)	
2	109 (16.4)	46 (12.7)	
3	273 (41.0)	198 (54.8)	
4	151 (22.7)	67 (18.6)	
5	81 (12.2)	38 (10.5)	
Priority - environment (%)			0.001
1-2	106 (15.9)	34 (9.4)	
3-4	196 (29.4)	126 (34.9)	
5-6	268 (40.2)	169 (46.8)	
7-8	96 (14.4)	32 (8.9)	
Priority - fairtrade (%)			<0.001
1-2	74 (11.1)	16 (4.4)	
3-4	153 (23.0)	107 (29.6)	
5-6	308 (46.2)	189 (52.4)	
7-8	131 (19.7)	49 (13.6)	

Table A 15: *Continued.*

	Consent	No consent	P-value
Priority - animal welfare (%)			0.013
1-2	29 (4.4)	15 (4.2)	
3-4	181 (27.2)	99 (27.4)	
5-6	278 (41.7)	181 (50.1)	
7-8	178 (26.7)	66 (18.3)	
Priority - waste (%)			0.789
1-2	23 (3.5)	11 (3.0)	
3-4	132 (19.8)	74 (20.5)	
5-6	333 (50.0)	189 (52.4)	
7-8	178 (26.7)	87 (24.1)	
Priority - local (%)			0.001
1-2	59 (8.9)	36 (10.0)	
3-4	214 (32.1)	150 (41.6)	
5-6	297 (44.6)	147 (40.7)	
7-8	96 (14.4)	28 (7.8)	
Priority - price (%)			0.396
1-2	17 (2.6)	8 (2.2)	
3-4	92 (13.8)	52 (14.4)	
5-6	320 (48.0)	155 (42.9)	
7-8	237 (35.6)	146 (40.4)	

Table A 15: *Continued.*

	Consent	No consent	P-value
Priority - taste (%)			0.353
1-2	12 (1.8)	4 (1.1)	
3-4	53 (8.0)	23 (6.4)	
5-6	289 (43.4)	175 (48.5)	
7-8	312 (46.8)	159 (44.0)	
Priority - ease (%)			0.552
1-2	30 (4.5)	13 (3.6)	
3-4	151 (22.7)	95 (26.3)	
5-6	359 (53.9)	190 (52.6)	
7-8	126 (18.9)	63 (17.5)	
Priority - health (%)			0.092
1-2	15 (2.3)	2 (0.6)	
3-4	91 (13.7)	56 (15.5)	
5-6	359 (53.9)	208 (57.6)	
7-8	201 (30.2)	95 (26.3)	
Consideration of future consequences (mean (SD))	24.56 (3.17)	24.91 (2.94)	0.085
Consideration of others (mean (SD))	11.05 (2.43)	11.31 (2.06)	0.084
Self deception (mean (SD))	34.47 (6.09)	34.83 (6.49)	0.385
Impression management (mean (SD))	36.80 (7.60)	38.54 (7.95)	0.001

Table A 15: *Continued.*

	Consent	No consent	P-value
Questionnaire complexity (%)			<0.001
very easy	227 (34.1)	72 (19.9)	
easy	278 (41.7)	177 (49.0)	
indifferent	144 (21.6)	105 (29.1)	
hard	15 (2.3)	7 (1.9)	
very hard	2 (0.3)	0 (0.0)	
Questionnaire length (%)			<0.001
very short	72 (10.8)	4 (1.1)	
short	74 (11.1)	21 (5.8)	
indifferent	433 (65.0)	282 (78.1)	
long	76 (11.4)	50 (13.9)	
very long	11 (1.7)	4 (1.1)	
Honest answers (%)			0.072
always	523 (78.5)	265 (73.4)	
mostly	103 (15.5)	80 (22.2)	
sometimes	28 (4.2)	13 (3.6)	
hardly	11 (1.7)	3 (0.8)	
never	1 (0.2)	0 (0.0)	
Familiarity collaborating supermarket (%)			<0.001
frequent buyer	100 (15.0)	2 (0.6)	
occasional buyer	44 (6.6)	8 (2.2)	
heard of	85 (12.8)	35 (9.7)	
never heard of	437 (65.6)	316 (87.5)	

Table A 15: *Continued.*

	Consent	No consent	P-value
Intention to use coupon = no (%)	177 (26.6)	351 (97.2)	<0.001
Batch (%)			0.004
1	61 (9.2)	42 (11.6)	
2	256 (38.4)	169 (46.8)	
3	349 (52.4)	150 (41.6)	
Version (%)			0.620
1	157 (23.6)	79 (21.9)	
2	167 (25.1)	102 (28.3)	
3	181 (27.2)	101 (28.0)	
4	161 (24.2)	79 (21.9)	
Use coupon = yes (%)	17 (2.6)	0 (0.0)	<0.001
Ordered veggies = yes (%)	8 (1.2)	0 (0.0)	<0.001
Ordered B-choice = yes (%)	4 (0.6)	0 (0.0)	<0.001

Note: N doesn't always add up to 1027. Not all respondents answered every question. Attitudes = 1: totally disagree – 5: totally agree. Priorities = 1: not important at all – 8: very important

References

REFERENCES

1. Bridges JFP, Jones C. Patient-based health technology assessment: A vision of the future. Int J Technol Assess Health Care. 2007;23(1):30–5.
2. Dirksen CD. The use of research evidence on patient preferences in health care decision-making: issues, controversies and moving forward. Expert Rev Pharmacoecon Outcomes Res. 2014 Dec 1;14(6):785–94.
3. Hansen HP, Lee A. Patient aspects and involvement in HTA: An academic perspective. Pharm Policy Law. 2011;13:123–8.
4. O'Mahony B, Kent A, Aymé S. Pfizer-sponsored satellite symposium at the European Haemophilia Consortium (EHC) Congress: Changing the policy landscape: Haemophilia patient involvement in healthcare decision-making Pfizer-sponsored satellite symposium at the European Haemophilia Consortium (EHC) Congress: Changing the policy landscape: Haemophilia patient involvement in healthcare decision-making. Eur J Haematol. 2014;92:1–8.
5. Ostermann J, Brown DS, de Bekker-Grob EW, Mühlbacher AC, Reed SD. Preferences for Health Interventions: Improving Uptake, Adherence, and Efficiency. Patient - Patient-Centered Outcomes Res. 2017 Aug 1;10(4):511–4.
6. Wale J, Scott AM, Hofmann B, Garner S, Low E, Sansom L. Why patients should be involved in health technology assessment. Int J Technol Assess Health Care. 2017 ed;33(1):1–4.
7. Soekhai V, Whichello C, Levitan B, Veldwijk J, Pinto CA, Donkers B, et al. Methods for exploring and eliciting patient preferences in the medical product lifecycle: a literature review. Drug Discov Today. 2019 Jul 1;24(7):1324–31.
8. Hauber AB, Fairchild AO, Johnson FR. Quantifying Benefit–Risk Preferences for Medical Interventions: An Overview of a Growing Empirical Literature. Appl Health Econ Health Policy. 2013 Aug 1;11(4):319–29.
9. Craig BM, Lancsar E, Mühlbacher AC, Brown DS, Ostermann J. Health Preference Research: An Overview. Patient - Patient-Centered Outcomes Res. 2017 Aug 1;10(4):507–10.
10. Thurstone LL. A law of comparative judgment. Psychol Rev. 1927;34(4):273–86.
11. Manski CF. The structure of random utility models. Theory Decis. 1977 Jul 1;8(3):229–54.
12. McFadden DL. Conditional Logit Analysis of Qualitative Choice Behavior. In: Frontiers in Econometrics [Internet]. P. Zarembka (ed.) Academic Press: New York; 1974 [cited 2021 Aug 11]. p. 105–42. Available from: https://eml.berkeley.edu/reprints/mcfadden/zarembka.pdf
13. Lancsar E, Louviere J. Conducting Discrete Choice Experiments to Inform Healthcare Decision Making. PharmacoEconomics. 2008 Aug 1;26(8):661–77.
14. Ryan M, Farrar S. Using conjoint analysis to elicit preferences for health care. BMJ. 2000 Jun 3;320(7248):1530–3.
15. Bridges JFP, Hauber AB, Marshall D, Lloyd A, Prosser LA, Regier DA, et al. Conjoint Analysis Applications in Health—a Checklist: A Report of the ISPOR Good Research Practices for Conjoint Analysis Task Force. Value Health. 2011 Jun 1;14(4):403–13.
16. Coast J, Al-Janabi H, Sutton EJ, Horrocks SA, Vosper AJ, Swancutt DR, et al. Using qualitative methods for attribute development for discrete choice experiments: issues and recommendations. Health Econ. 2012;21(6):730–41.

17. Johnson RF, Lancsar E, Marshall D, Kilambi V, Mühlbacher A, Regier DA, et al. Constructing experimental designs for discrete-choice experiments: report of the ISPOR Conjoint Analysis Experimental Design Good Research Practices Task Force. Value Health J Int Soc Pharmacoeconomics Outcomes Res. 2013 Feb;16(1):3–13.
18. Louviere JJ, Hensher DA, Swait JD. Stated Choice Methods: Analysis and Applications [Internet]. Cambridge: Cambridge University Press; 2000 [cited 2022 May 17]. Available from: https://www.cambridge.org/core/books/stated-choice-methods/0F20174043208130BA57E9E328EAFDCA
19. Hess S, Daly A. Handbook of Choice Modelling [Internet]. Handbook of Choice Modelling. Edward Elgar Publishing; 2014 [cited 2022 May 17]. Available from: https://www.elgaronline.com/view/edcoll/9781781003145/9781781003145.xml
20. Hensher DA, Rose JM, Greene WH. Applied choice analysis. 2015.
21. Hauber AB, González JM, Groothuis-Oudshoorn CGM, Prior T, Marshall DA, Cunningham C, et al. Statistical Methods for the Analysis of Discrete Choice Experiments: A Report of the ISPOR Conjoint Analysis Good Research Practices Task Force. Value Health. 2016 Jun 1;19(4):300–15.
22. Soekhai V, de Bekker-Grob EW, Ellis AR, Vass CM. Discrete Choice Experiments in Health Economics: Past, Present and Future. Pharmacoeconomics. 2019;37(2):201–26.
23. U.S. Food and Drug Administration (FDA). Patient Preference Information - Voluntary Submission, Review in Premarket Approval Applications, Humanitarian Device Exemption Applications, and De Novo Requests, and Inclusion in Decision Summaries and Device Labeling [Internet]. 2016 [cited 2021 Aug 11]. Available from: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/patient-preference-information-voluntary-submission-review-premarket-approval-applications
24. Levitan B, Hauber AB, Damiano MG, Jaffe R, Christopher S. The ball is in your court: agenda for research to advance the science of patient preferences in the regulatory review of medical devices in the United States. Patient - Patient-Centered Outcomes Res. 2017;
25. Medical Device Innovation Consortium (MDIC). A Framework for Incorporating Information on Patient Preferences Regarding Benefit and Risk into Regulatory Assessments of New Medical Technology.
26. Telser H, Zweifel P. Validity of discrete-choice experiments evidence for health risk reduction. Appl Econ. 2007 Jan 1;39(1):69–78.
27. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol. 2010 Jul 1;63(7):737–45.
28. Whitty JA, Walker R, Golenko X, Ratcliffe J. A Think Aloud Study Comparing the Validity and Acceptability of Discrete Choice and Best Worst Scaling Methods. PLOS ONE. 2014 Apr 23;9(4):e90635.
29. Krucien N, Gafni A, Pelletier-Fleury N. Empirical Testing of the External Validity of a Discrete Choice Experiment to Determine Preferred Treatment Option: The Case of Sleep Apnea. Health Econ. 2015;24(8):951–65.
30. Rakotonarivo OS, Schaafsma M, Hockley N. A systematic review of the reliability and validity of discrete choice experiments in valuing non-market environmental goods. J Environ Manage. 2016 Dec 1;183:98–109.

