Received: 25 April 2023 | Revised: 4 July 2025 | Accepted: 19 September 2025

DOI: 10.1111/jori.70026

ORIGINAL ARTICLE

Journal of Risk and Insurance

Heterogeneity in health insurance choice: An experimental investigation of consumer choice and feature preferences

Benedicta Hermanns[1,2] | **Nadja Kairies-Schwarz**[3,4] | **Johanna Kokot**[2] | **Markus Vomhof**[3,4]

[1]Institute for Digital Economics, Hamburg University of Technology, Hamburg, Germany

[2]Department of Socioeconomics, Hamburg Center for Health Economics, University of Hamburg, Hamburg, Germany

[3]Institute for Health Services Research and Health Economics, Centre for Health and Society, Medical Faculty and University Hospital Düsseldorf, Heinrich-Heine-University Düsseldorf, Düsseldorf, Germany

[4]Institute for Health Services Research and Health Economics, German Diabetes Center, Leibniz Center for Diabetes Research, Heinrich-Heine-University Düsseldorf, Düsseldorf, Germany

Correspondence

Johanna Kokot, University of Hamburg, Hamburg, Germany.
Email: johanna.kokot@uni-hamburg.de

Funding information

Deutsche Forschungsgemeinschaft, Grant/Award Number: FOR-655; Bundesministerium für Bildung und Forschung

Abstract

We investigate heterogeneity in health insurance choice using data from a controlled laboratory experiment. Participants make consecutive choices from sets of insurance plans that vary in premium, deductible, and complementary coverage of illnesses. We find that there is considerable heterogeneity in how much individuals are willing to pay for certain plan attributes. To better understand these differences, we account for individual risk preferences using a rank-dependent expected utility (RDEU) model and assess the welfare effects of plan choices. At the aggregate level, we find welfare losses under both the normative RDEU model and the descriptive EV model. At the individual level, however, the results are more differentiated: for some individuals, choices are consistent with their RDEU preferences, whereas for others, choices do not fit either model, suggesting either decision errors or reliance on heuristics.

KEYWORDS

consumer preferences, health insurance, heterogeneity, laboratory experiment, risk preferences

JEL CLASSIFICATION

C91, I13, D81, D83, G22

1 | INTRODUCTION

Based on the premise that consumers can best express their needs and preferences through their own decisions, many of the recent policy reforms in the United States (US) and in Europe have aimed to expand consumer options (Coughlin et al., 2008; Cronqvist & Thaler, 2004; Thomson et al., 2013). Investigating preferences for health insurance plans based on revealed choices relies on the assumption that the available plan sets can be observed (McFadden, 1974). In the United States, consumers predominantly select health insurance through their employers or centralized platforms, such as Medicare Part D or health insurance exchanges, where the available plans (i.e., choice sets) are known.[1] A large number of studies have investigated the rationality of insurance selection in such settings, and their findings suggest that consumers often appear to make suboptimal decisions (see, e.g., Abaluck & Gruber, 2011, 2016; Bhargava, Loewenstein & Sydnor, 2017; Handel, 2013; Heiss et al., 2013; Liu & Sydnor, 2022; McWilliams et al., 2011).

In this context, the heterogeneity of consumers' choices remains poorly understood, particularly regarding how their preferences for certain features of health insurance plans relate to their individual risk preferences and the resulting welfare effects. The existing evidence explains, to the best of our knowledge, only partial aspects of this relationship. Investigating individuals' preferences for specific plan features by examining how they acquire information, Schram and Sonnemans (2011) found that, regardless of the choice set, individuals were most interested in the insurance premium, followed by the deductible and complementary insurance. Building on this, Kairies-Schwarz et al. (2017) introduced a measure of decision quality based on experimentally elicited individual risk preferences to examine how consumers choose health insurance plans. However, they focused on decision quality at the aggregate level using the individual rank order of plans. Similarly, Jaspersen et al. (2022) investigated the relationship between insurance plan choices and individual risk preferences, also at the aggregate level. They found that while certain risk preference motives—utility curvature, probability weighting, and loss aversion—correlated with insurance choices, they could only explain a small part of the variation in insurance behavior. This suggests that there might be substantial heterogeneity in risk preference motives that the studies were not designed to capture. Rather than analyzing how well structural models predict insurance choices in aggregate, we propose that further individual-level analysis is needed (a) to better understand the heterogeneity of insurance choices and (b), following Harrison and Ng (2016), to provide a normative welfare assessment based on individually elicited risk preferences.

The aim of this study is therefore to investigate heterogeneity in health insurance choice by focusing on consumer preferences for certain features of health insurance plans and using the normative risk preference model of rank-dependent expected utility (RDEU) to assess the welfare implications of consumer choices. This model includes expected utility theory (EUT) as a special case but also captures important behavioral features observed in empirical research, such as nonlinear probability weighting in decision-making, which reflects individuals' different responses to probabilities and outcomes. This approach provides a more detailed understanding of heterogeneity in complex insurance decisions than is possible with descriptive expected value (EV) models or normative EUT models. For example, decisions that may appear suboptimal under a descriptive EV model can, when viewed through the RDEU model, be understood to have different welfare implications—differences that simpler models do not capture.

[1]In contrast, in many countries, such as Germany, health insurance choice is still decentralized, and it is difficult to infer how consumers shop for it and the choice sets upon which they base their final decisions.

We use a framed laboratory experiment with a sequential design.[2] The first part is inspired by Schram and Sonnemans (2011) and Kairies-Schwarz et al. (2017). In each round of the experiment, participants can acquire one or more of six different illnesses, each of which has a predetermined probability that is communicated to the participants. They must choose among health insurance plans in 12 varying choice sets, consisting of either six or 12 designed plans. The plans differ in premium, deductible (for basic coverage of three illnesses), and complementary coverage of three additional illnesses. For roughly half of the participants, we also introduce an ex-ante feature-based filter.[3] The filter allows participants to indicate which plan features they prefer and then highlights plans containing these features while preserving the full set of choices. This approach enables an analysis of how consumers arrive at their final choice of plan and whether the use of the filter correlates with the selected plan features. Additionally, it allows us to assess the robustness of these preferences.[4] In the second part of the experiment—similar to Harrison and Ng (2016), Kairies-Schwarz et al. (2017), Jaspersen et al. (2022), and Harrison, Kairies-Schwarz, et al. (2024)—we separately elicit individual risk preferences according to the RDEU model, thus addressing the fact that individuals differ with respect to the value and weighting functions of risk preferences (Payne et al., 1993). These exogenously elicited individual risk preferences can then be used as a prior for behavioral welfare analysis (Harrison, 2024).

The laboratory setting offers several advantages that facilitate the identification of individual preferences for plan features. First, we take advantage of a varying choice set, allowing options to change in complexity and composition across rounds. By systematically varying the cost and features of the plans, we can infer a participant's valuation of these features from the plans they select,[5] providing evidence on willingness to pay and preferred coverage. Second, the laboratory setting allows better control over choice motives, which can be difficult to disentangle when using revealed choices from field data. For instance, consumers may select a plan because they expect a claim will be more likely or because they are risk-averse, either of which may be reflected in preferences for features such as low deductibles (Ericson et al., 2021). In our stylized setting, all participants face identical probabilities of illness, which remain constant across rounds and are known to participants, allowing them to calculate the expected value of each plan. This design allows us to separate the effects of risk perceptions from preferences for a specific plan or certain plan features. Moreover, by separately eliciting each individual's risk preferences, we can characterize participants based on these behavioral priors,

[2]In doing so, we add to the literature on health economic experiments (Galizzi & Wiesen, 2017, 2018), particularly those involving health insurance (e.g., Biener & Zou, 2024; Harrison et al., 2023; Kairies-Schwarz et al., 2017, 2023; Krieger & Felder, 2013; Mimra et al., 2020; Samek & Sydnor, 2025; Schram & Sonnemans, 2011).

[3]As a treatment, our feature-based filter also relates to studies that explore the effects of decision support, such as those that show graphs of the financial consequences of different health insurance plans to consumers (see, e.g., Biener & Zou, 2024; Samek & Sydnor, 2025) or those that give a curated selection of plans (Gruber et al., 2021). In contrast to these studies, however, the purpose of our filter is not primarily to support decision-making (e.g., by reducing the size of the choice set) but to derive information about feature preferences independently of the plans as a whole.

[4]Abaluck and Adams-Prassl (2021) propose a new approach using a discrete choice model to test choice data for full information and show that it identifies features that are not immediately visible to consumers in search results. Although such approaches substantially improve our ability to infer preferences from observed data, as well as to account for missing choice sets and missing information on choice features, information about the search strategies themselves is missing.

[5]We focus on repeated independent insurance plan choices rather than on switching between plans. This allows us to map a wider range of plans and use the resulting information to determine preferences for plan features.

avoiding the need to infer risk preferences from observed insurance choices (Bhargava, Loewenstein & Sydnor, 2017).[6] Third, when working with field data, it can be difficult to control for external factors that can affect decisions. Choices may be influenced by inertia (e.g., Handel, 2013; Heiss et al., 2021), complexity or size of the choice set (Abaluck & Gruber, 2023; Besedeš et al., 2012a, 2012b; Biener & Zou, 2024; Iyengar & Kamenica, 2010; Schram & Sonnemans, 2011; Sinaiko & Hirth, 2011), a lack of understanding of the decision situation (Bhargava, Loewenstein & Benartzi, 2017), or information on the financial consequences of various plans (Samek & Sydnor, 2025). Temporary liquidity constraints are also difficult to control for, as data on individual financial circumstances are often lacking. Finally, there is evidence that consumers rely on simplified decision rules when making complex decisions (e.g., Ericson & Starc, 2012; Kamenica, 2008; Kairies-Schwarz et al., 2017). By using experimental data, we can control the decision environment—for example, by asking ex-ante comprehension questions and providing all participants with the same financial endowments, which are sufficient to purchase any of the available health insurance plans.

