

Discrete Choice Experiments on Preferences in the Health Care Sector

Disseration

zur Erlangung des Grades eines Doktors der Wirtschaftswissenschaft
der Rechts- und Wirtschaftswissenschaftlichen Fakultät
der Universität Bayreuth

Vorgelegt

von

Stefan Rudolf Gerhard Christian Rohrbacher

aus

Lichtenfels

Dekan/-in:	Prof. Dr. Claas Christian Germelmann
Erstberichterstatter/-in:	Prof. Dr. Martin Emmert
Zweitberichterstatter/-in:	Prof. Dr. Michael Lauerer
Tag der mündlichen Prüfung:	14.09.2026

Abstract

This dissertation investigates the methodology and application of discrete choice experiments (DCEs) in the health care sector to elicit preferences of patients and physicians, focusing on human immunodeficiency virus (HIV) treatment and total hip replacement (THR) surgery. Its objective is to understand how individuals value different healthcare attributes and to inform the development of patient-centred composite quality measures. Additionally, the research explores how the relative importance of attributes varies depending on the estimation approach used.

Grounded in random utility theory, DCEs provide a robust framework for understanding how individuals value different healthcare attributes by presenting hypothetical scenarios requiring trade-offs. The work reviews advanced statistical models for DCE analysis, including multinomial logit, random parameters logit, and latent class models among others, highlighting their strengths and limitations in capturing preference heterogeneity. It emphasises the importance of rigorous experimental design, including attribute selection, level coding, and efficient choice set construction, to ensure validity and reliability. It also critically assesses model selection criteria, advocating for the use of information-theoretic measures to balance model complexity and interpretability.

Empirical studies form the core of the dissertation, comprising three discrete choice experiments. The first investigates preferences of people living with HIV in Germany regarding long-acting antiretroviral therapies, revealing that dosing frequency and long-term side effect risks are paramount. The second study compares patient and physician preferences for hospital quality indicators related to THR, uncovering notable differences in attribute prioritisation, such as the emphasis on revision surgery rates by physicians versus postoperative complication rates by patients. The third study focuses on patient preferences for hospital choice in elective THR surgery, identifying quality of treatment and hospital case volume as the most influential factors.

The findings of the applied experiments have practical implications for healthcare decision-making, suggesting that incorporating patient and physician preferences into composite quality measures can enhance the relevance and acceptance of health report cards and treatment regimens. Furthermore, the comparison of model approaches indicates that the relative importance of attributes is insensitive across different single-class models, but sensitive in latent class models. The dissertation concludes by advocating for further research into the impact of variable types on model outcomes and the integration of preference-based weighting in health policy tools, thereby advancing patient-centred care and informed provider choice.

Acknowledgements

The completion of a doctoral thesis is the result of collective effort and support. First and foremost, I would like to express my deepest gratitude to my supervisor, Martin Emmert, whose guidance, support, and unwavering belief in my abilities have been invaluable throughout this process. His expertise, encouragement, and insightful feedback have shaped both this dissertation and my development as a researcher. Without his mentorship, this work would not have been possible. I would also like to extend my sincere thanks to the members of my dissertation committee, Michael Lauerer and Wilm Quentin, for the constructive comments, thoughtful suggestions, and the time devoted to reviewing this work.

To my colleagues Doris Buchner and Chris Scharrer, thank you for the collaborative and stimulating environment.

Moreover, I would like to acknowledge the participants in the studies, without whom this research would not have been possible. Your willingness to share your experiences and insights has greatly enriched the outcomes of this project.

Special thanks to friends and family who (sometimes unbeknownst) took a great share in support, namely my parents Gudrun and Gigs, my parents in law Monika and Zbyszek, my siblings Christiane and Bele, Kuba and Agnieszka, Kasia and Michal, Maria and Gerald, Ania and Tobi, Florina, Lutz, Philine and Thomas.

Most of all, however, I owe to my family, namely to my dear wife Edyta for her love, patience and believe in me as well as to our lovely children Hania and Janek for their joy with which they enrich my life.

Lichtenfels, 25th of September 2026

Contents

1	Introduction	1
2	Background	8
2.1	Statistical Methods	8
2.1.1	Random Utility Model	8
2.1.2	Logit Models as Standard Estimation Method	10
2.1.3	Independence of Irrelevant Alternatives	14
2.1.4	Probit Models	16
2.1.5	Heteroscedastic Multinomial Logit Model	17
2.1.6	Random Parameters Logit Model	19
2.1.7	Scaled Multinomial Logit Model	20
2.1.8	Generalised Multinomial Logit Model	21
2.1.9	Latent Class Multinomial Logit Model	22
2.2	Practical Implementation of Discrete Choice Experiments	24
2.2.1	Identification of Attributes and Levels	24
2.2.2	Experimental Design	26
2.2.3	Choice Set Design	28
2.2.4	Variable Coding	29
2.2.5	Utility Evaluation for Attribute Levels	33
2.2.6	Model Selection	34
2.2.7	Relative Importance of Attributes	36
3	Preferences of People Living with HIW for Long-Acting Anitretroviral Treatment in Germany: Evidence from a Discrete Choice Experiment	43
4	The Elicitation of Patient and Physician Preferences for Calculating Consumer-Based Composite Measures on Hospital Report Cards: Results of Two Discrete Choice Experiments	61

5	The Elicitation of Patient Preferences for Hip Replacement Surgery: A Discrete Choice Experiment	77
6	Discussion	90
6.1	Comparison of Experimental Designs	90
6.2	Estimation Results	92
6.3	Comparison of Relative Importance of Attributes	99
7	Conclusion	111

List of Figures

2.1	Predicted probability of a binary logit model	12
2.2	Alternative functional forms for the evaluation of attribute levels . . .	35
6.1	Relative importance of attributes for single-class models	105
6.2	Relative importance of attributes for latent-class models with two groups	106

List of Tables

2.1	Example of numerical, dummy, and effect coding with four levels and Level 1 as a benchmark	31
6.1	DCE 1 – Results of single-class regression models	94
6.2	DCE 2 – Results of single-class regression models	95
6.3	DCE 3 – Results of single-class regression models	96
6.4	DCE 4 – Results of single-class regression models	97
6.5	DCE 1 – Results of latent-class model with two classes	100
6.6	DCE 2 – Results of latent-class model with two classes	101
6.7	DCE 3 – Results of latent-class model with two classes	102
6.8	DCE 4 – Results of latent-class model with two classes	103

List of Publications

- Emmert, M., S. Rohrbacher, J. Jahn, K. Fernando, and M. Lauerer (2023). Preferences of people living with HIV for long-acting antiretroviral treatment in Germany: Evidence from a discrete choice experiment. *The Patient – Patient-Centered Outcomes Research* 16(5), 537–553.
- Emmert, M., S. Rohrbacher, F. Meier, L. Heppe, C. Drach, A. Schindler, U. Sander, C. Patzelt, C. Frömke, O. Schöffski, and M. Lauerer (2024). The elicitation of patient and physician preferences for calculating consumer-based composite measures on hospital report cards: Results of two discrete choice experiments. *The European Journal of Health Economics* 25(6), 1071–1085.
- Rohrbacher, S. and M. Emmert (2025). The elicitation of patient preferences for hip replacement surgery: A discrete choice experiment. *BMC Health Services Research* 25(1), 268.

Author's Contribution to the Essays

This cumulative doctoral thesis bases on three research papers. All essays were written in collaboration with multiple co-authors. This section outlines the settings and describes my individual contribution to each of the three articles.

“Preferences of people living with HIV for long-acting antiretroviral treatment in Germany: Evidence from a discrete choice experiment”

This article is joint work with four co-authors. I prepared the data and performed the statistical analyses. I made a substantial contribution to the interpretation of the results and the derivation of conclusions. I contributed about 25 percent of the original draft of the manuscript. Together with all co-authors, I reviewed and edited the final version.

“The elicitation of patient and physician preferences for calculating consumer-based composite measures on hospital report cards: Results of two discrete choice experiments”

This article is joint work with ten co-authors. I prepared the data and performed the statistical analyses for both experiments. I made a substantial contribution to the interpretation of the results and the derivation of conclusions. I contributed about 20 percent of the original draft of the manuscript. Together with all co-authors, I reviewed and edited the final version.

“The elicitation of patient preferences for hip replacement surgery: A discrete choice experiment”

This article is joint work with one co-author. I designed and directed this study as the principal investigator. I prepared the data and performed all estimations and analyses. Together with the co-author, I interpreted the results and derived the conclusions. I prepared the entire manuscript. Together with the co-author, I reviewed and edited the final version.

1 Introduction

The discipline of health economics is concerned with understanding how scarce resources are allocated in health systems to improve population health. Its fundamental tasks include the identification, measurement, assessment, and comparison of the costs and benefits of potential outcomes of policies or services. In health economics, utility serves as the theoretical and quantitative foundation for analysing how individuals and societies value health outcomes. It represents a preference-based measure of the desirability of different health states, linking economic theory to empirical assessments of quality of life and guiding decisions about resource allocation.

In classical utility theory, preference relations are the standard representation for utility (Jehle and Reny, 2001). Random utility theory extends the classical utility theory by adding a random component to utility. This allows for choice behaviour that may seem non-rational, which is often observed but not covered by classical utility theory. Thus, it provides a realistic, probabilistic framework to analyse and predict actual choice behaviour under uncertainty (Hensher et al., 2015).

The standard estimation technique for random utility theory is discrete choice experiments (DCEs). DCEs make it possible to estimate indirect utility functions, allowing one to derive utility values consistent with observed choices under uncertainty and trade-offs from data on either revealed or stated preferences. While revealed preferences are based on real-world data, stated preferences rely on survey data, with respondents choosing from hypothetical scenarios. There are several advantages of stated choice experiments. Among them are their ability to simulate actual decisions very closely, their ability to include alternatives that do not yet exist, or non-use values, and their easy implementation, whereas real-world data are often not available for the research question at hand (Rao, 2014; Hanley and Czajkowski, 2019).