31. Janssen EM, Marshall DA, Hauber AB, Bridges JFP. Improving the quality of discrete-choice experiments in health: how can we assess validity and reliability? Expert Rev Pharmacoecon Outcomes Res. 2017 Nov 2;17(6):531–42.
32. Quaife M, Terris-Prestholt F, Di Tanna GL, Vickerman P. How well do discrete choice experiments predict health choices? A systematic review and meta-analysis of external validity. Eur J Health Econ HEPAC Health Econ Prev Care. 2018 Nov;19(8):1053–66.
33. Ryan M, Bate A, Eastmond C, Ludbrook A. Use of discrete choice experiments to elicit preferences. Qual Health Care QHC. 2001 Sep;10(Suppl 1):i55–60.
34. Ryan M, Gerard K. Using discrete choice experiments to value health care programmes: current practice and future research reflections. Appl Health Econ Health Policy. 2003;2(1):55–64.
35. de Bekker-Grob EW, Hol L, Donkers B, Van Dam L, Habbema JDF, Van Leerdam ME, et al. Labeled versus Unlabeled Discrete Choice Experiments in Health Economics: An Application to Colorectal Cancer Screening. Value Health. 2010;13(2):315–23.
36. Kenny P, Hall J, Viney R, Haas M. Do Participants Understand a Stated Preference Health Survey? A Qualitative Approach to Assessing Validity. Int J Technol Assess Health Care. 2003 Dec;19(4):664–81.
37. Johnson FR, Yang JC, Reed SD. The Internal Validity of Discrete Choice Experiment Data: A Testing Tool for Quantitative Assessments. Value Health. 2019 Feb 1;22(2):157–60.
38. Özdemir S, Mohamed AF, Johnson FR, Hauber AB. Who pays attention in stated-choice surveys? Health Econ. 2010;19(1):111–8.
39. Hougaard JL, Tjur T, Østerdal LP. On the meaningfulness of testing preference axioms in stated preference discrete choice experiments. Eur J Health Econ. 2012 Aug 1;13(4):409–17.
40. Swait J, Louviere J. The role of the scale parameter in the estimation and comparison of multinational logit models. JMR J Mark Res Chic. 1993 Aug;30(3):305.
41. World Health Organization. HTA Definitions [Internet]. Health Technology Assessment. 2018. Available from: http://www.who.int/health-technology-assessment/about/Defining/en/
42. Boivin A, Green J, Meulen J van der, Légaré F. Why consider patients' preferences?: A discourse analysis of clinical practice guideline developers. Med Care [Internet]. 2009; Available from: http://journals.lww.com/lww-medicalcare/Abstract/2009/08000/Why_Consider_Patients__Preferences___A_Discourse.13.aspx
43. Montori VM, Brito JP, Murad MH. The optimal practice of evidence-based medicine: incorporating patient preferences in practice guidelines. Jama. 2013;310(23):2503–4.
44. Whitty JA, Fraenkel L, Saigal CS, Groothuis-Oudshoorn CGM, Regier DA, Marshall DA. Assessment of Individual Patient Preferences to Inform Clinical Practice. Patient - Patient-Centered Outcomes Res. 2017 Aug 1;10(4):519–21.
45. Domecq JP, Prutsky G, Elraiyah T, Wang Z, Nabhan M, Shippee N, et al. Patient engagement in research: a systematic review. BMC Health Serv Res. 2014;14(1):89.
46. Forsythe LP, Szydlowski V, Murad MH, Ip S, Wang Z, Elraiyah TA, et al. A Systematic Review of Approaches for Engaging Patients for Research on Rare Diseases. J Gen Intern Med. 2014 Aug;29(3):788–800.
47. Ho M, Saha A, McCleary KK, Levitan B, Christopher S, Zandlo K, et al. A Framework for Incorporating Patient Preferences Regarding Benefits and Risks into Regulatory Assessment of Medical Technologies. Value Health. 2016;19:746–50.

48. Johnson FR, Beusterien K, Özdemir S, Wilson L. Giving patients a meaningful voice in United States regulatory decision making: the role for health preference research. Patient - Patient-Centered Outcomes Res. 2017;10(4):523–6.
49. Mott DJ. Incorporating Quantitative Patient Preference Data into Healthcare Decision Making Processes: Is HTA Falling Behind? Patient - Patient-Centered Outcomes Res. 2018 Jun 1;11(3):249–52.
50. Food and Drug Administration. Patient Preference Information - Voluntary Submission, Review in Premarket Approval Applications, Humanitarian Device Exemption Applications, and De Novo Requests, and Inclusion in Decision Summaries and Device Labeling [Internet]. Health Center for Devices and Radiological. 2020 [cited 2021 Jun 10]. Available from: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/patient-preference-information-voluntary-submission-review-premarket-approval-applications
51. Bridges JFP; J C. Patient-based health technology assessment: A vision of the future. Int J Technol Assess Health Care. 2007;23:30–5.
52. Gagnon MP, Desmartis M, Lepage-Savary D, Gagnon J, St-Pierre M, Rhainds M, et al. Introducing patients' and the public's perspectives to health technology assessment: A systematic review of international experiences. Int J Technol Assess Health Care. 2011;27:31–42.
53. Facey KM, Hansen HP, Single ANV. Patient Involvement in Health Technology Assessment. 1st ed. Singapore: Adis; 2017.
54. Weernink MG, Janus SI, Van Til JA, Raisch DW, Van Manen JG, IJzerman MJ. A systematic review to identify the use of preference elicitation methods in healthcare decision making. Pharm Med. 2014;28(4):175–85.
55. Janssen EM, Marshall DA, Hauber AB, Bridges JFP. Improving the quality of discrete-choice experiments in health: how can we assess validity and reliability? Expert Rev Pharmacoecon Outcomes Res. 2017 Nov 2;17(6):531–42.
56. Facey K, Boivin A, Gracia J, Hansen HP, Scalzo AL, Mossman J, et al. Patients' perspectives in health technology assessment: a route to robust evidence and fair deliberation. Int J Technol Assess Health Care. 2010;26(3):334–40.
57. Facey KM, Bedlington N, Berglas S, Bertelsen N, Single ANV, Thomas V. Putting Patients at the Centre of Healthcare: Progress and Challenges for Health Technology Assessments. Patient - Patient-Centered Outcomes Res. 2018 Dec 1;11(6):581–9.
58. Moher D LA Tetzlaff J, Altman DG, The PRISMA Group. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med. 2009;6(6):e1000097.
59. Utens CMA, Dirksen CD, van der Weijden T, Joore MA. How to integrate research evidence on patient preferences in pharmaceutical coverage decisions and clinical practice guidelines: A qualitative study among Dutch stakeholders. Health Policy. 2016 Jan;120:120–8.
60. Brooker AS, Carcone S, Witteman W, Krahn M. Quantitative patient preference evidence for health technology assessment: A case study. Int J Technol Assess Health Care. 2013;29:290–300.
61. DeJean D, Giacomini M, Simeonov D, Smith A. Finding Qualitative Research Evidence for Health Technology Assessment. Qual Health Res. 2016 Aug;26:1307–17.

62. Rashid A, Thomas V, Shaw T, Leng G. Patient and Public Involvement in the Development of Healthcare Guidance: An Overview of Current Methods and Future Challenges. Patient. 2017;10:277–82.
63. Gagnon MP, Desmartis M, Gagnon J, St-Pierre M, Gauvin FP, Rhainds M, et al. Introducing the patient's perspective in hospital health technology assessment (HTA): the views of HTA producers, hospital managers and patients. Health Expect. 2014;17:888–900.
64. Buck D, Gamble C, Dudley L, Preston J, Hanley B, Williamson PR, et al. From plans to actions in patient and public involvement: Qualitative study of documented plans and the accounts of researchers and patients sampled from a cohort of clinical trials. BMJ Open [Internet]. 2014;4.
Available from: https://bmjopen.bmj.com/content/4/12/e006400
65. Mühlbacher AC, Kaczynski A. Making Good Decisions in Healthcare with Multi-Criteria Decision Analysis: The Use, Current Research and Future Development of MCDA. Appl Health Econ Health Policy. 2016;14:29–40.
66. Mühlbacher AC, Sadler A. The Probabilistic Efficiency Frontier: A Framework for Cost-Effectiveness Analysis in Germany Put into Practice for Hepatitis C Treatment Options. Value Health. 2;20:266–72.
67. Dipankui MT, Gagnon MP, Desmartis M, Légaré F, Piron F, Gagnon J, et al. Evaluation of Patient Involvement in a Health Technology Assessment. Int J Technol Assess Health Care. 2015;31:166–70.
68. Kreis J, Schmidt H. Public engagement in health technology assessment and coverage decisions: A study of experiences in France, Germany, and the United Kingdom. J Health Polit Policy Law. 2013;38:89–122.
69. Abelson J, Wagner F, DeJean D, Boesveld S, Gauvin FP, Bean S, et al. Public and Patient Involvement in Health Technology Assessment: A Framework for Action. Int J Technol Assess Health Care. 2016;32:256–64.
70. Utens CMA, van der Weijden T, Joore MA, Dirksen CD. The use of research evidence on patient preferences in pharmaceutical coverage decisions and clinical practice guideline development: exploratory study into current state of play and potential barriers. BMC Health Serv Res. 2014 Nov;14.
71. Wortley S, Wale J, Grainger D, Murphy P. Moving beyond the rhetoric of patient input in health technology assessment deliberations. Aust Health Rev. 2017;41:170–2.
72. Brereton L, Ingleton C, Gardiner C, Goyder E, Mozygemba K, Lysdahl KB, et al. Lay and professional stakeholder involvement in scoping palliative care issues: Methods used in seven European countries. Palliat Med. 2017;31:181–92.
73. Wahlster P, Brereton L, Burns J, Hofmann B, Mozygemba K, Oortwijn W, et al. An integrated perspective on the assessment of technologies: INTEGRATE-HTA. Int J Technol Assess Health Care. 2017;33:544–51.
74. Danner M, Vennedey V, Hiligsmann M, Fauser S, Gross C, Stock S. How Well Can Analytic Hierarchy Process be Used to Elicit Individual Preferences? Insights from a Survey in Patients Suffering from Age-Related Macular Degeneration. Patient. 2016 Oct;9:481–92.
75. Hailey D, Werkö S, Bakri R, Cameron A, Göhlen B, Myles S, et al. Involvement of consumers in health technology assessment activities by inahta agencies. Int J Technol Assess Health Care. 2013;29:79–83.

76. Single ANV, Scott AM, Wale J. Developing Guidance on Ethics for Patient Groups Collecting and Reporting Patient Information for Health Technology Assessments. Patient. 2016;9:1–4.
77. Kreis J, Puhan MA, Schunemann HJ, Dickersin K. Consumer involvement in systematic reviews of comparative effectiveness research. Health Expect. 2013 Dec;16:323–37.
78. Mott DJ, Najafzadeh M. Whose preferences should be elicited for use in health-care decision-making? A case study using anticoagulant therapy. Expert Rev Pharmacoecon Outcomes Res. 2016;16:33–9.
79. Thokala P, Devlin N, Marsh K, Baltussen R, Boysen M, Kalo Z, et al. Multiple criteria decision analysis for health care decision making - An introduction: Report 1 of the ISPOR MCDA Emerging Good Practices Task Force. Value Health. 2016;19:1–13.
80. Evers P, Greene L, Ricciardi M. The importance of early access to medicines for patients suffering from rare diseases. Regul Rapp. 2016;13ID-199:5–8.
81. Mühlbacher AC, Bridges JFP, Bethge S, Dintsios CM, Schwalm A, Gerber-Grote A, et al. Preferences for antiviral therapy of chronic hepatitis C: a discrete choice experiment. Eur J Health Econ. 2017;18:155–65.
82. Berglas S, Jutai L, MacKean G, Weeks L. Patients' perspectives can be integrated in health technology assessments: an exploratory analysis of CADTH Common Drug Review. Res Involv Engag. 2016;2:21.
83. Iskrov G, Stefanov R. Criteria for Drug Reimbursement Decision-Making: An Emerging Public Health Challenge in Bulgaria. Balk Med J. 2016 Jan;33:27–35.
84. Moreira T. Understanding the role of patient organizations in health technology assessment. Health Expect. 2015;18:3349–57.
85. Tordrup D, Mossman J, Kanavos P. Responsiveness of the Eq-5d to Clinical Change: Is the Patient Experience Adequately Represented? Int J Technol Assess Health Care. 2014 Jan;30:10–9.
86. Hämeen-Anttila K, Komulainen J, Enlund H, Mäkelä M, Mäkinen E, Rannanheimo P, et al. Incorporating patient perspectives in health technology assessments and clinical practice guidelines. Res Soc Adm Pharm. 2016;12:903–13.
87. Weeks L, Polisena J, Scott AM, Holtorf AP, Staniszewska S, Facey K. Evaluation of patient and public involvement initiatives in health technology assessment: a survey of international agencies. Int J Technol Assess Health Care. 2017;33:715–23.
88. Husereau D, Henshall C, Sampietro-Colom L, Thomas S. Changing Health Technology Assessment Paradigms. Int J Technol Assess Health Care. 2016;32:191–9.
89. Di Paolo A, Sarkozy F, Ryll B, Siebert U. Personalized medicine in Europe: not yet personal enough? BMC Health Serv Res. 2017 Apr 19;17(1):289.
90. Doctor J, MacEwan JP. Limitations of traditional health technology assessment methods and implications for the evaluation of novel therapies. Curr Med Res Opin. 2017 Sep 2;33(9):1635–42.
91. Mossman J, Baker MG, Kossler I. Patient Power as a Driver for Change: Reality or Rhetoric? Glob Policy. 2017 Mar;8:133–8.
92. Kievit W, Tummers M, Van Hoorn R, Booth A, Mozygemba K, Refolo P, et al. Taking patient heterogeneity and preferences into account in health technology assessments. Int J Technol Assess Health Care. 2017;33:562–9.

