UNIVERSITY OF WARSAW | FACULTY OF ECONOMIC SCIENCES

WORKING PAPERS
No. 4/2026 (498)

BEHAVIORAL BIASES IN STATED PREFERENCE VALUATION OF MORTALITY RISK REDUCTIONS: COST VECTOR, ANCHORING, AND SCOPE EFFECTS

WOJCIECH ZAWADZKI
HENRIK ANDERSSON
MIKOŁAJ CZAJKOWSKI
ARNE RISA HOLE

WARSAW 2026

ISSN 2957-0506

UNIVERSITY OF WARSAW
FACULTY OF ECONOMIC SCIENCES
WORKING PAPERS

Behavioral Biases in Stated Preference Valuation of Mortality Risk Reductions: Cost Vector, Anchoring, and Scope Effects

Wojciech Zawadzki[1*], *Henrik Andersson*[2 3], *Mikołaj Czajkowski*[1], *Arne Risa Hole*[4]

[1] *Faculty of Economic Sciences, University of Warsaw*
[2] *Swedish National Road and Transport Research Institute (VTI)*
[3] *Toulouse School of Economics, University of Toulouse Capitole*
[4] *Universitat Jaume I*
[*] *Corresponding author: wm.zawadzki@uw.edu.pl*

Abstract: This study investigates how behavioral biases influence stated preference valuation of mortality risk reductions, commonly summarized as the value of a statistical life (VSL). Using a discrete choice experiment (DCE) combined with a contingent valuation double-bounded dichotomous choice and an open-ended follow-up, we elicit individuals' willingness to pay (WTP) for cardiovascular mortality risk reductions. In a randomized design, we varied the cost attribute across three cost range treatments and manipulated information disclosure and feedback to examine three behavioral phenomena: cost vector effects (whether the range of costs presented affects WTP), scope insensitivity (whether WTP scales appropriately with the magnitude of the risk reduction), and anchoring (whether initial cost cues affect subsequent responses). Our results show that mean VSL estimates can vary by up to ~25% between cost treatments. Furthermore, WTP responses exhibit partial scope insensitivity – larger risk reductions do not proportionally increase WTP – indicating a deviation from theoretical expectations. Importantly, we find no strong evidence of anchoring: neither revealing all attribute levels upfront, nor starting with extreme cost levels, nor providing feedback on quiz questions significantly affected respondents' choices or WTP. Our findings underscore the need for careful survey design. Even if VSL distributions remain statistically similar across cost frames, substantial shifts in mean magnitudes could be consequential for policy. We call for standardized guidelines on cost attribute selection and survey protocols to mitigate bias, ensuring that stated preference methods yield reliable welfare estimates for health policy decisions.

Keywords: value of statistical life (VSL), stated preference (SP), contingent valuation (CV), behavioral biases, anchoring effect, scope insensitivity, discrete choice experiment (DCE), willingness to pay (WTP), mortality risk reduction, cardiovascular diseases

JEL codes: I12, D01, D61, Q51, C83, C93

Acknowledgements: Wojciech Zawadzki and Mikołaj Czajkowski gratefully acknowledge the support of the National Science Centre of Poland (Opus 2023/49/B/HS4/01655, Preludium 2023/49/N/HS4/02035). Henrik Andersson gratefully acknowledges funding from ANR, France, under grant ANR-17-EURE- 0010 (Investissements d'Avenir program). The authors gratefully acknowledge the valuable feedback received from participants at the Workshop on Mortality and Morbidity Risk Valuation (Paris, 2024), the International Choice Modelling Conference (Puerto Varas, 2024), the Society for Benefit-Cost Analysis Annual (online, 2024) and European (Warsaw, 2024) Conferences, and the Workshop on Non-Market Valuation (Nancy, 2025), which contributed to improving this work.

Funding/Support: This research was funded by the Excellence Initiative – Research University (IDUB), Action IV.4.1 (BOB-IDUB-622-124/2023), at the University of Warsaw.

Credit Author Statement: Wojciech Zawadzki: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data Curation, Writing - Original Draft, Writing - Review & Editing, Visualization, Project administration, Funding acquisition. Henrik Andersson: Conceptualization, Methodology, Investigation, Writing - Review & Editing, Visualization, Supervision. Mikołaj Czajkowski: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Writing - Original Draft, Writing - Review & Editing, Visualization, Supervision. Arne Risa Hole: Conceptualization, Methodology, Investigation, Writing - Review & Editing, Visualization, Supervision.

Conflict of Interest Statement: No potential competing interest was reported by the authors.

Ethics Approval Statement: The study received a positive ethics opinion from the Ethics Committee of the Faculty of Economic Sciences at the University of Warsaw, confirming that the research project complies with ethics principles.

Data Availability Statement: The data that support the findings of this study are available from the corresponding author, Wojciech Zawadzki, upon reasonable request.

Disclosure for Using AI Tools: During the preparation of this work, the authors used DeepL, ChatGPT, and Grammarly to improve the grammar and fluency of the text. After using these tools/services, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

1. Introduction

The challenge of assigning monetary values to reductions in mortality risk – often summarized as the value of statistical life (VSL) – has long occupied scholars and policymakers (Andersson & Treich, 2011; Banzhaf, 2022; Keller et al., 2021; Shah et al., 2015; Viscusi & Aldy, 2003). Stated preference (SP) methods, such as discrete choice experiments (DCEs) and contingent valuation (CV), have emerged as influential tools in this domain by eliciting individuals' willingness to pay (WTP) for non-market goods, including lifesaving health interventions (Alberini, 2019; Hanley & Czajkowski, 2019). Given the policy implications of VSL – whether for guiding public resource allocation, conducting cost-benefit analyses, or designing health regulations – securing valid and reliable estimates is paramount (Bishop & Boyle, 2019).

Despite their widespread use, SP methods are vulnerable to well-documented behavioral biases that can undermine the validity of VSL estimates. In particular, **cost vector effects** occur when the range or spacing of cost levels shapes WTP responses in ways unrelated to "true" preferences (Ahtiainen et al., 2023; Carlsson & Martinsson, 2008; Glenk et al., 2019). **Scope insensitivity** arises when respondents' WTP does not adequately scale with changes in the magnitude of the good being valued, challenging the fundamental principle that willingness to pay should reflect the scope of the improvement or risk reduction (Kahneman & Knetsch, 1992; Whitehead, 2016). Meanwhile, **anchoring effects** can emerge when participants' subsequent valuations are heavily influenced by initially presented values – whether they are explicit cost bids or implied numeric cues in earlier choice tasks (Ariely et al., 2003; Tversky & Kahneman, 1974). Although these three biases have been studied in DCEs and CV for various environmental and public goods, relatively few investigations have systematically examined their combined influence on the valuation of life-saving health interventions (Rice, 2013; Su et al., 2017).

In this paper, we address this gap by analyzing how cost vector selection, scope insensitivity, and anchoring potentially interact in the context of cardiovascular disease mortality risk reductions. We do so using a dual-method approach – combining a DCE with a CV double-bounded dichotomous choice (DBDC) format and a subsequent open-ended (OE) question – to elicit respondents' willingness to pay for reducing mortality risk. Through a randomized design that systematically varies the cost attribute levels across three cost treatments, we disentangle how each behavioral effect contributes to divergence in VSL estimates. Our findings highlight the importance of selecting sufficiently large and granular

cost vectors and show that even moderate changes in cost design can shift the mean VSL by up to 25%, although not always with statistical significance. Moreover, we find evidence that scope insensitivity is more pronounced in certain cost treatments, while anchoring does not appear to dominate final estimates once respondents receive consistent risk information.

By illustrating how cost vector range, scope, and anchoring effects can combine to influence VSL, our study underscores the necessity for clearer methodological guidelines in SP health valuation research. We propose that future studies adopt more systematic approaches to identifying cost vectors, use diagnostic checks for both anchoring and scope sensitivity, and employ experimental designs that mitigate potential distortions from learning and reference-dependent heuristics. These recommendations can help ensure that VSL estimates – increasingly integral to policy decisions – reliably capture individuals' true preferences for reductions in mortality risk.

Following this introduction, Section 2 offers a comprehensive literature review, outlining existing findings on cost vector effects, scope (in)sensitivity, and anchoring biases in SP studies. Section 3 then describes our study design and data collection procedures, detailing the questionnaire structure, the experimental treatments, and the sampling approach. Section 4 presents the econometric framework, including the specification of our mixed logit model with correlated random parameters and the methods used to derive WTP and VSL estimates. Section 5 reports empirical results from both the DCE and CV elicitation formats, examining cost vector and behavioral effects. In Section 6, we discuss these findings in the context of existing research and highlight implications for improving the reliability of health-focused stated preference methods. Finally, Section 7 concludes by summarizing our main contributions, noting study limitations, and suggesting avenues for future research.

2. Literature Review

2.1. Cost Vector and Attribute Range Effects

Within DCEs, each attribute is assigned discrete levels, which are then combined to form choice scenarios (Lancsar & Louviere, 2008; Mariel et al., 2021). Among these attributes, cost plays a pivotal role: it not only enables the estimation of WTP but also anchors the trade-off values for non-monetary attributes (Ryan et al., 2007). Cost vector effects occur when changes to the levels – or the spacing between those levels – of the cost attribute alter respondents' WTP

in ways that deviate from theoretical expectations (Carlsson & Martinsson, 2008; Glenk et al., 2019; Su et al., 2017).

Early studies produced mixed evidence. For example, Ryan and Wordsworth (2000) employed a split-sample approach varying the price attribute and other attributes, finding that while estimated attribute coefficients were mostly unaffected by changing cost vectors, mean WTP differed significantly in several cases. Conversely, Hanley et al. (2005) found no significant shift in preferences or WTP estimates across surveys with different cost vectors, suggesting that cost vector effects were neither pervasive nor universal.

However, a growing body of subsequent research underscores the possibility that cost vector design can systematically influence valuation outcomes. Carlsson and Martinsson (2008) observed that higher absolute cost levels can lead to elevated marginal WTP estimates, even if the relative differences remain constant. In a similar vein, Mørkbak et al. (2010) demonstrated that raising the maximum price level can increase the likelihood of a "choke price" effect, where respondents opt out more frequently or exhibit systematically higher or lower WTP. Kragt (2013) further argued that individuals might focus on relative rather than absolute price differences, effectively anchoring their decisions to the specific cost levels presented. More recently, Ahtiainen et al. (2023) showed that cost vector effects persist even when the status quo alternative is associated with a positive cost, affecting the proportion of status quo choices and overall WTP distributions.

Mitigation strategies have been proposed. Börger et al. (2025) found that clarifying opt-out reminders and including cheap talk scripts reduced the impact of different cost vectors on respondents' choices. Meanwhile, Glenk et al. (2024) suggested a diagnostic check wherein WTP estimates must align with the highest cost level to ensure that respondents' valuations reflect genuine trade-offs rather than artifacts of the experimental design. Taken together, these studies imply that insufficiently high cost levels may fail to "choke off" demand, while excessively large intervals between cost levels can distort WTP distribution and reduce the accuracy of final estimates.

Although best practices have begun to crystallize – such as using wide but context-appropriate cost ranges – there remains no universally agreed-upon standard for cost attribute design (Aravena et al., 2014; Lancsar & Louviere, 2008). This is due in part to the situational nature of DCEs: attributes and levels must remain realistic and salient within a specific research context, making it difficult to prescribe a one-size-fits-all rule. Nonetheless, the literature

consistently emphasizes that cost vectors should be chosen with care. Poorly selected cost attributes can systematically bias WTP estimates and, by extension, policy advice derived from these estimates. The following sections investigate additional sources of potential bias – scope insensitivity and anchoring – and their role in shaping health-related valuations.

2.2. Scope Insensitivity

Scope insensitivity – or scope neglect – refers to a discrepancy wherein WTP does not increase proportionally with the scale of the good or service being valued (Carson, 1997; Kahneman & Knetsch, 1992). In the context of mortality risk reduction, scope insensitivity arises when individuals express nearly the same WTP for different magnitudes of risk reduction, thus failing to reflect theoretically consistent preferences. Since VSL is premised on the idea that WTP should scale – at least weakly – with the size of the risk reduction, scope insensitivity can undermine the reliability of these estimates (Andersson & Treich, 2011; Bishop & Boyle, 2019).