The methodology of DCEs is rooted in the principles of random utility theory and consumer choice theory, which were developed by McFadden (1973; 1981) and Lancaster (1966), respectively. It is based on the assumptions that alternatives can be described by characteristics and that their utility for individuals depends on the levels of these

attributes (de Bekker-Grob, 2009). Being developed originally for market research, DCEs gained popularity in health economics over time and evolved into a significant approach in the domain (de Bekker-Grob, 2009; de Bekker-Grob et al., 2012; Soekhai et al., 2019; Nouwens et al., 2025). Other fields of application include, e.g. environmental economics (e.g. Lanz and Provins, 2013), transport economics (e.g. Bhat, 1995), and labour market research (e.g. Woodgate et al., 2025).

Discrete choice experiments have emerged as a pivotal methodological approach in health economics for quantifying preferences across a wide spectrum of health care services, interventions, and policies. They allow the systematic analysis of how individuals value different characteristics. This is achieved by presenting participants with hypothetical scenarios involving two or more alternatives, each described by a different combination of characteristics, and asking them to make choices between them. Within this framework, individuals select the alternative that offers the greatest utility. By observing multiple choices across scenarios, researchers can infer the relative utilities of different attributes, thereby gaining a comprehensive understanding of individual preferences. This way, DCEs make it possible to reveal respondents' preferences and derive their utility for single attributes. For most types of estimation models, it is possible to deduce the marginal rate of substitution, which captures the trade-off between attribute levels, and conduct standard welfare analysis (de Bekker-Grob, 2009; Cameron and Trivedi, 2005; Greene, 2012; Eggers et al., 2022). If the model is connected to a monetary cost or travel distance, one can construct respondents' willingness to pay (e.g. Wong et al., 2016; Böhn et al., 2020) or willingness to travel (e.g. Shalowitz et al., 2018).

One advantage of DCEs is that they allow for tasks that are reasonably straightforward and resemble realistic situations for participants (Mangham et al., 2009). Unlike more traditional methods such as rating and ranking, which ask respondents to evaluate alternatives directly based on their overall preferences, DCEs force participants to make trade-offs between competing attributes (Eggers et al., 2022). This way, DCEs are able to mimic real-life choices more accurately compared to rating or ranking, as individuals select among realistic alternatives rather than assigning abstract scores.

This design avoids scale bias and provides stronger inferential validity for behavioural predictions and policy analysis. Furthermore, DCEs yield parameter estimates that are consistent with economic demand theory, enabling formal welfare measures that are not possible through simple ranking exercises.

However, numerous approaches to conducting conjoint analysis exist, and not all of them are suitable for addressing every research question. Without consensus-based methodological standards, the acceptance and reliability of these methods may remain limited (Bridges et al., 2011). Therefore, checklists and guidelines for good research practice (e.g. Lancsar and Louviere, 2008; Bridges et al., 2011; Johnson et al., 2013; Hauber et al., 2016) ensure the validity and credibility of the experiments.

This dissertation deals with the quantification of the preferences of patients and physicians in the health care sector by using DCEs based on stated preferences. Its contribution comprises the elicitation of preferences in the health sector for patients with HIV, for patients before and after total hip replacement (THR), and for physicians related to THR. It is built around three articles published previously: Emmert et al. (2023) study the preferences of patients with HIV for current and novel long-acting antiretroviral therapy; Emmert et al. (2024) compare the preferences of physicians and patients after THR surgery for important hospital characteristics; and Rohrbacher and Emmert (2025) analyse the preferences for hospital properties of patients searching for information on THR treatment. While the first article may help in improving the quality and acceptance of future therapy regimens, the second and third articles may support the development and enhancement of health report cards on THR treatment. Additionally, I analyse the estimation approaches used in the three studies in order to identify superior and inferior modelling strategies.

The following text provides an in-depth exploration of the theoretical foundations, practical implementations, and current challenges related to DCEs in health economics. It is a compilation of the three abovementioned studies applying DCEs in health economics and addresses the following objectives:

1. To provide an overview of the current state of statistical methods and the appro-

priate implementation of experiments for analysing DCEs in health economics.

2. To critically assess the implementation and results of three recent DCE studies in health economics: Emmert et al. (2023), Emmert et al. (2024), and Rohrbacher and Emmert (2025).
3. To compare the sensitivity of relative attribute importance across different model approaches on the basis of these three studies.

This dissertation contains three parts. The first part (Chapter 2) provides an overview of the latest estimation methods for DCEs and current good practice in the experimental design of DCEs. The second part (Chapters 3–5) presents the actual substance of the dissertation: It describes four applications of DCEs developed for real choice problems in health services research based on the publications Emmert et al. (2023), Emmert et al. (2024), and Rohrbacher and Emmert (2025). The third and last part (Chapter 6) presents the critical discussion and comparison of the studies presented in the previous part and draws conclusions concerning the stated objectives.

References

- Bhat, C. R. (1995). A heteroscedastic extreme value model of intercity travel mode choice. *Transportation Research Part B: Methodological* 29(6), 471–483.
- Bridges, J. F., A. B. Hauber, D. Marshall, A. Lloyd, L. A. Prosser, D. A. Regier, F. R. Johnson, and J. Mauskopf (2011). Conjoint analysis applications in health-a checklist: A report of the ISPOR good research practices for conjoint analysis task force. *Value in Health* 14(4), 403–413.
- Bühn, S., J. Holstiege, and D. Pieper (2020). Are patients willing to accept longer travel times to decrease their risk associated with surgical procedures? A systematic review. *BMC Public Health* 20(1), 253.
- Cameron, A. C. and P. K. Trivedi (2005). *Microeconometrics: Methods and Applications*. Cambridge University Press.

- de Bekker-Grob, E. (2009). *Discrete choice experiments in health care: Theory and applications*. Doctoral thesis, Erasmus University Rotterdam.
- de Bekker-Grob, E. W., M. Ryan, and K. Gerard (2012). Discrete choice experiments in health economics: A review of the literature. *Health Economics* 21(2), 145–172.
- Eggers, F., H. Sattler, T. Teichert, and F. Völckner (2022). Choice-based conjoint analysis. In *Handbook of Market Research*, pp. 781–819. Heidelberg: Springer.
- Emmert, M., S. Rohrbacher, J. Jahn, K. Fernando, and M. Lauerer (2023). Preferences of people living with HIV for long-acting antiretroviral treatment in Germany: Evidence from a discrete choice experiment. *The Patient – Patient-Centered Outcomes Research* 16(5), 537–553.
- Emmert, M., S. Rohrbacher, F. Meier, L. Heppe, C. Drach, A. Schindler, U. Sander, C. Patzelt, C. Frömke, O. Schöffski, and M. Lauerer (2024). The elicitation of patient and physician preferences for calculating consumer-based composite measures on hospital report cards: Results of two discrete choice experiments. *The European Journal of Health Economics* 25(6), 1071–1085.
- Greene, W. H. (2012). *Econometric Analysis* (7th ed.). Prentice Hall: Pearson.
- Hanley, N. and M. Czajkowski (2019). The role of stated preference valuation methods in understanding choices and informing policy. *Review of Environmental Economics and Policy* 13(2), 248–266.
- Hauber, A. B., J. M. González, C. G. Groothuis-Oudshoorn, T. Prior, D. A. Marshall, C. Cunningham, M. J. IJzerman, and J. F. Bridges (2016). Statistical methods for the analysis of discrete choice experiments: A report of the ISPOR conjoint analysis good research practices task force. *Value in Health* 19(4), 300–315.
- Hensher, D. A., J. M. Rose, and W. H. Greene (2015). *Applied Choice Analysis* (2nd ed.). Cambridge University Press.
- Jehle, G. A. and P. J. Reny (2001). *Advanced Microeconomic Theory* (2nd ed.). Boston: Addison-Wesley.

- Johnson, F. R., E. Lancsar, D. Marshall, V. Kilambi, A. Mühlbacher, D. A. Regier, B. W. Bresnahan, B. Kanninen, and J. F. Bridges (2013). Constructing experimental designs for discrete-choice experiments: Report of the ISPOR conjoint analysis experimental design good research practices task force. *Value in Health* 16(1), 3–13.
- Lancaster, K. J. (1966). A new approach to consumer theory. *Journal of Political Economy* 74(2), 132–157.
- Lancsar, E. and J. Louviere (2008). Conducting discrete choice experiments to inform healthcare decision making: A user’s guide. *PharmacoEconomics* 26, 661–677.
- Lanz, B. and A. Provins (2013). Valuing local environmental amenity with discrete choice experiments: Spatial scope sensitivity and heterogeneous marginal utility of income. *Environmental and Resource Economics* 56(1), 105–130.
- Mangham, L. J., K. Hanson, and B. McPake (2009). How to do (or not to do)... designing a discrete choice experiment for application in a low-income country. *Health Policy and Planning* 24(2), 151–158.
- McFadden, D. (1973). Conditional logit analysis of qualitative choice behaviour. In P. Zarembka (Ed.), *Frontiers in Econometrics*, pp. 105–142. New York, NY, USA: Academic Press New York.
- McFadden, D. (1981). Econometric models of probabilistic choice. *Structural Analysis of Discrete Data with Econometric Applications* 198272, 198–272.
- Nouwens, S. P. H., S. M. Marceta, M. Bui, D. M. A. H. van Dijk, C. G. M. Groothuis-Oudshoorn, J. Veldwijk, J. A. van Til, and E. W. de Bekker-Grob (2025). The evolving landscape of discrete choice experiments in health economics: A systematic review. *PharmacoEconomics* 43, 879–936.
- Rao, V. R. (2014). *Applied Conjoint Analysis*. Heidelberg: Springer.
- Rohrbacher, S. and M. Emmert (2025). The elicitation of patient preferences for hip replacement surgery: A discrete choice experiment. *BMC Health Services Research* 25(1), 268.

- Shalowitz, D. I., E. Nivasch, R. A. Burger, and M. M. Schapira (2018). Are patients willing to travel for better ovarian cancer care? *Gynecologic Oncology* 148(1), 42–48.
- Soekhai, V., E. W. de Bekker-Grob, A. R. Ellis, and C. M. Vass (2019). Discrete choice experiments in health economics: Past, present and future. *PharmacoEconomics* 37, 201–226.
- Wong, S. F., R. Norman, T. L. Dunning, D. M. Ashley, M. Khasraw, T. M. Hayes, I. Collins, and P. K. Lorgelly (2016). A discrete choice experiment to examine the preferences of patients with cancer and their willingness to pay for different types of health care appointments. *Journal of the National Comprehensive Cancer Network* 14(3), 311–319.
- Woodgate, R. L., C. A. Isaak, J. Witt, P. Tennent, and A. Bell (2025). The employment preferences of young people in Canada: A discrete choice experiment. *BMC Public Health* 25(1), 715.