Our results show no significant differences in the plan features selected by participants between treatments with and without the feature-based filter. We therefore pool the data from both treatments. On average, individuals choose plans with significantly higher premiums and significantly more complementary coverage of illnesses than the plan with the best EV. However, we observe considerable heterogeneity in plan choices. To further examine this heterogeneity, we account for individual risk preferences. Following Harrison and Ng (2016) and Harrison, Kairies-Schwarz, et al. (2024), we evaluate plan choices made overall and by class using a welfare measure based on individual risk preferences under an RDEU model. At the aggregate level, both the normative RDEU model and the descriptive EV model indicate welfare losses, although they are smaller than those under random choice. At the individual level, the picture is more differentiated. Some individuals incur welfare losses according to the EV model, but little or almost no loss under the RDEU model, suggesting that their choices may be wrongly assessed if viewed only from an EV perspective. Other individuals incur low welfare losses under both models, suggesting a good fit with EV. There are also individuals whose choices appear favorable under the EV model but result in higher welfare losses under the RDEU model; for these individuals, choices that might appear satisfactory from an EV perspective may instead reflect decision errors or heuristics, such as choosing the plan with the lowest premium. Finally, a substantial proportion of individuals show welfare losses under both models, suggesting either decision errors or non-identifiable individual decision strategies.

To further link individual preferences for plan attributes with risk preferences, we use a latent class model that identifies five classes of participants based on preferences for plan attributes, and we analyze welfare losses at the class level. The results show that for some individuals, WTP for specific plan features is consistent with their elicited RDEU risk preferences, while for others it is not. This discrepancy may be the result of decision errors or a reliance on heuristics.

The remainder of this paper is structured as follows: Section 2 introduces our experimental design. Section 3 presents the results. Section 4 provides robustness checks. Section 5 discusses the results, as well as the limitations of our approach, and presents our conclusions.

[6]There are also approaches that aim to identify risk preferences from observed choices; see for example, Barseghyan et al. (2021) in automobile collision insurance and Ericson et al. (2021) in health insurance. The latter authors present a method that identifies heterogeneity in risk perceptions and preferences using health insurance choice data from the Massachusetts Health Insurance Exchange. However, it still considers risk preferences based on the expected utility theory, assuming risk type distributions.

TABLE 1 Features of health insurance plans.

Insurance type	Basic			Complementary		
Deductible (in talers)	0, 10, 20, 30					
Illness	A	B	C	D	E	F
Treatment costs without coverage (in talers)	60	40	20	2000	70	40
Probability of illness occurrence	5%	20%	50%	1%	10%	30%

2 | EXPERIMENTAL DESIGN

Our experiment followed a sequential design consisting of two parts. In the first part, participants were presented with decision scenarios that required them to choose a health insurance plan from a menu of such plans.[7] In the second part, they were presented with lottery choices to elicit their individual risk preferences. The instructions for participants are given in Appendix A.

2.1 | Part 1—Health insurance choice sets

Participants were provided with an initial endowment of 2300 talers, with 100 units of this lab currency equaling 0.50 euros. From this endowment, they could purchase health insurance to insure themselves against six possible illnesses (A, B, C, D, E, and F), each of which differed in its probability of occurrence and costs in the event of a claim. The costs in talers and the probabilities of occurrence are shown in Table 1. Other monetary and nonmonetary costs that accompany an illness, such as lost wages or pain, are not considered here. The treatment costs and probabilities of each illness remained unchanged throughout the experiment.

The decision situation was as follows: Participants had to choose from menus of stylized health insurance plans in 12 independent decision rounds.[8] The decision framework was similar to that in Schram and Sonnemans (2011) and Kairies-Schwarz et al. (2017).[9] In each round, participants were presented with a different menu of choices. The first two choice sets were menus with six plans. In the consecutive rounds, the menus alternated in their degree of complexity (six or 12 plans). This was intended to jointly test the influence of menu size and feature complexity on the choice of plan. We also did this so as not to provoke a status quo bias. At the end of the experiment, one decision

[7]We opted for a health framing instead of neutral framing, as many studies have done before us. For an overview, see Galizzi and Wiesen (2017). In a setting similar to ours, Kairies-Schwarz et al. (2017) observed higher decision quality in health framing.

[8]We acknowledge that this design differs from the reality of knowing whether one has actually become ill, thus potentially reducing external validity. However, we expect that there are individual differences in the way participants react to an illness that they know they have acquired. In such cases, it would not be possible to distinguish between someone who chooses a plan because they prefer it and someone who chooses a plan because they suffered a loss (occurrence of illness) in the previous decision round. Therefore, we did not provide feedback after each decision. This means that our participants' health insurance choices are independent insofar as they are made by participants given the same health status.

[9]Compared to Kairies-Schwarz et al. (2017), we increased the complexity of the decision scenario by adding an illness and offering more decision options. In addition, to allow for heterogeneity in the decisions, we adjusted the sets of plans per decision in such a way that different preferences should lead to different choices of plans. Regarding the possible preferences, we followed Kairies-Schwarz et al. (2017).

round was randomly determined to be relevant for payment. For this round, health status (i.e., whether a participant had one or more illnesses) was determined.

2.1.1 | Plan design

The menu of health insurance plans took the form of a table displaying plans that varied along three dimensions: premium, deductible for basic insurance, and complementary insurance. Illnesses A, B, and C were covered by basic insurance, and D, E, and F could be covered by complementary insurance. Deductibles for illnesses covered by basic insurance started at 0 talers and went up to 30 talers in increments of 10. In case these illnesses or a combination of them occurred, the participant had to bear the treatment costs incurred up to the amount of the deductible, with health insurance paying the amount that exceeded this. In total, out of the initial endowment, a participant had to pay the premium, the potential treatment costs up to the amount of the deductible under basic insurance, and the potential costs of illnesses D, E, and F if not covered by complementary insurance. A sample screenshot of the decision scenario is provided in the participant instructions in Appendix A.

Without considering the premium, there were 32 ways to compose plans differently with our six illnesses. A fair premium for illnesses D, E, and F and the deductibles was then calculated for each of these 32 variants. A random component was added to the premium to induce a ranking of plans to ensure that plans had different levels of attractiveness. The resulting order of plans was basically determined randomly.[10] Ultimately, the 32 plans differed in rank-ordering according to the expected value (EV).[11] We based our parameters on objective EV because it represents (a) the risk-neutral variant of expected utility theory and (b) a special case of the rank-dependent expected utility theory (RDEU) when the value function is linear and there is no probability weighting. Table B1 in the Appendix provides an overview of the plans. All plans that were offered in a decision round were selected from these 32 plans and alternated among the decision rounds to prevent participants from remembering plans from the previous decision. The order of the plans was drawn at random in advance and was identical for all participants. Table B2 in the Appendix provides an overview of all plans in all choice sets in the order displayed.

2.1.2 | Experimental conditions

To investigate initial preferences for plan features, we implemented two treatments: a baseline condition with insurance choice (IC) and a treatment variation with a feature-based filter (IC filter). The use of the filter was voluntary. This allows us to observe different types of users and their search behavior. With the filter, participants had the opportunity to indicate the characteristics of their preferred plan (i.e., their maximum deductible and the illnesses they desired to have covered by complementary insurance) in each decision round. Plans that

[10]However, the exception to this was that the 10 best plans (by EV rank: 1–10) did not cover the 2000 taler treatment cost feature, whereas the 10 worst plans (by EV rank: 23–32) did. This rule served to ensure that players had an actual trade-off in terms of rank-dependent expected utility (RDEU). For example, if the best EV plan covered the 2000-taler cost feature, this would be a dominant plan that would be easy for participants to identify.

[11]Note that we used EV as an objective measure to determine the rank-ordering of plans because we did not have individual risk preference measures before the experiment. This also most likely resembles a realistic health insurance choice scenario in which the underlying choice set is the same for all individuals.

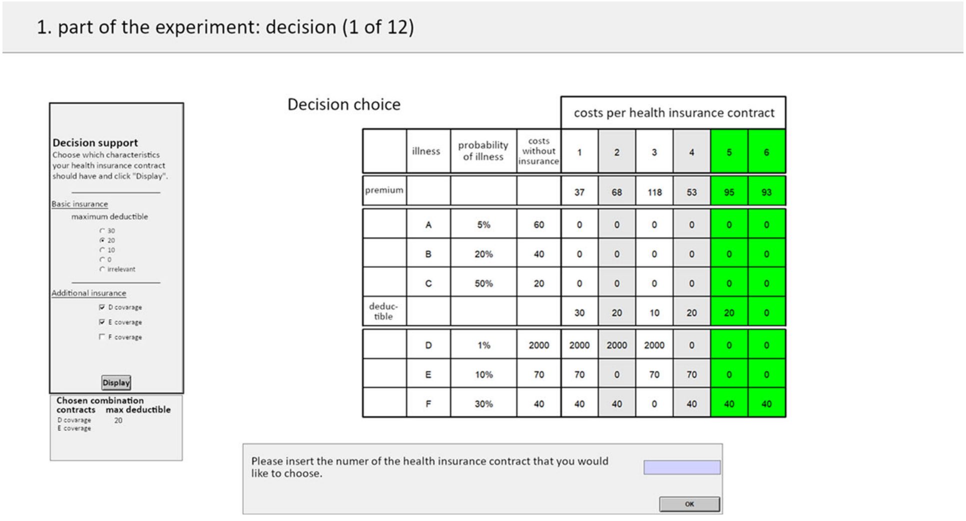

FIGURE 1 Screenshot of choice set with a feature-based filter. [Color figure can be viewed at wileyonlinelibrary.com]

matched the filter selection were highlighted in green, and the other plans remained as a choice option.[12] Participants could change their feature selection as often as they wished during a decision round, with other plans being highlighted accordingly. There was also a possibility that no plan in the selection fulfilled all the conditions from the filter, resulting in no plan being displayed in green. Figure 1 provides a screenshot of the decision situation with the feature-based filter. The filter treatment allows us to fill an information gap by enabling the identification of preferences for plan features that cannot be clearly inferred from revealed choices in field data. We can then infer the importance of the features based on the observed choices.