The literature provides diverse explanations for scope insensitivity. Cognitive constraints – such as difficulty conceptualizing small probability changes – are frequently cited (Kahneman et al., 1999). Respondents may also succumb to embedding effects, where they focus on the general idea of contributing to a "good cause" or "saving lives" and fail to distinguish among varying levels of the good (Kahneman & Knetsch, 1992). As a result, WTP responses can converge around a similar value regardless of the actual change in risk. Psychological factors like "warm glow" (i.e., the satisfaction of giving) and diminishing marginal utility for risk reductions may also contribute to non-proportional responses (Chilton & Hutchinson, 2000; Czajkowski & Hanley, 2009).

In empirical research, failures to demonstrate scope sensitivity can take two forms:

1. **Weak scope insensitivity**, where WTP increases in response to larger quantities of the good but not in a linear or near-proportional manner (Goldberg & Roosen, 2007).
2. **Strong scope insensitivity**, where WTP remains essentially unchanged despite large differences in the quantity or quality of the good being valued (Carson, 1997; Whitehead, 2016).

Studies on mortality risk reduction have repeatedly documented these patterns. Kartman et al. (1996) and Hammitt and Graham (1999) showed that WTP for preventing a fatal health outcome often fails to scale with meaningful changes in mortality probability. Goldberg and

Roosen (2007) found subadditive relationships between the scope of health risk reductions and stated WTP, and Andersson et al. (2016) demonstrated that significant variation in risk levels could produce widely disparate and often incongruent VSL estimates.

Methodological approaches have been proposed to mitigate or detect scope insensitivity. Constructing internally valid "within-respondent" tests – where the same individual values multiple versions of the good – can reveal whether WTP increases with scope (Bateman et al., 2006; Lindhjem et al., 2011). Others recommend using risk changes expressed in terms comprehensible to lay audiences – such as higher percentage changes – so that respondents can more readily perceive the difference in risk (Herrera-Araujo et al., 2022). Moreover, anchoring efforts or diagnostic checks, such as forcing respondents to consider different magnitudes explicitly or validating responses with debriefing questions, may assist in distinguishing genuine preferences from a lack of comprehension (Bateman et al., 2004; Whitehead, 2016).

Within health economics, understanding the scope sensitivity of VSL is critical. A failure to detect proportional changes in stated WTP for risk reductions may misguide policy. For instance, interventions that produce larger risk reductions could appear less cost-effective if WTP remains flat, while small interventions might be overvalued. Consequently, addressing scope insensitivity requires careful experimental design, transparent presentation of risk changes, and, possibly, the inclusion of learning or feedback mechanisms to ensure respondents have adequate context for expressing coherent preferences.

2.3. Anchoring Effects

Originally described by Tversky and Kahneman (1974), the anchoring effect posits that individuals exposed to an initial reference value – whether a price, statistic, or another numerical cue – tend to anchor subsequent judgments around that reference, even when it is arbitrary. In SP contexts, this phenomenon can lead to starting-point bias or reference-point dependence, where valuation responses are systematically influenced by numerical prompts that precede the main WTP questions (Ariely et al., 2003; Furnham & Boo, 2011).

Anchoring is particularly relevant to studies using CV, where an initial bid in a dichotomous choice format may shape respondents' subsequent willingness to pay (Boyle et al., 1985; Flachaire & Hollard, 2007). In double-bounded dichotomous choice (DBDC) settings, for instance, participants are first asked to accept or reject an initial bid, and then presented with a follow-up bid based on their response. The outcome of the first question can

influence the second, potentially leading to systematically biased valuations – either overstated or understated (Herriges & Shogren, 1996; Onwujekwe & Nwagbo, 2002).

In DCEs, anchoring can manifest through repeated choice tasks or instructional choice sets, which give respondents initial cost or attribute levels that shift how they perceive subsequent alternatives (Ladenburg & Olsen, 2008; Meyerhoff & Glenk, 2015). Cost vector design can itself act as an anchor: respondents may interpret higher cost levels as indicators of a more valuable or premium good, thus inflating WTP (Carlsson & Martinsson, 2008; Kragt, 2013). Alternatively, if cost levels are set too low, participants might anchor on those amounts and discount the good's potential value (Sun et al., 2019).

Several theoretical explanations underscore why anchoring is so pervasive. The coherent arbitrariness hypothesis (Ariely et al., 2003) argues that consumers form preferences only once they are prompted to do so; because non-market goods are not traded in typical markets, respondents rely heavily on contextual cues – like initial cost bids – in constructing their valuations. Meanwhile, Bayesian updating approaches posit that participants may incorporate anchors as prior information, updating their "beliefs" in line with the anchor when knowledge about the good is otherwise limited (Bateman et al., 2008; Nguyen et al., 2015).

Mitigation strategies include cheap talk scripts – which warn respondents about hypothetical or anchoring biases – opt-out reminders, and careful ordering or randomization of cost levels (Börger et al., 2025; List & Gallet, 2001). Some studies also recommend eliminating instructional choice sets or ensuring that training tasks do not inadvertently bias real choice tasks (Meyerhoff & Glenk, 2015). Despite these techniques, anchoring remains difficult to eradicate completely, especially in health valuation scenarios where respondents often lack robust benchmarks for comparing risk reductions or health outcomes (Bestard & Font, 2021; Rice, 2013).

Given that anchoring often interacts with other biases – such as scope insensitivity and cost vector effects – untangling its specific influence can be challenging. Nevertheless, understanding anchoring's role is vital to designing SP surveys that capture genuine trade-offs. As with scope insensitivity, properly calibrated experimental designs and diagnostics, alongside transparency about the potential for anchoring, can help mitigate distortions in WTP estimates and foster more reliable policy recommendations.

3. Study Design and Data Collection

3.1. Survey Instrument Development

Designing a SP survey for non-market valuation demands rigorous preparation and adherence to recognized best practices (Johnston et al., 2017; Lancsar et al., 2017; Lancsar & Louviere, 2008; Mariel et al., 2021). In this study, we developed a comprehensive questionnaire that integrates both a DCE and CV components – namely, a double-bounded dichotomous choice (DBDC) followed by an open-ended (OE) question. Our primary goal was to investigate how cost vector selection, scope (in)sensitivity, and anchoring might influence valuations of a hypothetical health intervention that reduces cardiovascular disease mortality risk.

The questionnaire begins by explaining the purpose of the study and highlighting its potential policy relevance. Respondents are informed that aggregated findings will be made available to governmental and research institutions, thereby emphasizing consequentiality (Carson & Groves, 2007; Vossler et al., 2012). After confirming that their responses will remain anonymous, the survey collects standard demographic information (sex, age, education, region, and household size) and household income data. These questions establish the socio-economic context necessary for interpreting WTP differences and controlling for heterogeneity in subsequent analyses.

We then assess respondents' current health status using the EQ-5D-5L module, which captures dimensions like mobility, self-care, usual activities, pain/discomfort, and anxiety/depression. Additional questions probe the frequency of doctor visits and the presence of chronic or cardiovascular conditions, including any familial history. By prompting participants to reflect on their own health, we aim to heighten engagement with the concept of mortality risk. Next, the survey offers a concise overview of cardiovascular diseases in Poland – emphasizing their leading share of total mortality – and briefly outlines common risk factors (e.g., genetic predisposition, diet, smoking).

To aid respondents in grasping mortality risk, we employ a visual grid that depicts black squares (deaths) and white squares (survivors) out of 1,000 people over the next 10 years. The average mortality risk in Poland, set at 47 deaths per 1,000 people per decade, is introduced using this visual aid. Respondents must complete three comprehension questions (e.g., translating risk figures into proportions or comparing risk levels) to ensure they understand the basic concepts. A treatment at this stage randomly assigns respondents to either receive immediate feedback on their answers or not, allowing us to observe if corrective

feedback influences subsequent decision-making or risk perception. To further anchor mortality risk in a familiar context, we also present information on statistical life expectancy derived from gender- and age-specific actuarial life tables, as expressing benefits in terms of life expectancy gains appears the most persuasive (Grisolía et al., 2018). Overall, this approach ensures that participants are informed about their baseline risk and that comprehension is enhanced through the use of a visual aid (Corso et al., 2001; Marceta et al., 2025; OECD, 2025).

After the comprehension module, participants are prompted to estimate their own mortality risk relative to the national average. They can declare a higher, lower, or similar risk level – or indicate uncertainty – while being reminded of modifiable and non-modifiable factors that could influence actual risk. This approach encourages introspection about personal health circumstances before respondents move on to evaluate hypothetical interventions.

Participants then engage in the DCE, evaluating a hypothetical drug designed to prevent some deaths from cardiovascular diseases. The drug is presented as having no side effects, receiving endorsements from official health agencies, and being capable of reducing mortality risk and eliminating hospitalizations due to cardiovascular events over the next decade. Each DCE choice task offers participants two different drug profiles to compare, alongside a status quo option that incurs zero cost and offers no improvement in mortality risk. For simplicity, the survey incorporates only two attributes: (1) the scale of mortality risk reduction, quantified as 1, 2, or 4 fewer deaths per 1,000 people over a span of 10 years, and (2) the annual cost associated with each option.

This simplified design was deliberately chosen to isolate the effect of cost vectors in relation to mortality risk reductions. Although DCEs often include additional attributes, such as morbidity outcomes, implementation lags, or quality-of-life impacts, introducing these elements could obscure the interpretation of cost vector effects. For this reason, we opted for a deliberately minimal attribute structure, ensuring that no other factors shape respondents' preferences. This approach allows for a clearer assessment of how cost-vector selection influences the resulting VSL estimates.

Regarding cost vectors, the survey utilizes three categories – low, medium, and high – with each category escalating in increments (by 50, 100, and 200, respectively), such as 50–300 PLN for low, 100–600 PLN for medium, and 200–1,200 PLN for high. Consequently, the selected cost vectors vary by the starting point, choke-off point, and intervals. Random assignment ensures that each participant encounters only one of these cost vector treatments.

Additionally, the survey randomly either discloses or withholds all potential attribute levels before the DCE begins, allowing for an analysis of potential anchoring effects caused by early exposure to the cost information.

The chosen cost vector levels were determined through pretesting with focus groups and are informed by prior research on the valuation of cardiovascular risk reduction in Poland (Bartczak et al., 2021). While the majority of respondents during qualitative pretesting expressed a WTP to reduce mortality risk by 1 in 1,000, ranging from approximately 100 to 150 PLN, some reported valuations approaching zero or exceeding these amounts. This substantial preference heterogeneity, combined with the study design depicting reductions of 1, 2, or 4 deaths per 1,000, justified the inclusion of such a diverse set of cost vectors.

Ideally, a study would incorporate a wide-ranging cost vector with a low defined starting point, a high choke-off point, and granularly spaced intermediate values. However, achieving this would require a substantially larger experimental design and, consequently, a much larger sample size. Researchers therefore face a trade-off, often settling on a cost vector with relatively few (several) elements – a modest number given that most DCEs include several other attributes, each with multiple levels. As such, the cost vector can appear somewhat artificial, particularly in light of the broader absence of standardized guidelines for selecting cost elements in SP studies. Because cost vectors are typically designed a priori, without precise knowledge of the underlying demand curve, decisions regarding the initial and final values largely reflect the researcher's own judgement. Our study thus provides empirical evidence on how variations in these cost vectors – arguably imprecisely chosen – can influence final outcomes.

We employ a Bayesian D-efficient experimental design (Huber & Zwerina, 1996; Sandor & Wedel, 2001) to generate 12 choice tasks per respondent, divided into blocks. The design maximizes the precision of parameter estimation by incorporating prior information (initially set to zero, then updated with pilot data). Each participant is randomly allocated to one design block, ensuring balanced representation across cost treatments.