93. Brereton L, Wahlster P, Mozygemba K, Lysdahl KB, Burns J, Polus S, et al. Stakeholder involvement throughout health technology assessment: An example from palliative care. Int J Technol Assess Health Care. 2017;33:552–61.
94. Scott AM, Wale JL. Patient advocate perspectives on involvement in HTA: an international snapshot. Res Involv Engag. 2017;3:2.
95. Douglas CMW, Wilcox E, Burgess M, Lynd LD. Why orphan drug coverage reimbursement decision-making needs patient and public involvement. Health Policy. 2015;119:588–96.
96. Lopes E, Street J, Carter D, Merlin T. Involving patients in health technology funding decisions: stakeholder perspectives on processes used in Australia. Health Expect. 2016;19:331–44.
97. Gagnon MP, Desmartis M, Gagnon J, St-Pierre M, Rhainds M, Coulombe M, et al. Framework for User Involvement in Health Technology Assessment at the Local Level: Views of Health Managers, User Representatives, and Clinicians. Int J Technol Assess Health Care. 2015;31:68–77.
98. Cassels A. Regional perspectives Patient speaking for patients: What constitutes genuine patient input into pharmaceutical policymaking? Int J Health Gov. 2016;21:89–95.
99. Mühlbacher AC, Juhnke C, Beyer AR, Garner S. Patient-Focused Benefit-Risk Analysis to Inform Regulatory Decisions: The European Union Perspective. Value Health. 2016 Sep;19:734–40.
100. Whitty JA. An international survey of the public engagement practices of health technology assessment organizations. Value Health. 2013 Jan;16:155–63.
101. Janssen IM, Scheibler F, Gerhardus A. Importance of hemodialysis-related outcomes: Comparison of ratings by a self-help group, clinicians, and health technology assessment authors with those by a large reference group of patients. Patient Prefer Adherence. 2016;10:2491–500.
102. Ducey A, Ross S, Pott T, Thompson C. The moral economy of health technology assessment: an empirical qualitative study. Evid Polciy. 2017 Jan;13:7–27.
103. MacLeod TE, Harris AH, Mahal A. Stated and Revealed Preferences for Funding New High-Cost Cancer Drugs: A Critical Review of the Evidence from Patients, the Public and Payers. Patient. 2016;9:201–22.
104. Morel T, Cano SJ. Measuring what matters to rare disease patients - reflections on the work by the IRDiRC taskforce on patient-centered outcome measures. Orphanet J Rare Dis. 2017 Nov;12.
105. Low E. Potential for patients and patient-driven organizations to improve evidence for health technology assessment. Int J Technol Assess Health Care. 2015;31:226–7.
106. Narbutas Š, York K, Stein BD, Magsanoc-Alikpala K, Majima Y, Kalo Z, et al. Overview on patient centricity in cancer care. Front Pharmacol [Internet]. 2017;8. Available from: http://www.embase.com/search/results?subaction=viewrecord&from=export&id=L618536766 https://fjfsdata01prod.blob.core.windows.net/articles/files/294396/pubmed-zip/.versions/1/.package-entries/fphar-08-00698/fphar-08-00698.pdf?sv=2015-12-11&sr=b&sig=ogEsXlDcoqNQlpkzsdDEbDGT9ucMYta9EOuQiuBdbXg%3D&se=2018-04-03T12%3A09%3A40Z&sp=r&rscd=attachment%3B%20filename%2A%3DUTF--8%27%27fphar-08-00698.pdf
107. Burke W, Trinidad SB, Press NA. Essential elements of personalized medicine. Urol Oncol-Semin Orig Investig. 2014 Feb;32:193–7.

108. Mühlbacher AC. Patient-centric HTA: Different strokes for different folks. Expert Rev Pharmacoecon Outcomes Res. 2015;15:591–7.
109. Beresniak A, Medina-Lara A, Auray JP, De Wever A, Praet JC, Tarricone R, et al. Validation of the underlying assumptions of the quality-adjusted life-years outcome: results from the ECHOUTCOME European project. Pharmacoeconomics. 2015 Jan;33:61–9.
110. Gagnon MP, Candas B, Desmartis M, Gagnon J, La Roche D, Rhainds M, et al. Involving patient in the early stages of health technology assessment (HTA): a study protocol. BMC Health Serv Res. 2014;14:273.
111. Kleme J, Pohjanoksa-Mäntylä M, Airaksinen M, Enlund H, Kastarinen H, Peura P, et al. Patient perspective in health technology assessment of pharmaceuticals in Finland. Int J Technol Assess Health Care. 2014;30:306–11.
112. Drummond M, Tarricone R, Torbica A. Assessing the added value of health technologies: Reconciling different perspectives. Value Health. 2013;16:S7–13.
113. Li H, Ngorsuraches S. Revisit what is next for pharmacoeconomics and outcomes research in Asia. Value Health Reg Issues. 2014;3:1–4.
114. Dalle Fratte CF, Passerini A, Vivori C, Dalla Palma P, Guarrera GM. The relevance of citizen involvement in Health Technology Assessment. A concrete application in the assessment of HPV co-testing in the Autonomous Province of Trento. Epidemiol Biostat Public Health. 2015;12.
115. Kennedy-Martin T, Paczkowski R, Rayner S. Utility values in diabetic kidney disease: A literature review. Curr Med Res Opin. 2015;31:1271–82.
116. Payakachat N, Ali MM, Tilford JM. Can The EQ-5D Detect Meaningful Change? A Systematic Review. Pharmacoeconomics. 2015;33:1137–54.
117. Weernink MGM, van Til JA, Groothuis-Oudshoorn CGM, Ijzerman MJ. Patient and Public Preferences for Treatment Attributes in Parkinson's Disease. Patient. 2017 Dec;10:763–72.
118. Facey K, Granados A, Guyatt G, Kent A, Shah N, Van Der Wilt GJ, et al. Generating health technology assessment evidence for rare diseases. Int J Technol Assess Health Care. 2014;30:416–22.
119. Mühlbacher AC, Bethge S, Kaczynski A. Treatment after Acute Coronary Syndrome: Analysis of Patient's Priorities with Analytic Hierarchy Process. Int J Technol Assess Health Care. 2016;32:284–91.
120. Morgan H, Hoddinott P, Thomson G, Crossland N, Farrar S, Yi D, et al. Benefits of Incentives for Breastfeeding and Smoking cessation in pregnancy (BIBS): a mixed-methods study to inform trial design. Health Technol Assess. 2015 Apr;19:1–522, vii–viii.
121. Chen RC. Comparative effectiveness research in oncology: The promise, challenges, and opportunities. Semin Radiat Oncol. 2014;24:1–4.
122. Regier DA, Bentley C, Mitton C, Bryan S, Burgess MM, Chesney E, et al. Public engagement in priority-setting: Results from a pan-Canadian survey of decision-makers in cancer control. Soc Sci Med. 2014 Dec;122:130–9.
123. Versteegh MM, Brouwer WBF. Patient and general public preferences for health states: A call to reconsider current guidelines. Soc Sci Med. 2016 Sep 1;165:66–74.
124. Cubi-Molla P, Shah K, Burström K. Experience-Based Values: A Framework for Classifying Different Types of Experience in Health Valuation Research. Patient - Patient-Centered Outcomes Res. 2018;11(3):253–70.
125. Drummond M, Brixner D, Gold M, Kind P, McGuire A, Nord E, et al. Toward a consensus on the QALY. Value Health. 2009;12:S31–5.

126. Menzel P, Dolan P, Richardson J, Olsen JA. The role of adaptation to disability and disease in health state valuation: a preliminary normative analysis. Soc Sci Med. 2002;55(12):2149–58.
127. Nord E, Daniels N, Kamlet M. QALYs: some challenges. Value Health. 2009;12:S10–5.
128. Ubel PA, Loewenstein G, Jepson C. Whose quality of life? A commentary exploring discrepancies between health state evaluations of patients and the general public. Qual Life Res. 2003;12(6):599–607.
129. The University of Sheffield. Extending the QALY [Internet]. Available from: https://scharr.dept.shef.ac.uk/e-qaly/
130. Marsh KD, Sculpher M, Caro JJ, Tervonen T. The Use of MCDA in HTA: Great Potential, but More Effort Needed. Value Health. 2018 Apr 1;21(4):394–7.
131. Campillo-Artero C, Puig-Junoy J, Culyer AJ. Does MCDA Trump CEA? Appl Health Econ Health Policy. 2018 Apr;16(2):147–51.
132. Angelis A, Kanavos P. Comment on: "Does MCDA Trump CEA?" Appl Health Econ Health Policy [Internet]. 2018 Nov 21; Available from: https://doi.org/10.1007/s40258-018-0445-z
133. Timmermans DRM. What Clinicians Can Offer: Assessing and Communicating Probabilities for Individual Patient Decision Making. Horm Res Paediatr. 1999;51(Suppl. 1):58–66.
134. Austin JC, Hippman C, Honer WG. Descriptive and numeric estimation of risk for psychotic disorders among affected individuals and relatives: Implications for clinical practice. Psychiatry Res. 2012 Mar 30;196(1):52–6.
135. Montgomery A, Fahey T. How do patients' treatment preferences compare with those of clinicians? Qual Health Care QHC. 2001 Sep;10(Suppl 1):i39–43.
136. Rosen AB, Tsai JS, Downs SM. Variations in Risk Attitude across Race, Gender, and Education. Med Decis Making. 2003 Nov 1;23(6):511–7.
137. Cher DJ, Miyamoto J, Lenert LA. Incorporating risk attitude into Markov-process decision models: importance for individual decision making. Med Decis Mak Int J Soc Med Decis Mak. 1997 Sep;17(3):340–50.
138. Woodward RS, Schnitzler MA, Kvols LK. Reduced uncertainty as a diagnostic benefit: an initial assessment of somatostatic receptor scintigraphy's value in detecting distant metastases of carcinoid liver tumours. Health Econ. 1998 Mar;7(2):149–60.
139. Dieteren CM, Brouwer WBF, van Exel J. How do combinations of unhealthy behaviors relate to attitudinal factors and subjective health among the adult population in the Netherlands? BMC Public Health. 2020 Apr 3;20(1):441.
140. Himmler S, van Exel J, Brouwer W. Did the COVID-19 pandemic change the willingness to pay for an early warning system for infectious diseases in Europe? Eur J Health Econ [Internet]. 2021 Jul 20 [cited 2021 Sep 13]; Available from: https://doi.org/10.1007/s10198-021-01353-6
141. Barfoed BL, Paulsen MS, Christensen PM, Halvorsen PA, Kjær T, Larsen ML, et al. Associations between patients' risk attitude and their adherence to statin treatment – a population based questionnaire and register study. BMC Fam Pract. 2016 Dec;17(1):28.
142. Ostermann J, Brown DS, de Bekker-Grob EW, Mühlbacher AC, Reed SD. Preferences for Health Interventions: Improving Uptake, Adherence, and Efficiency. Patient - Patient-Centered Outcomes Res. 2017 Aug 1;10(4):511–4.