2.2 | Part 2—Risk elicitation and definition of welfare

In Part 2 of the experiment, we measure individual risk preferences based on the concept of rank-dependent expected utility (RDEU) introduced by Quiggin (1982). RDEU has become one of the most relevant empirical generalizations of expected utility theory (EUT). Although the utility function in RDEU is specified similarly to EUT and includes EUT as a special case, RDEU introduces an important modification: nonlinear probability weighting. This accounts for the possibility that individuals do not necessarily evaluate risky decisions with given probabilities based solely on their objective values. Instead, it is assumed that individuals first rank outcomes according to their preferences and then assign subjective probability weights to them. The RDEU model derives the probability weights from the entire distribution of ranked outcomes, reflecting the subjective distortions of objective probabilities. The RDEU model thus

[12]We are aware that this scenario does not correspond to many real-world decision-support tools for choosing health insurance, which reduce the number of alternatives within the choice set. However, by keeping all plans within the menu, we can investigate whether participants ultimately choose a plan that corresponds to their indicated feature preferences. We are also able to compare insurance choices to the treatment with no filter.

requires introducing a probability weighting function. Despite its flexible approach, the RDEU model allows for welfare analysis, as exemplified in the work of Harrison and Ng (2016).

The RDEU model accommodates behavioral observations such as probability weighting and decreasing sensitivity to changes in probability—that is, individuals tend to respond less to variations in probabilities when these probabilities are moderate compared to when they are extreme. In particular, individuals often overestimate the probability of events that are unlikely while underestimating the probability of events that are more common. These characteristics are particularly relevant in insurance contexts because the consequences of the insurance choice involve risk and uncertainty, and plans often contain features directly related to risk and uncertainty, such as deductibles or additional coverage against events that are nontrivial, rare, or extreme. As a result, the RDEU model provides a flexible framework for analyzing heterogeneity in complex insurance decisions. In particular, it can help clarify why decisions that appear, for example, suboptimal under a descriptive EV model may be optimal under a normative RDEU model. Furthermore, the results of Jaspersen et al. (2022) suggest that while there are modest correlations between insurance levels and factors such as utility curvature, loss aversion, and probability weighting, these elements can predict only a small part of the variations in insurance behavior. This suggests that there might be substantial heterogeneity in preference motives, which the RDEU model is well equipped to capture.[13]

In the experiment, we follow an elicitation procedure of RDEU risk preferences that is similar to the one used by Kairies-Schwarz et al. (2017). Because the consequences of insurance plan choices in the experiment are framed as losses, we consider RDEU here as being defined only over losses. Considering a complete ranking of all (negative) outcomes of a prospect f, that is, $0 \geq x_{\mathrm{j}} \geq \ldots \geq x_n$, and associated probabilities $p_1, \ldots, p_n$ the value of the prospect is calculated as:

$$RDEU(f) = \sum_{j=1}^{n} \pi_j(p) \cdot U(x_j). \tag{1}$$

The utility function $U(x_j)$ is evaluated by a probability weighting function $\pi(x_j)$. The weighting function is strictly increasing in probabilities between [0, 1], and $w(0) = 0$ and $w(1) = 1$ must hold. Equation 2 shows how we calculate decision weights while accounting for rank dependence[14]:

$$\pi_j(p) = w(p_j + \ldots + p_n) - \mathrm{w}(p_{j+1} + \ldots + p_n). \tag{2}$$

In the experiment, we applied the trade-off method (Wakker & Deneffe, 1996) and implemented a bisection procedure, which is similar to the approach taken by Abdellaoui (2000) and Abdellaoui et al. (2007), to facilitate the decision process. As previously mentioned, we focused only on the loss domain relevant to the choices of insurance in Part 1 of the experiment. We therefore did not need to elicit the degree of loss aversion in this setting. We scaled the lotteries to fit the decisions made in Part 1 and to ensure that scaling effects would not bias our results.

[13]The findings of Harrison and Swarthout (2023) moreover support the notion of using RDEU, e.g. instead of cumulative prospect theory. Having developed a battery test that considers a wide range of parameters in common utility models such as EUT, RDEU, Cumulative Prospect Theory, or Dual Theory, they show that the RDEU model best characterizes individual and pooled decisions.

[14]Note that the weight is defined as $\pi_1(p) = w(1) - \mathrm{w}(p_{j+1} + \ldots + p_n)$ and $\pi_n(p) = w(p_n) - \mathrm{w}(0)$ for the largest and smallest outcome, respectively.

Participants had to choose between 70 pairs of lotteries, providing a large and detailed dataset for analysis. In 38 decision rounds, participants had the choice between two lotteries to calculate the value function parameters. In the remaining 32 decision rounds, they had the choice between a safe option and a lottery, allowing us to calculate the weighting function parameters. See Appendix D for a detailed example of the decision situation. Participants received an initial endowment of 4800 talers in Part 2. One of the participants' decisions was selected randomly, and the outcome of the decision was determined and subtracted from the initial endowment.

2.3 | Experimental procedure

The experiment was programmed with z-Tree (Fischbacher, 2007) and conducted through the Essen Laboratory for Experimental Economics (elfe) in 2016. It involved 253 individuals recruited from the general student population of the University of Duisburg-Essen using the Online Recruitment System for Economic Experiments (Greiner, 2015). In total, 117 individuals participated in the IC filter and 136 individuals participated in the IC treatment. Overall, we conducted seven sessions at the laboratory, each lasting approximately 135 min, to collect data.

The experimental procedure for all sessions was as follows: Upon arrival, participants were randomly assigned to individual cubicles and given written instructions. After reading these, participants had the opportunity to ask questions. To prepare participants for the complexity of the insurance plans, we asked them eight comprehension questions focusing on the premium, the probabilities of an illness occurring, the total cost, the individual medical costs, and the amount of the deductible (see Appendix A for an overview of the specific questions). The experiment did not start until all participants had answered the comprehension questions correctly. Participants were provided with a simple calculator and pen and paper to assist during the experiment. At the end of the experiment, participants were asked to complete a short questionnaire on demographics, including age, gender, and their university major. They were also asked whether they had paid special attention during the experiment to any particular features of the plans.[15] Our participant pool was 52% female with an average age of 23.8 years (std. dev. 0.39). In total, 44.3% were enrolled in economics, 18.2% in a STEM subject, and 17.0% in humanities. Participants earned an average of 25.42 euros and received their payment privately in cash at the end of the experimental session.

3 | RESULTS

3.1 | Descriptive analysis

We first examine whether there are differences in the choice of plans between the treatment with a feature-based filter (IC filter) and the treatment without a filter (IC). Table 2 shows the mean premium, mean deductible, and proportion of plans with complementary insurance chosen by participants in the IC group, the IC filter group, and both groups combined. The lack of

[15]Details in Table G1 in Appendix G.

TABLE 2 Mean features of actual choices and best EV plans.

	Actual choices					
	IC filter	IC	*p*-value (MWU)	Overall	Best EV plans	*p*-value (WSR)
Premium	83.375 (2.105)	79.165 (1.962)	0.145	81.112 (1.439)	42.917 (3.227)	<0.001
Deductible size	10.150 (0.336)	11.072 (0.356)	0.179	10.646 (0.248)	10.833 (3.362)	0.173
Coverage of illness						
D (2000 talers with 1% probability of occurrence)	57.83% (0.033)	55.27% (0.030)	0.525	56.46% (0.022)	33.33% (0.142)	<0.001
E (70 talers with 10% probability of occurrence)	50.64% (0.017)	48.77% (0.017)	0.380	49.64% (0.012)	16.67% (0.112)	<0.001
F (40 talers with 30% probability of occurrence)	67.31% (0.014)	63.24% (0.014)	0.079	65.12% (0.010)	41.67% (0.149)	<0.001
N	117	136		253		

Note: For the actual choices, we calculate, for each participant, the mean value of each feature across all 12 plan decisions, and then calculate the mean of these values, for each feature, across all participants. For the best EV plans, we calculate the mean value of each feature across all 12 best-value plans, one for each of the 12 choice sets. The percentages shown for illnesses D, E, and F represent, in the left half of the table, the percentage of actual choices in which coverage for each illness was selected and, in the right half of the table, the percentage of best EV plans in which coverage for each illness would be included. Standard errors are in parentheses. We use the Mann-Whitney U test (MWU) for between-treatment comparisons and a one-sided Wilcoxon signed-rank test (WSR) for comparisons between overall actual choices and best EV plans.

significant differences in the features chosen by participants in the two treatments (see Table 2) is consistent with findings by Samek and Sydnor (2025) and Biener and Zou (2024), who report that decision aids, such as consequence graphs, that do not restrict choice sets have little or no effect on plan choices. For all further analyses, we therefore pool the results from both treatments.

Because the plan features do not stand alone but are always part of a stylized plan, the average values and proportions of the plan features can only be interpreted relative to one another. To provide context, we compare the observed values to those that would result from selecting the best plan based on EV in each decision.[16] All values are reported in the experimental currency, talers. The mean premium of the plans chosen in our experiment is 81.11 talers, which is nearly twice that of the best average EV plan. In contrast, the mean deductible of the plans chosen by our participants is 10.65 talers, which is not significantly different from that of the best EV plan ($p = 0.173$). Complementary insurance for illnesses D, E, and F is chosen relatively frequently: illness D is covered in 56.46% of selected plans, illness E in 49.64%, and illness F in 65.12%. All of these percentages are significantly higher than those observed for the best EV plans.