Immediately following the DCE, respondents answer a double-bounded dichotomous choice (DBDC) question about their maximum WTP to reduce mortality risk by 1 in 1,000. Three initial bids (200, 400, or 600 PLN) are randomly assigned; if the participant accepts the bid, the follow-up doubles the amount; if the participant rejects, the follow-up halves it. An open-ended (OE) question then invites participants to state the exact amount they would be

willing to pay. These CV tasks provide an additional data source for WTP, enabling comparisons with the DCE-derived estimates and checks for anchoring and scope effects.

Finally, the survey concludes with debriefing questions to gauge the perceived realism and clarity of the scenarios, as well as the respondents' sense of survey consequentiality. We also confirm whether participants felt they understood the risk attributes well enough to provide meaningful valuations.

This layered design aims to isolate and identify potential behavioral biases related to cost vector selection, scope insensitivity, and anchoring within a single study. By combining DCE tasks, CV questions, and strategic treatments (e.g., disclosing or withholding attribute levels in advance, offering or denying feedback on comprehension), we gather a rich set of observations that illuminate how respondents construct WTP for mortality risk reductions – and what might systematically distort those valuations.

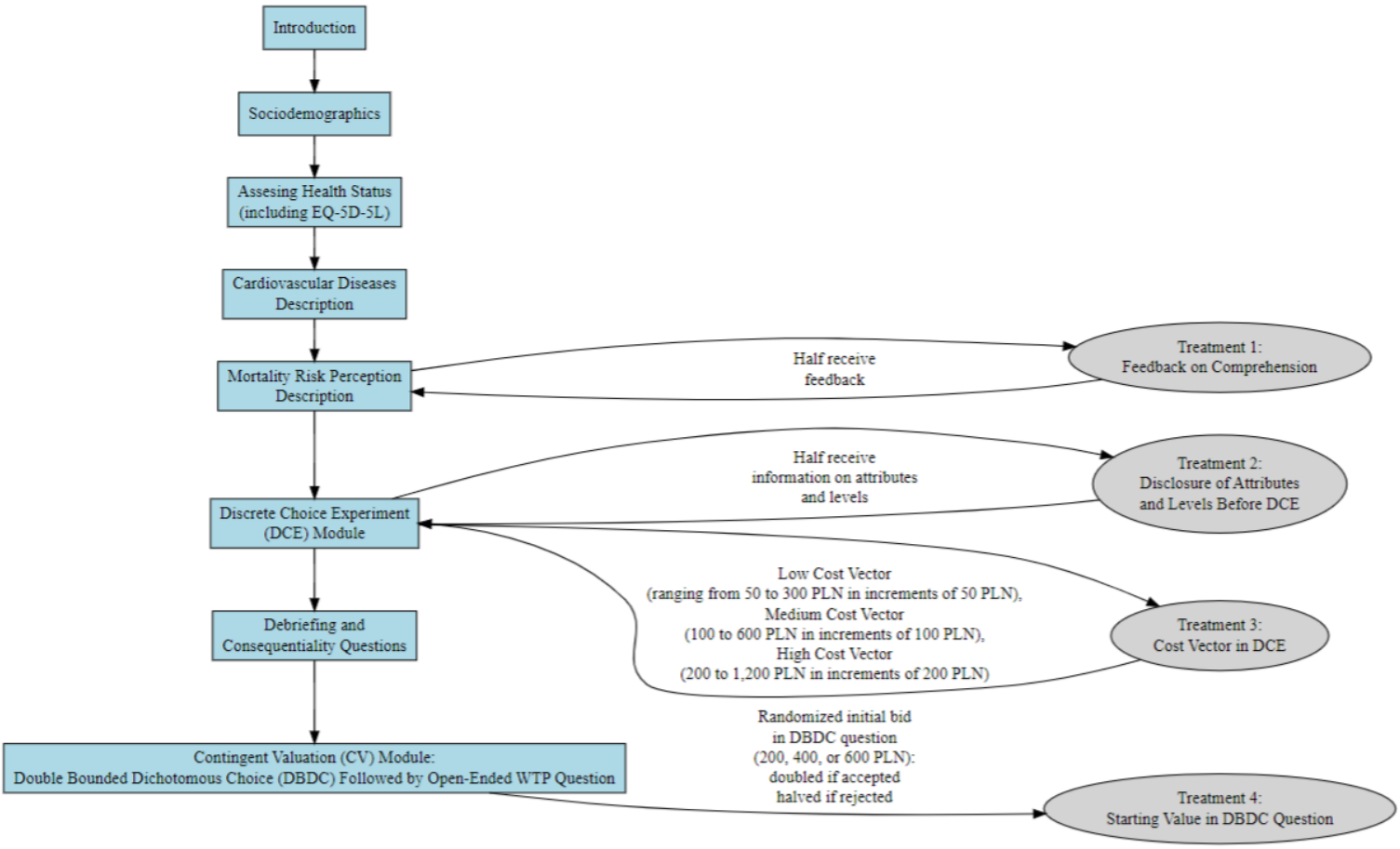
3.2. Experimental Treatments

Building on the survey instrument design, our study incorporates several randomized treatments to examine cost vector effects, potential anchoring, and feedback mechanisms. These treatments include: (1) **varying the cost attribute** across three levels, (2) **pre-disclosing or withholding** the full set of attribute levels, and (3) **providing or denying feedback** on respondents' comprehension-question performance. These randomized assignments were block-managed to ensure balanced representation and adequate sample sizes within each treatment combination. By systematically assigning participants to these different conditions, we can isolate how each factor influences stated preferences for mortality risk reductions. Figure 1 depicts how treatments and modules were integrated into the survey sequence.

A primary experimental treatment involves assigning respondents to one of three cost vectors in the DCE. These vectors – low, medium, and high – differ in both their starting point and the increments between levels:

- **Low-Cost Vector (LC):** Annual costs ranging from 50 to 300 PLN, in 50 PLN increments.
- **Medium-Cost Vector (MC):** Annual costs ranging from 100 to 600 PLN, in 100 PLN increments.
- **High-Cost Vector (HC):** Annual costs ranging from 200 to 1,200 PLN, in 200 PLN increments.

Figure 1. Flowchart of Questionnaire Structure and Randomized Treatment Allocation



During pilot testing, these ranges were selected for realism and salience, drawing on existing research regarding cardiovascular disease interventions (Bartczak et al., 2021) and qualitative feedback from focus groups. Each participant was randomly assigned to exactly one of these three cost sets. This randomization allows us to assess whether different cost vectors elicit divergent WTP responses – and, by extension, whether cost levels themselves serve as anchors. A more detailed discussion on the selection of cost vector elements is presented in Section 3.1.

Half of the respondents in each cost-treatment group were shown a screen listing all possible levels of the risk-reduction and cost attributes prior to seeing any DCE choice tasks. The other half received no preview of attribute levels. This design feature tests for anchoring effects by examining whether early exposure to the full range of costs systematically shifts WTP or preference structures. If disclosing high cost figures acts as an anchor, we would expect these participants to exhibit higher WTP values than those who do not see the preview screen.

Because understanding mortality risk is cognitively demanding, respondents first completed three comprehension questions about the risk illustrations. After each question, half of the sample received immediate feedback (e.g., "Correct, that's the right answer" or "Incorrect, the correct number is…"), whereas the remaining half did not. We hypothesize that respondents who received feedback might refine their understanding of probabilities and thus display more internally consistent risk valuations. Conversely, those who learned they misunderstood a question might experience frustration or mistrust in the survey, potentially dampening or biasing their subsequent WTP.

Overall, the combination of cost vector adjustments, anchoring manipulations, and feedback interventions within a randomized controlled design offers a robust framework for disentangling how each factor – alone and in concert – may shape respondents' preferences and stated willingness to pay for life-saving health interventions.

3.3. Sampling and Administration

The study sample was administered through a computer-assisted web interviewing (CAWI) platform in March 2023, yielding 1,299 usable responses from Polish residents. This approach provided a cost-effective and flexible means of reaching a broad cross-section of the population while offering advanced tools for randomization and response tracking.

To ensure demographic diversity, we implemented quota-based recruitment, aligning the sample with national statistics for sex, age, education, and region. Table 1 presents the sociodemographic breakdown, confirming that the final sample closely reflects the broader Polish adult population. In total, 416 respondents were assigned to the low-cost (LC) vector, 433 to the medium-cost (MC) vector, and 450 to the high-cost (HC) vector, with further splits within each group for attribute disclosure and comprehension feedback treatments. Additional details on sample characteristics and stratification are provided in Table 1, confirming that each treatment group mirrors the broader Polish population across key sociodemographic variables.

Potential participants received an online invitation explaining the general nature of the study (i.e., research on health and mortality risk) and assuring them of data confidentiality. Respondents who agreed to participate accessed the questionnaire via a secure link, where the survey's internal randomization engine assigned them to the relevant treatments. Progress bars and time reminders helped maintain engagement and alert participants to the approximate length of the survey.

Table 1. Representativeness of the sample

		Full sample	Low cost	Medium cost	High cost
N	Total sample	1299	416	433	450
Sex	Female	53.27%	54.81%	53.35%	51.78%
	Male	46.73%	45.19%	46.65%	48.22%
Age group	18-24	9.70%	9.38%	9.24%	10.44%
	25-44	40.03%	39.42%	41.80%	38.89%
	45-64	32.10%	32.93%	31.64%	31.78%
	65+	18.17%	18.27%	17.32%	18.89%
Education	Incomplete Primary or No School Education	0.46%	0.72%	0.46%	0.22%
	Completed Primary Education	2.00%	2.40%	2.54%	1.11%
	Lower Secondary Education	1.08%	0.96%	1.15%	1.11%
	Basic Vocational Education	11.70%	12.50%	10.62%	12.00%
	Secondary or Post-secondary Non-tertiary Education	38.57%	35.10%	39.03%	41.33%
	Higher Education	46.19%	48.32%	46.19%	44.22%
Region (voivodeship)	dolnośląskie	7.16%	7.21%	6.93%	7.33%
	kujawsko-pomorskie	5.16%	5.53%	5.54%	4.44%
	lubelskie	5.77%	5.29%	5.31%	6.67%
	lubuskie	2.16%	2.40%	1.85%	2.22%
	łódzkie	6.62%	6.25%	7.39%	6.22%
	małopolskie	6.93%	6.49%	6.93%	7.33%
	mazowieckie	15.09%	15.87%	16.17%	13.33%
	opolskie	2.62%	2.16%	3.23%	2.44%
	podkarpackie	4.93%	6.25%	3.70%	4.89%
	podlaskie	3.23%	3.61%	2.77%	3.33%
	pomorskie	7.08%	6.25%	9.01%	6.00%
	śląskie	14.47%	12.98%	13.86%	16.44%
	świętokrzyskie	2.23%	1.68%	2.77%	2.22%
	warmińsko-mazurskie	4.08%	4.09%	3.93%	4.22%
	wielkopolskie	8.70%	10.34%	7.39%	8.44%
	zachodniopomorskie	3.77%	3.61%	3.23%	4.44%
Household size	1	13.47%	13.94%	15.01%	11.56%
	2	36.80%	36.30%	36.49%	37.56%
	3	23.71%	23.56%	23.09%	24.44%
	4	16.32%	15.14%	15.47%	18.22%
	5+	9.70%	11.06%	9.93%	8.22%

To enhance the quality of data collected, several mechanisms were systematically implemented throughout the survey process. First, completion time thresholds were established to flag responses submitted too quickly, which could indicate minimal engagement with the survey questions, often referred to as "speeding." In addition to monitoring completion times, attention and comprehension checks were integrated throughout the survey. These were not limited only to the main risk-comprehension tasks; reminders were also interspersed to prompt respondents to focus on the trade-offs required by the survey tasks, aiming to mitigate careless or inattentive responses. Furthermore, technical support was available, offering clarification on question wording and study objectives to help respondents resolve misunderstandings promptly, thereby maintaining the integrity and accuracy of the data collected.