2 Background

This section surveys recent methodological advancements in the field of DCEs. Section 2.1 systematically reviews contemporary statistical techniques employed in DCE analysis. Section 2.2 examines the current state of practical implementation and experimental design of DCEs.

2.1 Statistical Methods

This part reviews the foundations of various modelling approaches for estimating DCEs. It begins by presenting the random utility model as the theoretical basis for eliciting preferences by using DCEs, and introduces logit models as the primary estimation method for DCEs. Subsequently, the underlying model assumptions and associated issues are discussed, and a range of advanced modelling approaches are presented. These approaches allow for the relaxation or modification of the restrictive assumptions of traditional logit specifications.

2.1.1 Random Utility Model

The theoretical fundament for linking discrete choice experiments to a decision-maker's choice is the random utility model (RUM) (McFadden, 1981; Walker and Ben-Akiva, 2002). The RUM allows a consumer to choose between alternatives according to the highest satisfaction or utility (Cameron and Trivedi, 2005; Greene, 2012). The assumption is that specific characteristics of an alternative – so-called attributes – determine the individual utility evaluation, which forms the basis for the observable choice behaviour (Eggers et al., 2022). Starting with the binary case of two alternatives, the RUM specifies the utility U_{ij} of alternatives $j \in \{0, 1\}$ for individual i as

$$\begin{aligned} U_{i0} &= V_{i0} + \varepsilon_{i0}, \\ U_{i1} &= V_{i1} + \varepsilon_{i1}, \end{aligned} \tag{2.1}$$

where V_{ij} comprises the deterministic components of utility, while ε_{ij} describes the random component of utility (McFadden, 1974; Eggers et al., 2022). The deterministic components may comprise choice-specific attributes, as well as individual-specific features like gender or age, whereas the random components represent the stochastic elements of the individuals that are unknown to the observer (Greene, 2012). The consumer chooses the alternative that provides a higher utility. Denote the consumer's choice of alternative 1 as $y_i = 1$; this implies $U_{i1} > U_{i0}$. Due to the random components of utility, the choice is a random event with

$$\begin{aligned}
\Pr[y_i = 1] &= \Pr[U_{i1} > U_{i0}] \\
&= \Pr[V_{i1} + \varepsilon_{i1} > V_{i0} + \varepsilon_{i0}] \\
&= \Pr[V_{i1} - V_{i0} > \varepsilon_{i0} - \varepsilon_{i1}] \\
&= F(V_{i1} - V_{i0}),
\end{aligned} \tag{2.2}$$

where $F(\cdot)$ is the cumulative distribution function of $(\varepsilon_{i0} - \varepsilon_{i1})$ and $(V_{i1} - V_{i0})$ collects all observable differences between both alternatives (Cameron and Trivedi, 2005; Greene, 2012). Hence, the consumer is more likely to choose the alternative with the higher utility, which requires that the difference among deterministic components exceeds the residual of errors (Eggers et al., 2022). Note that $(V_{i1} - V_{i0})$ reduces to the differences between choice-specific features, because individual-specific features are constant for the choice among two alternatives (Greene, 2012; Eggers et al., 2022).

The RUM is easily generalised for more than two alternatives. In this case, it becomes a multinomial model of utility with a choice between m alternatives, where the utility of alternative j for individual i is described as

$$U_{ij} = V_{ij} + \varepsilon_{ij}, \quad j = 1, \dots, m. \tag{2.3}$$

Now, each alternative is compared against all other alternatives. As before, the RUM assumes that the alternative with the highest utility is chosen. Therefore, equation (2.2)

for m alternatives turns into

$$\begin{aligned}
\Pr[y_i = j] &= \Pr[U_{ij} > U_{ik}, \quad \text{all } k \neq j] \\
&= \Pr[U_{ij} - U_{ik} > 0, \quad \text{all } k \neq j] \\
&= \Pr[V_{ij} - V_{ik} > \varepsilon_{ik} - \varepsilon_{ij}, \quad \text{all } k \neq j].
\end{aligned} \tag{2.4}$$

Also, in the multinomial case, individual i chooses alternative j over all alternatives k if the utility of j is highest, which requires a positive residual from $(V_{ij} - V_{ik})$ that is larger than the differences between errors $(\varepsilon_{ik} - \varepsilon_{ij})$ (Cameron and Trivedi, 2005; Train, 2009; Greene, 2012).

The deterministic components of utility comprise the impact of the alternative's attributes and corresponding levels translated into partworth utilities. In its general form, this function can be written as

$$V_{ij} = \Psi_i[f_{i1}(v_{ij1}), f_{i2}(v_{ij2}), \dots, f_{in}(v_{ijn})], \tag{2.5}$$

where v_{ija} stands for the partworth utility of attribute $a = 1, 2, \dots, n$ of alternative $j = 1, 2, \dots, m$ for individual i , f_{ia} represents the evaluation function of attribute a , and Ψ_i is the linkage function to combine partworth utilities across attributes. The evaluation function f_{ia} characterises the isolated assessment of the attribute, whereas the linkage function Ψ_i defines how to combine the partworth utilities of all attributes (Teichert, 2001). In the most simple way, Ψ_i may manifest in additive form such that

$$V_{ij} = f_{i1}(v_{ij1}) + f_{i2}(v_{ij2}) + \dots + f_{in}(v_{ijn}), \tag{2.6}$$

as proposed by, e.g. Rao (2014) or Eggers et al. (2022). Here the deterministic component of utility is simply the sum of all underlying partworth utilities.

2.1.2 Logit Models as Standard Estimation Method

The RUM requires estimation methods suitable for the discrete outcome of the dependent variable. In this class, the logit model and its variations happened to become the

workhorse for estimating the RUM (de Bekker-Grob et al., 2012; Soekhai et al., 2019; Nouwens et al., 2025), due to conceptional and computational advantages (Cameron and Trivedi, 2005; Greene, 2012).

Let us start once again with the binary case. The logit model arises if both ε_{i0} and ε_{i1} are assumed to be independently identically distributed (i.i.d.) with a type 1 extreme value distribution (EV1) (McFadden, 1973). In this case, the difference between the random variables, $\varepsilon_i = (\varepsilon_{i0} - \varepsilon_{i1})$, has a density of

$$f(\varepsilon_i) = e^{-\varepsilon_i} \exp(-e^{-\varepsilon_i}), \quad -\infty < \varepsilon_i < \infty, \quad (2.7)$$

and the corresponding cumulated distribution function (cdf) is $F(\varepsilon_i) = \exp(-e^{-\varepsilon_i})$. Given this, the difference $(\varepsilon_0 - \varepsilon_1)$ can be shown to be logistic-distributed such that $F(\cdot)$ in equation (2.2) becomes the logistic cumulated distribution function (Cameron and Trivedi, 2005).¹ Accordingly, for $(V_1 - V_0) = \mathbf{x}'_i \beta$, equation (2.2) takes the form of the logit model, which reads as

$$\Pr[y_i = 1] = F(\mathbf{x}'_i \beta) = \frac{e^{\mathbf{x}'_i \beta}}{1 + e^{\mathbf{x}'_i \beta}}, \quad (2.8)$$

where $\mathbf{x}_i$ is a vector of explanatory variables of individual i and β is the vector with the corresponding estimation parameters. The assumed EV1 distribution is right-skewed over $(-\infty, \infty)$, with most mass between -2 and 5. In its standard form without an additional scaling parameter², it has a non-zero mean $E(\varepsilon) = 0.57221$ and variance $\text{Var}(\varepsilon) = \frac{\pi^2}{6}$. Compared to the normal distribution, it has less weight in the tails (Cameron and Trivedi, 2005; Hensher et al., 2015). The difference between two EV1-distributed random variables, as described in equations (2.2) and (2.4), has a logistic distribution, which is symmetric, like the normal distribution, and thus is considered an appropriate approximation of the normal distribution (Greene, 2012).

¹Cameron and Trivedi (2005) also provide a detailed derivation of this result.

²Hensher et al. (2015) discuss the impact of various scale parameters on the distribution.

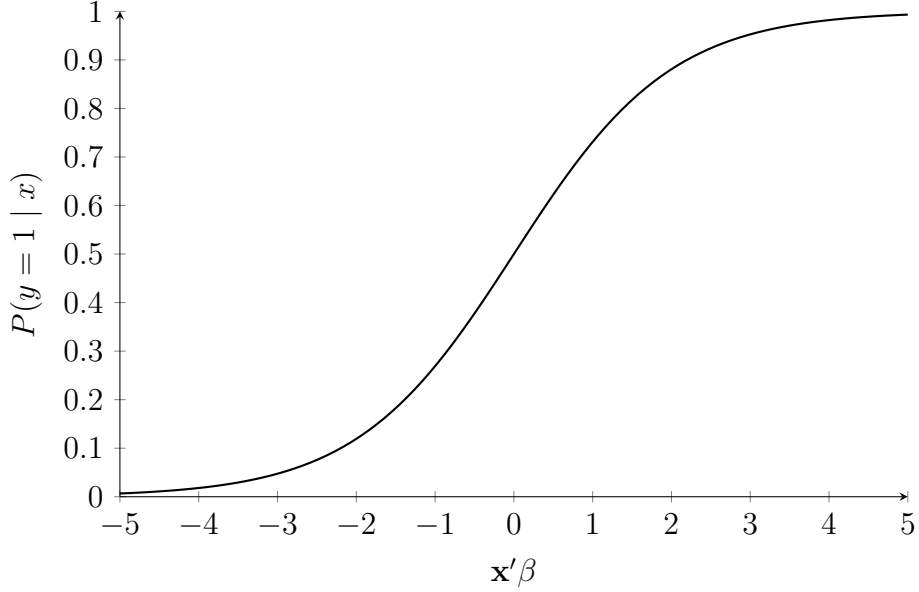


Figure 2.1: Predicted probability of a binary logit model

Although the assumption of EV1 errors may seem to be rather unusual at first, the logit model comprises several advantages. First of all, unlike linear probability models it takes the discreteness of the dependent variable into account. Moreover, it ensures that the predicted probabilities remain between zero and one, as can be seen from Figure 2.1 (Cameron and Trivedi, 2005; Greene, 2012). Due to its closed form, the estimation is easily implemented with maximum likelihood methods (Cameron and Trivedi, 2005; Greene, 2012). Therefore, the logit model is a good representation for the binary RUM.