143. Russo S, Monzani D, Pinto CA, Vergani L, Marton G, Falahee M, et al. Taking into Account Patient Preferences: A Consensus Study on the Assessment of Psychological Dimensions Within Patient Preference Studies. Patient Prefer Adherence. 2021 Jun 18;15:1331–45.
144. Russo S, Jongerius C, Faccio F, Pizzoli SFM, Pinto CA, Veldwijk J, et al. Understanding Patients' Preferences: A Systematic Review of Psychological Instruments Used in Patients' Preference and Decision Studies. Value Health. 2019 Apr 1;22(4):491–501.
145. Prosser LA, Wittenberg E. Do Risk Attitudes Differ across Domains and Respondent Types? Med Decis Making. 2007 mei;27(3):281–7.
146. Weber EU, Blais AR, Betz NE. A domain-specific risk-attitude scale: measuring risk perceptions and risk behaviors. J Behav Decis Mak. 2002;15(4):263–90.
147. Huls SPI, van Osch SMC, Brouwer WBF, van Exel J, Stiggelbout AM. Psychometric evaluation of the Health-Risk Attitude Scale (HRAS-13): assessing the reliability, dimensionality and validity in the general population and a patient population. Psychol Health. 2020 Dec 1;1–17.
148. Zhang DC, Highhouse S, Nye CD. Development and validation of the General Risk Propensity Scale (GRiPS). J Behav Decis Mak. 2019;32(2):152–67.
149. Charness G, Garcia T, Offerman T, Villeval MC. Do measures of risk attitude in the laboratory predict behavior under risk in and outside of the laboratory? J Risk Uncertain [Internet]. 2020 Jul 18 [cited 2020 Jul 28]; Available from: https://doi.org/10.1007/s11166-020-09325-6
150. European Medicines Agency (EMA). The patient's voice in the evaluation of medicines [Internet]. 2013 [cited 2021 Aug 11]. Available from: https://www.ema.europa.eu/en/documents/report/report-workshop-patients-voice-evaluation-medicines_en.pdf
151. Harrison M, Rigby D, Vass C, Flynn T, Louviere J, Payne K. Risk as an attribute in discrete choice experiments: a systematic review of the literature. The Patient. 2014;7(2):151–70.
152. Tsuge T, Kishimoto A, Takeuchi K. A Choice Experiment Approach to the Valuation of Mortality. J Risk Uncertain. 2005 Jul;31(1):73–95.
153. Huls SPI, Whichello CL, van Exel J, Uyl-de Groot CA, de Bekker-Grob EW. What Is Next for Patient Preferences in Health Technology Assessment? A Systematic Review of the Challenges. Value Health. 2019 Nov 1;22(11):1318–28.
154. Whichello C, van Overbeeke E, Janssens R, Schölin Bywall K, Russo S, Veldwijk J, et al. Factors and Situations Affecting the Value of Patient Preference Studies: Semi-Structured Interviews in Europe and the US. Front Pharmacol [Internet]. 2019 [cited 2021 Aug 11];0. Available from: https://www.frontiersin.org/articles/10.3389/fphar.2019.01009/full
155. Vass C, Rigby D, Payne K. "I Was Trying to Do the Maths": exploring the impact of risk communication in discrete choice experiments. Patient-Patient-Centered Outcomes Res. 2019;12(1):113–23.
156. Nguyen F, Carrere MO, Moumjid N. Framing Effects of Risk Communication in Health-Related Decision Making; Learning from a Discrete Choice Experiment [Internet]. Rochester, NY: Social Science Research Network; 2009 Nov [cited 2022 Mar 2]. Report No.: ID 1513293. Available from: https://papers.ssrn.com/abstract=1513293
157. de Bekker-Grob EW, Swait JD, Kassahun HT, Bliemer MCJ, Jonker MF, Veldwijk J, et al. Are Healthcare Choices Predictable? The Impact of Discrete Choice Experiment Designs and Models. Value Health. 2019 Sep 1;22(9):1050–62.

158. Veldwijk J, van der Heide I, Rademakers J, Schuit AJ, de Wit GA, Uiters E, et al. Preferences for Vaccination: Does Health Literacy Make a Difference? Med Decis Making. 2015 Nov 1;35(8):948–58.
159. Visser LA, Huls SPI, Uyl-de Groot CA, de Bekker-Grob EW, Redekop WK. An implantable device to treat multiple sclerosis: A discrete choice experiment on patient preferences in three European countries. J Neurol Sci. 2021 Jul 24;117587.
160. Ancillotti M, Eriksson S, Andersson DI, Godskesen T, Nihlén Fahlquist J, Veldwijk J. Preferences regarding antibiotic treatment and the role of antibiotic resistance: A discrete choice experiment. Int J Antimicrob Agents. 2020 Dec 1;56(6):106198.
161. Arslan IG, Huls SPI, Bekker-Grob EW de, Rozendaal R, Persoons MCT, Hell MES van, et al. Patients', healthcare providers', and insurance company employees' preferences for knee and hip osteoarthritis care: a discrete choice experiment. Osteoarthritis Cartilage [Internet]. 2020 Jul 14 [cited 2020 Aug 5];28(10). Available from: https://www.oarsijournal.com/article/S1063-4584(20)31062-1/abstract
162. Anderhub V, Güth W, Gneezy U, Sonsino D. On the Interaction of Risk and Time Preferences: An Experimental Study. Ger Econ Rev. 2001 Aug 1;2(3):239–53.
163. Gafni A, Torrance GW. Risk Attitude and Time Preference in Health. Manag Sci. 1984;30(4):440–51.
164. Conroy RM, Smyth O, Siriwardena R, Fernandes P. Health anxiety and characteristics of self-initiated general practitioner consultations. J Psychosom Res. 1999 Jan 1;46(1):45–50.
165. Charpentier CJ, Aylward J, Roiser JP, Robinson OJ. Enhanced Risk Aversion, But Not Loss Aversion, in Unmedicated Pathological Anxiety. Biol Psychiatry. 2017 Jun 15;81(12):1014–22.
166. ChoiceMetrics. Ngene User Manual [Internet]. 2018 [cited 2021 Aug 26]. Available from: http://www.choice-metrics.com/NgeneManual120.pdf
167. Islam A, Smyth R, Tan HA, Wang LC. Survey measures versus incentivized measures of risk preferences: Evidence from sex workers' risky sexual transactions. Soc Sci Med. 2019 oktober;238:112497.
168. European qualifications framework (EQF) [Internet]. Cedefop. 2009 [cited 2021 Jul 29]. Available from: https://www.cedefop.europa.eu/en/events-and-projects/projects/european-qualifications-framework-eqf
169. Netherlands Qualification Framework (NLQF). Niveaus [Internet]. [cited 2021 Jun 18]. Available from: https://www.nlqf.nl/nlqf-niveaus
170. Central Bureau of Statistics (CBS). Opleidingsniveau [Internet]. Centraal Bureau voor de Statistiek. [cited 2021 Jun 18]. Available from: https://www.cbs.nl/nl-nl/nieuws/2019/33/verschil-levensverwachting-hoog-en-laagopgeleid-groeit/opleidingsniveau
171. Ishikawa H, Nomura K, Sato M, Yano E. Developing a measure of communicative and critical health literacy: a pilot study of Japanese office workers. Health Promot Int. 2008 Sep 1;23(3):269–74.
172. de Bekker-Grob EW, Donkers B, Bliemer MCJ, Veldwijk J, Swait JD. Can healthcare choice be predicted using stated preference data? Soc Sci Med. 2020 Feb 1;246:112736.
173. de Bekker-Grob EW, Donkers B, Veldwijk J, Jonker MF, Buis S, Huisman J, et al. What Factors Influence Non-Participation Most in Colorectal Cancer Screening? A Discrete Choice Experiment. Patient - Patient-Centered Outcomes Res. 2021 Mar 1;14(2):269–81.
174. van der Vaart R, Drossaert CHC, Taal E, ten Klooster PM, Hilderink-Koertshuis RTE, Klaase JM, et al. Validation of the Dutch functional, communicative and critical health literacy scales. Patient Educ Couns. 2012 Oct 1;89(1):82–8.

175. Wångdahl JM, Mårtensson LI. The Communicative and Critical Health Literacy Scale – Swedish Version. Scand J Public Health. 2014 Feb 1;42(1):25–31.
176. McNaughton CD, Cavanaugh KL, Kripalani S, Rothman RL, Wallston KA. Validation of a Short, 3-Item Version of the Subjective Numeracy Scale. Med Decis Mak Int J Soc Med Decis Mak. 2015 Nov;35(8):932–6.
177. Hess S. Latent class structures: taste heterogeneity and beyond. In: Handbook of Choice Modelling [Internet]. 2014 [cited 2021 May 6]. Available from: https://doi.org/10.4337/9781781003152
178. Greene WH, Hensher DA. A latent class model for discrete choice analysis: contrasts with mixed logit. Transp Res Part B Methodol. 2003 Sep 1;37(8):681–98.
179. Swait J. A structural equation model of latent segmentation and product choice for cross-sectional revealed preference choice data. J Retail Consum Serv. 1994 Oct 1;1(2):77–89.
180. Bansback N, Harrison M, Sadatsafavi M, Stiggelbout A, Whitehurst DGT. Attitude to health risk in the Canadian population: a cross-sectional survey. CMAJ Open. 2016 Jun 3;4(2):E284–91.
181. Wakker P, Deneffe D. Eliciting von Neumann-Morgenstern Utilities When Probabilities Are Distorted or Unknown. Manag Sci. 1996 Aug 1;42(8):1131–50.
182. Bleichrodt H, Pinto JL, Wakker PP. Making Descriptive Use of Prospect Theory to Improve the Prescriptive Use of Expected Utility. Manag Sci. 2001 Nov 1;47(11):1498–514.
183. Bleichrodt H, Abellan-Perpiñan JM, Pinto-Prades JL, Mendez-Martinez I. Resolving Inconsistencies in Utility Measurement Under Risk: Tests of Generalizations of Expected Utility. Manag Sci. 2007 Mar 1;53(3):469–82.
184. Lejuez CW, Read JP, Kahler CW, Richards JB, Ramsey SE, Stuart GL, et al. Evaluation of a behavioral measure of risk taking: The Balloon Analogue Risk Task (BART). J Exp Psychol Appl. 2002;8(2):75–84.
185. Peters E, Hart PS, Fraenkel L. Informing Patients: The Influence of Numeracy, Framing, and Format of Side Effect Information on Risk Perceptions. Med Decis Making. 2011 mei;31(3):432–6.
186. de Bekker-Grob EW, Ryan M, Gerard K. Discrete choice experiments in health economics: a review of the literature. Health Econ. 2012;21(2):145–72.
187. Cornelsen L, Quaife M, Lagarde M, Smith RD. Framing and signalling effects of taxes on sugary drinks: A discrete choice experiment among households in Great Britain. Health Econ. 2020;29(10):1132–47.
188. Genie MG, Krucien N, Ryan M. Weighting or aggregating? Investigating information processing in multi-attribute choices. Health Econ. 2021;30(6):1291–305.
189. Gonzalez Sepulveda JM, Johnson FR, Marshall DA. Incomplete information and irrelevant attributes in stated-preference values for health interventions. Health Econ. 2021;30(11):2637–48.
190. Mansfield C, Sutphin J, Boeri M. Assessing the impact of excluded attributes on choice in a discrete choice experiment using a follow-up question. Health Econ. 2020;29(10):1307–15.
191. Hausman J, Ruud PA. Specifying and testing econometric models for rank-ordered data. J Econom. 1987 Jan 1;34(1–2):83–104.
192. Lancsar E, Louviere J, Donaldson C, Currie G, Burgess L. Best worst discrete choice experiments in health: methods and an application. Soc Sci Med 1982. 2013 Jan;76(1):74–82.
193. Palma MA. Improving the prediction of ranking data. Empir Econ. 2017 Dec 1;53(4):1681–710.

194. Vermeulen B, Goos P, Vandebroek M. Rank-order choice-based conjoint experiments: Efficiency and design. J Stat Plan Inference. 2011 Aug 1;141(8):2519–31.
195. Hess S, Rose JM. Allowing for intra-respondent variations in coefficients estimated on repeated choice data. Transp Res Part B Methodol. 2009 Jul;43(6):708–19.
196. Lancsar E, Louviere J. Conducting discrete choice experiments to inform healthcare decision making: a user's guide. PharmacoEconomics. 2008;26(8):661–77.
197. Louviere JJ, Street D, Burgess L, Wasi N, Islam T, Marley AAJ. Modeling the choices of individual decision-makers by combining efficient choice experiment designs with extra preference information. J Choice Model. 2008 Jan 1;1(1):128–64.
198. Beggs S, Cardell S, Hausman JA. Assessing the potential demand for electric cars. J Econom. 1981 Sep 1;17(1):1–19.
199. Chapman R, Staelin R. Exploiting Rank Ordered Choice Set Data Within the Stochastic Utility Model. J Mark Res. 1982 Aug 1;19.
200. Punj GN, Staelin R. The Choice Process for Graduate Business Schools. J Mark Res. 1978 Nov 1;15(4):588–98.
201. Ben-Akiva M, Morikawa T, Shiroishi F. Analysis of the reliability of preference ranking data. J Bus Res. 1991 Nov 1;23(3):253–68.
202. Bradley M, Daly A. Use of the logit scaling approach to test for rank-order and fatigue effects in stated preference data. Transportation. 1994 May 1;21(2):167–84.
203. Fok D, Paap R, Dijk BV. A Rank-Ordered Logit Model with Unobserved Heterogeneity in Ranking Capabilities. J Appl Econom. 2012;27(5):831–46.
204. Marley AAJ, Louviere JJ. Some probabilistic models of best, worst, and best–worst choices. J Math Psychol. 2005 Dec 1;49(6):464–80.
205. Finn A, Louviere JJ. Determining the Appropriate Response to Evidence of Public Concern: The Case of Food Safety. J Public Policy Mark. 1992;11(2):12–25.
206. Flynn TN. Valuing citizen and patient preferences in health: recent developments in three types of best–worst scaling. Expert Rev Pharmacoecon Outcomes Res. 2010 Jun 1;10(3):259–67.
207. Delle Site P, Kilani K, Gatta V, Marcucci E, de Palma A. Estimation of consistent Logit and Probit models using best, worst and best–worst choices. Transp Res Part B Methodol. 2019 Oct 1;128:87–106.
208. Dyachenko T, Reczek RW, Allenby GM. Models of Sequential Evaluation in Best-Worst Choice Tasks. Mark Sci. 2014 Sep 8;33(6):828–48.
209. Hawkins GE, Marley AAJ, Heathcote A, Flynn TN, Louviere JJ, Brown SD. The best of times and the worst of times are interchangeable. Decision. 2014;1(3):192–214.
210. Hawkins GE, Islam T, Marley AAJ. Like It or Not, You Are Using One Value Representation. Decision [Internet]. 2019 Jul [cited 2020 Apr 23];6(3). Available from: https://oce-ovid-com.eur.idm.oclc.org/article/01762424-201907000-00004/HTML
211. Ghijben P, Lancsar E, Zavarsek S. Preferences for Oral Anticoagulants in Atrial Fibrillation: a Best–Best Discrete Choice Experiment. PharmacoEconomics. 2014 Nov 1;32(11):1115–27.
212. Scarpa R, Notaro S, Louviere J, Raffaelli R. Exploring Scale Effects of Best/Worst Rank Ordered Choice Data to Estimate Benefits of Tourism in Alpine Grazing Commons. Am J Agric Econ. 2011;93(3):813–28.