Before full-scale data collection commenced, the survey was subjected to multiple rounds of qualitative testing to ensure its robustness and clarity. Initially, four medical experts conducted an expert review to verify the accuracy of the descriptions of cardiovascular diseases and the dimensions of associated risks. Following this, three focus groups consisting of individuals with varied demographic profiles were convened to work through a preliminary version of the questionnaire. Their feedback provided valuable insights on the clarity of the questions and the plausibility of the cost levels presented. Additionally, individual interviews were conducted with a language specialist and two general respondents. These verbal-protocol sessions were instrumental in refining the wording of questions related to risks and costs. These comprehensive and iterative steps ensured that the final instrument was clear and understandable to non-experts, minimized cognitive load, and avoided the use of overly technical language or ambiguous phrases. Notably, the survey instrument and research idea were also consulted with health economics and choice experiment experts at several workshops and conferences (citation to post-conference materials – the abstract – is hidden to preserve anonymity).

In keeping with standard ethical guidelines for survey-based research, all participants provided informed consent at the outset. They were reminded that their participation was voluntary and that they could exit at any stage without penalty. Each respondent was compensated through the survey panel provider according to standard rates – compensation levels considered minor but sufficient to encourage thoughtful participation.

Upon survey completion, responses were compiled and subjected to basic filtering. Observations flagged for excessive speeding or major inconsistencies in demographic data were removed prior to analysis. After final checks, 1,299 valid responses remained, forming the dataset used for the subsequent econometric modeling (Sections 4 and 5).

Overall, this sampling and administration process aligns with best practices in stated preference research by ensuring a balanced, sufficiently large participant pool, employing rigorous randomization, and implementing careful data-quality protocols. The resulting dataset provides a robust foundation for investigating how cost vector design, anchoring, and scope-related issues impact respondents' willingness to pay for a hypothetical cardiovascular risk-reducing intervention.

4. Econometric Framework

SP research often relies on DCEs to estimate individuals' valuations of non-market goods. This study analysis follows a utility-maximization approach rooted in Lancaster's consumer theory (Lancaster, 1966) and random utility theory (McFadden, 1973). Within this framework, an individual i considering a set of alternatives in a choice situation t is assumed to select the alternative j that maximizes the individual's overall utility, denoted by:

$$U_{ijt} = X_{ijt}\beta_i + \varepsilon_{ijt},$$

where X_{ijt} is a vector of observed attributes describing alternative j, β_i is a vector of preference parameters specific to individual i, and ε_{ijt} is a random error term capturing unobserved utility components.

To accommodate heterogeneity in preferences and relax the independence-from-irrelevant-alternatives (IIA) assumption inherent in the simple multinomial logit, we adopt a mixed logit (also known as a random parameters logit) specification (Train, 2009). In this model, the parameter vector β_i is drawn from a specified probability distribution $f(\beta \mid \theta)$, where θ represents distributional parameters (e.g., means, standard deviations, or correlations). For each choice situation t, the probability that individual i selects alternative j from a set of J alternatives is written as:

$$p_{ijt}(\theta) = \int \left[\frac{exp\left(X_{ijt}\beta\right)}{\sum_{j=1}^{J} exp\left(X_{ijt}\beta\right)} f(\beta_n|\theta) \right] d\beta_n.$$

Because there is no closed-form solution for this integral, we approximate it through simulation. Specifically, we draw from the distribution $f(\beta \mid \theta)$ 10,000 times (scrambled Sobol draws) and average the resulting choice probabilities to obtain simulated likelihood estimates (Czajkowski & Budziński, 2019; Hole, 2007b).

Following standard practice in health and environmental valuation (Hensher & Greene, 2003; Mariel et al., 2021), we allow the cost parameter to follow a lognormal distribution to ensure that marginal utilities for cost lie on the expected side of zero (i.e., a negative coefficient for a price-like attribute). Meanwhile, the coefficients for mortality risk reductions and alternative-specific constants (ASCs) are assumed to follow a normal distribution.

To account for correlations among random coefficients, reflecting the possibility that unobserved factors influencing one parameter may be related to unobserved factors affecting

others (Mariel & Artabe, 2020), we allow the random coefficients to be correlated. Therefore, our approach can be summarized as the mixed logit model with correlated random parameters.

From the estimated parameter distributions, we derive WTP for specific attribute changes using a ratio of coefficients. In a simple two-attribute model (risk reduction and cost), the marginal WTP for a one-unit change in risk reduction is:

$$WTP = -\frac{\beta_{Risk}}{\beta_{Cost}},$$

where β_{Risk} is the estimated coefficient for mortality risk reduction and β_{Cost} is the estimated coefficient for the cost attribute. Negative signs reflect the fact that cost typically has a negative sign in utility functions. Because the distribution of these parameters is random, we obtain a distribution of WTP estimates, which we characterize by means, standard deviations, and confidence intervals. We rely on simulation methods, specifically the Krinsky-Robb procedure (Krinsky & Robb, 1991), for constructing these intervals (Hole, 2007a).

To translate WTP for a given mortality risk reduction into the VSL, we follow a common approach in the VSL literature (Andersson & Treich, 2011; Kniesner & Viscusi, 2019). Specifically, if $WTP(\Delta\text{Risk})$ is the annual willingness to pay for lowering mortality risk by $\Delta Risk$, we convert it into a VSL by dividing by the change in risk and adjusting for the relevant time horizon. For an annual cost that persists over 10 years, for example, the VSL for a risk reduction from, say, 1/1,000 to 4/1,000 (i.e., a $\Delta Risk$=3/1,000 shift) is often expressed as:

$$VSL = \frac{WTP(\Delta\text{Risk})}{\Delta\text{Risk}} \times 10.$$

This scaled figure represents the implied monetary value placed on preventing one statistical death, based on the incremental gain in survival probability and the payment period.

In sum, our econometric framework is designed to capture individual heterogeneity in preferences for reducing cardiovascular mortality risk. By combining a mixed logit approach with robust simulation methods for obtaining WTP and VSL estimates, we can assess the degree to which cost vector design, anchoring, and scope insensitivity might distort or stabilize these valuations. The next section reports the main results, beginning with DCE estimates before examining the CV responses.

5. Results

5.1. Discrete Choice Experiment Findings

Our main goal in this section is to determine whether different cost vectors in the DCE produce systematically different valuations and, if so, how this variation affects VSL estimates. We therefore estimated three mixed logit models with correlated random parameters – one each for respondents who faced the low-cost (LC), medium-cost (MC), and high-cost (HC) vectors. An alternative-specific constant (ASC) was included to capture any baseline preference for an option that offers mortality risk reduction over the status quo. Meanwhile, we decomposed the risk attribute into two binary indicators, representing reductions of two and four deaths per 1,000 people over 10 years; a reduction of one death per 1,000 served as the baseline.

Table 2. Mixed Logit Results with Correlated Random Parameters and Derived VSL Estimates for Low Cost, Medium Cost, and High Cost Samples

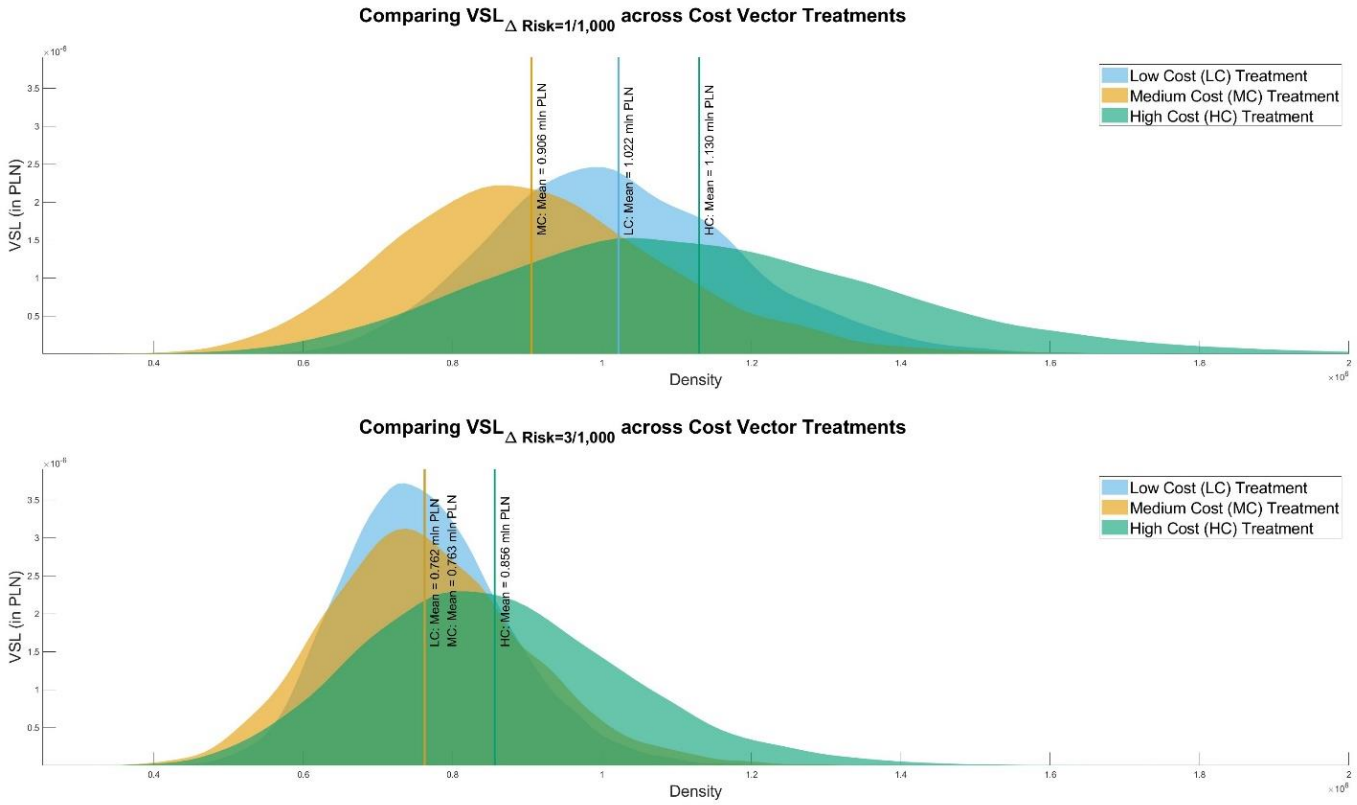
MXL results						
	Low cost vector (LC)		**Medium cost vector (MC)**		**High cost vector (HC)**	
Attributes	Mean (μ)	Standard Deviation (σ)	Mean (μ)	Standard Deviation (σ)	Mean (μ)	Standard Deviation (σ)
ASC	4.45***	5.32***	4.27***	4.87***	4.95***	6.41***
(norm dist.)	(0.45)	(0.53)	(0.42)	(0.71)	(0.52)	(0.66)
Risk==2	1.97***	2.06***	1.71***	1.98***	1.51***	2.01***
(norm dist.)	(0.17)	(0.19)	(0.17)	(0.20)	(0.17)	(0.18)
Risk==4	4.72***	5.11***	4.39***	5.00***	3.34***	4.34***
(norm dist.)	(0.36)	(0.39)	(0.37)	(0.42)	(0.30)	(0.35)
Annual Cost/100 in PLN	0.10	2.18***	-0.29**	2.75***	-0.74***	2.60***
(log-norm dist.)	(0.14)	(0.16)	(0.15)	(0.18)	(0.14)	(0.20)
VSL estimates						
VSL estimates	Mean (std. dev)	CI 2.5% CI 97.5%	Mean (std. dev)	CI 2.5% CI 97.5%	Mean (std. dev)	CI 2.5% CI 97.5%
$VSL_{\Delta Risk=1}$ (PLN)	1,021,993*** (165,791)	726,581 1,378,214	905,684*** (184,412)	578,997 1,297,386	1,130,028*** (267,752)	665,903 1,719,090
$VSL_{\Delta Risk=3}$ (PLN)	762,240*** (108,595)	571,776 995,184	762,871*** (133,090)	523,717 1,051,394	856,401*** (175,152)	549,631 1,236,756

Notes: Significance levels: *** for p<0.01, ** for p<0.05, and * for p<0.1. $VSL_{\Delta Risk=1}$ indicates the value of statistical life based on the WTP for a mortality risk reduction from 1 to 2 per 1,000 people. $VSL_{\Delta Risk=3}$ is calculated for a reduction from 1 to 4 per 1,000. The confidence intervals were estimated using the Krinsky-Robb simulation procedure. Standard errors in parentheses.