As discussed above, it is possible to derive the multinomial logit (MNL) model from the multinomial RUM with plausible and interpretable assumptions about the error terms (Croissant, 2020). Consider once again i.i.d. EV1 error components for all options such that

$$f(\varepsilon_{ij}) = e^{-\varepsilon_{ij}} \exp(-e^{-\varepsilon_{ij}}), \quad -\infty < \varepsilon_{ij} < \infty, \quad (2.9)$$

where ε_{ij} is the random component of individual i choosing alternative j . Given this,

equation (2.4) can be evaluated as

$$\Pr[y_i = j] = \frac{e^{V_{ij}}}{e^{V_1} + e^{V_2} + \dots + e^{V_m}}, \quad (2.10)$$

where $V_{ij} = \mathbf{x}'_{ij}\beta$.³ This corresponds to an extension of the binary logit model to the multinomial case, which estimates the probability that individual i chooses alternative j among m alternatives. The properties of the binary model also apply for the multinomial model. Due to the closed form, the estimation of the MNL model remains tractable through maximum likelihood estimation and does not require computationally demanding simulation methods like most alternative model approaches. Moreover, the discreteness and range of the dependent variable are respected and the model is consistent with the utility-maximising behaviour of the RUM (Cameron and Trivedi, 2005; Greene, 2012; Hensher et al., 2015). Therefore, also in the multinomial case, the logit model is a good representation of the RUM.

Thus far, only alternative-varying regressors are assumed for the sake of simplicity, such that

$$V_{ij} = \mathbf{x}'_{ij}\beta \quad (2.11)$$

in equation (2.10). In this case, $\mathbf{x}_{ij}$ takes different values for different alternatives j and the estimated parameter β is constant across alternatives. However, with no loss of generality it is possible to implement alternative-invariant regressors such as, e.g. socioeconomic characteristics. In this case, $\mathbf{x}_i$ does not change across alternatives, whereas now the parameters β_j vary with alternative j . The extreme case of solely alternative-invariant regressors would give

$$V_{ij} = \mathbf{x}'_i\beta_j. \quad (2.12)$$

Finally, it is also possible to combine alternative-varying and alternative-invariant regressors such that

$$V_{ij} = \mathbf{x}'_{ij}\beta + \mathbf{x}'_i\gamma_j \quad (2.13)$$

³For a detailed derivation of this result, see, e.g. Croissant (2020).

in equation (2.10), where β is the estimate of the alternative-varying variables $\mathbf{x}_{ij}$ and γ_j is the estimate of the alternative-invariant variables $\mathbf{x}_i$ (Cameron and Trivedi, 2005; Greene, 2012).⁴

To summarise, the logit model is a useful approach due to mathematical convenience. Its major advantage is that it keeps its closed form even in the multinomial case. Therefore, estimation remains tractable within maximum likelihood estimation methods. Other approaches discussed in the following chapter cannot be solved in closed form, meaning that their estimation relies on computationally more complex solution methods. Therefore, the logit model is an obvious choice to use to estimate the RUM and a good starting point; as a result, it finds widespread application among health-related discrete choice experiments (de Bekker-Grob et al., 2012; Soekhai et al., 2019; Nouwens et al., 2025). However, the closed form arises from the rather strong assumption of i.i.d. EV1 error components. The postulation of independent errors is quite restrictive, as it may be easily violated if two alternatives are similar. This violation is also known as the assumption of independence of irrelevant alternatives or the red bus-blue bus problem (Cameron and Trivedi, 2005). The following subsections discuss the relaxation of these restrictions and the corresponding models.

2.1.3 Independence of Irrelevant Alternatives

The MNL model is a reasonable starting point, but it may be too restrictive due to the strong assumption that unobserved components are independent and identically distributed and follow a type 1 extreme value distribution (see, e.g. Cameron and Trivedi, 2005; Greene, 2012; Hensher et al., 2015). This assumption implies that random com-

⁴Please note that the naming of the models is inconsistent in the literature due to historical reasons. Sometimes, the multinomial logit model is considered only with alternative-invariant regressors, whereas the model with only alternative-varying regressors is called the conditional logit model; meanwhile, the model with both kinds of regressors is called the mixed logit model (Cameron and Trivedi, 2005; Greene, 2012). However, to stay in line with the recent literature in health economics and to avoid ambiguities, I call all three kinds of models multinomial logit models in this text despite being aware of this discussion.

ponents are uncorrelated and of equal size across alternatives (Hensher et al., 2015). The hypothesis of such errors results in the independence of irrelevant alternatives (IIA) hypothesis.

The IIA may lead to implausible behaviour from the MNL model. For illustration, let us consider two alternatives p and q out of m possible alternatives. Given this, equation (2.10) implies

$$\Pr[y_i = q] = \frac{e^{V_{iq}}}{\sum_{k=1}^m e^{V_k}}, \quad (2.14)$$

$$\Pr[y_i = p] = \frac{e^{V_{ip}}}{\sum_{k=1}^m e^{V_k}}, \quad (2.15)$$

for both alternatives. The ratio of the probabilities of the alternatives is then

$$\frac{\Pr[y_i = q]}{\Pr[y_i = p]} = \frac{e^{V_{iq}}}{e^{V_{ip}}} = e^{V_q - V_p}. \quad (2.16)$$

Here, the probability ratio depends only on the characteristics of both alternatives but is independent of the probabilities of the remaining alternatives (Greene, 2012; Croissant, 2020). As an extreme example, let us consider the travelling options of a car and a red bus with initially equal probabilities of 0.5. Let us now introduce a blue bus, which is identical to the red bus apart from the colour. Under the IIA hypothesis and continuing with equal likelihood, the probabilities of each alternative would now be 0.33 each, implying that taking the bus (independent of colour) would now have a higher probability of 0.67, whereas the probability of taking the car would shrink to 0.33. In reality, one would expect that the probability for using the car would remain at 0.5, whereas the probability for taking the bus would split up between blue and red such that the cumulated sum of probabilities for taking any bus would be 0.5 (Cameron and Trivedi, 2005). This weakness usually appears in settings with three or more alternatives.

The IIA itself is not a problem, but it may be violated in applied data (Croissant, 2020). However, different models can be generated by different assumptions about

the joint distribution of the errors. This helps to create models that circumvent the IIA hypothesis (Cameron and Trivedi, 2005). Furthermore, the MNL model implicitly assumes that individuals and their choices are homogeneous. However, preference may be structurally heterogeneous across individuals or groups. Therefore, the following subsections present various approaches to relax the assumptions of the MNL model to correct for the IIA hypothesis and allow unobserved preference heterogeneity. Please note that the discussion is limited to models mainly applied in health-related studies (de Bekker-Grob et al., 2012; Soekhai et al., 2019; Nouwens et al., 2025). For a more detailed (though not complete) list, see, e.g. Hensher et al. (2015).

2.1.4 Probit Models

The previous subsection presented the logit model as a standard estimation approach for the RUM. Logit models rely on the assumption that errors follow a type 1 extreme value distribution. At first glance, an obvious alternative would be the probit model, where the errors are independent and normally distributed. For the binary case, with the normalisation of the variance of $(\varepsilon_{i0} - \varepsilon_{i1})$ to unity, the cdf $F(\cdot)$ in equation (2.2) becomes the standard normal cdf with $\varepsilon_i \sim N[0, 1]$. Inserting the standard normal distribution function into $F(\cdot)$ of equation (2.2) then gives us the binary probit model

$$\Pr[y_i = 1] = F(\mathbf{x}_i' \beta) = \int_{-\infty}^{\mathbf{x}_i' \beta} \phi(z) dz, \quad (2.17)$$

where $\phi(\cdot)$ is the standard density function (Cameron and Trivedi, 2005; Greene, 2012). In the binary case, the probit model can be solved by maximum likelihood estimation just like the logit model. Empirically, there is little difference between the models (Cameron and Trivedi, 2005).

In the multinomial case, the cdf $F(\cdot)$ becomes the multivariate normal cdf in equation (2.4) with zero mean and a covariance matrix Σ such that

$$\varepsilon_i \sim N[0, \Sigma]. \quad (2.18)$$

Here, Σ is in general not restricted to be diagonal, as it is in the case of the MNL model. Hence, it can circumvent the problem of the IIA. However, practical obstacles appear. Unlike in the bivariate case, in the multinomial case the estimation of Σ is unfortunately cumbersome. Unlike the MNL model, the multinomial probit (MNP) model lacks a closed-form solution for the probability of choosing alternative j ($\Pr[y_i = j]$), even if all errors are assumed to be uncorrelated (Cameron and Trivedi, 2005; Greene, 2012). Therefore, there is no analytical solution, which causes computational difficulties for the implementation of the MNP model. Due to the multivariate normal probabilities, models with more than two alternatives require simulation-based methods such as Monte Carlo simulations, which are computationally intensive, time consuming, and only approximative. Correct estimation for corner solutions of p_{ij} is especially challenging, even with modern computational capacities (Cameron and Trivedi, 2005; Greene, 2012; Hensher et al., 2015). Additionally, restrictions on Σ are necessary for identification, as different MNP models arise from different specifications of Σ (Cameron and Trivedi, 2005; Greene, 2012; Hensher et al., 2015). Successful implementation may rely on further restrictions on Σ and β ; otherwise, the estimation may be highly imprecise (Cameron and Trivedi, 2005).

To summarise, solving the MNP model becomes complex in the multinomial case. Identification remains a serious problem within the MNP model (Greene, 2012). Given that in the binary case logit and probit models provide in general similar results, one may ask if it is worth counting on the assumption of a standard normal distribution in the multinomial case or if one should instead adopt EV1 errors and use the reliable and computationally less burdensome MNL model or avoid the IIA with other model specifications (Cameron and Trivedi, 2005).