213. Xie F, Pullenayegum E, Gaebel K, Oppe M, Krabbe PFM. Eliciting preferences to the EQ-5D-5L health states: discrete choice experiment or multiprofile case of best–worst scaling? Eur J Health Econ. 2014 Apr;15(3):281–8.
214. Akaichi F, Nayga RM, Gil JM. Are Results from Non-hypothetical Choice-based Conjoint Analyses and Non-hypothetical Recoded-ranking Conjoint Analyses Similar? Am J Agric Econ. 2013;95(4):949–63.
215. Caparrós A, Oviedo JL, Campos P. Would You Choose Your Preferred Option? Comparing Choice and Recorded Ranking Experiments. Am J Agric Econ. 2008;90(3):843–55.
216. Chang JB, Lusk JL, Norwood FB. How Closely Do Hypothetical Surveys and Laboratory Experiments Predict Field Behavior? Am J Agric Econ. 2009;91(2):518–34.
217. Petrolia DR, Interis MG, Hwang J. Single-Choice, Repeated-Choice, and Best-Worst Scaling Elicitation Formats: Do Results Differ and by How Much? Environ Resour Econ Off J Eur Assoc Environ Resour Econ. 2018;69(2):365–93.
218. Yangui A, Akaichi F, Costa-Font M, Gil JM. Comparing results of ranking conjoint analyses, best-worst scaling and discrete choice experiments in a nonhypothetical context. Aust J Agric Resour Econ. 2019;63(2):221–46.
219. Krucien N, Sicsic J, Ryan M. For better or worse? Investigating the validity of best–worst discrete choice experiments in health. Health Econ. 2019;28(4):572–86.
220. Giergiczny M, Dekker T, Hess S, Chintakayala PK. Testing the stability of utility parameters in repeated best, repeated best-worst and one-off best-worst studies. Eur J Transp Infrastruct Res [Internet]. 2017 Sep 1 [cited 2020 Jul 21];17(4). Available from: https://journals.open.tudelft.nl/ejtir/article/view/3209
221. Lancsar E, Ride J, Black N, Burgess L, Peeters A. Social acceptability of standard and behavioural economic inspired policies designed to reduce and prevent obesity. Health Econ. 2021;
222. Train K. Discrete Choice Methods with Simulation [Internet]. SUNY-Oswego, Department of Economics; 2003 [cited 2020 Jul 31]. Available from: https://econpapers.repec.org/bookchap/oettbooks/emetr2.htm
223. Louviere JJ, Islam T, Wasi N, Street D, Burgess L. Designing Discrete Choice Experiments: Do Optimal Designs Come at a Price? J Consum Res. 2008 Aug 1;35(2):360–75.
224. McKelvey RD, Zavoina W. A statistical model for the analysis of ordinal level dependent variables. J Math Sociol. 1975 Jan 1;4(1):103–20.
225. Krinsky I, Robb AL. On Approximating the Statistical Properties of Elasticities. Rev Econ Stat. 1986;68(4):715–9.
226. Park T, Loomis JB, Creel M. Confidence Intervals for Evaluating Benefits Estimates from Dichotomous Choice Contingent Valuation Studies. Land Econ. 1991;67(1):64–73.
227. Rose JM, Bliemer MCJ. Sample size requirements for stated choice experiments. Transportation. 2013 Sep 1;40(5):1021–41.
228. Vanniyasingam T, Cunningham CE, Foster G, Thabane L. Simulation study to determine the impact of different design features on design efficiency in discrete choice experiments. BMJ Open. 2016 Jul 1;6(7):e011985.
229. Shafir E. Choosing versus rejecting: why some options are both better and worse than others. Mem Cognit. 1993 Jul;21(4):546–56.
230. Dhar R, Wertenbroch K. Consumer choice between hedonic and utilitarian goods. J Mark Res. 2000 Feb;37(1):60–71.

231. Meloy MG, Russo JE. Binary choice under instructions to select versus reject. Organ Behav Hum Decis Process. 2004 maart;93(2):114–28.
232. Laran J, Wilcox K. Choice, Rejection, and Elaboration on Preference-Inconsistent Alternatives. J Consum Res. 2011 Aug 1;38(2):229–41.
233. Brockner J, Higgins ET. Regulatory Focus Theory: Implications for the Study of Emotions at Work. Organ Behav Hum Decis Process. 2001 Sep 1;86(1):35–66.
234. Higgins ET, Shah J, Friedman R. Emotional responses to goal attainment: Strength of regulatory focus as moderator. J Pers Soc Psychol. 1997;72(3):515–25.
235. Brazell JD, Diener CG, Karniouchina E, Moore WL, Séverin V, Uldry PF. The no-choice option and dual response choice designs. Mark Lett. 2006 Dec 1;17(4):255–68.
236. de Bekker-Grob EW, Ryan M, Gerard K. Discrete choice experiments in health economics: a review of the literature. Health Econ. 2012;21(2):145–72.
237. Clark MD, Determann D, Petrou S, Moro D, de Bekker-Grob EW. Discrete Choice Experiments in Health Economics: A Review of the Literature. PharmacoEconomics. 2014 Sep 1;32(9):883–902.
238. Lancsar E, Swait J. Reconceptualising the External Validity of Discrete Choice Experiments. PharmacoEconomics. 2014 Oct 1;32(10):951–65.
239. Quaife M, Terris-Prestholt F, Di Tanna GL, Vickerman P. How well do discrete choice experiments predict health choices? A systematic review and meta-analysis of external validity. Eur J Health Econ. 2018 Nov 1;19(8):1053–66.
240. Haghani M, Bliemer MCJ, Rose JM, Oppewal H, Lancsar E. 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. J Choice Model. 2021 Aug 1;100309.
241. Haghani M, Bliemer MCJ, Rose JM, Oppewal H, Lancsar E. Hypothetical bias in stated choice experiments: Part II. Conceptualisation of external validity, sources and explanations of bias and effectiveness of mitigation methods. J Choice Model. 2021 Dec 1;41:100322.
242. European Cancer Information System. Estimates of cancer incidence and mortality in 2020 [Internet]. 2020 [cited 2022 Jan 6]. Available from: https://ecis.jrc.ec.europa.eu/
243. Cardoso R, Guo F, Heisser T, Hackl M, Ihle P, De Schutter H, et al. Colorectal cancer incidence, mortality, and stage distribution in European countries in the colorectal cancer screening era: an international population-based study. Lancet Oncol. 2021 Jul 1;22(7):1002–13.
244. Elmunzer BJ, Singal AG, Sussman JB, Deshpande AR, Sussman DA, Conte ML, et al. Comparing the effectiveness of competing tests for reducing colorectal cancer mortality: a network meta-analysis. Gastrointest Endosc. 2015 maart;81(3):700-709.e3.
245. RIVM. Colorectal cancer screening programme [Internet]. 2011 [cited 2021 Oct 11]. Available from: https://www.rivm.nl/en/colorectal-cancer-screening-programme
246. Benning TM, Dellaert BGC, Dirksen CD, Severens JL. Preferences for potential innovations in non-invasive colorectal cancer screening: A labeled discrete choice experiment for a Dutch screening campaign. Acta Oncol. 2014 Jul 1;53(7):898–908.
247. van Dam L, Hol L, de Bekker-Grob EW, Steyerberg EW, Kuipers EJ, Habbema JDF, et al. What determines individuals' preferences for colorectal cancer screening programmes? A discrete choice experiment. Eur J Cancer. 2010 Jan 1;46(1):150–9.

248. Veldwijk J, Lambooij MS, Kallenberg FGJ, van Kranen HJ, Bredenoord AL, Dekker E, et al. Preferences for genetic testing for colorectal cancer within a population-based screening program: a discrete choice experiment. Eur J Hum Genet. 2016 Mar;24(3):361–6.
249. Heidenreich S, Rutten LJF, Miller-Wilson LA, Jimenez-Moreno C, Chua GN, Fisher DA. Colorectal cancer screening preferences among physicians and individuals at average risk: A discrete choice experiment. Cancer Med [Internet]. 2022 Mar 21 [cited 2022 Mar 29]; Available from: http://onlinelibrary.wiley.com/doi/10.1002/cam4.4678
250. Ladabaum U, Dominitz JA, Kahi C, Schoen RE. Strategies for Colorectal Cancer Screening. Gastroenterology. 2020 Jan 1;158(2):418–32.
251. Bevolkingsonderzoek. Bevolkingsonderzoek darmkanker een groot succes [Internet]. 2019 [cited 2022 Jun 18]. Available from: https://www.bevolkingsonderzoeknederland.nl/nieuws/bevolkingsonderzoek-darmkanker-een-groot-succes/
252. Schreuders EH, Ruco A, Rabeneck L, Schoen RE, Sung JJY, Young GP, et al. Colorectal cancer screening: a global overview of existing programmes. Gut. 2015 Oct 1;64(10):1637–49.
253. Lansdorp-Vogelaar I, Knudsen AB, Brenner H. Cost-effectiveness of Colorectal Cancer Screening. Epidemiol Rev. 2011 Jul 1;33(1):88–100.
254. Quintero E, Castells A, Bujanda L, Cubiella J, Salas D, Lanas Á, et al. Colonoscopy versus Fecal Immunochemical Testing in Colorectal-Cancer Screening. N Engl J Med. 2012 Feb 23;366(8):697–706.
255. Hiligsmann M, Durme C van, Geusens P, Dellaert BG, Dirksen CD, Weijden T van der, et al. Nominal group technique to select attributes for discrete choice experiments: an example for drug treatment choice in osteoporosis. Patient Prefer Adherence. 2013 Feb 7;7:133–9.
256. de Bekker-Grob EW, Donkers B, Jonker MF, Stolk EA. Sample Size Requirements for Discrete-Choice Experiments in Healthcare: a Practical Guide. The Patient. 2015 Oct;8(5):373–84.
257. Ding M. An Incentive-Aligned Mechanism for Conjoint Analysis. J Mark Res. 2007 May 1;44(2):214–23.
258. Ding M, Grewal R, Liechty J. Incentive-Aligned Conjoint Analysis. J Mark Res. 2005 Feb 1;42(1):67–82.
259. Dong S, Ding M, Huber J. A simple mechanism to incentive-align conjoint experiments. Int J Res Mark. 2010 Mar;27(1):25–32.
260. de Bekker-Grob EW, Veldwijk J, Jonker M, Donkers B, Huisman J, Buis S, et al. The impact of vaccination and patient characteristics on influenza vaccination uptake of elderly people: A discrete choice experiment. Vaccine. 2018 Mar 7;36(11):1467–76.
261. Cedefop. European qualifications framework (EQF) [Internet]. CEDEFOP. 2009 [cited 2022 Jun 20]. Available from: https://www.cedefop.europa.eu/en/projects/european-qualifications-framework-eqf
262. Netherlands Qualification Framework (NLQF). Niveaus [Internet]. nd [cited 2022 Jun 20]. Available from: https://nlqf.nl/nlqf-niveaus
263. Centraal Bureau voor de Statistiek (CBS). Opleidingsniveau [Internet]. Centraal Bureau voor de Statistiek. [cited 2022 Jun 20]. Available from: https://www.cbs.nl/nl-nl/nieuws/2019/33/verschil-levensverwachting-hoog-en-laagopgeleid-groeit/opleidingsniveau
264. Glanz K, Rimer BK, Viswanath K. Health Behavior: Theory, Research, and Practice. 5th ed. Wiley; 2015.
265. Pachur T, Spaar M. Domain-specific preferences for intuition and deliberation in decision making. J Appl Res Mem Cogn. 2015 Sep 1;4(3):303–11.