Table 2 presents the results of these models and reports the VSL calculations. To account for scope sensitivity, we first simulated WTP for two- and four-death risk reductions in each cost treatment, then derived separate VSL measures for each reduction level by dividing the simulated WTP by the change in risk (1/1,000 or 3/1,000) and multiplying by 10 to reflect the 10-year time horizon. This approach follows standard practice (Andersson & Treich, 2011; Kniesner & Viscusi, 2019; Robinson et al., 2019). The VSL estimates are visualized in Figure 2.

Under full scope sensitivity, all welfare measures would remain consistent across different levels of mortality risk reduction; deviations from that pattern can indicate cost vector effects or scope insensitivity.

Figure 2. Visual Comparison of VSL estimates – Density Ridgeline Plots



As shown, VSL estimates differ by up to 48% across treatments, ranging from a low of about 762,000 PLN (LC $VSL_{\Delta Risk=3}$) to a high of just over 1,130,000 PLN (HC $VSL_{\Delta Risk=1}$). Notably, mean VSL in the HC sample is about 12% higher than in the LC sample (for both $VSL_{\Delta Risk=1}$ and $VSL_{\Delta Risk=3}$). Meanwhile, the $VSL_{\Delta Risk=1}$ estimate in MC is roughly 10% lower than its LC counterpart, whereas $VSL_{\Delta Risk=3}$ in LC and MC are almost identical. The largest difference in mean VSL estimates across treatments (approximately 25%) occurs between the MC and HC groups (for $VSL_{\Delta Risk=1}$). While these discrepancies appear substantial, statistical tests (Poe et al., 2005) indicate these differences are not significant at conventional levels.

A notable pattern is that in each model, the higher-risk-reduction scenario ($VSL_{\Delta Risk=3}$) yields a smaller implied value of life than the lower-risk-reduction scenario ($VSL_{\Delta Risk=1}$). Differences range from 16% in MC to 25% in LC. This result conflicts with the assumption of

perfect scope sensitivity, as we would theoretically expect similar marginal rates of substitution for risk reductions of one additional death versus three.

Using the Poe et al. (2005) convolutions test, we find no statistically significant differences among VSL estimates across cost vectors (LC vs. MC, LC vs. HC, and MC vs. HC). In other words, although there are differences in mean values, those differences appear indistinguishable from sampling variation. By contrast, when comparing internal scope sensitivity ($VSL_{\Delta Risk=1}$ vs. $VSL_{\Delta Risk=3}$ within the same cost treatment), the test yields smaller p-values (0.09 in LC, 0.27 in MC, and 0.19 in HC), suggesting partial scope insensitivity – particularly in the LC treatment.

We also estimated a second model incorporating covariates of means for the continuous risk attribute (Supplementary Materials Table S1). While no significant differences arose between the MC and LC samples, the HC sample differed significantly from LC, reinforcing the observed disparity. Neither the disclosure of attribute levels nor including extreme cost levels (lowest or highest) in the first choice task influenced preferences in a statistically significant way. Similarly, whether participants had been informed of correct answers to risk-comprehension questions did not alter the basic relationship between cost and mortality risk.

These extended results further highlight notable heterogeneity: respondents' demographic factors, health status, and risk perceptions can affect WTP for mortality risk reduction, yet gender differences were not significant despite the known higher cardiovascular risk among males. Overall, these findings affirm that while cost vector ranges may shift mean VSL estimates, the variation is often not strong enough to be considered statistically meaningful, though the interplay with scope sensitivity underscores the complexity of health-related DCE design.

5.2. Contingent Valuation Findings

Following the DCE, respondents completed a CV module comprising a double-bounded dichotomous choice (DBDC) sequence and an open-ended (OE) question. In the DBDC format, each respondent was presented with an initial bid of 200, 400, or 600 PLN, drawn at random. Those who accepted the initial bid were offered a follow-up bid set at twice the initial amount; those who rejected it received a follow-up bid at half the initial amount. Table 3 documents the proportions of "yes-yes," "yes-no," "no-yes," and "no-no" across the low-cost (LC), medium-cost (MC), and high-cost (HC) DCE treatments, as well as mean and median responses to the

OE question. Overall, the data do not reveal consistent or strong patterns linking the assigned DCE cost vector to acceptance rates or the final WTP amounts. While some discrepancies appear at certain bid levels (e.g., slightly higher "yes-yes" acceptance rates in the HC group for the 200 PLN and 600 PLN bids), Two-Proportion Z-tests fail to confirm their statistical significance in most comparisons (Supplementary Materials Table S2).

Table 3. DBDC and OE Responses by Cost Treatment

		Full sample	Low cost DCE (LC)	Medium cost DCE (MC)	High cost DCE (HC)
N	Total sample	1299	416	433	450
OE WTP	mean	378.40	355.75	361.29	415.79
	median	150	150	200	200
N	X=200	488	156	175	157
		37.57%	37.50%	40.42%	34.89%
OE WTP X=200	mean	324.67	320.81	260.16	400.41
	median	150	100	200	200
DBDC X=200	Yes (200 PLN) - Yes (400 PLN)	34.84%	30.77%	33.14%	40.76%
	Yes (200 PLN) - No (400 PLN)	21.93%	19.87%	26.86%	18.47%
	No (200 PLN) - Yes (100 PLN)	11.27%	15.38%	10.29%	8.28%
	No (200 PLN) - No (100 PLN)	31.97%	33.97%	29.71%	32.48%
N	X=400	423	123	137	163
		32.56%	29.57%	31.64%	36.22%
OE WTP X=400	mean	420.99	463.17	395.92	410.22
	median	200	200	300	200
DBDC X=400	Yes (400 PLN) - Yes (800 PLN)	25.06%	25.20%	22.63%	26.99%
	Yes (400 PLN) - No (800 PLN)	19.39%	19.51%	22.63%	16.56%
	No (400 PLN) - Yes (200 PLN)	13.71%	15.45%	12.41%	13.50%
	No (400 PLN) - No (200 PLN)	41.84%	39.84%	42.34%	42.94%
N	X=600	388	137	121	130
		29.87%	32.93%	27.94%	28.89%
OE WTP X=600	mean	399.55	299.10	468.35	441.36
	median	140	150	300	150
DBDC X=600	Yes (600 PLN) - Yes (1200 PLN)	18.56%	14.60%	18.18%	23.08%
	Yes (600 PLN) - No (1200 PLN)	15.21%	12.41%	18.18%	15.38%
	No (600 PLN) - Yes (300 PLN)	14.18%	16.79%	13.22%	12.31%
	No (600 PLN) - No (300 PLN)	52.06%	56.20%	50.41%	49.23%

In addition to the DBDC questions, each participant provided an exact OE WTP response. Statistical tests (Permutation tests for medians, Two-Sample T-tests for means, and Mann-Whitney U tests for distributional differences) showed no significant variation in OE-based WTP across the three cost-vector groups (Supplementary Materials Table S3). Taken together, these findings suggest that CV responses remain stable regardless of whether participants previously encountered relatively low, medium, or high cost vectors in the DCE.

While the DCE results (Section 5.1) indicated moderate differences in mean VSL across cost treatments (up to 25% in some comparisons), the CV-based WTP shows far less variation. This discrepancy is notable because it suggests that when respondents value a single, narrowly defined risk reduction – rather than choosing among a few alternatives – they may be less sensitive to the ranges of cost levels shown in a prior exercise. In other words, the CV elicitation format may mitigate certain design-based biases that appear in the DCE, such as cost vector effects.

Nonetheless, caution should be exercised in directly comparing raw CV-based and DCE-based estimates. The DCE tasks value different magnitudes of mortality risk reduction (1 vs. 2 vs. 4 fewer deaths per 1,000 people), whereas the CV tasks focus on a single measure: reducing one death per 1,000. Moreover, the structure of a DBDC sequence differs fundamentally from the repeated choice scenarios in a DCE. Still, the broad takeaway is that CV responses do not exhibit the same degree of variability associated with cost levels that we observed in the DCE.

Although anchoring and framing effects can arise in contingent valuation (e.g., via initial bid values or prior exposure to cost information), we observe no systematic shifts in acceptance rates or OE WTP that would point to strong anchoring from the DCE phase. Even where the initial DBDC bid was comparatively high, subsequent bids and final OE amounts do not differ significantly between respondents who encountered high DCE cost vectors and those who saw low or medium ones. Likewise, disclosing or withholding the full set of attribute levels before the DCE – another possible source of anchoring – does not appear to have carried over into the CV responses.

In summary, the CV component yields relatively uniform WTP estimates that are largely insensitive to the earlier DCE cost range manipulations, with no strong evidence of anchoring or framing bias. This does not, however, negate the importance of careful DCE design; as the earlier sections show, multi-attribute choice tasks and cost vector selection can create notable variation in implied VSL. The next section interprets these findings in a broader methodological context, discussing possible explanations and implications for both research and policy.

6. Discussion

SP methods have long been used to investigate WTP for health improvements (Bridges, 2003; Diener et al., 1998; Klose, 1999; Shiell & Gold, 2003). In particular, DCEs play a critical role in valuing mortality risk reductions, despite ongoing debates concerning their reliability

(Doucouliagos et al., 2012; Weinstein et al., 2001). Our findings contribute to this discussion by showing that preferences related to cardiovascular disease mortality risk appear generally consistent across different cost vector designs, even though mean VSL estimates can vary by up to 25%. Statistical tests, based on Poe et al. (2005), indicate these observed differences are not significant at conventional levels, suggesting that, in practice, respondents' underlying preferences remain relatively stable across low, medium, and high cost treatments. However, because health policy practitioners often prioritize point estimates of marginal willingness to pay to inform decision-making, substantially different mean values may arise even when the underlying VSL/WTP distributions are similar, potentially leading to misguided policy decisions (Chandoevwit & Wasi, 2020).

Although scope sensitivity is theoretically essential for health risk valuation – since WTP should rise in proportion to larger risk reductions – our results underscore a persistent partial scope insensitivity in certain conditions. Specifically, we observe that respondents' WTP for reducing mortality risk from one to four deaths per 1,000 is less than proportional when compared to smaller increments, particularly in the low-cost (LC) DCE version. These findings align with earlier literature that identifies difficulty among respondents in scaling willingness to pay for changes in small probabilities (Andersson et al., 2016; Kahneman & Knetsch, 1992). Despite our observation of weak scope sensitivity – WTP did rise for bigger risk reductions – the lack of full proportionality exposes a challenge in eliciting consistent values for life-threatening health risks.

Anchoring effects are another major concern in SP studies (Ariely et al., 2003; Tversky & Kahneman, 1974), yet we detect little evidence that varying the extremes of the cost vector or providing attribute disclosures affects respondents' final valuations. Neither the inclusion of extremely high or low cost levels in the first choice set nor showing respondents all possible attribute levels in advance significantly influenced estimated preferences. This suggests that participants may rely more on other cues – such as their own health status or perceived baseline mortality risk – rather than latching onto extreme cost amounts as anchors, at least for the type of health-risk reductions considered here.

Overall, our findings confirm that while there is some variability in mean VSL estimates across cost vectors, these differences often do not rise to the level of statistical significance. The more pressing concern appears to be scope insensitivity within specific treatments, highlighting the difficulty that respondents have in valuing changes in small mortality risks.

The observed partial scope insensitivity and variability in VSL across cost treatments emphasize the importance of careful design choices in stated preference research. First and foremost, **cost vector spacing** emerges as a critical factor. Our results suggest that larger intervals between cost levels can obscure the precise revelation of WTP, potentially inflating or deflating final welfare measures. Conversely, smaller increments may improve granularity but risk failing to suppress demand if the highest cost level is set too low. Hence, researchers should balance the need for finer cost spacing with ensuring that at least one cost level is sufficiently large to capture the "choke price" for most respondents.