2.1.5 Heteroscedastic Multinomial Logit Model

Another way to relax the IIA hypothesis is to allow for heteroscedasticity among error components, as suggested, e.g. by Bhat (1995). This can be achieved by assuming that the random components have a type 1 extreme value distribution and are independent,

but not identically distributed. As a result, the alternative-specific scale factor θ_j enters the density function, such that it is now written as

$$f(\varepsilon_{ij}) = \frac{1}{\theta_j} e^{-\frac{\varepsilon_{ij}}{\theta_j}} \exp(-e^{-\frac{\varepsilon_{ij}}{\theta_j}}), \quad -\infty < \varepsilon_{ij} < \infty. \quad (2.19)$$

The corresponding variance of any alternative now depends on its scaling factor θ_j , such that $\text{Var}(\varepsilon_{ij}) = \frac{\pi^2 \theta_j^2}{6}$. Thus, the variance may vary across alternatives. Given this, θ_j would enter the RUM in equation (2.4), such that

$$U_{ij} = V_{ij} + \frac{\varepsilon_{ij}}{\theta_j}. \quad (2.20)$$

With no loss of generality, the previous equation rewrites as

$$\tilde{U}_{ij} = \theta_j V_{ij} + \varepsilon_{ij}, \quad (2.21)$$

where the unobservable $\tilde{U}_{ij} = \theta_j U_{ij}$ can be treated just like U_{ij} in equation (2.4), since the relation between U_{ij} and V_{ij} remains the same. Therefore, it is straightforward to implement the assumption of heteroscedastic errors within the theoretical framework of the RUM. The resulting heteroscedastic MNL model is usually called the heteroscedastic extreme value (HEV) model due to its assumptions about the extreme value distribution. The HEV model collapses to the MNL model with $\theta_j = 1, \forall j$ (Bhat, 1995; Hensher et al., 2015).

Although the HEV model lacks a closed form, it is parsimonious compared to the MNP model, so that the computational burden is notably lower and a robust approximation is feasible (Bhat, 1995; Hensher et al., 1998; Croissant, 2020). It nests the more restrictive MNL model and allows for differential cross-elasticities between alternatives, which relaxes the IIA assumption similarly to the MNP model (Hensher et al., 1998). It avoids the drawbacks of the IIA properties by allowing for differences in the scaling parameter θ_j . Thus, the error components can capture unobserved alternative-specific characteristics (Bhat, 1995). It defines the relative weights of the deterministic and

random components in estimating the choice probability, as can be seen in equation (2.21). Thus, a change in the systemic component of one alternative affects the differential between this alternative and another such that the ratio of probabilities depends on all scaling parameters θ_j (Bhat, 1995; Hensher et al., 1998; Greene, 2012). As in the MNL model, in the HEV model the probabilities for all alternatives sum to one (Bhat, 1995).

2.1.6 Random Parameters Logit Model

Another widely used variant is the random parameters logit (RPL) model. The RPL model differs from the MNL model in that it assumes that at least some coefficients contain a random component across individuals (Cameron and Trivedi, 2005; Train, 2009; Fiebig et al., 2010; Greene, 2012; Hensher et al., 2015). This implies individual-specific coefficients

$$\beta_i = \beta + \eta_i, \quad (2.22)$$

which consist of the mean attribute utility weights in the population β and an individual-specific random effect η_i . This way, the RPL model allows for unobserved heterogeneity among the population (Croissant, 2020). The individual parameter β_i is now part of the deterministic component

$$V_{ij} = \mathbf{x}_{ij}'\beta_i, \quad (2.23)$$

and the RUM accordingly comprises the random components such that

$$U_{ij} = (\beta + \eta_i)x_{ij} + \varepsilon_{ij}. \quad (2.24)$$

The introduction of random parameters has the appealing property of inducing correlation across alternatives (Cameron and Trivedi, 2005; Fiebig et al., 2010). This relaxes the IIA hypothesis while maintaining the properties of the EV1 distribution.

Although in most applications η_i is assumed to be multivariate-normal-distributed, it is not restricted to normality, meaning that other distributions may find appropriate usage (Fiebig et al., 2010; Greene, 2012; Sarrias and Daziano, 2017). At the same time, ε_{ij} remains i.i.d. EV1-distributed (Fiebig et al., 2010). Hence, for $\eta_i = 0 \forall i$, the RPL model reduces to the MNL model (Hensher et al., 1998).

As for the MNP model, there is no closed form for the RPL model. However, there does exist a closed form conditional on the individual-specific component η_i (Cameron and Trivedi, 2005). Therefore, simulation methods are required but approximation is less cumbersome for the RPL model compared to the MNP model. Thus, ensuring model flexibility seems to be easier using the RPL model compared to the MNP model (Cameron and Trivedi, 2005; Fiebig et al., 2010).

2.1.7 Scaled Multinomial Logit Model

Alternatively, rather than using the RPL model, one can extend the MNL model with a scaling factor, which creates the scaled MNL (SMNL) model. The SMNL model is written as

$$U_{ij} = \sigma_i V_{ij} + \varepsilon_{ij}. \quad (2.25)$$

Here, unlike the HEV model, the scaling factor σ_i captures individual-specific heterogeneity, just like in the RPL model, and not alternative-specific heterogeneity, as in the HEV model. The scaling factor allows for the variance of the error across individuals, reflecting that some people make choices more decisively, while for others choice depends more on the random component. Since σ_i is an estimated weight and does not depend on a random process, the SMNL model provides a much more parsimonious form compared to the RPL model yet allows for heterogeneity among the population, which relaxes the IIA. For computational convenience, σ_i is often assumed to be log-normal-distributed (Fiebig et al., 2010; Hensher et al., 2015). With equal weights $\sigma_i = 1 \forall i$, the SMNL model collapses to the MNL model.

Note that scale heterogeneity is equivalent to a particular type of heterogeneity in the utility weights. Random coefficients and scale heterogeneity are thus only two different ways of specifying the distribution of coefficient heterogeneity. Consequently, one can think of the SMNL model as a simplified special case of the RPL model (Fiebig et al., 2010). Louviere et al. (2008) argue that most of the heterogeneity in discrete choice models is captured better by the SMNL model than by the RPL model. Therefore, in many cases it may be sufficient to rely on the more parsimonious and computationally more efficient SMNL model (Fiebig et al., 2010).

2.1.8 Generalised Multinomial Logit Model

In an approach to harmonise the manifold models, Fiebig et al. (2010) suggest nesting the RPL model and the SMNL model within a generalised MNL (GMNL) model. This makes it possible to capture various kinds of heterogeneity simultaneously within one model. The RUM extends to

$$U_{ij} = [\sigma_i \beta + \gamma \eta_i + (1 - \gamma) \sigma_i \eta_i] x_{ij} + \varepsilon_{ij}, \quad (2.26)$$

where σ_i captures scale heterogeneity, as in the SMNL model; η_i captures individual taste heterogeneity, as in the RPL model; and γ is a weighting parameter indicating how the variance in unobserved preference heterogeneity varies with scale. The GMNL model nests various models, namely, the RPL, SMNL, MNL, GMNL1, and GMNL2 models: If $\sigma_i = 1 \forall i$, the GMNL model reduces to the RPL model; if $\eta_i = 0 \forall i$ (or $Var(\eta_i) = 0$), the GMNL model reduces to the SMNL model; and if $\sigma_i = 1 \forall i$ and $\eta_i = 0 \forall i$, the GMNL model reduces to the MNL model. Furthermore, one can derive two extreme cases of the GMNL model, where the residual taste heterogeneity η_i either varies with scale σ_i (GMNL1) or is independent of it (GMNL2). This is governed by the parameter γ : $\gamma \rightarrow 1$ leads to GMNL1; $\gamma \rightarrow 0$, leads to GMNL2 (Fiebig et al., 2010; Hensher et al., 2015).

The advantage of the GMNL model is its combined presentation of scaled heterogeneity and individual taste heterogeneity. This way, it is possible to utilise a very flexible dis-

tribution of unobserved heterogeneity (Fiebig et al., 2010; Greene, 2012; Hensher et al., 2015). It also makes it possible to disentangle heterogeneity in individual coefficients from scale heterogeneity (Hess and Rose, 2012).

As in the case of the models discussed above, the computation of the GMNL model requires simulation-based methods.⁵ The generalisation of the GMNL model comes with more flexibility but also with a higher degree of complexity. Due to the distribution flexibility, estimated distributions of unobserved heterogeneity may not be concave, meaning that the estimation may converge to local maxima (Sarrias and Daziano, 2017). Therefore, the GMNL model should be used with caution.

2.1.9 Latent Class Multinomial Logit Model

Thus far, I introduced unobserved heterogeneity on the individual level of participants. However, a popular alternative exists in the form of the latent class multinomial logit (LC) model. Just like the RPL model, the LC model is used widely in practice to accommodate preference heterogeneity across respondents (de Bekker-Grob et al., 2012; Soekhai et al., 2019; Nouwens et al., 2025). Unlike the RPL model, which assumes that the random parameter follows a continuous distribution, the LC model assumes a discrete mixing distribution to capture unobserved heterogeneity (Shen, 2009; Hensher et al., 2015; Sarrias and Daziano, 2017). The discrete distribution appears through the assignment of respondents to a predefined number of classes C . The individuals are then allocated to the different classes according to their response behaviour and eventually to socioeconomic factors. For class $c \in C$, the estimates of individual i become class-specific, such that

$$\beta_i = \beta_c \quad \text{with probability } w_{ic} \text{ for } c = 1, \dots, C, \quad (2.27)$$

where $\sum_c w_{ic} = 1$ and $w_{ic} > 0$ holds for the class assignment probabilities w_{ic} . The weights w_{ic} themselves are estimated by a semi-parametric multinomial logit format

⁵For a detailed description of the computational implementation, please refer to Fiebig et al. (2010), Hensher et al. (2015), or Sarrias and Daziano (2017).

such that

$$\Pr[w_{ic} = c] = \frac{e^{\gamma_c}}{\sum_{c=1}^C e^{\gamma_c}}, \quad c = 1, \dots, C, \quad (2.28)$$

where the parameters of the first class γ_1 are set to zero (Shen, 2009; Sarrias and Daziano, 2017). Given this, the estimation problem within one class becomes

$$\Pr[y_i = j | c] = \frac{e^{V_{ij|c}}}{\sum_{k=1}^m e^{V_{k|c}}}, \quad (2.29)$$

where $V_{ij|c} = \mathbf{x}'_{ij}\beta_c$ implies that within class c , the probability of individual i choosing alternative j is equivalent to this probability in the MNL model described in equation (2.10). The difference is that the choice is conditional on class assignment c . The class assignment in equation (2.28) is estimated simultaneously with equation (2.29) (Shen, 2009). This is possible due to the analytical tractability of the LC model, which does not require simulation methods (Sarrias and Daziano, 2017). Instead, the source of heterogeneity enters the model through the a priori choice of the number of classes by the researcher. In consequence, the model is less flexible than the RPL and GMNL models (Sarrias and Daziano, 2017). Additionally, since heterogeneity depends in the LC model solely on the predefined number of classes, the model actually avoids misspecification problems in the distribution of individual heterogeneity (Sarrias and Daziano, 2017). Shen (2009) summarises the advantages of the LC model, stating that it is ‘relatively simple, reasonably plausible and statistically testable’. Its disadvantages are that the researcher may choose the number of classes ad hoc and that it is less flexible than the RPL and GMNL models (Shen, 2009; Hensher et al., 2015; Sarrias and Daziano, 2017).