Across all decision rounds, the expected costs range from an average minimum of 75.17 talers to an average maximum of 157.22 talers. The plans chosen by participants result in an average expected cost of 103.52 talers. Compared to the best EV plan in each round, this corresponds to an average increase in expected costs of 37.7%, indicating considerable efficiency losses under a descriptive EV model.

Result 1. *Overall, individuals choose plans with significantly higher premiums and significantly more complementary coverage of illnesses than would be the case if they had selected the plans with the best EV. As a result, participants face an average expected cost increase of 37.7%, indicating considerable efficiency losses under a descriptive EV model.*

3.2 | Heterogeneity analysis

To examine insurance choices in more detail and assess heterogeneity at the individual level,[17] we employ a mixed logit regression. This model considers all available plan options (both selected and unselected) in each decision round, and accounts for individual differences in preferences for plan features. The estimates presented in Table 3 provide information on these preferences and further illustrate the heterogeneity among participants.

First, the signs of the coefficients point in the expected direction: A higher premium is associated with a lower probability of the plan being selected, and the same is true if a plan includes a deductible. Moreover, plans that cover illnesses D, E, and/or F were more likely to be selected. Second, the standard deviations are significant for the premium and all complementary insurance options. For the deductibles, only the standard deviation for the deductible of 20 talers is significant, albeit only marginally so. This indicates a particularly high degree of heterogeneity across participants in terms of their preferences for coverage against illness, whereas their preferences for avoiding deductibles appear less heterogeneous.

[16]Similar to Jaspersen et al. (2022), Table C1 in Appendix C provides a comparison of actual choices to a random choice. It suggests that choices rely on a planned decision process rather than on random choice because, for all plan features, individuals' choices deviate significantly from the random choices.

[17]Figure C1 in Appendix C provides an overview of selected plans by decision rounds.

TABLE 3 Preference heterogeneity—Mixed logit regression.

	Mean	Std. dev.	WTP	EV
Premium	−0.054***	0.026***		
	(0.004)	(0.003)		
Illness D (2000 talers with 1% probability of occurrence)	3.977***	3.857***	73.37	20
	(0.709)	(0.572)	[44.18, 102.56]	
Illness E (70 talers with 10% probability of occurrence)	1.617***	0.926***	29.83	7
	(0.134)	(0.219)	[24.33, 35.32]	
Illness F (40 talers with 30% probability of occurrence)	2.319***	0.850***	42.78	12
	(0.131)	(0.204)	[36.28, 49.28]	
Deductible of 10	−0.224**	0.106	−4.14	−6.2
	(0.074)	(0.162)	[−6.71, −1.56]	
Deductible of 20	−1.302***	0.654*	−24.03	−12.4
	(0.114)	(0.261)	[−28.45, −19.61]	
Deductible of 30	−1.456***	0.391	−26.86	−14.8
	(0.114)	(0.381)	[−31.67, −22.05]	
n	25,806	25,806		
N	253	253		

Note: Mixed logit regression. Mean, standard deviation (std. dev.), and willingness to pay (WTP). Negative WTP values indicate a willingness to accept (WTA). The EV of each plan feature is added for comparison with WTP/WTA. Standard errors clustered at the individual level are in parentheses. 95% confidence intervals of WTP/WTA values are in brackets.

*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$.

By calculating individuals' willingness to pay (WTP), we can provide more detailed information about their preferences for plan features. Our results indicate that the WTP to cover each of the three illnesses is more than three times higher than the EV. In other words, the participants were, on average, willing to insure against the illnesses at a higher cost than would have been the case for a risk-neutral, rational decision-maker. Negative values for the WTP and EV of the deductibles indicate a willingness to accept (WTA).

For the lowest deductible (i.e., of 10 talers), the confidence interval of the WTA includes the corresponding EV and is, therefore, not significantly different from it, whereas the WTAs for the higher deductibles are about twice as high as the EV. Our results suggest that participants, on average, disliked high deductibles and needed to be compensated by a reduction in the premium that was greater than the EV. In the case of the deductible of 20 talers, the required compensation had to be greater, on average, than the deductible amount itself. This raises the question of whether individual risk preferences, as captured by the RDEU model, can better explain this behavior.

Result 2. *Individuals' WTP for illness coverage is generally higher than the corresponding EV, whereas the WTP for deductibles is negative. At the same time, there is considerable heterogeneity across all attributes.*

3.3 | Welfare analysis

To better understand how individuals' plan choices relate to their individual risk preferences, we now turn to a welfare analysis. We evaluate whether individuals' selected plans are consistent with their RDEU preferences. For comparison, we also assess plan choices using the EV model as a benchmark.

We deduce RDEU preferences by fitting a value and weighting function to the data from Part 2 of the experiment using nonlinear least squares. In our value function, $\theta < 1$ implies risk-seeking behavior, while $\theta > 1$ implies risk-averse behavior. The observed overall median value of $\theta = 1.00$ implies risk neutrality on average. The median value of s is 0.720 ($s < 1$), which suggests a more pronounced concave portion of the weighting function relative to $s = 1$ and implies an overweighting of probabilities. The median value of the parameter r is slightly below 1 ($r = 0.985$), indicating almost no overweighting of small probabilities compared to large ones.[18] Prior studies indicate that participants either tend to underweight very low probabilities or are willing to pay much more than the expected value (Jaspersen et al., 2022; McClelland et al., 1993).

Following the approach of Harrison and Ng (2016) and Harrison, Kairies-Schwarz, et al. (2024), we calculate the certainty equivalents (CEs) for each possible plan based on the RDEU preferences of each participant. We then compute the difference in CE between the chosen plan and the individual best plan for that decision, and express this relative to the difference between the individual best and the worst plan for each decision (see Appendix E for details). Under this measure, choosing the best plan leads to a relative welfare loss of 0, whereas choosing the worst plan leads to a relative welfare loss of 1. We then compute the average relative welfare loss for each participant across their 12 decisions. Using the same procedure, we also calculate the efficiency measure based on the descriptive EV model.[19]

To contextualize our results, we use random choice as a benchmark, similar to the approach of Jaspersen et al. (2022). We derive the mean efficiency of random choice for both the EV and RDEU models. In this case, we assume that one of the six or 12 available plans is randomly chosen in each decision round. The relative welfare loss for random choice is then calculated as the expected relative welfare loss, obtained by summing the losses for all possible plans in each decision, with each plan weighted equally (1/6 or 1/12).

Overall, we find that neither model fully explains the observed behavior. The average relative welfare loss from the RDEU perspective is 0.41 (std. dev. 0.28), whereas from the EV perspective it is slightly lower at 0.35 (std. dev. 0.20), indicating that EV performs slightly better on average. However, a substantial number of individuals appear better off when evaluated under their fitted RDEU preferences. In both models, actual choices are superior to random choice in terms of efficiency: mean losses are lower than those under random choice (RDEU: 0.41 vs. 0.51; EV: 0.35 vs. 0.47).

[18]For an analysis of the relationship between the RDEU model risk preference parameters and plan features, particularly with respect to coverage for illness D, please refer to Appendix D. This analysis indicates that risk-seeking individuals cover illness D slightly less frequently than risk-averse individuals. Moreover, the parameter s tends to be lower in decisions where illness D is covered compared to those in which it is not; however, at the threshold value of 1, only minor differences in coverage rates are observed.

[19]In this case, the CE of a plan is uniformly the same for all participants due to the descriptive nature of the EV model. Consequently, all participants share the same best and the same worst EV plans within a decision.

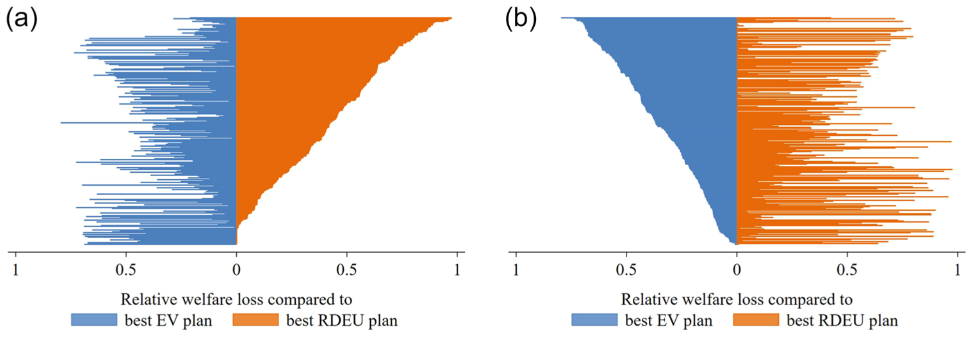

FIGURE 2 Relative welfare loss compared to best plan. *Note*: The figures illustrate mean relative welfare losses over 12 decisions of each participant by the descriptive EV perspective and the normative RDEU perspective, sorted by the size of losses according to the best RDEU plan (2a) and the EV best plan (2b). $N = 253$. (a) RDEU perspective. (b) EV perspective. [Color figure can be viewed at wileyonlinelibrary.com]

Result 3a. *Neither the normative RDEU model nor the descriptive EV model fully explains the observed behavior. However, in both models, welfare losses in terms of efficiency are smaller than those expected under random choice.*

To provide a more detailed analysis, Figure 2 presents the average relative welfare losses per individual compared to both the best EV and best RDEU plan.[20] The two subfigures display the same underlying choices but are sorted differently: Figure 2a is sorted by losses under the RDEU perspective, and Figure 2b is sorted by losses under the EV perspective. Although the EV model yields slightly lower average losses, the distribution of losses under RDEU is more dispersed, with a higher median and a longer upper tail.