Second, **feedback mechanisms** – such as informing participants of correct or incorrect responses to risk-comprehension tests – did not substantially shift valuations in our sample. Although providing feedback can help maintain engagement and clarify complex probabilities, it appears that it may not drastically alter WTP estimates in health contexts like ours. Nevertheless, carefully implemented comprehension checks remain advisable as a means of ensuring that respondents understand the nature of the good being valued.

Third, while **framing and anchoring** are well-known concerns, our study finds that disclosing attribute levels upfront or including extreme cost values in the initial choice set does not automatically lead to biased results. This outcome may reflect the intrinsic salience of health issues: respondents might rely on personal experiences and risk perceptions rather than purely external numeric cues. It would therefore be premature to conclude that anchoring poses no risk in other, less salient contexts. Researchers should continue to randomize the presentation of attribute information and test for anchoring effects where possible.

In practical terms, we recommend that practitioners designing SP studies for valuing health risks adopt the following principles:

1. **Ensure Wide yet Realistic Cost Ranges** – incorporate at least one high cost level sufficient to "choke off" demand if respondents' true WTP is below that threshold.
2. **Use Refined Cost Increments** – finer granularity (i.e., smaller intervals between cost levels) can help reveal variations in WTP, although sample sizes may need to be increased to maintain statistical power.
3. **Include Comprehension Checks** – this helps confirm that respondents understand key elements (e.g., the probability changes), but do not assume that feedback alone will correct biases.

4. **Test for Scope Sensitivity Explicitly** – present multiple magnitudes of risk reduction and analyze whether WTP scales proportionally. Use follow-up checks or alternative elicitation formats to ensure respondents grasp small probability changes.
5. **Evaluate Anchoring Risks** – randomize the order in which cost levels or attributes are introduced, and, where feasible, compare sub-samples exposed to different framing strategies.

By incorporating these considerations, researchers can reduce the influence of cost vector effects, mitigate scope insensitivity, and address anchoring concerns. While our findings demonstrate that cost vector selection does not entirely undermine the consistency of VSL estimates, misalignment in design choices can significantly alter mean welfare measures. Continued investigation into best practices – particularly in health applications where mortality risk is complex – will advance both methodological rigor and policy relevance in stated preference studies.

7. Conclusions

This study investigated the robustness of VSL estimates derived from SP methods – particularly DCEs – in the context of cardiovascular disease mortality risk reductions. By systematically varying cost vectors, examining potential anchoring effects, and testing for scope sensitivity, we shed new light on how survey design factors and cognitive biases may shape respondents' WTP. Our main findings highlight that although mean VSL values can differ by up to 25% across low, medium, and high cost vectors, these differences generally fail to reach statistical significance. Moreover, neither early disclosure of attribute levels nor feedback about comprehension significantly altered participants' preferences. Together, these results suggest that, while cost vector design is indeed relevant, the underlying preferences are relatively stable when facing small changes in mortality risk. However, scope insensitivity remains a challenge, indicating that respondents do not always scale their WTP in proportion to larger risk reductions.

From a broader perspective, this study underscores the importance of controlling for potential behavioral biases in health risk valuation. Cost vector effects can subtly shift average WTP if intervals are not carefully calibrated, and scope insensitivity undermines the very premise that WTP for mortality risk reduction should correlate with changes in risk magnitude.

By identifying where and how these biases emerge, researchers can refine SP methods to produce more reliable estimates for health-related policy applications.

While our findings offer valuable insights, the study also has certain limitations that point toward fruitful avenues for further research. First, we focused on a specific type of risk: mortality from cardiovascular diseases. Although this choice allowed for a more in-depth analysis of cost vector design and anchoring, future studies might explore broader health contexts or other risk factors to assess whether similar patterns hold. Second, despite experimenting with three different cost ranges, we did not test every possible spacing, nor did we incorporate morbidity attributes – an aspect that can influence valuation of health interventions (Gentry & Viscusi, 2016). Future research might extend this work by incorporating morbidity risk or additional attributes to capture a more comprehensive view of health outcomes.

Third, our design did not employ pivot-based approaches, which can tailor cost levels to each respondent's baseline conditions, potentially reducing irrelevant variation in cost vectors (Hess & Rose, 2009; Rose & Hess, 2009). Investigating whether pivot designs mitigate the scope insensitivity observed here is a logical next step. Finally, we focused on conventional presentation formats for small probability changes, but growing evidence suggests that re-expressing these as percentages or using visual cues may improve comprehension (Herrera-Araujo et al., 2022). Implementing such strategies – and systematically comparing them to more traditional formats – could offer further insights into how individuals process small but critical shifts in health risks.

In sum, while our analysis lends support to the consistency of core preferences in stated choice surveys, it also underscores the pressing need for standardized guidelines on cost vector selection, attention to scope sensitivity, and continued innovation in survey design. By building on this body of work, researchers and practitioners can refine the methods used to value life-saving health interventions, ultimately leading to more robust policy recommendations and better public health outcomes.

References

Ahtiainen, H., Pouta, E., Zawadzki, W., Tienhaara, A., 2023. Cost vector effects in discrete choice experiments with positive status quo cost. Journal of Choice Modelling, 47, 100401. https://doi.org/10.1016/j.jocm.2023.100401

Alberini, A., 2019. Revealed versus stated preferences: what have we learned about valuation and behavior? Review of Environmental Economics and Policy. https://doi.org/10.1093/reep/rez010

Andersson, H., Hole, A.R., Svensson, M., 2016. Valuation of small and multiple health risks: A critical analysis of SP data applied to food and water safety. Journal of Environmental Economics and Management, 75, 41-53. https://doi.org/10.1016/j.jeem.2015.11.001

Andersson, H., Treich, N., 2011. The value of a statistical life. In A handbook of transport economics. Edward Elgar Publishing. https://doi.org/10.4337/9780857930873.00025

Aravena, C., Martinsson, P., Scarpa, R., 2014. Does money talk?—The effect of a monetary attribute on the marginal values in a choice experiment. Energy Economics, 44, 483-491. https://doi.org/10.1016/j.eneco.2014.02.017

Ariely, D., Loewenstein, G., Prelec, D., 2003. "Coherent arbitrariness": Stable demand curves without stable preferences. The Quarterly Journal of Economics, 118(1), 73-106. https://doi.org/10.1162/00335530360535153

Banzhaf, H.S., 2022. The value of statistical life: A meta-analysis of meta-analyses. Journal of Benefit-Cost Analysis, 13(2), 182-197. https://doi.org/10.1017/bca.2022.9

Bartczak, A., Budziński, W., Chilton, S., McDonald, R., Nielsen, J.S., 2021. Altruism and efficient allocations in three-generation households. Journal of Risk and Uncertainty, 62(2), 113-135. https://doi.org/10.1007/s11166-021-09350-z

Bateman, I., Munro, A., Rhodes, B., Starmer, C.V., Sugden, R., 2006. Anchoring and yea-saying with private goods: An experiment. In Using experimental methods in environmental and resource economics (pp. 1-19). Edward Elgar Publishing. https://doi.org/10.4337/9781847203045.00006

Bateman, I.J., Carson, R.T., Day, B., Dupont, D., Louviere, J.J., Morimoto, S., Scarpa, R., Wang, P., 2008. Choice set awareness and ordering effects in discrete choice experiments in discrete choice experiments.

Bateman, I.J., Cole, M., Cooper, P., Georgiou, S., Hadley, D., Poe, G.L., 2004. On visible choice sets and scope sensitivity. Journal of Environmental Economics and Management, 47(1), 71-93. https://doi.org/10.1016/S0095-0696(03)00057-3

Bestard, A.B., Font, A.R., 2021. Attribute range effects: Preference anomaly or unexplained variance? Journal of Choice Modelling, 41, 100321. https://doi.org/10.1016/j.jocm.2021.100321

Bishop, R.C., Boyle, K.J., 2019. Reliability and validity in nonmarket valuation. Environmental and Resource Economics, 72, 559-582. https://doi.org/10.1007/s10640-017-0215-7

Börger, T., Glenk, K., Meyerhoff, J., Rehdanz, K., 2025. Mitigating cost-vector effects in stated choice experiments using cheap talk and opt-out reminders. Journal of the Association of Environmental and Resource Economists, 12(3), 701-737. https://doi.org/10.1086/731886

Boyle, K.J., Bishop, R.C., Welsh, M.P., 1985. Starting point bias in contingent valuation bidding games. Land Economics, 61(2), 188-194. https://doi.org/10.2307/3145811

Bridges, J., 2003. Stated preference methods in health care evaluation: an emerging methodological paradigm in health economics. Applied Health Economics and Health Policy, 2(4), 213-224.

Carlsson, F., Martinsson, P., 2008. How much is too much? An investigation of the effect of the number of choice sets, context dependence and the choice of bid vectors in choice experiments. Environmental and Resource Economics, 40, 165-176. https://doi.org/10.1007/s10640-007-9146-z

Carson, R.T., 1997. Contingent valuation surveys and tests of insensitivity to scope, in: Determining the value of non-marketed goods: economic, psychological, and policy relevant aspects of contingent valuation methods (pp. 127-163). Springer.

Carson, R.T., Groves, T., 2007. Incentive and informational properties of preference questions. Environmental and Resource Economics, 37(1), 181-210. https://doi.org/10.1007/s10640-007-9124-5

Chandoevwit, W., Wasi, N., 2020. Incorporating discrete choice experiments into policy decisions: case of designing public long-term care insurance. Social Science & Medicine, 258, 113044. https://doi.org/10.1016/j.socscimed.2020.113044

Chilton, S., Hutchinson, W., 2000. A note on the warm glow of giving and scope sensitivity in contingent valuation studies. Journal of Economic Psychology, 21(4), 343-349. https://doi.org/10.1016/S0167-4870(00)00009-X

Corso, P.S., Hammitt, J.K., Graham, J.D., 2001. Valuing mortality-risk reduction: using visual aids to improve the validity of contingent valuation. Journal of Risk and Uncertainty, 23(2), 165-184. https://doi.org/10.1023/A:1011184119153

Czajkowski, M., Budziński, W., 2019. Simulation error in maximum likelihood estimation of discrete choice models. Journal of Choice Modelling, 31, 73-85. https://doi.org/10.1016/j.jocm.2019.04.003

Czajkowski, M., Hanley, N., 2009. Using labels to investigate scope effects in stated preference methods. Environmental and Resource Economics, 44(4), 521-535. https://doi.org/10.1007/s10640-009-9299-z

Diener, A., O'Brien, B., Gafni, A., 1998. Health care contingent valuation studies: a review and classification of the literature. Health Economics, 7(4), 313-326. https://doi.org/10.1002/(SICI)1099-1050(199806)7:4%3C313::AID-HEC350%3E3.0.CO;2-B

Doucouliagos, C., Stanley, T.D., Giles, M., 2012. Are estimates of the value of a statistical life exaggerated? Journal of Health Economics, 31(1), 197-206. https://doi.org/10.1016/j.jhealeco.2011.10.001

Flachaire, E., Hollard, G., 2007. Starting point bias and respondent uncertainty in dichotomous choice contingent valuation surveys. Resource and Energy Economics, 29(3), 183-194. https://doi.org/10.1016/j.reseneeco.2007.05.003

Furnham, A., Boo, H.C., 2011. A literature review of the anchoring effect. The journal of socio-economics, 40(1), 35-42. https://doi.org/10.1016/j.socec.2010.10.008

Gentry, E.P., Viscusi, W.K., 2016. The fatality and morbidity components of the value of statistical life. Journal of Health Economics, 46, 90-99. https://doi.org/10.1016/j.jhealeco.2016.01.011

Glenk, K., Meyerhoff, J., Akaichi, F., Martin-Ortega, J., 2019. Revisiting cost vector effects in discrete choice experiments. Resource and Energy Economics, 57, 135-155. https://doi.org/10.1016/j.reseneeco.2019.05.001