Recently, several studies have attempted to combine the RPL model and the LC model to capture both individual- and group-specific heterogeneity (Bujosa et al., 2010; Keane and Wasi, 2013; Greene and Hensher, 2013). Therein, random individual parameters may appear within each class. This way, it is possible to implement a considerable extension of flexibility while computation remains relatively manageable (Greene and

Hensher, 2013). The advantage of this approach is that it has the most flexible heterogeneity distribution among the considered models, which is beneficial, especially in datasets with a very complex structure of heterogeneity (Greene and Hensher, 2013; Keane and Wasi, 2013). This allows the simultaneous estimation on the one hand of segments of consumers with almost lexicographic behaviour and on the other hand of segments of consumers whose choice is influenced so little by observed attributes that it seems random. However, as in the standard LC model, here too the selection of classes is done ad hoc by the researcher. This implies the need for intensive testing for the appropriate number of classes. Moreover, the model with combined unobserved heterogeneity quickly becomes quite complex, such that parameters proliferate rapidly with additional types. This leads to difficulties in interpretation, and solving the model becomes computationally demanding (Keane and Wasi, 2013). These disadvantages may be the reason that this model is rarely used in studies on health-related preferences (de Bekker-Grob et al., 2012; Soekhai et al., 2019; Nouwens et al., 2025).

2.2 Practical Implementation of Discrete Choice Experiments

Having established the statistical framework for discrete choice experiments (DCEs), this section proceeds to outline the key steps involved in their applied implementation. The discussion begins with a systematic approach to the identification and selection of attributes and levels, ensuring their relevance and appropriateness for the research context. Subsequently, the section addresses the principles of experimental and choice task design, followed by an examination of variable coding and the utility evaluation of attribute levels. The section concludes with guidance on the selection of appropriate model specifications and the measurement of attribute importance, with a view to ensuring robust and policy-relevant results.

2.2.1 Identification of Attributes and Levels

To begin with, it is necessary to identify the relevant determinants for consumers' preferences. This is one of the most important issues, as otherwise the validity of the

whole study can be questioned (Eggers et al., 2022).

Several requirements exist that attributes should fulfil (Green and Srinivasan, 1978, 1990; Orme, 2002; Rao, 2014): First, attributes should affect responders' utility. This can be achieved by identifying relevant attributes through previous qualitative surveys such as, e.g. focus groups or interviews, or based on reliable information retrieved from, e.g. literature reviews or existing health outcome measures (Bridges et al., 2011; Coast et al., 2012; Helter and Boehler, 2016; Eggers et al., 2022). Second, attributes should be chosen such that they discriminate and differentiate between the reviewed alternatives. This enables to identify valid preference patterns (Doherty et al., 2013; Helter and Boehler, 2016). Third, attributes should be independent and not inter-related. If attributes are correlated, certain combinations may be more likely or less realistic. Moreover, two attributes may capture the same effect, such that it is difficult to disentangle the impact on each. This may undermine the validity of the research (Eggers et al., 2022). Fourth, the number of attributes should be manageable for the participants in order to keep the complexity of the experiment low. Otherwise, respondents may lose focus or answer inconsistently due to confusion, causing the data quality of answers to suffer (Kessels et al., 2011; Coast et al., 2012).

After all relevant attributes are identified, it is necessary to determine the corresponding attribute levels. Levels should be precise statements with unambiguous meaning, and clear wording referring to specific values is preferred over vague formulations (Bridges et al., 2011; Helter and Boehler, 2016; Eggers et al., 2022). Next, levels should be generally acceptable. If levels are considered as unrealistic or unacceptable, respondents may answer randomly and in a biased way, which may affect the validity and credibility of the survey (Eggers et al., 2022). To avoid ambiguous or inconsistent answers, levels should be mutually exclusive (Orme, 2002; Coast et al., 2012). Finally, the level span should be moderately larger than it is in reality in order to maximise sensitivity to preferences while remaining sufficiently realistic (Eggers et al., 2022).

With regard to the number of attributes, several recommendations should be taken into account. First, the number of levels should remain limited in order to keep the

complexity low (Eggers et al., 2022). For example, Bridges et al. (2011) suggest limiting the number of levels to 3–4 levels, as otherwise the risk of cognitive overload and survey fatigue among respondents increases. Second, the number of levels should be balanced across attributes. Otherwise, the ‘number-of-levels effect’ may arise, which implies an artificially higher relevance for attributes with more levels (Eggers and Sattler, 2009). Third, when setting the number of attributes, one should also consider the linearity or non-linearity of the utility function (Lancsar and Louviere, 2008; Eggers et al., 2022).

2.2.2 Experimental Design

The experimental design is the main mechanism used to determine the combinations of attribute levels that participants are exposed to (Eggers et al., 2022). The experimental design contains the factorial design, i.e. how combinations of attribute levels are allocated in the survey, and the choice design, i.e. how the stimuli are distributed across choice sets. Several aspects affect the optimal design concerning factorial design and choice design. The major criteria to evaluate the efficiency of an experimental design are level balance and the orthogonality of levels with respect to factorial design, as well as minimal overlap and utility balance, which are related to choice design (Huber and Zwerina, 1996; Rao, 2014; Eggers et al., 2022).

Level balance ensures that all attribute levels appear equally across all choice sets such that each level is represented with similar frequency. This helps to avoid level bias due to levels being over- or under-represented, which improves the reliability of parameter estimations (Huber and Zwerina, 1996; Rao, 2014; Eggers et al., 2022). Orthogonality implies that attribute levels are uncorrelated, providing independence between attribute levels in the choice sets. An orthogonal design ensures that the effect of each attribute level can be estimated without confounding effects from other attributes (Huber and Zwerina, 1996; Rao, 2014; Eggers et al., 2022).

Combining all potential stimuli results in a full factorial design. Full factorials are by definition balanced and orthogonal (Eggers et al., 2022). However, full factorial designs quickly become very difficult to handle for participants, as the corresponding

survey size grows exponentially with every additional level. To address this, one may reduce the design to smaller factorial designs that still allow the estimation of the main effects while reducing the burden on respondents, using designs such as, e.g. fractional factorial designs (Rao, 2014).

Unfortunately, level balance and orthogonality often conflict in creating an efficient factorial design (Huber and Zwerina, 1996). Therefore, it is necessary to evaluate the efficiency of a design. In order to assess the efficiency of a design, one can rely on D-efficiency as a measure (Bridges et al., 2011). A D-efficient design minimises the asymptotic variance-covariance matrix of the estimated parameters, which corresponds to more precise preference estimates for a given sample size. Thus, higher D-efficiency values indicate statistically more powerful designs (Bridges et al., 2011; Rao, 2014).

Minimal overlap aims at organising choice sets in a way that ensures distinct attribute information within choice sets. It requires maximal distance between alternatives, i.e. attributes that have different levels (Huber and Zwerina, 1996; Eggers et al., 2022). Otherwise, differences of attributes within a task may be too small to carry meaningful information about the value of attributes (Rao, 2014). Hence, among the available options, equal attribute levels across alternatives should be avoided.

Utility balance means that alternatives should appear equally attractive such that dominant or dominated alternatives are avoided (Huber and Zwerina, 1996; Bridges et al., 2011; Rao, 2014; Eggers et al., 2022). If a choice set contains unbalanced alternatives, it may lose any new insight, as choice may become predictable. However, identifying dominating and dominated attributes requires a priori knowledge about the respondents' preferences (Eggers et al., 2022). Appropriate utility balancing may increase the efficiency of a design by 10–50% (Huber and Zwerina, 1996). Today, standard software, such as, e.g. Sawtooth, executes the complex process of creating optimal choice design and provides a direct analysis of the design's efficiency (Eggers et al., 2022).

2.2.3 Choice Set Design

The presentation of choice tasks to participants affects the performance of any survey. It is possible to vary the tasks in multiple ways by altering the number of alternatives, attributes, or choice tasks (Bridges et al., 2011). First, it is necessary to address how many alternatives are included in one choice set (Rao, 2014). The answer to this question depends on the trade-off between increasing information from more alternatives and the corresponding increase in complexity (Eggers et al., 2022). However, several arguments indicate that the number of alternatives should not be set too high. First, each choice task should be manageable and avoid complexity, which favours showing a smaller number of alternatives per choice set (Batsell and Louviere, 1991). Additionally, Eggers et al. (2022) recommend adjusting the choice set size to make it similar to the realistic size of compared alternatives. Finally, the number of alternatives should not exceed the number of attribute levels since using a number of alternatives that is a subset of the number of levels provides statistical benefits (Zeithammer and Lenk, 2009). Moreover, Zeithammer and Lenk (2009) indicate that pairs are often the optimal subsets for tasks. Correspondingly, for discrete choice experiments in the context of health economics, most often the number of alternatives equals two (de Bekker-Grob et al., 2012; Soekhai et al., 2019; Nouwens et al., 2025).

Moreover, one can also affect the survey by choosing the number of attributes per alternative. In order to keep the task manageable for respondents, the number of attributes per alternative should be set appropriately (Rao, 2014). In the majority of health-related studies, the number of attributes ranges from four to six (de Bekker-Grob et al., 2012; Soekhai et al., 2019; Nouwens et al., 2025).