266. Fagerlin A, Zikmund-Fisher BJ, Ubel PA, Jankovic A, Derry HA, Smith DM. Measuring numeracy without a math test: development of the Subjective Numeracy Scale. Med Decis Mak Int J Soc Med Decis Mak. 2007 Oct;27(5):672–80.
267. Zikmund-Fisher BJ, Smith DM, Ubel PA, Fagerlin A. Validation of the Subjective Numeracy Scale: Effects of Low Numeracy on Comprehension of Risk Communications and Utility Elicitations. Med Decis Making. 2007 Sep 1;27(5):663–71.
268. Scarpa R, Ferrini S, Willis K. Performance of Error Component Models for Status-Quo Effects in Choice Experiments. In: Scarpa R, Alberini A, editors. Applications of Simulation Methods in Environmental and Resource Economics [Internet]. Dordrecht: Springer Netherlands; 2005 [cited 2021 Dec 15]. p. 247–73. (The Economics of Non-Market Goods and Resources). Available from: https://doi.org/10.1007/1-4020-3684-1_13
269. Bierlaire M. A short introduction to PandasBiogeme. :22.
270. Hess S, Rose JM. Can scale and coefficient heterogeneity be separated in random coefficients models? Transportation. 2012 Nov 1;39(6):1225–39.
271. RIVM. Monitor bevolkingsonderzoek darmkanker 2020 [Internet]. 2021 [cited 2022 Apr 21]. Available from: https://www.rivm.nl/documenten/monitor-bevolkingsonderzoek-darmkanker-2020
272. Beck MJ, Fifer S, Rose JM. Can you ever be certain? Reducing hypothetical bias in stated choice experiments via respondent reported choice certainty. Transp Res Part B Methodol. 2016 Jul 1;89:149–67.
273. Penn JM, Hu W. Understanding Hypothetical Bias: An Enhanced Meta-Analysis. Am J Agric Econ. 2018;100(4):1186–206.
274. Backstrom M. Higher-Order Factors in a Five-Factor Personality Inventory and its Relation to Social Desirability. J Psychol Assess 2007. 2007;23(2):63–70.
275. Milfont TL. The effects of social desirability on self-reported environmental attitudes and ecological behaviour. Environmentalist. 2009 Sep;29(3):263–9.
276. Paulhus DL. Two-component models of socially desirable responding. J Pers. 1984 Mar;46(3):598–609.
277. Cerri J, Testa F, Rizzi F, Frey M. Factorial surveys reveal social desirability bias over self-reported organic fruit consumption. Br Food J. 2019 Jan 1;121(4):897–909.
278. Krumpal I. Determinants of social desirability bias in sensitive surveys: a literature review. Qual Quant. 2013 Jun 1;47(4):2025–47.
279. Murphy JJ, Allen PG, Stevens TH, Weatherhead D. A Meta-Analysis of Hypothetical Bias in Stated Preference Valuation. Environ Resour Econ. 2005;30:313–25.
280. Johansson-Stenman O, Svedsäter H. Self-image and valuation of moral goods: Stated versus actual willingness to pay. J Econ Behav Organ. 2012 Dec 1;84(3):879–91.
281. Konopka R, Wright MJ, Avis M, Feetham PM. If you think about it more, do you want it more? The case of fairtrade. Eur J Mark. 2019 Jan 1;53(12):2556–81.
282. Horiuchi Y, Markovich Z, Yamamoto T. Does Conjoint Analysis Mitigate Social Desirability Bias? SSRN. 2018;MIT Political Science Department Research Paper No. 2018-15 Political Analysis, forthcoming.:29.
283. Norwood FB, Lusk JL. Social Desirability Bias in Real, Hypothetical, and Inferred Valuation Experiments. Am J Agric Econ. 2011;93(2):528–34.
284. Cummings RG, Taylor LO. Unbiased Value Estimates for Environmental Goods: A Cheap Talk Design for the Contingent Valuation Method. Am Econ Rev. 1999 Jun 1;89(3):649–65.

285. Penn J, Hu W. Cheap talk efficacy under potential and actual Hypothetical Bias: A meta-analysis. J Environ Econ Manag. 2019 Jul 1;96:22–35.
286. Chowdhury S, Meenakshi JV, Tomlins KI, Owori C. Are Consumers in Developing Countries Willing to Pay More for Micronutrient-Dense Biofortified Foods? Evidence from a Field Experiment in Uganda. Am J Agric Econ. 2011;93(1):83–97.
287. Andorfer VA, Liebe U. Do information, price, or morals influence ethical consumption? A natural field experiment and customer survey on the purchase of Fair Trade coffee. Soc Sci Res. 2015 Jul 1;52:330–50.
288. Cerri J, Thøgersen J, Testa F. Social desirability and sustainable food research: A systematic literature review. Food Qual Prefer. 2019 Jan 1;71:136–40.
289. Huitink M, Poelman MP, van den Eynde E, Seidell JC, Dijkstra SC. Social norm nudges in shopping trolleys to promote vegetable purchases: A quasi-experimental study in a supermarket in a deprived urban area in the Netherlands. Appetite. 2020 Aug 1;151:104655.
290. Giordano C, Piras S, Boschini M, Falasconi L. Are questionnaires a reliable method to measure food waste? A pilot study on Italian households. Br Food J. 2018 Jan 1;120(12):2885–97.
291. Quested TE, Palmer G, Moreno LC, McDermott C, Schumacher K. Comparing diaries and waste compositional analysis for measuring food waste in the home. J Clean Prod. 2020 Jul 20;262:121263.
292. Silva A, Nayga RM, Campbell BL, Park JL. Can perceived task complexity influence cheap talk's effectiveness in reducing hypothetical bias in stated choice studies? Appl Econ Lett. 2012 Nov 1;19(17):1711–4.
293. Moser R, Raffaelli R, Notaro S. Testing hypothetical bias with a real choice experiment using respondents' own money. Eur Rev Agric Econ. 2014 Feb 1;41(1):25–46.
294. de-Magistris T, Pascucci S. The effect of the solemn oath script in hypothetical choice experiment survey: A pilot study. Econ Lett. 2014 May 1;123(2):252–5.
295. Bello M, Abdulai A. Impact of Ex-Ante Hypothetical Bias Mitigation Methods on Attribute Non-Attendance in Choice Experiments. Am J Agric Econ. 2016;98(5):1486–506.
296. Bello M, Abdulai A. Measuring heterogeneity, survey engagement and response quality in preferences for organic products in Nigeria. Appl Econ. 2016 Mar 15;48(13):1159–71.
297. Lin W, Ortega DL, Caputo V. Are Ex-Ante Hypothetical Bias Calibration Methods Context Dependent? Evidence from Online Food Shoppers in China. J Consum Aff. 2019;53(2):520–44.
298. Gschwandtner A, Burton M. Comparing treatments to reduce hypothetical bias in choice experiments regarding organic food. Eur Rev Agric Econ. 2020 Jun 15;47(3):1302–37.
299. Engelen C. Kromkommer – Gekke groente of gekke kwaliteitseisen? [Internet]. 2016 [cited 2022 Apr 19]. Available from: https://www.kromkommer.com/gekkekwaliteitseisen/
300. Keuringsdienst van Waarde. Wat de klasse zegt over groente en fruit [Internet]. npo3.nl. 2018 [cited 2022 Apr 19]. Available from: https://www.npo3.nl/keuringsdienst-van-waarde/wat-de-klasse-zegt-over-groente-en-fruit
301. Brink E, van Rossum C, Postma-Smeets A, Stafleu A, Wolvers D, van Dooren C, et al. Development of healthy and sustainable food-based dietary guidelines for the Netherlands. Public Health Nutr. 2019 Sep;22(13):2419–35.

302. Centre for the Promotion of Imports from developing countries (CBI). What requirements must fresh fruit or vegetables comply with to be allowed on the European market? [Internet]. Ministry of Foreign Affairs; 2022 [cited 2022 Jun 1]. Available from: https://www.cbi.eu/market-information/fresh-fruit-vegetables/buyer-requirements
303. Tanner C, Wölfing Kast S. Promoting sustainable consumption: Determinants of green purchases by Swiss consumers. Psychol Mark. 2003;20(10):883–902.
304. Meijers M, van Dam Y. Sustainable food purchases in the Netherlands: The influence of consumer characteristics. J Chain Netw Sci. 2012 Jan 1;12:181–98.
305. Vilar R, Milfont TL, Araújo R de CR, Coelho GL de H, Soares AKS, Gouveia VV. Consideration of future consequences (CFC): Validation and proposition of an ultra-short scale. Curr Psychol [Internet]. 2020 Jun 16 [cited 2021 Aug 20]; Available from: https://doi.org/10.1007/s12144-020-00840-y
306. Joireman J, Shaffer MJ, Balliet D, Strathman A. Promotion Orientation Explains Why Future-Oriented People Exercise and Eat Healthy: Evidence From the Two-Factor Consideration of Future Consequences-14 Scale. Pers Soc Psychol Bull. 2012 Oct 1;38(10):1272–87.
307. Weinberger DA, Schwartz GE. Distress and Restraint as Superordinate Dimensions of Self-Reported Adjustment: A Typological Perspective. J Pers. 1990;58(2):381–417.
308. Paulhus DL. Measurement and Control of Response Bias [Internet]. In Robinson, J. P., Shaver, P. R., Wrightsman, L. S. (Eds.), Measures of personality and social psychological attitudes (pp. 17-59). 1991. Available from: http://www.sjdm.org/dmidi/Balanced_Inventory_of_Desirable_Responding.html
309. Hart CM, Ritchie TD, Hepper EG, Gebauer JE. The Balanced Inventory of Desirable Responding Short Form (BIDR-16). SAGE Open. 2015 Oct 1;5(4):2158244015621113.
310. Brewer MB, Crano WD. Research Design and Issues of Validity. In: Judd CM, Reis HT, editors. Handbook of Research Methods in Social and Personality Psychology [Internet]. 2nd ed. Cambridge: Cambridge University Press; 2014 [cited 2022 May 25]. p. 11–26. Available from: https://www.cambridge.org/core/books/handbook-of-research-methods-in-social-and-personality-psychology/research-design-and-issues-of-validity/A11466F0A7FDA0259AA8F10415A38BD8
311. Hole AR. Fitting Mixed Logit Models by Using Maximum Simulated Likelihood. Stata J. 2007 Sep 1;7(3):388–401.
312. Dit zijn de gezondheidsvoordelen van een paprika en dit betekent de kleur [Internet]. RTL Nieuws. 2022 [cited 2022 May 30]. Available from: https://www.rtlnieuws.nl/lifestyle/eigen-huis-tuin/artikel/5284303/de-gezondheidsvoordelen-van-een-paprika-en-het-verschil
313. Carlsson F, Frykblom P, Johan Lagerkvist C. Using cheap talk as a test of validity in choice experiments. Econ Lett. 2005 Nov 1;89(2):147–52.
314. Tonsor GT, Shupp RS. Cheap Talk Scripts and Online Choice Experiments: "Looking Beyond the Mean". Am J Agric Econ. 2011;93(4):1015–31.
315. List JA, Sinha P, Taylor MH. Using Choice Experiments to Value Non-Market Goods and Services: Evidence from Field Experiments. BE J Econ Anal Policy [Internet]. 2006 Jan 13 [cited 2022 Jun 3];6(2). Available from: https://www.degruyter.com/document/doi/10.2202/1538-0637.1132/html
316. Fifer S, Rose J, Greaves S. Hypothetical bias in Stated Choice Experiments: Is it a problem? And if so, how do we deal with it? Transp Res Part Policy Pract. 2014 Mar 1;61:164–77.
317. Carlsson F, Daruvala D, Jaldell H. Do you do what you say or do you do what you say others do? J Choice Model. 2010 Jan 1;3(2):113–33.