We also observe mixed behavioral patterns. Figure 2a indicates that some individuals incur almost no losses under RDEU (lower part); for these individuals, the RDEU model appears to fit well. In contrast, Figure 2b shows individuals with very low losses under EV but simultaneously high losses under RDEU (lower part). Although EV is a specific case of RDEU, and the RDEU model can capture EV-maximizing preferences, such a discrepancy may result from individuals relying on heuristics (e.g., choosing the cheapest plan; see, e.g., Ericson & Stark, 2012; Kairies-Schwarz et al., 2017) or making mistakes in the insurance choices or in the risk elicitation tasks. Only a small subset of individuals shows low losses under both models, indicating that their choices are evaluated as efficient by both RDEU and EV perspectives.[21] Finally, several individuals show substantial welfare losses under both models, again suggesting possible reliance on heuristics or decision errors.

Result 3b. *The individual distributions of welfare losses reflect different behavioral patterns: Some participants' decisions can be well explained by the RDEU model; for some participants, by preferences consistent with EV (as a special case of RDEU). However, some*

[20]The classification of participants as EV- or RDEU-consistent depends on the selected relative welfare loss thresholds. See Appendix E, Table E1, for detailed distributions of relative welfare losses.

[21]This pattern is illustrated in Figure E1 in Appendix E. The figure displays the distribution of loss differences between the EV model and the RDEU model.

participants show small losses under EV but high losses under RDEU, suggesting the use of heuristics or decision errors. Finally, a substantial number of participants exhibit substantial welfare losses under both models, further indicating heuristics or decision errors.

3.4 | Subgroup analysis

To explore the relationship between plan feature preferences and welfare losses, we introduce a latent class logit model to identify subgroups with homogeneous preferences.[22] This allows us to examine variation in welfare losses across these subgroups. As with the mixed logit model (see again Table 3, Section 3.2), the latent class logit model considers all available alternatives within each choice set in addition to the chosen plan.

Employing the Bayesian information criterion (BIC) to select the optimal class size, which is the class size that minimizes the BIC, leads to an optimal number of nine or more classes.[23] This high number indicates a high degree of heterogeneity and a variety of decision patterns. However, the use of nine classes results in a small number of observations per class. Therefore, in this analysis, we discuss the results for a reduced number of classes. To determine the optimal reduction in the number of classes, we employed the elbow method (see Nylund-Gibson & Choi, 2018). Figure F1 in Appendix F visualizes statistical fits for various class sizes and shows that the improvement in model fit plateaus with five classes. When we reduce the number of classes considered in the analyses to this number, the model fit is marginally poorer than with nine (see Table F2 in Appendix F).[24] Nevertheless, with an average probability of 0.94 (std. dev. 0.10) for the highest individual class membership, the model with five classes still appears to distinguish very well between the several classes of preferences. In the following, the allocation of individuals to the classes is based on their highest class membership probability.

Figure 3 graphically illustrates the results of the latent class logit model, presenting these as the WTP estimates for a particular feature in each class.[25] The class proportions range from 13.8% to 25.3%. Again, we compare the WTP estimates with the EV of each feature. We sort the classes according to the summed absolute distances between the WTP estimates (dots) and the EV (blue crosses) of each feature, starting from the smallest (Class 1) to the largest (Class 5) distance. The higher the class numbering, the higher the overall WTP for coverage compared to a risk-neutral decision-maker.

As a result of this sorting, Class 1 and Class 5 show the most pronounced differences in their WTP and WTA values for the respective plan features.[26] This is mainly because of different preferences for complementary insurance. The WTP and WTA values of Class 1 members are closest to the EV. In contrast, members of Class 5 were willing to pay much more to insure against all illnesses than those in other classes. This is especially pronounced for illness D (low probability of high costs), for which the WTP was around 10 times higher than the EV. Classes 2 to 4 are located between Classes 1 and 5, with Class 2 placing a positive value on

[22]We used the Stata routine lclogit2 to estimate the latent class model. See Yoo (2020) for a detailed model description.

[23]For 10 or more classes, only a few models converged. Thus, the BIC and AIC might be minimized at nine or more classes. Fit statistics for two to nine classes can be found in Table F1 in Appendix F.

[24]Results for nine classes can be found in Table F4 and Figure F2 in Appendix F.

[25]Regression results are presented in Table F3 in Appendix F. We do not find statistically significant differences between the classes regarding the experimental treatments with or without an IC filter.

[26]For nine classes, we find a similar range (see Figure F2 in Appendix F).

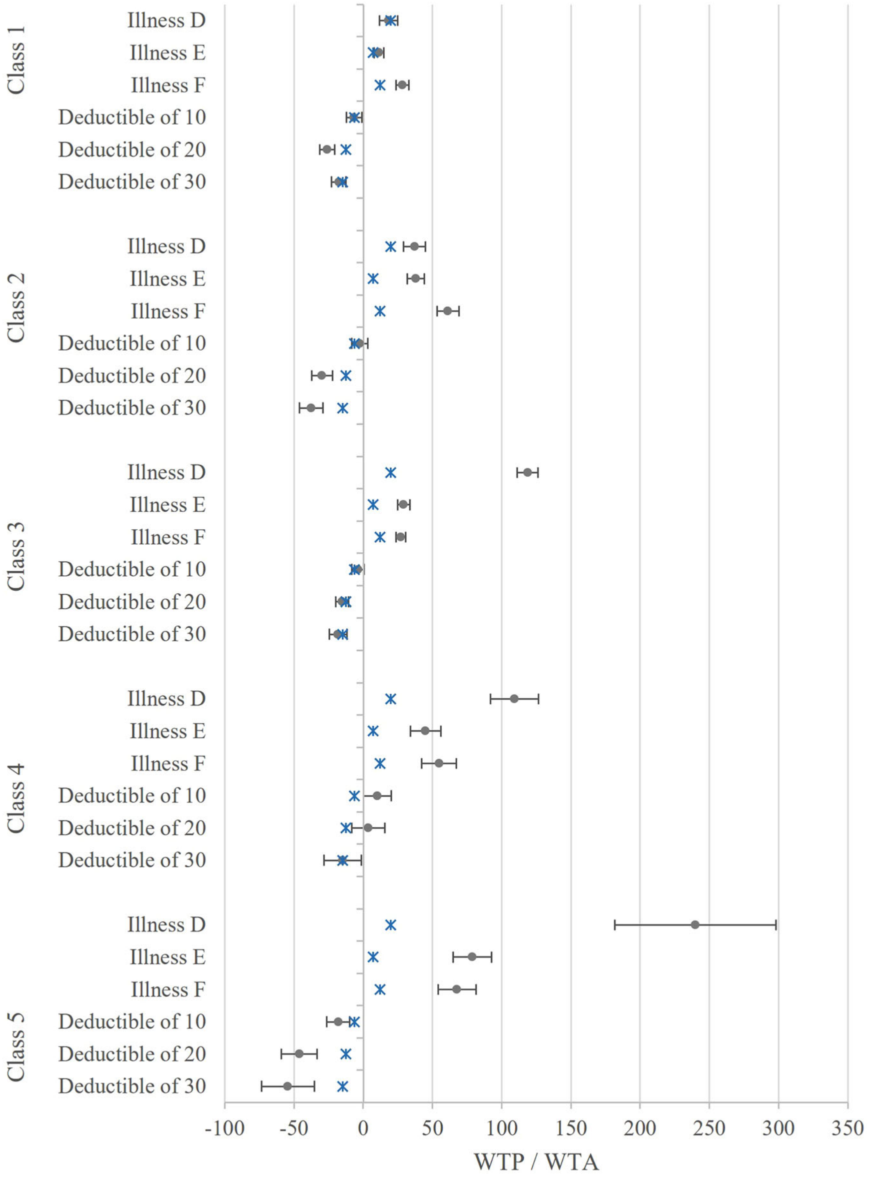

FIGURE 3 Willingness to pay by classes and features. *Note*: Error bars represent 95% confidence intervals. Blue stars represent EV of features. Illness D has costs of 2000 talers with a 1% probability of occurrence, illness E has costs of 70 talers with a 10% probability of occurrence, and illness F has costs of 40 talers with a 30% probability of occurrence. [Color figure can be viewed at wileyonlinelibrary.com]

additional insurance for illness F (relatively high probability of occurrence), whereas Class 3 rather focused on insurance for illness D.