Another important role for the construction of the questionnaire is played by the number of choice sets (Bridges et al., 2011; Rao, 2014; Eggers et al., 2022). On the one hand, more task sets increase the reliability of the estimated parameters. On the other hand, more task sets also increase the risk of fatigue among participants, which in turn may affect the data quality of the responses (Eggers et al., 2022). Several studies indicate a decreasing marginal benefit of additional tasks, again indicating that the

number of task sets should remain limited (see, e.g. Teichert, 2001; Bech et al., 2011). Consequently, the majority of studies in health economics consider between 9 and 16 choice sets, followed by less than 9 choice sets, whereas only a small share use more than 16 choice sets (de Bekker-Grob et al., 2012; Soekhai et al., 2019; Nouwens et al., 2025).

Sometimes, researchers may be interested in incorporating no-choice options or status-quo options (Bridges et al., 2011; Eggers et al., 2022). The advantage of doing this is that it increases the realism of the task, as respondents actually may not approve of any of the offered alternatives (Haaijer et al., 2001) or may be indifferent (Bridges et al., 2011). However, the usage of such options may also be inappropriate depending on whether the studied decision actually allows for such behaviour (Bridges et al., 2011). For example, such an option may be available for purchases of potentially unnecessary products but not in the case of a hospital choice for a compulsory surgery. Regarding studies related to health issues, about one third (Soekhai et al., 2019) to half (Nouwens et al., 2025) allow for no-choice or status-quo options.

It is also possible to expose participants only to a subset of tasks. This approach becomes interesting in the presence of large numbers of attributes (Rao, 2014).⁶ However, if the choice tasks are not too complex and participants are not prone to employing simplifying heuristics, it is considered good practice in research on health economics to apply full profiles (Bridges et al., 2011).

2.2.4 Variable Coding

Up to now, this chapter has presented ways to design the experiment and the survey. In the following subsections, I want to discuss how to connect the survey design and the statistical methodology. Namely, this includes how to code variables, how the shape of variables translates into utility evaluation to fill the RUM, how to select the appropriate model, and how to measure the importance of attributes.

⁶Rao (2014) provides a more detailed discussion on partial profile designs.

Various variable coding schemes allow for different approaches to the estimation of choice models. In a somewhat naive approach, one may first consider the attributes as quantitative data and dedicate numerical values to the single levels. Doing so assumes that levels are ordered and the distances between levels are of equal size. This rather strong assumption may, for example, hold for truly quantitative, ordinal, or binary variables, but with sufficient caution it is also possible to apply it to simplify categorical variables.

The first column of Table 2.1 provides an example of the numerical coding of an arbitrary attribute with four levels. This case requires only one variable $x_{num} \in \{1, 2, 3, 4\}$, where each entry represents one level. The entries $\{1, 2, 3, 4\}$ may represent true numerical values as well as an ordinal hierarchy. In the latter case, the entries have no numerical meaning and are ad hoc, meaning that they can easily be replaced by any other sequence of numbers with comparable characteristics with respect to order and distance, such as, e.g. $\{5, 6, 7, 8\}$ or $\{-0.48, -0.36, -0.24, -0.12\}$.

The advantage of this numerical coding is its simplicity and that it suffices for the identification and weighting of attributes with reliable estimation results while keeping estimation requirements low, as it requires only one variable per attribute (Mühlbacher et al., 2017). It finds application in several studies (see, e.g. Mühlbacher and Bethge, 2015, 2016; Mühlbacher et al., 2015, 2017; Emmert et al., 2024). However, in the case of unordered nominal levels, such as, e.g. administering medication by pill or subcutaneous or intramuscular injection, or considering a car, bus, or train as a means of transport, the assumption reaches its limits of use, as estimation results become less interpretable. The same applies if the data remain ordered but equidistance is not given, or if linearity is not given. In such cases, it is recommended to refrain from numerical coding and turn towards alternative kinds of coding. Yet, even if the levels are neither ordered nor equidistant, numerical coding may still provide a good first glimpse of the importance and significance of the attributes.

Using non-linear coding schemes is one way to overcome the shortcomings of numerical coding. For the sake of brevity, I restrict the discussion to dummy coding and effect

	Numerical coding	Dummy coding			Effect coding		
	x_{num}	x_{d2}	x_{d3}	x_{d4}	x_{e2}	x_{e3}	x_{e4}
Level 1	1	0	0	0	-1	-1	-1
Level 2	2	1	0	0	1	0	0
Level 3	3	0	1	0	0	1	0
Level 4	4	0	0	1	0	0	1

Table 2.1: Example of numerical, dummy, and effect coding with four levels and Level 1 as a benchmark

coding, as these are the major coding schemes used (Hensher et al., 2015).⁷ Both allow for a non-linear relationship between attribute levels and utility. Additionally, they require the creation of a set of new variables.

Dummy coding involves the generation of dummy variables for each attribute level. Transforming attributes through dummy coding enables to estimate partworth utilities. In this case, it is necessary to create $l_a - 1$ dummy variables for every attribute, where l_a describes the number of levels for attribute a . Each new variable represents one level and takes the value 1 in its presence and 0 otherwise. Additionally, it is required to declare one level per attribute as a benchmark to which all other levels are compared. Therefore, estimation results vary depending on the chosen benchmark, which is a set of levels across all attributes. Thus, the interpretation of dummy-coded variables is also relative to the benchmark, all else being equal. The advantage of dummy coding over numerical coding is the ability to estimate partworth utilities, which do not assume any functional relationship between attribute levels and therefore allow for piecewise, non-linear relations between attribute levels and utility (Hensher et al., 2015; Eggers et al., 2022). Thus, it is not necessary to assume any kind of functional form of utility as e.g. with numerical coding. In the middle column, Table 2.1 illustrates how the four levels are coded in the three dummy variables x_{d2} , x_{d3} , and x_{d4} , which represent levels

⁷Hensher et al. (2015) also discuss further coding schemes, such as, e.g. orthogonal polynomial coding. However, I skip them here, as they rarely find application.

2, 3, and 4, respectively. Level 1 serves as a benchmark; hence, it is not coded itself, and if it is present, x_{d2} , x_{d3} , and x_{d4} all take the value 0. In the case of the presence of Level 2, Level 3, or Level 4, x_{d2} , x_{d3} , or x_{d4} take the value 1, respectively, and are zero otherwise.

As mentioned above, a major drawback of dummy coding is the fact that estimates are relative to the base levels. Thus, estimate results are sensitive to the choice of benchmark. Effect coding can overcome this drawback by choosing a different reference point, which is independent of the choice of base levels. Again, this requires to transform the initial variable into $l_a - 1$ variables for every attribute level. The difference between dummy coding and effect coding emerges in the treatment of the base level. As shown in the right column of Table 2.1, effect coding can be employed by setting the reference level to -1 and not to 0, as in the case of dummy coding. In other words, in the presence of the benchmark level, all new variables x_{e2} , x_{e3} , and x_{e4} representing levels 2–4 take the value -1 . As explained above, this technique allows the estimation of partworth utilities. Effect-coded variables are zero-centred, i.e. the sum of the utilities of all attribute levels equals zero. In consequence, the estimated attribute levels are not relative to the (ad hoc) benchmark but in reference to a unique utility value that is distinct from base levels. If all attributes are effect-coded, this unique utility coincides with the grand mean, i.e. the average of the means of all attributes (Hensher et al., 2015). Yet, the relative size of level estimates within each attribute is the same as with dummy coding (Hauber et al., 2016). One positive effect of zero-centred variables is that it enables to construct a utility value for the benchmark levels such that the sum of all level utilities equals zero (Hensher et al., 2015; Eggers et al., 2022). In dummy coding however, the benchmark includes multiple attribute levels, meaning that it is impossible to disentangle the marginal utilities of single base levels from the entire benchmark (Hensher et al., 2015). Another advantage is that it allows to compare the utility of all attribute levels against the same reference point such that positive (negative) partworth utility values indicate higher (lower) preference compared to the reference point, i.e. compared to the grand mean if all attributes are effect-coded (Hauber et al., 2016; Eggers et al., 2022).

2.2.5 Utility Evaluation for Attribute Levels

In the following, I want to specify how to implement the various coding schemes in the RUM framework. Remember from equation (2.6) that it is possible to write the utility V_{ij} as the sum of all partworth utilities $f_{ia}(v_{ija})$ such that

$$V_{ij} = \sum_{a=1}^n f_{ia}(v_{ija}),$$

where f_{ia} defines the functional form determining how v_{ija} – the partworth utility of attribute a – translates into utility. Once again, variables may manifest as quantitative or categorical variables. Quantitative variables are modelled in a linear or non-linear way. The linear form is reached by defining f_{ia} to be a linear function. Correspondingly, the partworth utility for attribute a is

$$v_{ija} = \beta_a x_{ija}. \tag{2.30}$$

Thus, the partworth utility v_{ija} depends on the estimate β_a and the numeric value of the attribute x_{ija} . The utility for attribute a increases or decreases proportionally with its numeric value depending on the sign of β_a (Rao, 2014; Hensher et al., 2015; Eggers et al., 2022). The linear model is the most parsimonious approach, with only one parameter per attribute (Eggers et al., 2022). It may be appropriate to assume a monotonous relationship between the attribute and utility, e.g. the robustness of an artificial hip joint or the weight of an insulin pump.

For a non-linear relationship between the attribute and utility, the linear model may be misspecified. Instead, one could apply the ideal point model, which means that there is a quadratic term in f_{ia} such that the partworth utility function becomes

$$v_{ija} = \beta_{a1} x_{ija} + \beta_{a2} x_{ija}^2. \tag{2.31}$$

In the ideal point model, the values of the estimates β_{a1} and β_{a2} define at which value of x_{ija} the turning point in utility appears. This way, it incorporates diminishing (or

increasing) marginal utilities (Rao, 2014; Hensher et al., 2015; Eggers et al., 2022). One reasonable example may be the optimal frequency of medical consultations. At low frequencies, it may be beneficial to have medical checks more often in order to reduce the uncertainty about the current health state. However, at some point the utility may decrease with frequency, as the costs in terms of money and time may outweigh the benefits of additional checks.