Glenk, K., Meyerhoff, J., Colombo, S., Faccioli, M., 2024. Enhancing the face validity of choice experiments: A simple diagnostic check. Ecological Economics, 221, 108160. https://doi.org/10.1016/j.ecolecon.2024.108160

Goldberg, I., Roosen, J., 2007. Scope insensitivity in health risk reduction studies: A comparison of choice experiments and the contingent valuation method for valuing safer food. Journal of Risk and Uncertainty, 34(2), 123-144. https://doi.org/10.1007/s11166-007-9006-9

Grisolía, J.M., Longo, A., Hutchinson, G., Kee, F., 2018. Comparing mortality risk reduction, life expectancy gains, and probability of achieving full life span, as alternatives for presenting CVD mortality risk reduction: a discrete choice study of framing risk and health behaviour change. Social Science & Medicine, 211, 164-174. https://doi.org/10.1016/j.socscimed.2018.06.011

Hammitt, J.K., Graham, J.D., 1999. Willingness to pay for health protection: inadequate sensitivity to probability? Journal of Risk and Uncertainty, 18(1), 33-62. https://doi.org/10.1023/A:1007760327375

Hanley, N., Adamowicz, W., Wright, R.E., 2005. Price vector effects in choice experiments: an empirical test. Resource and Energy Economics, 27(3), 227-234. https://doi.org/10.1016/j.reseneeco.2004.11.001

Hanley, N., Czajkowski, M., 2019. The role of stated preference valuation methods in understanding choices and informing policy. Review of Environmental Economics and Policy. https://doi.org/10.1093/reep/rez005

Hensher, D.A., Greene, W.H., 2003. The mixed logit model: the state of practice. Transportation, 30, 133-176. https://doi.org/10.1023/A:1022558715350

Herrera-Araujo, D., Rheinberger, C.M., Hammitt, J.K., 2022. Valuing non-marginal changes in mortality and morbidity risk. Journal of Health Economics, 84, 102627. https://doi.org/10.1016/j.jhealeco.2022.102627

Herriges, J.A., Shogren, J.F., 1996. Starting point bias in dichotomous choice valuation with follow-up questioning. Journal of Environmental Economics and Management, 30(1), 112-131. https://doi.org/10.1006/jeem.1996.0008

Hess, S., Rose, J.M., 2009. Should reference alternatives in pivot design SC surveys be treated differently? Environmental and Resource Economics, 42(3), 297-317. https://doi.org/10.1007/s10640-008-9244-6

Hole, A.R., 2007a. A comparison of approaches to estimating confidence intervals for willingness to pay measures. Health Economics, 16(8), 827-840. https://doi.org/10.1002/hec.1197

Hole, A.R., 2007b. Fitting mixed logit models by using maximum simulated likelihood. The Stata Journal, 7(3), 388-401. https://doi.org/10.1177/1536867X0700700306

Huber, J., Zwerina, K., 1996. The importance of utility balance in efficient choice designs. Journal of Marketing Research, 33(3), 307-317. https://doi.org/10.1177/002224379603300305

Johnston, R.J., Boyle, K.J., Adamowicz, W., Bennett, J., Brouwer, R., Cameron, T.A., Hanemann, W.M., Hanley, N., Ryan, M., Scarpa, R., 2017. Contemporary guidance for stated preference studies. Journal of the Association of Environmental and Resource Economists, 4(2), 319-405. https://doi.org/10.1086/691697

Kahneman, D., Knetsch, J.L., 1992. Valuing public goods: the purchase of moral satisfaction. Journal of Environmental Economics and Management, 22(1), 57-70. https://doi.org/10.1016/0095-0696(92)90019-S

Kahneman, D., Ritov, I., Schkade, D., 1999. Economic preferences or attitude expressions?: an analysis of dollar responses to public issues. Journal of Risk and Uncertainty, 19(1), 203-235. https://doi.org/10.1023/A:1007835629236

Kartman, B., Stålhammar, N.O., Johannesson, M., 1996. Valuation of health changes with the contingent valuation method: a test of scope and question order effects. Health Economics, 5(6), 531-541. https://doi.org/10.1002/(SICI)1099-1050(199611)5:6%3C531::AID-HEC235%3E3.0.CO;2-J

Keller, E., Newman, J.E., Ortmann, A., Jorm, L.R., Chambers, G.M., 2021. How much is a human life worth? A systematic review. Value in Health, 24(10), 1531-1541. https://doi.org/10.1016/j.jval.2021.04.003

Klose, T., 1999. The contingent valuation method in health care. Health Policy, 47(2), 97-123.

Kniesner, T.J., Viscusi, W.K., 2019. The value of a statistical life. In Oxford research encyclopedia of economics and finance. https://doi.org/10.1093/acrefore/9780190625979.013.138

Kragt, M.E., 2013. The effects of changing cost vectors on choices and scale heterogeneity. Environmental and Resource Economics, 54, 201-221. https://doi.org/10.1007/s10640-012-9587-x

Krinsky, I., Robb, A.L., 1991. Three Methods for Calculating the Statistical Properties of Elasticities: A Comparison. Empirical Economics, 16(2). https://doi.org/10.1007/BF01193491

Ladenburg, J., Olsen, S.B., 2008. Gender-specific starting point bias in choice experiments: Evidence from an empirical study. Journal of Environmental Economics and Management, 56(3), 275-285. https://doi.org/10.1016/j.jeem.2008.01.004

Lancaster, K.J., 1966. A new approach to consumer theory. Journal of Political Economy, 74(2), 132-157. https://doi.org/10.1086/259131

Lancsar, E., Fiebig, D.G., Hole, A.R., 2017. Discrete choice experiments: a guide to model specification, estimation and software. Pharmacoeconomics, 35, 697-716. https://doi.org/10.1007/s40273-017-0506-4

Lancsar, E., Louviere, J., 2008. Conducting discrete choice experiments to inform healthcare decision making: a user's guide. Pharmacoeconomics, 26, 661-677. https://doi.org/10.2165/00019053-200826080-00004

Lindhjem, H., Navrud, S., Braathen, N.A., Biausque, V., 2011. Valuing mortality risk reductions from environmental, transport, and health policies: A global meta-analysis of stated preference studies. Risk Analysis: An International Journal, 31(9), 1381-1407. https://doi.org/10.1111/j.1539-6924.2011.01694.x

List, J.A., Gallet, C.A., 2001. What experimental protocol influence disparities between actual and hypothetical stated values? Environmental and Resource Economics, 20(3), 241-254.

Marceta, S.M., de Bekker-Grob, E.W., Veldwijk, J., 2025. Visualizing risk: Risk graphics' impact on patient understanding and choices in discrete choice experiments. Social Science & Medicine, 118504. https://doi.org/10.1016/j.socscimed.2025.118504

Mariel, P., Artabe, A., 2020. Interpreting correlated random parameters in choice experiments. Journal of Environmental Economics and Management, 103, 102363. https://doi.org/10.1016/j.jeem.2020.102363

Mariel, P., Hoyos, D., Meyerhoff, J., Czajkowski, M., Dekker, T., Glenk, K., Jacobsen, J.B., Liebe, U., Olsen, S.B., Sagebiel, J., 2021. Environmental valuation with discrete choice experiments: Guidance on design, implementation and data analysis. Springer Nature. https://doi.org/10.1007/978-3-030-62669-3

McFadden, D., 1973. Conditional logit analysis of qualitative choice behavior, in: Zarembka P. (Ed.), Frontiers in Econometrics (pp. 105-142). Academics Press.

Meyerhoff, J., Glenk, K., 2015. Learning how to choose—effects of instructional choice sets in discrete choice experiments. Resource and Energy Economics, 41, 122-142. https://doi.org/10.1016/j.reseneeco.2015.04.006

Mørkbak, M.R., Christensen, T., Gyrd-Hansen, D., 2010. Choke price bias in choice experiments. Environmental and Resource Economics, 45, 537-551. https://doi.org/10.1007/s10640-009-9327-z

Nguyen, T.C., Robinson, J., Kaneko, S., Nguyen, T.C., 2015. Examining ordering effects in discrete choice experiments: A case study in Vietnam. Economic Analysis and Policy, 45, 39-57. https://doi.org/10.1016/j.eap.2015.01.003

OECD, 2025. Mortality Risk Valuation in Policy Assessment: A Global Meta-Analysis of Value of Statistical Life Studies.

Onwujekwe, O., Nwagbo, D., 2002. Investigating starting-point bias: a survey of willingness to pay for insecticide-treated nets. Social Science & Medicine, 55(12), 2121-2130. https://doi.org/10.1016/S0277-9536(01)00355-0

Poe, G.L., Giraud, K.L., Loomis, J.B., 2005. Computational methods for measuring the difference of empirical distributions. American Journal of Agricultural Economics, 87(2), 353-365. https://doi.org/10.1111/j.1467-8276.2005.00727.x

Rice, T., 2013. The behavioral economics of health and health care. Annual Review of Public Health, 34(1), 431-447. https://doi.org/10.1146/annurev-publhealth-031912-114353

Robinson, L.A., Hammitt, J.K., O'Keeffe, L., 2019. Valuing mortality risk reductions in global benefit-cost analysis. Journal of Benefit-Cost Analysis, 10(S1), 15-50. https://doi.org/10.1017/bca.2018.26

Rose, J.M., Hess, S., 2009. Dual-response choices in pivoted stated choice experiments. Transportation Research Record, 2135(1), 25-33. https://doi.org/10.3141/2135-04

Ryan, M., Gerard, K., Amaya-Amaya, M., 2007. Using discrete choice experiments to value health and health care (Vol. 11). Springer Science & Business Media.

Ryan, M., Wordsworth, S., 2000. Sensitivity of willingness to pay estimates to the level of attributes in discrete choice experiments. Scottish Journal of Political Economy, 47(5), 504-524. https://doi.org/10.1111/1467-9485.00176

Sandor, Z., Wedel, M., 2001. Designing conjoint choice experiments using managers' prior beliefs. Journal of Marketing Research, 38(4), 430-444. https://doi.org/10.1509/jmkr.38.4.430.18904

Shah, K.K., Tsuchiya, A., Wailoo, A.J., 2015. Valuing health at the end of life: a stated preference discrete choice experiment. Social Science & Medicine, 124, 48-56. https://doi.org/10.1016/j.socscimed.2014.11.022

Shiell, A., Gold, L., 2003. If the price is right: vagueness and values clarification in contingent valuation. Health Economics, 12(11), 909-919. https://doi.org/10.1002/hec.777

Su, L., Adam, B.D., Lusk, J.L., Arthur, F., 2017. Anchoring, information, and fragility of choice experiments: An application to consumer willingness to pay for rice with improved storage management. Journal of Agricultural and Resource Economics, 255-274. http://dx.doi.org/10.22004/ag.econ.258001

Sun, S., Hu, W., Qing, P., 2019. Comparative price in choice experiments: the effects of price information on willingness to pay.