Most participants preferred to avoid deductibles for the illnesses covered by basic insurance, as indicated by their WTA values. It appears that participants in Classes 2 to 4 disregarded the low deductible of 10 talers because their WTA/WTP values were not significantly different from 0. For Class 4, this was also the case for the deductible of 20 talers. In contrast, participants in Class 5

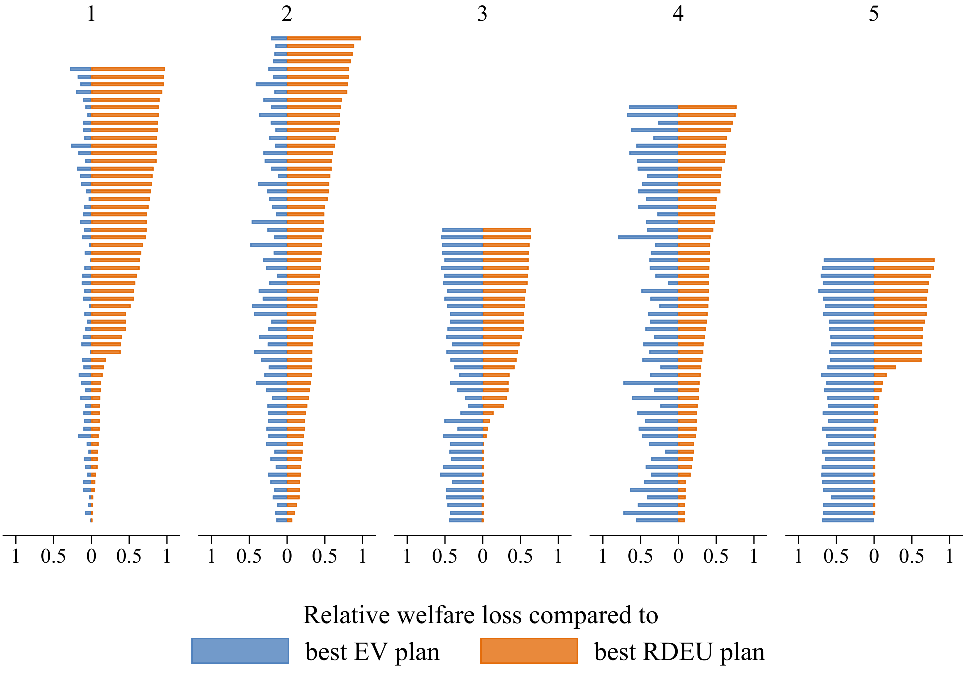

FIGURE 4 Relative welfare losses compared to best plan by class. *Note*: The figure illustrates mean relative welfare losses over 12 decisions of each participant by the descriptive EV perspective and the normative RDEU perspective divided by class, sorted by the size of losses according to the best RDEU plan. [Color figure can be viewed at wileyonlinelibrary.com]

required the greatest compensation for deductibles because their WTA for the deductible was significantly greater than the corresponding EV. Participants' approach to deductibles may reflect factors more complex than considerations related to insuring against illnesses D, E, and F in our experiment, potentially including a general tendency to avoid deductibles.

> **Result 4a.** *We identify five homogeneous classes of participants who differ in their preferences for plan features when choosing health insurance. These classes range from* (*a*) *those who exhibit WTP and WTA that are close to the expected values to* (*b*) *those with a high WTP for insurance against all illnesses.*

To better understand which individuals' insurance decisions align more closely with EV or RDEU preferences, we next examine the distribution of welfare losses across individuals. Figure 4 presents the relative welfare losses of each individual under both models (EV, in blue; RDEU, in orange), sorted by class.[27] This visualization allows us to identify distinct types of individuals and to link these patterns to specific plan features (Figure 3) and to RDEU parameters (Table F5 in Appendix F).[28]

[27]Figures F3 and F4 in Appendix F present the average relative welfare losses and the distribution of average relative welfare losses for each class from both the EV and RDEU perspectives.

[28]To assess the classification of participants as EV- or RDEU-consistent, one would need to define thresholds for relative welfare loss (e.g., <10%). Appendix F Table F7 provides detailed proportions of participants within each class whose average welfare losses fall below a range of cutoffs.

Across all classes, we find that the class assignment, based solely on plan choices, also reflects the structure of welfare losses under the EV model quite well. Within each class, EV losses are relatively homogeneous but increase from Class 1 to Class 5. Under the RDEU model, however, within-class variation in relative welfare losses is more pronounced. This indicates that even among individuals with similar preferences for plan features, actual choices do not always align with their elicited RDEU preferences.

Within each class, we observe both well-aligned decision-makers (lower bars, with minimal welfare loss) and individuals with substantial deviations (upper bars, with larger losses). In Classes 3 and 5, many individuals show relatively low RDEU losses, suggesting that their high WTP for certain plan attributes is consistent with their elicited risk preferences. In contrast, in Class 1, many individuals appear to choose EV-maximizing plans, but their relatively low observed WTP for plan attributes is not consistent with their elicited risk preferences. For these individuals, plans with higher coverage might have yielded lower RDEU welfare losses. This discrepancy may reflect the use of heuristics or errors in either the insurance choice task or risk elicitation tasks.[29]

These patterns suggest that individual decision-making is partly driven by dynamics not fully captured by the RDEU model and that no single theory universally applies to all individuals (Harrison & Rutström, 2009). Taken together, our results suggest that while the RDEU model does not improve welfare assessments on average, it does capture important aspects of behavior for specific subgroups.

Result 4b. *Under the RDEU model, we observe greater within-class variation in welfare losses. This suggests that while some individual plan choices reflect RDEU-consistent preferences (including sensitivity to rare, high-cost events), there remains substantial evidence of heuristic-driven behavior or decision errors.*

4 | ROBUSTNESS CHECKS

As noted in Section 3.1, we find no differences in the plan features chosen by participants in the groups with or without a filter. However, data on participants' use of the filter make it possible to analyze which of the features that are selectable in the filter appear to have been most important for the participants when making their ultimate choice of plan. This can help us better understand the thought process behind plan selection and thus validate feature preferences. The advantage of the experiment is that we can exclude all other aspects, such as different health states, liquidity constraints, or expectations that may also impact choices and are not easily observable in the field. Because each plan had a particular set of features, simply analyzing the selected plan would not provide insights into how individuals arrived at their decision.

Overall, we observe that the filter was used by 88.9% of all participants in at least one decision round. With a mean of 6.28 rounds (std. dev. 0.37) per participant, the filter was used in about half of the 12 decision rounds. By using the filter, participants could reduce the

[29]For instance, the plan with the lowest premium was chosen in 10.34% of all decisions. This choice was most common in Class 1, where it was selected in 26.94% of cases. In Class 5, it was never selected. Overall, 61.78% of all lowest-premium choices came from participants in Class 1. In three out of the 12 rounds, the lowest-premium plan also had the best or nearly the best expected value.

complexity of choosing a health insurance plan out of a large choice set. Based on the participants' final filter setting, we see that, for low-complexity decisions, they used the filter to reduce the number of plans highlighted in a set within a decision round from 6 to 3.77 on average (std. dev. 1.32, median 4). For high-complexity decisions, the reduction was from 12 to 7.42 plans (std. dev. 2.46, median 7).[30] Within any given decision round, some participants used the filter several times—that is, they tested how their choice of features would affect which plans would be highlighted. We observe that participants selected different filter variations up to nine times, with a mean of 1.62 (std. dev. 0.04) for the decisions in which the filter was used at least once. That is, some participants did not use the filter at all, others used it once to decide on a plan, and others tried several different filter configurations before deciding on a plan.[31]

Table 4 presents the frequency of filter use, proportions of preferred features in the filter, and proportions of these preferred features that were in the plans ultimately chosen by members of different classes of participants. We observe that the different classes were heterogeneous in their filter use: the filter was used by members of Class 1 in only a small proportion of the decisions (28.8%), which is about half of what we observe for the other classes (53.2% to 61.5%). Because members of Class 1 did not seem to place strong emphasis on plan features, participants may have chosen not to use the filter because they felt they did not need it to make their choices. Instead, they may have been guided mainly by the premium, which was not included in the filter.

The first panel of Table 4 provides information on the composition of the plans selected by participants. The preferences for plan features by class are in concordance with the findings from our latent class analysis: A high proportion of plans chosen by members of Class 1 have coverage for illness F (low costs with a high probability of occurrence). For Class 2, a high proportion of plans chosen by participants include coverage both for illness F and illness E (medium costs with medium probability of occurrence). Meanwhile for Class 3, a high proportion of the plans include coverage of illness D (high costs with low probability of occurrence). The plans chosen by members of Class 4 have the highest average deductible. In Classes 4 and 5, a high proportion of plans cover all illnesses. The proportion of plans with coverage for illness D is highest in Class 5, with this feature being included in 99.8% of participants' plan choices. Given that a high level of coverage is associated with a higher premium, the increasing proportion of coverage by class number is also reflected in a trend towards higher average premiums.

Overall, across all classes, we find that the probability of illness occurrence played a role in participants' decisions. There is a preference for coverage against illness F, which has a 30% probability, over illness E, which has a 10% probability. This preference is consistent with the findings of Jaspersen et al. (2022), who observed increased demand for insurance with higher loss probabilities. In our experiment, we observe a heterogeneous pattern in the overweighting of probabilities, which is particularly evident in the lower probability range for about half of the participants. For Classes 1 and 2, there is higher demand for coverage against illnesses E and F, which have higher probabilities of occurrence than illness D. In contrast, Classes 3, 4, and 5 exhibit a different pattern, showing a stronger preference for coverage against illness D compared to E or F. These findings highlight differences that are more subtle than those identified

[30]All of the highlighted plans correspond to the preferences selected in the filter. This suggests that the participants rarely used the filter to narrow down the selection in such a way that only one or two plans would remain. Instead, for most decisions, there was still room to decide among a large variety of plans.

[31]For our analysis, we defined the last selection in the filter as the relevant selection.

TABLE 4 Chosen and preferred features by class.