318. Ladenburg J, Olsen SB. Augmenting short Cheap Talk scripts with a repeated Opt-Out Reminder in Choice Experiment surveys. Resour Energy Econ. 2014 Aug 1;37:39–63.
319. Varela E, Mahieu PA, Giergiczny M, Riera P, Soliño M. Testing the single opt-out reminder in choice experiments: An application to fuel break management in Spain. J For Econ. 2014 Aug 1;20(3):212–22.
320. Howard G, Roe BE, Nisbet EC, Martin JF. Hypothetical Bias Mitigation Techniques in Choice Experiments: Do Cheap Talk and Honesty Priming Effects Fade with Repeated Choices? J Assoc Environ Resour Econ. 2017 Jun;4(2):543–73.
321. Carson RT, Flores NE, Martin KM, Wright JL. Contingent Valuation and Revealed Preference Methodologies: Comparing the Estimates for Quasi-Public Goods. Land Econ. 1996;72(1):80–99.
322. Loomis J. What's to Know About Hypothetical Bias in Stated Preference Valuation Studies? J Econ Surv. 2011;25(2):363–70.
323. Lee J, Hwang U. Hypothetical Bias in Risk Preferences as a Driver of Hypothetical Bias in Willingness to Pay: Experimental Evidence. Environ Resour Econ. 2016 Dec 1;65(4):789–811.
324. Lusk JL, Schroeder TC. Are Choice Experiments Incentive Compatible? A Test with Quality Differentiated Beef Steaks. Am J Agric Econ. 2004;86(2):467–82.
325. Schläpfer F. Contingent valuation: A new perspective. Ecol Econ. 2008 Feb 1;64(4):729–40.
326. Özdemir S, Johnson FR, Hauber AB. Hypothetical bias, cheap talk, and stated willingness to pay for health care. J Health Econ. 2009 Jul 1;28(4):894–901.
327. Özdemir S. Improving the Validity of Stated-Preference Data in Health Research: The Potential of the Time-to-Think Approach. Patient - Patient-Centered Outcomes Res. 2015 Jun 1;8(3):247–55.
328. Regier DA, Veenstra DL, Basu A, Carlson JJ. Demand for Precision Medicine: A Discrete-Choice Experiment and External Validation Study. PharmacoEconomics. 2020 Jan 1;38(1):57–68.
329. Whichello C, Bywall KS, Mauer J, Stephen W, Cleemput I, Pinto CA, et al. An overview of critical decision-points in the medical product lifecycle: Where to include patient preference information in the decision-making process? Health Policy Amst Neth. 2020 Dec;124(12):1325–32.
330. Marsh K, van Til JA, Molsen-David E, Juhnke C, Hawken N, Oehrlein EM, et al. 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 Health J Int Soc Pharmacoeconomics Outcomes Res. 2020 Jul;23(7):831–41.
331. Staniszewska S, Hill EM, Grant R, Grove P, Porter J, Shiri T, et al. Developing a Framework for Public Involvement in Mathematical and Economic Modelling: Bringing New Dynamism to Vaccination Policy Recommendations. Patient - Patient-Centered Outcomes Res [Internet]. 2021 Jan 19 [cited 2021 Jun 10]; Available from: https://doi.org/10.1007/s40271-020-00476-x
332. Bouvy JC, Cowie L, Lovett R, Morrison D, Livingstone H, Crabb N. Use of Patient Preference Studies in HTA Decision Making: A NICE Perspective. Patient - Patient-Centered Outcomes Res. 2020 Apr 1;13(2):145–9.
333. Ghabri S, Mühlbacher A. Opportunities and Challenges Incorporating Patient Preference Information in Health Technology Assessment Frameworks. Value & Outcomes Spotlight. 2020;6(4):28–9.

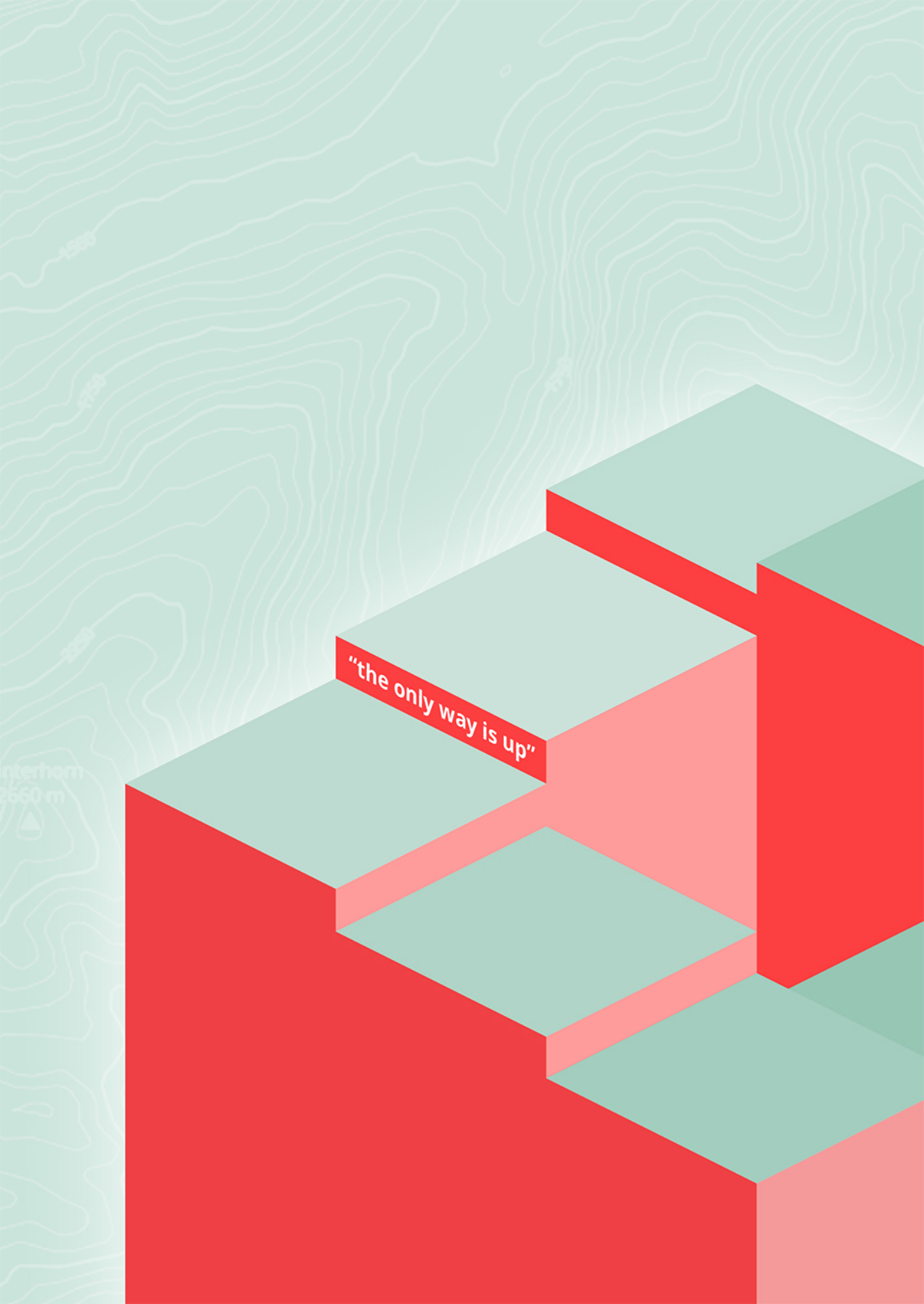
334. Linthicum MT, dosReis S, Slejko JF, Mattingly TJ, Bright JL. The Importance of Collaboration in Pursuit of Patient-Centered Value Assessment. Patient - Patient-Centered Outcomes Res [Internet]. 2020 Aug 28 [cited 2021 Jun 10]; Available from: https://doi.org/10.1007/s40271-020-00446-3
335. Marsh K, Bekker-Grob E de, Cook N, Collacott H, Danyliv A. How to integrate evidence from patient preference studies into health technology assessment: a critical review and recommendations. Int J Technol Assess Health Care [Internet]. 2021 ed [cited 2021 Oct 13];37(1). Available from: https://www.cambridge.org/core/journals/international-journal-of-technology-assessment-in-health-care/article/how-to-integrate-evidence-from-patient-preference-studies-into-health-technology-assessment-a-critical-review-and-recommendations/F78C266FDDA12A1280EF2740D40CA8A5
336. Russell J, Fudge N, Greenhalgh T. The impact of public involvement in health research: what are we measuring? Why are we measuring it? Should we stop measuring it? Res Involv Engagem. 2020 Oct 27;6(1):63.
337. Shields GE, Brown L, Wells A, Capobianco L, Vass C. Utilising Patient and Public Involvement in Stated Preference Research in Health: Learning from the Existing Literature and a Case Study. Patient - Patient-Centered Outcomes Res [Internet]. 2020 Aug 4 [cited 2021 Jun 10]; Available from: https://doi.org/10.1007/s40271-020-00439-2
338. Whitty JA, de Bekker-Grob EW, Cook NS, Terris-Prestholt F, Drummond M, Falchetto R, et al. Patient Preferences in the Medical Product Lifecycle. Patient - Patient-Centered Outcomes Res. 2020 Feb 1;13(1):7–10.
339. Cook NS, Cave J, Holtorf AP. Patient Preference Studies During Early Drug Development: Aligning Stakeholders to Ensure Development Plans Meet Patient Needs. Front Med [Internet]. 2019 [cited 2021 Jun 10];6. Available from: https://www.frontiersin.org/articles/10.3389/fmed.2019.00082/full
340. Janssens R, Huys I, van Overbeeke E, Whichello C, Harding S, Kübler J, et al. Opportunities and challenges for the inclusion of patient preferences in the medical product life cycle: a systematic review. BMC Med Inform Decis Mak. 2019 Oct 4;19(1):189.
341. van Overbeeke E, Whichello C, Janssens R, Veldwijk J, Cleemput I, Simoens S, et al. Factors and situations influencing the value of patient preference studies along the medical product lifecycle: a literature review. Drug Discov Today. 2019 Jan 1;24(1):57–68.
342. Helgesson G, Ernstsson O, Åström M, Burström K. Whom should we ask? A systematic literature review of the arguments regarding the most accurate source of information for valuation of health states. Qual Life Res. 2020 Jun 1;29(6):1465–82.
343. Mason RJ, Searle KM, Bombard Y, Rahmadian A, Chambers A, Mai H, et al. Evaluation of the impact of patient involvement in health technology assessments: A scoping review. Int J Technol Assess Health Care. 2020 Jun;36(3):217–23.
344. van Overbeeke E, Forrester V, Simoens S, Huys I. Use of Patient Preferences in Health Technology Assessment: Perspectives of Canadian, Belgian and German HTA Representatives. Patient - Patient-Centered Outcomes Res. 2021 Jan 1;14(1):119–28.
345. Whichello C, Levitan B, Juhaeri J, Patadia V, DiSantostefano R, Pinto CA, et al. Appraising patient preference methods for decision-making in the medical product lifecycle: an empirical comparison. BMC Med Inform Decis Mak. 2020 Jun 19;20(1):114.

346. Janssens R, Russo S, van Overbeeke E, Whichello C, Harding S, Kübler J, et al. Patient Preferences in the Medical Product Life Cycle: What do Stakeholders Think? Semi-Structured Qualitative Interviews in Europe and the USA. Patient - Patient-Centered Outcomes Res. 2019 Oct 1;12(5):513–26.
347. Gagnon MP, Dipankui MT, Poder TG, Payne-Gagnon J, Mbemba G, Beretta V. Patient and public involvement in health technology assessment: update of a systematic review of international experiences. Int J Technol Assess Health Care [Internet]. 2021 ed [cited 2022 Jan 11];37(1). Available from: https://www.cambridge.org/core/journals/international-journal-of-technology-assessment-in-health-care/article/patient-and-public-involvement-in-health-technology-assessment-update-of-a-systematic-review-of-international-experiences/38E1E2EAE4195F6242E2F3D7A2EBF668
348. Bowling A, Ebrahim S. Measuring patients' preferences for treatment and perceptions of risk. Qual Health Care QHC. 2001 Sep;10(Suppl 1):i2–8.
349. Howard K, Salkeld G. Does Attribute Framing in Discrete Choice Experiments Influence Willingness to Pay? Results from a Discrete Choice Experiment in Screening for Colorectal Cancer. Value Health. 2009;12(2):354–63.
350. Veldwijk J, Essers BAB, Lambooij MS, Dirksen CD, Smit HA, de Wit GA. Survival or Mortality: Does Risk Attribute Framing Influence Decision-Making Behavior in a Discrete Choice Experiment? Value Health J Int Soc Pharmacoeconomics Outcomes Res. 2016 Apr;19(2):202–9.
351. Veldwijk J, Lambooij MS, Bekker-Grob EW de, Smit HA, Wit GA de. The Effect of Including an Opt-Out Option in Discrete Choice Experiments. PLOS ONE. 2014 Nov 3;9(11):e111805.
352. Heidenreich S, Phillips-Beyer A, Flamion B, Ross M, Seo J, Marsh K. Benefit–Risk or Risk–Benefit Trade-Offs? Another Look at Attribute Ordering Effects in a Pilot Choice Experiment. Patient - Patient-Centered Outcomes Res. 2021 Jan 1;14(1):65–74.
353. Campbell D, Erdem S. Including Opt-Out Options in Discrete Choice Experiments: Issues to Consider. Patient - Patient-Centered Outcomes Res. 2019 Feb 1;12(1):1–14.
354. Determann D, Gyrd-Hansen D, de Wit GA, de Bekker-Grob EW, Steyerberg EW, Lambooij MS, et al. Designing Unforced Choice Experiments to Inform Health Care Decision Making: Implications of Using Opt-Out, Neither, or Status Quo Alternatives in Discrete Choice Experiments. Med Decis Making. 2019 Aug 1;39(6):681–92.
355. Mulhern B, Norman R, Street DJ, Viney R. One Method, Many Methodological Choices: A Structured Review of Discrete-Choice Experiments for Health State Valuation. PharmacoEconomics. 2019 Jan 1;37(1):29–43.
356. Jonker MF, De Bekker-Grob EW, Veldwijk J, Goossens LMA, Ruttten-van Mölkeneen M. Improved External Validity of DCE Uptake Predictions Based on a Dual-Response None Option Format? In: The Patient - Patient-Centered Outcomes Research [Internet]. 2021 [cited 2022 Jun 11]. p. 863–8. Available from: https://doi.org/10.1007/s40271-021-00532-0
357. Buckell J, Hess S. Stubbing out hypothetical bias: improving tobacco market predictions by combining stated and revealed preference data. J Health Econ. 2019 May 1;65:93–102.
358. Vass C, Boeri M, Karim S, Marshall D, Craig B, Ho KA, et al. Accounting for Preference Heterogeneity in Discrete-Choice Experiments: An ISPOR Special Interest Group Report. Value Health. 2022 May 1;25(5):685–94.
359. Krueger R, Bierlaire M, Daziano RA, Rashidi TH, Bansal P. Evaluating the predictive abilities of mixed logit models with unobserved inter- and intra-individual heterogeneity. J Choice Model. 2021 Dec 1;41:100323.