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

Tversky, A., Kahneman, D., 1974. Judgment under Uncertainty: Heuristics and Biases: Biases in judgments reveal some heuristics of thinking under uncertainty. Science, 185(4157), 1124-1131. https://doi.org/10.1126/science.185.4157.1124

Viscusi, W.K., Aldy, J.E., 2003. The value of a statistical life: a critical review of market estimates throughout the world. Journal of Risk and Uncertainty, 27, 5-76. https://doi.org/10.1023/A:1025598106257

Vossler, C.A., Doyon, M., Rondeau, D., 2012. Truth in consequentiality: theory and field evidence on discrete choice experiments. American Economic Journal: Microeconomics, 4(4), 145-171. https://doi.org/10.1257/mic.4.4.145

Weinstein, M.C., Toy, E.L., Sandberg, E.A., Neumann, P.J., Evans, J.S., Kuntz, K.M., Graham, J.D., Hammitt, J.K., 2001. Modeling for health care and other policy decisions: uses, roles, and validity. Value in Health, 4(5), 348-361. https://doi.org/10.1046/j.1524-4733.2001.45061.x

Whitehead, J.C., 2016. Plausible responsiveness to scope in contingent valuation. Ecological Economics, 128, 17-22. https://doi.org/10.1016/j.ecolecon.2016.03.011

Supplementary Materials

Table S1. Mixed logit model with correlated random parameters in preference-space with covariates of means – part A

	Main effects		Interactions (Covariates of mean)		
Attributes	Mean (μ)	Standard Deviation (σ)	Medium cost vector in DCE (bsl: Low cost vector in DCE)	High cost vector in DCE (bsl: Low cost vector in DCE)	No disclosure of attributes (bsl: Disclosure of attributes)
Not SQ (norm. dist.)	-0.74 (1.53)	5.38*** (0.31)	-0.40 (0.55)	0.38 (0.63)	0.73 (0.46)
Risk (norm dist.)	1.59*** (0.38)	1.36*** (0.07)	-0.09 (0.14)	-0.32** (0.14)	0.16 (0.11)
Annual Cost/100 in PLN (log-norm dist.)	-1.71*** (0.49)	2.09*** (0.10)	-0.11 (0.14)	-0.35* (0.19)	0.12 (0.12)

	Interactions (Covariates of mean) - continued				
Attributes	Only the lowest cost in 1st CS (bsl: Not facing both the lowest and the highest elements of the cost vector in the first choice task)	Only the highest cost in 1st CS (bsl: Not facing both the lowest and the highest elements of the cost vector in the first choice task)	The lowest and highest cost in the 1st CS (bsl: Not facing both the lowest and the highest elements of the cost vector in the first choice task)	Information on correct answer X Incomplete comprehension 2/3 (bsl: Answering correctly three comprehension questions when getting informed about the correct answer)	Information on correct answer X Incomplete comprehension 1/3 (bsl: Answering correctly three comprehension questions when getting informed about the correct answer)
Not SQ (norm. dist.)	0.15 (0.63)	0.78 (0.58)	0.94 (0.69)	1.73** (0.86)	1.50* (0.88)
Risk (norm dist.)	0.15 (0.15)	0.12 (0.14)	0.24 (0.17)	-0.47** (0.20)	-1.12*** (0.21)
Annual Cost/100 in PLN (log-norm dist.)	-0.09 (0.14)	0.07 (0.16)	0.13 (0.24)	-0.47*** (0.18)	-1.18*** (0.29)

	Interactions (Covariates of mean) - continued				
Attributes	Information on correct answer X Incomplete comprehension 0/3 (bsl: Answering correctly three comprehension questions when getting informed about the correct answer)	No information on correct answer X Incomplete comprehension 2/3 (bsl: Answering correctly three comprehension questions without getting informed about the correct answer)	No information on correct answer X Incomplete comprehension 1/3 (bsl: Answering correctly three comprehension questions without getting informed about the correct answer)	No information on correct answer X Incomplete comprehension 0/3 (bsl: Answering correctly three comprehension questions without getting informed about the correct answer)	Male (bsl: Female)
Not SQ (norm. dist.)	0.03 (0.83)	-0.35 (0.80)	1.05 (0.92)	0.81 (0.83)	-0.62 (0.49)
Risk (norm dist.)	-1.48*** (0.20)	-0.78*** (0.20)	-1.00*** (0.21)	-1.33*** (0.20)	-0.19 (0.12)
Annual Cost/100 in PLN (log-norm dist.)	-1.96*** (0.25)	-0.52** (0.25)	-2.05*** (0.24)	-1.91*** (0.26)	-0.02 (0.13)

Table S1. Mixed logit model with correlated random parameters in preference-space with covariates of means – part B

Interactions (Covariates of mean) - continued					
Attributes	Age 25-44 (bsl: age 18-24)	Age 45-64 (bsl: age 18-24)	Age 65+ (bsl: age 18-24)	Lower Secondary Education or lower (bsl: Basic Vocational Education or Secondary and Post-secondary Non-tertiary Education)	Higher education (bsl: Basic Vocational Education or Secondary and Post-secondary Non-tertiary Education)
Not SQ (norm. dist.)	0.04 (0.79)	-0.64 (0.87)	-2.57** (1.00)	0.15 (1.61)	-0.02 (0.53)
Risk (norm dist.)	-0.08 (0.19)	-0.14 (0.21)	0.08 (0.25)	0.01 (0.37)	0.24* (0.12)
Annual Cost/100 in PLN (log-norm dist.)	0.38* (0.22)	0.47* (0.25)	-0.39* (0.22)	0.75** (0.35)	0.17 (0.17)

Interactions (Covariates of mean) - continued					
Attributes	Countryside (bsl: living in the city up to 500k residents)	Big city 500k+ (bsl: living in the city up to 500k residents)	House hold size==1 (bsl: house hold size >=3)	HH==2 (bsl: house hold size >=3)	EQ5D5L HUS <0.8.0.9) (bsl: health utility score >=0.9 perfect health)
Not SQ (norm. dist.)	-0.63 (0.55)	-0.65 (0.63)	1.42* (0.80)	0.99* (0.55)	0.31 (0.76)
Risk (norm dist.)	-0.03 (0.13)	0.35** (0.16)	-0.19 (0.19)	0.11 (0.13)	0.12 (0.18)
Annual Cost/100 in PLN (log-norm dist.)	-0.14 (0.16)	0.40*** (0.12)	0.43** (0.17)	-0.02 (0.13)	0.44*** (0.16)

Interactions (Covariates of mean) - continued					
Attributes	EQ5D5L <0.8 (bsl: health utility score >=0.9 perfect health)	No doctor appointments (bsl: at least once a year, but no more frequent than once every two weeks)	Frequent doctor appointments (bsl: at least once a year, but no more frequent than once every two weeks)	No alcohol consumption (bsl: at least once a year, but no more frequent than once every two weeks)	Frequent alcohol consumption (bsl: at least once a year, but no more frequent than once every two weeks)
Not SQ (norm. dist.)	-1.00 (0.79)	-0.82 (1.35)	3.12 (1.96)	0.20 (0.63)	0.76 (0.59)
Risk (norm dist.)	-0.33* (0.20)	-0.04 (0.32)	0.18 (0.44)	-0.13 (0.16)	0.13 (0.15)
Annual Cost/100 in PLN (log-norm dist.)	-0.48* (0.26)	0.70* (0.43)	-0.03 (0.38)	0.53*** (0.18)	0.74*** (0.16)

Interactions (Covariates of mean) - continued					
Attributes	No smoking (bsl: at least once a year, but no more frequent than once every two weeks)	Frequent smoking (bsl: at least once a year, but no more frequent than once every two weeks)	No drugs (bsl: at least once a year, but no more frequent than once every two weeks)	Frequent drug use (bsl: at least once a year, but no more frequent than once every two weeks)	0 min of activity (bsl: 1-300 min of physical activity per week)
Not SQ (norm. dist.)	-0.95 (0.89)	0.19 (0.93)	2.77** (1.15)	2.88 (2.33)	-0.94 (0.84)
Risk (norm dist.)	0.18 (0.22)	0.28 (0.23)	-0.03 (0.29)	0.18 (0.55)	0.29 (0.21)
Annual Cost/100 in PLN (log-norm dist.)	0.14 (0.23)	0.24 (0.21)	0.87*** (0.32)	-0.89* (0.50)	0.81*** (0.24)

Table S1. Mixed logit model with correlated random parameters in preference-space with covariates of means – part C

Interactions (Covariates of mean) - continued					
Attributes	300 min+ of activity (bsl: 1-300 min of physical activity per week)	BMI (25.35> (bsl: BMI under 25; underweight)	BMI 35+ (bsl: BMI under 25; underweight)	Chronic diseases (bsl: not having any chronic diseases)	Cardiovascular diseases (bsl: not having any cardiovascular diseases)
Not SQ (norm. dist.)	-0.95 (0.72)	0.58 (0.65)	0.73 (0.80)	0.73 (0.50)	-0.02 (0.63)
Risk (norm dist.)	0.27 (0.20)	-0.16 (0.16)	-0.04 (0.18)	-0.02 (0.12)	0.09 (0.15)
Annual Cost/100 in PLN (log-norm dist.)	0.39** (0.17)	-0.08 (0.13)	0.27 (0.16)	0.00 (0.16)	-0.34** (0.15)

Interactions (Covariates of mean) - continued					
Attributes	Cardiovascular diseases in family (bsl: not having any cardiovascular diseases in family)	Risk perception – lower (bsl: stating that risk perception is the same as average in population)	Risk perception – higher (bsl: stating that risk perception is the same as average in population)	Risk perception - I don't know (bsl: stating that risk perception is the same as average in population)	Income >10k PLN/month (bsl: income <10k PLN/month)
Not SQ (norm. dist.)	1.71*** (0.49)	0.55 (0.56)	-1.46* (0.86)	-0.54 (0.74)	-1.23** (0.60)
Risk (norm dist.)	0.17 (0.12)	-0.07 (0.13)	0.35 (0.23)	-0.59*** (0.17)	0.10 (0.15)
Annual Cost/100 in PLN (log-norm dist.)	0.36*** (0.12)	0.49*** (0.12)	0.04 (0.22)	0.59** (0.24)	-0.59*** (0.13)

Note: *** for p-value < 0.01, ** for p-value < 0.05, * for p-value < 0.1. The EQ-5D-5L utility tariffs were calculated using the value set derived by Golicki et al. (2019). bsl – baseline level

Golicki, D., Jakubczyk, M., Graczyk, K., Niewada, M., 2019. Valuation of EQ-5D-5L health states in Poland: the first EQ-VT-based study in Central and Eastern Europe. Pharmacoeconomics, 37, 1165-1176. https://doi.org/10.1007/s40273-019-00811-7

Table S2. The results (p-values) from the Two-Proportion Z-test comparing the proportions "yes-yes," "yes-no," "no-yes," and "no-no" responses from the WTP DBDC question

Two-proportions z-test (p-values)		Low Cost DCE | Medium Cost DCE	Low Cost DCE | High Cost DCE	Medium Cost DCE | High Cost DCE
DBDC X=200	Yes (200 PLN) - Yes (400 PLN)	0.64	0.07	0.15
	Yes (200 PLN) - No (400 PLN)	0.13	0.75	0.07
	No (200 PLN) - Yes (100 PLN)	0.16	0.05	0.53
	No (200 PLN) - No (100 PLN)	0.41	0.78	0.59
DBDC X=400	Yes (400 PLN) - Yes (800 PLN)	0.63	0.73	0.38
	Yes (400 PLN) - No (800 PLN)	0.54	0.52	0.19
	No (400 PLN) - Yes (200 PLN)	0.48	0.64	0.78
	No (400 PLN) - No (200 PLN)	0.68	0.60	0.92
DBDC X=600	Yes (600 PLN) - Yes (1200 PLN)	0.49	0.12	0.36
	Yes (600 PLN) - No (1200 PLN)	0.22	0.50	0.54
	No (600 PLN) - Yes (300 PLN)	0.33	0.22	0.77
	No (600 PLN) - No (300 PLN)	0.26	0.19	0.82

Note: High p-values indicate that we cannot reject the null hypothesis that proportions are equal across the compared formats

Table S3. The results (p-values) from the Permutation test, Two-Sample T-test, and Mann-Whitney U test from the WTP OE question

Differences between WTP OE across treatments	Permutation test	Two-sample t-test	Mann-Whitney U Test (Wilcoxon Rank-Sum Test)
X = 200 | X = 400	0.9465	0.1210	0.2438
X = 200 | X = 600	0.6798	0.1475	0.4040
X = 400 | X = 600	0.3792	0.7371	0.9108
Low cost DCE | Medium cost DCE	1.0000	0.9282	0.9731
Low cost DCE | High cost DCE	0.5490	0.2963	0.2166
Medium cost DCE | High cost DCE	0.8891	0.3545	0.2023

Note: High p-values indicate that we cannot reject the null hypothesis that WTP responses are equal across the compared formats.

University of Warsaw
Faculty of Economic Sciences
44/50 Długa St.
00-241 Warsaw
www.wne.uw.edu.pl
ISSN 2957-0506