	Class 1	Class 2	Class 3	Class 4	Class 5	Overall
I. Plans chosen by participants						
Mean premium	52.156 (0.742)	72.745 (1.204)	95.066 (1.329)	90.345 (1.955)	115.993 (0.690)	81.112 (1.439)
Illness D (1%) covered	17.36% (0.018)	32.29% (0.020)	96.79% (0.011)	71.06% (0.023)	99.76% (0.002)	56.46% (0.022)
Illness E (10%) covered	39.03% (0.018)	64.58% (0.021)	32.48% (0.017)	51.06% (0.025)	57.38% (0.019)	49.64% (0.012)
Illness F (30%) covered	62.08% (0.022)	79.30% (0.015)	55.34% (0.012)	56.36% (0.019)	69.05% (0.013)	65.12% (0.010)
Mean deductible	11.125 (0.547)	8.828 (0.497)	11.303 (0.361)	13.197 (0.511)	8.405 (0.316)	10.646 (0.248)
n	720	768	468	660	420	3,036
N	60	64	39	55	35	253
II. Stated preferences						
Coverage for illness D preferred	11.67% (0.054)	16.63% (0.043)	92.36% (0.029)	58.69% (0.081)	97.67% (0.018)	50.58% (0.042)
Coverage for illness E preferred	16.63% (0.073)	32.40% (0.055)	8.49% (0.049)	21.72% (0.048)	30.54% (0.085)	23.08% (0.029)
Coverage for illness F preferred	66.52% (0.099)	71.57% (0.058)	12.38% (0.048)	56.68% (0.066)	33.55% (0.090)	50.69% (0.038)
Mean preferred max deductible	21.239 (1.711)	20.189 (1.306)	26.408 (1.438)	20.938 (1.745)	25.694 (1.460)	22.527 (0.724)
n	76	246	155	134	124	735
N	15	34	21	19	15	104
Proportion of decisions made using the filter	0.288 (0.064)	0.586 (0.054)	0.615 (0.057)	0.532 (0.080)	0.574 (0.085)	0.524 (0.031)

Note: The proportion of chosen plans that cover illnesses D, E, or F, as well as the mean premium and deductible of all chosen plans, is shown for the complete sample ($n = 3036$). The proportion of filter use in which preferred coverage for illnesses D, E, or F was selected as the final filter setting, as well as the mean of the selected maximum deductible, is shown for decisions in which participants in the IC filter treatment used the filter ($n = 735$). For the mean preferred deductibles, the selected maximum deductible is used; choosing nothing or "irrelevant" is treated as a maximum deductible of 30 talers. Standard errors are in parentheses.

by Jaspersen et al. (2022), underscoring the importance of considering heterogeneity in decision-making.

The second panel of Table 4 provides information on the features selected by participants who used the filter. We observe that the feature preferred most by these participants, measured in terms of the features they selected in the filter, was that which was ultimately included in their chosen plan. We also observe a prioritization of features insofar as participants focused on particular features in the filter, which then mainly drove their choice of plan. For Classes 3 and 4, it seems that insuring against illness D was the main driver of participants' plan choice, whereas for Classes 1 and 2, the main driver appeared to be insuring against illness F. It is striking that the feature preferences of the subset of participants who used the filter mirror the actual plan choices made by all members of each class. For example, even if the filter was used in only about 30% of the choices made by members of Class 1, this subset still appears representative of the average of all further choices made by participants in both the IC and IC filter treatments.

In the post-experimental questionnaire, we asked participants about the role that specific plan features played in their decisions (results are provided in Table G1 in Appendix G). For example, we asked how often a participant considered premiums and deductibles when making their decisions. Additionally, we asked how often they considered whether a plan covered the cost of 2000 talers when making their decisions. Our results reveal that Class 1 has the highest percentage of members (98.3%) who always or frequently considered the premium. In contrast, 50.0% of Class 1 members reported never having considered the coverage of the cost of 2000 talers, whereas 94.3% of Class 5 members always considered it. Notably, Class 3, which is the class that paid the least attention to deductibles, is the only one with WTP values for deductibles instead of WTA values. This suggests that participants were aware of their preferences and based their decisions on this.

5 | DISCUSSION AND CONCLUSION

In this study, we investigated heterogeneity in preferences for health insurance features using choice data from a controlled laboratory experiment with a sequential design. Participants first made decisions regarding health insurance plans and then completed a series of lotteries to elicit individual risk preferences. To validate feature preferences, a voluntary feature-based filter was provided in one treatment group. The filter allowed participants to choose a maximum deductible and additional coverage for other illnesses and then highlighted the corresponding plans while preserving the entire choice set. Risk preferences were elicited using the RDEU model in the loss domain, similar to the approach taken by Wakker and Deneffe (1996).

Our study has four main findings. First, our results suggest that, on average, participants are willing to pay more than the EV to insure against losses, particularly low-probability, high-cost risks. This stands in contrast to empirical evidence. A meta-analysis by Mankaï et al. (2024) reports that, on average across studies, stated WTP is lower than expected losses, though some experimental studies find WTP above the EV. For example, Zimmer et al. (2018) find that individuals are willing to pay more than the EV for default-free insurance, but that WTP drops sharply and falls below the EV when even small default risks are introduced. Our study contributes to this debate by (a) supporting existing experimental findings that there are many individuals who are willing to pay more than the EV to insure against low-probability, high-cost risks and (b) providing evidence that this willingness is highly heterogeneous across individuals.

Second, our welfare analysis shows that welfare losses occur under both the EV and the RDEU models, indicating that neither fully explains observed behavior at the aggregate level. However, both models yield more efficient outcomes than random choice. In contrast to the findings of Jaspersen et al. (2022), who reported that neither EV nor RDEU outperforms random selection in predicting insurance choices, our results provide a descriptive rather than predictive or prescriptive perspective, focusing on behavioral welfare evaluation.

Third, at the individual level, the distribution of welfare losses shows distinct behavioral patterns. The decisions of some individuals are well described by the RDEU model, whereas a smaller group displays decision patterns consistent with simple EV maximization, which is a special case of RDEU. However, several individuals show small losses under EV but substantial losses under RDEU, suggesting the use of heuristics or the presence of decision errors. Other individuals show substantial welfare losses under both models, likewise pointing to heuristics or decision errors. These findings suggest that while the RDEU model (which includes EV as a special case) can capture the decision patterns of some participants very well it does not fit all behaviors observed. This is consistent with previous research showing that, when faced with complex health insurance choices, many consumers tend to focus on salient plan features and use simplifying heuristics, such as choosing the plan with the lowest premium (e.g., Besedeš et al., 2012a, 2012b; Ericson & Starc, 2012; Kairies-Schwarz et al., 2017), or they make erroneous decisions (Handel & Kolstad, 2015; Sinaiko & Hirth, 2011).

Fourth, we observe substantial heterogeneity in plan feature preferences across five classes of participants (identified using a latent class model), as well as in how well individuals' choices fit with their elicited risk preferences. Under RDEU, we find considerable variation in welfare losses within each class. This suggests that while the plan choices of some individuals are consistent with sensitivity to rare, high-cost events, others rely more heavily on heuristics or make decision errors.

Furthermore, our robustness checks show that this heterogeneity in WTP for certain plan features is also reflected in participants' use of the feature-based filter. These results may provide a basis for improving the design of insurance choice environments—for example, by offering tools that help match individuals with plans that reflect their feature preferences.

Our study has some important limitations that should be considered when interpreting its findings. As with any laboratory experiment, concerns regarding external validity apply. In this study, two aspects are particularly relevant. First, while the sets of health insurance choices used in our experimental design provided a well-controlled decision environment, the results may not be readily generalizable beyond the specific sets included in the study. Second, the participants were university students, who generally have higher levels of education, stronger cognitive abilities, and fewer financial constraints than the general population. In addition, because the sample was drawn from a German university, a majority of participants (56%) had limited experience with health insurance decisions, as they could remain under family coverage until the age of 25. Moreover, plan features such as deductibles and illness-specific coverage are not common in Germany's statutory health insurance system. Nonetheless, many participants were approaching an age at which they would soon confront their first real-world insurance choices. To increase external validity, future research could follow the approach of Samek and Sydnor (2025), who combined a laboratory experiment using a student sample and a parallel crowdsourcing experiment using a representative sample. Such studies could also be conducted in countries in which individuals are more accustomed to choosing health insurance plans with deductibles and varying options.

Two further limitations concern the elicitation of risk preferences. First, the methods used to elicit individual risk preferences with the RDEU model are more complex and time-consuming than those employed to elicit preferences under EUT, which may have affected the reliability of the measurements. Second, our evaluation of plan choices is based on an exogenous measure of risk preferences. Although this avoids the difficulty of inferring preferences from observed insurance choices, applying exogenously elicited risk preferences to complex insurance presents its own challenges. Our findings suggest that such a transfer of preferences across decision domains does not hold universally: some individuals' insurance choices are consistent with their elicited risk preferences, while others are not. A possible interpretation is that, in the more complex insurance setting, some individuals relied on simplifying heuristics or that the greater complexity contributed to decision errors. Although we cannot rule out the possibility that heuristics or errors may also have affected choices in the lottery tasks, the higher cognitive demands of the insurance task probably heightened this tendency. Further research is needed to clarify the relationship between risk preferences in insurance decision-making.

In conclusion, our results indicate that preferences for various features in health insurance plans and their welfare implications are highly heterogeneous. This should be carefully considered in policy reforms that seek to facilitate individual choice. Uniform approaches that treat all individuals the same would miss the mark. A more effective approach would be to tailor choice environments—for example, by curating plan sets based on individuals' risk preferences—in ways that better match plan offerings with individuals' true preferences and ultimately lead to improved welfare outcomes.

ACKNOWLEDGMENTS

Open Access funding enabled and organized by Projekt DEAL. We are grateful for the valuable comments and suggestions from Katharina Blankart, Stefan Boes, Martin Dufwenberg, Jonathan Federle, Massimo Finocchiaro, Mathias Kifmann, Georg Kirchsteiger, Petra Steinorth, Justin Sydnor, Stefan Trautmann, Joachim Winter, Peter Zweifel, and participants of the CEAR/MRIC Behavioral Insurance Workshop 2022 in Munich, the ECORES summer school 2022 in Leuven, the EuHEA PhD student-supervisor conference 2021 online, the annual meeting of the German Society for Health Economics 2021 online, and the allocation workshop of the German Society for Health Economics 2021 in Mannheim. We gratefully acknowledge financial support provided by the German Federal Ministry of Education and Research.

ORCID

Benedicta Hermanns https://orcid.org/0000-0002-3696-5707
Johanna Kokot https://orcid.org/0000-0002-2267-9785

REFERENCES

Abaluck, J., & Adams-Prassl, A. (2021). What do consumers consider before they choose? Identification from asymmetric demand responses. *Quarterly Journal of Economics*, *136*(3), 1611–1663.