360. Grandin T, Johnson C. Animals in Translation: Using the Mysteries of Autism to Decode Animal Behavior [Internet]. 1st edition. Orlando: Harcourt; 2006. 358 p. Available from: https://abcnews.go.com/Primetime/Technology/story?id=549279&page=1
361. Himmler S, Soekhai V, van Exel J, Brouwer W. 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. J Choice Model. 2021 Mar 1;38:100265.
362. Milte R, Ratcliffe J, Chen G, Lancsar E, Miller M, Crotty M. Cognitive Overload? An Exploration of the Potential Impact of Cognitive Functioning in Discrete Choice Experiments with Older People in Health Care. Value Health. 2014 Jul 1;17(5):655–9.
363. Lancsar E, Louviere J. Deleting 'irrational' responses from discrete choice experiments: a case of investigating or imposing preferences? Health Econ. 2006;15(8):797–811.
364. Miguel FS, Ryan M, Amaya-Amaya M. 'Irrational' stated preferences: a quantitative and qualitative investigation. Health Econ. 2005;14(3):307–22.
365. Ryan M, Watson V, Entwistle V. Rationalising the 'irrational': a think aloud study of discrete choice experiment responses. Health Econ. 2009;18(3):321–36.
366. Smets K. Poor Homo Economicus [Internet]. Medium. 2022 [cited 2022 May 6]. Available from: https://koenfucius.medium.com/poor-homo-economicus-55a2893eca5
367. Tervonen T, Schmidt-Ott T, Marsh K, Bridges JF, Quaife M, Janssen E. Assessing rationality in discrete choice experiments in health: an investigation into the use of dominance tests. Value Health. 2018;21(10):1192–7.
368. Özdemir S, Finkelstein EA. Cognitive Bias: The Downside of Shared Decision Making. JCO Clin Cancer Inform. 2018 Dec;(2):1–10.
369. Bruch E, Swait J. Choice Set Formation in Residential Mobility and Its Implications for Segregation Dynamics. Demography. 2019 Aug 21;56(5):1665–92.
370. Veldwijk J, Swait JD. The Role of Attribute Screening and Choice Set Formation in Health Discrete Choice Experiments: Modeling the Impact of Benefit and Risk Attributes. Value Health [Internet]. 2022 May 20 [cited 2022 May 27]; Available from: https://www.sciencedirect.com/science/article/pii/S1098301522001061
371. International Choice Modelling Conference. Conference programme - 7th International Choice Modelling Conference. In 2022. Available from: http://www.icmconference.org.uk/ICMC_2022_files/ICMC_2022_programme.pdf
372. Veldwijk J, Johansson JV, Donkers B, de Bekker-Grob EW. Mimicking Real-Life Decision Making in Health: Allowing Respondents Time to Think in a Discrete Choice Experiment. Value Health. 2020 Jul 1;23(7):945–52.
373. Aguiar M, Harrison M, Munro S, Burch T, Kaal KJ, Hudson M, et al. Designing Discrete Choice Experiments Using a Patient-Oriented Approach. Patient - Patient-Centered Outcomes Res [Internet]. 2020 Jul 17 [cited 2021 Jun 10]; Available from: https://doi.org/10.1007/s40271-020-00431-w
374. Lips P, Timmers L, Bal R, Delnoij D. Involvement of Patients and Medical Professionals in the Assessment of Relative Effectiveness: A Need for Closer Cooperation [Internet]. 2022 [cited 2022 May 11]. Available from: https://reader.elsevier.com/reader/sd/pii/S1098301522001899?token=AC63998DEC6F54DD6C89C32DCF315556077860134FF874AD5C1D5255F1A31CF0E623EF793AB19EEBD849A48F3865E0B2&originRegion=eu-west-1&originCreation=20220511073009

375. Poder TG, Safianyk C, Fournier MF, Ganache I, Touré M, Pomey MP, et al. Patients, users, caregivers, and citizens' involvement in local health technology assessment unit in Quebec: a survey. Int J Technol Assess Health Care [Internet]. 2021 ed [cited 2022 Jun 10];37(1). Available from: https://www.cambridge.org/core/journals/international-journal-of-technology-assessment-in-health-care/article/patients-users-caregivers-and-citizens-involvement-in-local-health-technology-assessment-unit-in-quebec-a-survey/D86A322D1E0C741E660DBC14624D18F2
376. Hannigan A. Public and patient involvement in quantitative health research: A statistical perspective. Health Expect. 2018;21(6):939–43.
377. Ní Shé É, Cassidy J, Davies C, De Brún A, Donnelly S, Dorris E, et al. Minding the gap: identifying values to enable public and patient involvement at the pre-commencement stage of research projects. Res Involv Engagem. 2020 Dec;6(1):46.
378. Hahn DL, Hoffmann AE, Felzien M, LeMaster JW, Xu J, Fagnan LJ. Tokenism in patient engagement. Fam Pract. 2017 Jun 1;34(3):290–5.
379. O'Shea E, Ogbebor F, Queally M, Murphy E. Knowledge of public patient involvement among health economists in Ireland: a baseline audit [Internet]. HRB Open Research; 2019 [cited 2022 Jun 10]. Available from: https://hrbopenresearch.org/articles/2-4
380. Mercer RE, Chambers A, Mai H, McDonald V, McMahon C, Chan KKW. Are We Making a Difference? A Qualitative Study of Patient Engagement at the pan-Canadian Oncology Drug Review: Perspectives of Patient Groups. Value Health. 2020 Sep 1;23(9):1157–62.
381. Hawton A, Boddy K, Kandiyali R, Tatnell L, Gibson A, Goodwin E. Involving Patients in Health Economics Research: "The PACTS Principles". Patient - Patient-Centered Outcomes Res [Internet]. 2020 Oct 12 [cited 2021 Jun 10]; Available from: https://doi.org/10.1007/s40271-020-00461-4
382. Al-Janabi H, Coles J, Copping J, Dhanji N, McLoughlin C, Murphy J, et al. Patient and Public Involvement (PPI) in Health Economics Methodology Research: Reflections and Recommendations. Patient - Patient-Centered Outcomes Res [Internet]. 2020 Sep 17 [cited 2021 Jun 10]; Available from: https://doi.org/10.1007/s40271-020-00445-4
383. Sarri G, Freitag A, Szegvari B, Mountian I, Brixner D, Bertelsen N, et al. The Role of Patient Experience in the Value Assessment of Complex Technologies – Do HTA Bodies Need to Reconsider How Value is Assessed? Health Policy. 2021 May 1;125(5):593–601.
384. U.S. Food and Drug Administration (FDA). Patient-Focused Drug Development: Collecting Comprehensive and Representative Input [Internet]. Center for Drug Evaluation and Research; 2020 Jun [cited 2022 Jun 11]. Available from: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/patient-focused-drug-development-collecting-comprehensive-and-representative-input
385. Lambooij MS, Harmsen IA, Veldwijk J, de Melker H, Mollema L, van Weert YW, et al. Consistency between stated and revealed preferences: a discrete choice experiment and a behavioural experiment on vaccination behaviour compared. BMC Med Res Methodol. 2015 Mar 12;15(1):19.
386. Meghani SH, Chittams J, Hanlon AL, Curry J. Measuring preferences for analgesic treatment for cancer pain: how do African-Americans and Whites perform on choice-based conjoint (CBC) analysis experiments? BMC Med Inform Decis Mak. 2013 Oct 18;13:118.
387. Mohammadi T, Bansback N, Marra F, Khakban A, Campbell JR, FitzGerald JM, et al. Testing the External Validity of a Discrete Choice Experiment Method: An Application to Latent Tuberculosis Infection Treatment. Value Health. 2017 Jul 1;20(7):969–75.

388. Natter M, Feurstein M. Real world performance of choice-based conjoint models. Eur J Oper Res. 2002 Mar 1;137(2):448–58.
389. Ryan M, Watson V. Comparing welfare estimates from payment card contingent valuation and discrete choice experiments. Health Econ. 2009;18(4):389–401.
390. Salampessy BH, Veldwijk J, Jantine Schuit A, van den Brekel-Dijkstra K, Neslo REJ, Ardine de Wit G, et al. The Predictive Value of Discrete Choice Experiments in Public Health: An Exploratory Application. Patient - Patient-Centered Outcomes Res. 2015 Dec 1;8(6):521–9.
391. Zipursky RB, Cunningham CE, Stewart B, Rimas H, Cole E, Vaz SM. Characterizing outcome preferences in patients with psychotic disorders: a discrete choice conjoint experiment. Schizophr Res. 2017 Jul 1;185:107–13.
392. Adamowicz W, Louviere J, Williams M. Combining Revealed and Stated Preference Methods for Valuing Environmental Amenities. J Environ Econ Manag. 1994 May 1;26(3):271–92.
393. Cameron TA, Poe GL, Ethier RG, Schulze WD. Alternative Non-market Value-Elicitation Methods: Are the Underlying Preferences the Same? J Environ Econ Manag. 2002 Nov 1;44(3):391–425.
394. Carlsson F, Martinsson P. Do Hypothetical and Actual Marginal Willingness to Pay Differ in Choice Experiments?: Application to the Valuation of the Environment. J Environ Econ Manag. 2001 Mar 1;41(2):179–92.
395. Kesternich I, Heiss F, McFadden D, Winter J. Suit the action to the word, the word to the action: Hypothetical choices and real decisions in Medicare Part D. J Health Econ. 2013 Dec 1;32(6):1313–24.
396. Whitehead JC. Environmental Risk and Averting Behavior: Predictive Validity of Jointly Estimated Revealed and Stated Behavior Data. Environ Resour Econ. 2005 Nov 1;32(3):301–16.
397. Linley WG, Hughes DA. Decision-Makers' Preferences for Approving New Medicines in Wales: A Discrete-Choice Experiment with Assessment of External Validity. PharmacoEconomics. 2013 Apr;31(4):345–55.
398. Vossler CA, Watson SB. Understanding the consequences of consequentiality: Testing the validity of stated preferences in the field. J Econ Behav Organ. 2013 Feb;86:137–47.
399. de Corte K, Cairns J, Grieve R. Stated versus revealed preferences: An approach to reduce bias. Health Econ. 2021;30(5):1095–123.
400. Mark TL, Swait J. Using stated preference and revealed preference modeling to evaluate prescribing decisions. Health Econ. 2004;13(6):563–73.
401. Mueller S, Osidacz P, Francis IL, Lockshin L. Combining discrete choice and informed sensory testing in a two-stage process: Can it predict wine market share? Food Qual Prefer. 2010 Oct 1;21(7):741–54.
402. Kruk ME, Paczkowski M, Mbaruku G, de Pinho H, Galea S. Women's Preferences for Place of Delivery in Rural Tanzania: A Population-Based Discrete Choice Experiment. Am J Public Health. 2009 Sep;99(9):1666–72.
403. Bosworth R, Taylor LO. Hypothetical Bias in Choice Experiments: Is Cheap Talk Effective at Eliminating Bias on the Intensive and Extensive Margins of Choice? BE J Econ Anal Policy [Internet]. 2012 Dec 12 [cited 2022 Jun 3];12(1). Available from: https://www.degruyter.com/document/doi/10.1515/1935-1682.3278/html
404. Broadbent CD. Evaluating mitigation and calibration techniques for hypothetical bias in choice experiments. J Environ Plan Manag. 2014 Dec 2;57(12):1831–48.


"the only way is up"