Abaluck, J., & Gruber, J. (2011). Heterogeneity in choice inconsistencies among the elderly: Evidence from prescription drug plan choice. *American Economic Review*, *101*(3), 377–381.

Abaluck, J., & Gruber, J. (2016). Evolving choice inconsistencies in choice of prescription drug insurance. *American Economic Review*, *106*(8), 2145–2184.

Abaluck, J., & Gruber, J. (2023). When less is more: Improving choices in health insurance markets. *The Review of Economic Studies*, *90*(3), 1011–1040.

Abdellaoui, M. (2000). Parameter-free elicitation of utility and probability weighting functions. *Management Science*, *46*(11), 1497–1512.

Abdellaoui, M., Bleichrodt, H., & Paraschiv, C. (2007). Loss aversion under prospect theory: A parameter-free measurement. *Management Science*, *53*(10), 1659–1674.

Barseghyan, L., Coughlin, M., Molinari, F., & Teitelbaum, J. C. (2021). Heterogeneous choice sets and preferences. *Econometrica*, *89*(5), 2015–2048.

Besedeš, T., Deck, C., Sarangi, S., & Shor, M. (2012a). Age effects and heuristics in decision making. *Review of Economics and Statistics*, *94*(2), 580–595.

Besedeš, T., Deck, C., Sarangi, S., & Shor, M. (2012b). Decision-making strategies and performance among seniors. *Journal of Economic Behavior & Organization*, *81*(2), 524–533.

Bhargava, S., Loewenstein, G., & Benartzi, S. (2017). The costs of poor health (plan choices) & prescriptions for reform. *Behavioral Science & Policy*, *3*(1), 1–12.

Bhargava, S., Loewenstein, G., & Sydnor, J. (2017). Choose to lose: Health plan choices from a menu with dominated option. *The Quarterly Journal of Economics*, *132*(3), 1319–1372.

Biener, C., & Zou, L. (2024). More options, more problems? Lost in the health insurance maze. *Journal of Risk and Insurance*, *91*(1), 5–35.

Coughlin, T. A., Long, S. K., Triplett, T., Artiga, S., Lyons, B., Duncan, R. P., & Hall, A. G. (2008). Florida's medicaid reform: Informed consumer choice?: After Florida implemented reforms, a sizable minority of recipients were not even aware that they were enrolled in a reform plan, and many did not understand how their plan worked. *Health Affairs*, *27*(6), w523–w532.

Cronqvist, H., & Thaler, R. H. (2004). Design choices in privatized social-security systems: Learning from the Swedish experience. *American Economic Review*, *94*(2), 424–428.

Ericson, K. M., Kircher, P., Spinnewijn, J., & Starc, A. (2021). Inferring risk perceptions and preferences using choice from insurance menus: Theory and evidence. *The Economic Journal*, *131*(634), 713–744.

Ericson, K. M., & Starc, A. (2012). Heuristics and heterogeneity in health insurance exchanges: Evidence from the Massachusetts connector. *American Economic Review*, *102*(3), 493–497.

Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, *10*(2), 171–178.

Galizzi, M. M., & Wiesen, D. (2017). Behavioural experiments in health: An introduction. *Health Economics*, *26*(S3), 3–5.

Galizzi, M. M., & Wiesen, D. (2018). Behavioural experiments in health. In J. Hamilton (Ed.), *Oxford research encyclopedia of economics and finance*. Oxford University Press.

Greiner, B. (2015). Subject pool recruitment procedures: Organizing experiments with ORSEE. *Journal of the Economic Science Association*, *1*, 114–125.

Gruber, J., Handel, B. R., Kina, S. H., & Kolstad, J. T. (2021). Managing intelligence: Skilled experts and decision support in markets for complex products. *NBER Working Paper*, No. w27038.

Handel, B. R. (2013). Adverse selection and inertia in health insurance markets: When nudging hurts. *American Economic Review*, *103*(7), 2643–2682.

Handel, B. R., & Kolstad, J. T. (2015). Health insurance for "humans": Information frictions, plan choice, and consumer welfare. *American Economic Review*, *105*(8), 2449–2500.

Harrison, G. W. (2024). Risk preferences and risk perceptions in insurance experiments: Some methodological challenges. *The Geneva Risk and Insurance Review*, *49*(1), 127–161.

Harrison, G. W., Kairies-Schwarz, N., & Han, J. (2024). Deductibles and health care utilization: An experiment on the role of forward-looking behavior. *Journal of Economic Behavior & Organization*, *224*, 717–748.

Harrison, G. W., & Ng, J. M. (2016). Evaluating the expected welfare gain from insurance. *Journal of Risk and Insurance*, *83*(1), 91–120.

Harrison, G. W., & Rutström, E. E. (2009). Expected utility theory and prospect theory: One wedding and a decent funeral. *Experimental Economics*, *12*(2), 133–158.

Harrison, G. W., & Swarthout, J. T. (2023). Cumulative prospect theory in the laboratory: A reconsideration. In G. W. Harrison & D. Ross (Eds.), *Models of risk preferences: Descriptive and normative challenges* (pp. 107–192). Emerald Publishing Limited.

Heiss, F., Leive, A., McFadden, D., & Winter, J. (2013). Plan selection in Medicare Part D: Evidence from administrative data. *Journal of Health Economics*, *32*(6), 1325–1344.

Heiss, F., McFadden, D., Winter, J., Wuppermann, A., & Zhou, B. (2021). Inattention and switching costs as sources of inertia in medicare part D. *American Economic Review*, *111*(9), 2737–2781.

Iyengar, S. S., & Kamenica, E. (2010). Choice proliferation, simplicity seeking, and asset allocation. *Journal of Public Economics*, *94*(7), 530–539.

Jaspersen, J. G., Ragin, M. A., & Sydnor, J. R. (2022). Predicting insurance demand from risk attitudes. *Journal of Risk and Insurance*, *89*(1), 63–96.

Kairies-Schwarz, N., Kokot, J., Vomhof, M., & Weßling, J. (2017). Health insurance choice and risk preferences under cumulative prospect theory—An experiment. *Journal of Economic Behavior & Organization*, *137*, 374–397.

Kairies-Schwarz, N., Rieger-Fels, M., & Waibel, C. (2023). Cost-sharing or rebate: The impact of health insurance design on reducing inefficient care. *Frontiers in Behavioral Economics*, *2*, 1043188.

Kamenica, E. (2008). Contextual inference in markets: On the informational content of product lines. *American Economic Review*, *98*(5), 2127–2149.

Krieger, M., & Felder, S. (2013). Can decision biases improve insurance outcomes? An experiment on status quo bias in health insurance choice. *International Journal of Environmental Research and Public Health*, *10*(6), 2560–2577.

Liu, C., & Sydnor, J. (2022). Dominated options in health insurance plans. *American Economic Journal: Economic Policy*, *14*(1), 277–300.

Mankaï, S., Marchand, S., & Le, N. H. (2024). Valuing insurance against small probability risks: A meta-analysis. *Journal of Behavioral and Experimental Economics*, *109*, 102181.

McClelland, G. H., Schulze, W. D., & Coursey, D. L. (1993). Insurance for low-probability hazards: A bimodal response to unlikely events. In *Making decisions about liability and insurance* (pp. 95–116). Springer.

McFadden, D. L. (1974). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers in econometrics* (pp. 105–142). Academic Press.

McWilliams, J. M., Afendulis, C. C., McGuire, T. G., & Landon, B. E. (2011). Complex medicare advantage choices may overwhelm seniors—Especially those with impaired decision making. *Health Affairs*, *30*(9), 1786–1794.

Mimra, W., Nemitz, J., & Waibel, C. (2020). Voluntary pooling of genetic risk: A health insurance experiment. *Journal of Economic Behavior & Organization*, *180*, 864–882.

Nylund-Gibson, K., & Choi, A. Y. (2018). Ten frequently asked questions about latent class analysis. *Translational Issues in Psychological Science*, *4*(4), 440–461.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1993). *The adaptive decision maker*. Cambridge University Press.

Quiggin, J. (1982). A theory of anticipated utility. *Journal of Economic Behavior & Organization*, *3*(4), 323–343.

Samek, A., & Sydnor, J. (2025). The impact of consequence information on insurance choice. *Review of Economics and Statistics*, 1–46.

Schram, A., & Sonnemans, J. (2011). How individuals choose health insurance: An experimental analysis. *European Economic Review*, *55*(6), 799–819.

Sinaiko, A. D., & Hirth, R. A. (2011). Consumers, health insurance and dominated choices. *Journal of Health Economics*, *30*(2), 450–457.

Thomson, S., Busse, R., Crivelli, L., van de Ven, W., & Van de Voorde, C. (2013). Statutory health insurance competition in Europe: A four-country comparison. *Health Policy*, *109*(3), 209–225.

Wakker, P., & Deneffe, D. (1996). Eliciting von Neumann-Morgenstern utilities when probabilities are distorted or unknown. *Management Science*, *42*(8), 1131–1150.

Yoo, H. I. (2020). lclogit2: An enhanced command to fit latent class conditional logit models. *The Stata Journal*, *20*(2), 405–425.

Zimmer, A., Gründl, H., Schade, C. D., & Glenzer, F. (2018). An incentive-compatible experiment on probabilistic insurance and implications for an insurer's solvency level. *Journal of Risk and Insurance*, *85*(1), 245–273.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Hermanns, B., Kairies-Schwarz, N., Kokot, J., & Vomhof, M. (2025). Heterogeneity in health insurance choice: An experimental investigation of consumer choice and feature preferences. *Journal of Risk and Insurance*, 1–26. https://doi.org/10.1111/jori.70026