For categorical variables, the approach differs slightly, as multiple dummy variables need to represent the variable: For categorical variables with l_a levels, $l_a - 1$ dummy variables are introduced according to the given coding scheme. In this case, the partworth function model

$$v_{ija} = \sum_{l=1}^{l_a-1} \beta_{al} x_{ijal} \quad (2.32)$$

captures utility most appropriately (Rao, 2014; Hensher et al., 2015; Eggers et al., 2022). It sums up the potential partworth utilities for all levels l . However, since only one level is chosen and all other dummies remain at a value of zero, v_{ija} corresponds to the estimate of the particular level.⁸ An example for the appropriate usage of categorical variables may be various modes of administering medicine, but they can also be applied to initially quantitative, numeric attributes (Rao, 2014; Eggers et al., 2022). The advantage of categorical variables is their flexibility in constituting non-linear and non-monotonous relationships without further assumptions a priori. The disadvantage is the laborious data preparation. Figure 2.2 depicts the functional forms discussed.

2.2.6 Model Selection

The selection of the appropriate model is an obviously important topic. Multinomial models are usually estimated using maximum likelihood. For this, the log likelihood

⁸For an extension of the partworth utility model with interaction terms, see, e.g. Hensher et al. (2015) or Eggers et al. (2022).

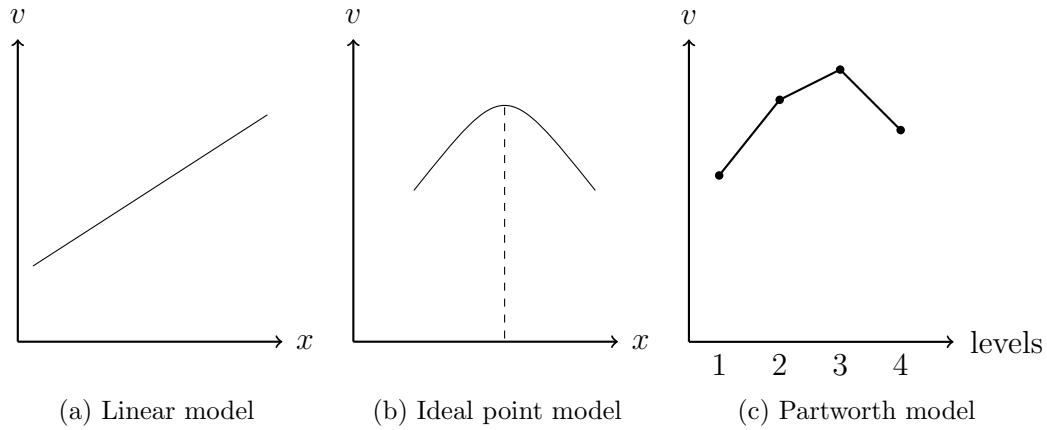


Figure 2.2: Alternative functional forms for the evaluation of attribute levels

function is the basis for any testing (Cameron and Trivedi, 2005; Greene, 2012; Rao, 2014; Hensher et al., 2015). A useful starting point is the likelihood ratio index (LRI) suggested by McFadden (1973)⁹:

$$LRI = 1 - \frac{\ln L}{\ln L_0}, \quad (2.33)$$

where $\ln L$ and $\ln L_0$ are the log likelihoods for the fitted model and the intercept-only null model, respectively. The larger the LRI , the better the fit of the estimated model to the given data (Cameron and Trivedi, 2005; Rao, 2014). Unfortunately, the values of the LRI have no natural interpretation, as depending on the data and number of explanatory variables, the lower and upper limits of the LRI approach 0 and 1 but never reach these values (Hauber et al., 2016). Therefore, the LRI is not a normalised measure and must be used with caution (Cameron and Trivedi, 2005; Greene, 2012).

Nested models can be tested with several statistical tests for significant differences, such as the likelihood ratio test or the Wald test (Cameron and Trivedi, 2005; Hensher et al., 2015; Croissant, 2020). With these, it is possible to test whether variables present only in one of the compared models play a significant role.

If models are not nested, these tests are not applicable and one must turn towards other methods. Information criteria such as the Akaike information criterion (AIC),

⁹Some studies refer to the LRI as *pseudo* $-R^2$.

Bayesian information criterion (BIC), or conditional AIC (CAIC) are those most often used for comparing non-nested models (Nylund et al., 2007; Rao, 2014). They are defined as

$$AIC = -2 \ln L + 2q, \quad (2.34)$$

$$BIC = -2 \ln L + (\ln N)q, \quad (2.35)$$

$$CAIC = -2 \ln L + (1 + \ln N)q, \quad (2.36)$$

as described, e.g. by Cameron and Trivedi (2005). All these criteria decrease with the value of the log likelihood function $\ln L$ and increase with the number of parameters q . The model with the lowest value is the most preferred for all three measures. In all, the second term on the right side represents a penalty for the used degrees of freedom. This punishment shall ensure that models remain as parsimonious as possible to describe the data. Otherwise, the best model would be a perfect replication of the world, which is not expedient (Cameron and Trivedi, 2005; Nylund et al., 2007). However, the BIC and CAIC have stronger penalties when the number of observations N increases, whereas the AIC does not. Thus, the BIC and CAIC are more restrictive of large models, while the AIC favours large models. This may be problematic, especially with respect to LC models, since the AIC may prefer models that may not provide a reasonable interpretation. Therefore, Nylund et al. (2007) recommend not using the AIC and instead refer to the BIC and CAIC.

2.2.7 Relative Importance of Attributes

A convenient measure to capture the importance of attributes is the relative range of the partworth utility functions. For this, one computes the difference between the most and least preferred attribute levels relative to the sum of all attributes' ranges (see, e.g. Rao, 2014; Hensher et al., 2015; Eggers et al., 2022). Consider n the total number of attributes, where a describes a single attribute and w_a is its relative importance. If β_a comprises all level coefficients of one attribute, then $\max(\beta_a)$ and $\min(\beta_a)$ describe

the most and least preferred level, respectively. Consequently, the range of attribute a is $(\max(\beta_a) - \min(\beta_a))$, and the relative attribute importance reads as

$$w_a = \frac{\max(\beta_a) - \min(\beta_a)}{\sum_{k=1}^n (\max(\beta_k) - \min(\beta_k))}. \quad (2.37)$$

This way, the sum of all weights are normalised to 1, with every w_a relative to the sum of all weights. Some scholars apply different methods of normalisation, e.g. measuring the weights relative to the largest range of attributes (see, e.g. Mühlbacher et al., 2017, 2018). Yet, the numerical differences between both approaches remain negligible, and the qualitative results are comparable.

References

- Batsell, R. R. and J. J. Louviere (1991). Experimental analysis of choice. *Marketing Letters* 2(3), 199–214.
- Bech, M., T. Kjaer, and J. Lauridsen (2011). Does the number of choice sets matter? Results from a web survey applying a discrete choice experiment. *Health Economics* 20(3), 273–286.
- Bhat, C. R. (1995). A heteroscedastic extreme value model of intercity travel mode choice. *Transportation Research Part B: Methodological* 29(6), 471–483.
- Bridges, J. F., A. B. Hauber, D. Marshall, A. Lloyd, L. A. Prosser, D. A. Regier, F. R. Johnson, and J. Mauskopf (2011). Conjoint analysis applications in health-a checklist: A report of the ISPOR good research practices for conjoint analysis task force. *Value in Health* 14(4), 403–413.
- Bujosa, A., A. Riera, and R. L. Hicks (2010). Combining discrete and continuous representations of preference heterogeneity: A latent class approach. *Environmental and Resource Economics* 47(4), 477–493.
- Cameron, A. C. and P. K. Trivedi (2005). *Microeconometrics: Methods and Applications*. Cambridge University Press.

- Coast, J., H. Al-Janabi, E. J. Sutton, S. A. Horrocks, A. J. Vosper, D. R. Swancutt, and T. N. Flynn (2012). Using qualitative methods for attribute development for discrete choice experiments: Issues and recommendations. *Health Economics* 21, 730–741.
- Croissant, Y. (2020). Estimation of random utility models in R: The mlogit package. *Journal of Statistical Software* 95(11), 1–41.
- de Bekker-Grob, E. W., M. Ryan, and K. Gerard (2012). Discrete choice experiments in health economics: A review of the literature. *Health Economics* 21(2), 145–172.
- Doherty, E., D. Campbell, S. Hynes, and T. van Rensburg (2013). Examining labelling effects within discrete choice experiments: An application to recreational site choice. *Journal of Environmental Management* 125, 94–104.
- Eggers, F. and H. Sattler (2009). Hybrid individualized two-level choice-based conjoint (HIT-CBC): A new method for measuring preference structures with many attribute levels. *International Journal of Research in Marketing* 26(2), 108–118.
- Eggers, F., H. Sattler, T. Teichert, and F. Völckner (2022). Choice-based conjoint analysis. In *Handbook of Market Research*, pp. 781–819. Heidelberg: Springer.
- Emmert, M., S. Rohrbacher, F. Meier, L. Heppe, C. Drach, A. Schindler, U. Sander, C. Patzelt, C. Frömke, O. Schöffski, and M. Lauerer (2024). The elicitation of patient and physician preferences for calculating consumer-based composite measures on hospital report cards: Results of two discrete choice experiments. *The European Journal of Health Economics* 25(6), 1071–1085.
- Fiebig, D., M. Keane, J. Louviere, and N. Wasi (2010). The generalized multinomial logit model: Accounting for scale and coefficient heterogeneity. *Marketing Science* 29(3), 393–421.
- Green, P. E. and V. Srinivasan (1978). Conjoint analysis in consumer research: Issues and outlook. *Journal of Consumer Research* 5(2), 103–123.

- Green, P. E. and V. Srinivasan (1990). Conjoint analysis in marketing: New developments with implications for research and practice. *Journal of Marketing* 54(4), 3–19.
- Greene, W. H. (2012). *Econometric Analysis* (7th ed.). Prentice Hall: Pearson.
- Greene, W. H. and D. A. Hensher (2013). Revealing additional dimensions of preference heterogeneity in a latent class mixed multinomial logit model. *Applied Economics* 45(14), 1897–1902.
- Haaijer, R., W. A. Kamakura, and M. Wedel (2001). The ‘no-choice’ alternative in conjoint choice experiments. *International Journal of Market Research* 43(1), 93–106.
- Hauber, A. B., J. M. González, C. G. Groothuis-Oudshoorn, T. Prior, D. A. Marshall, C. Cunningham, M. J. IJzerman, and J. F. Bridges (2016). Statistical methods for the analysis of discrete choice experiments: A report of the ISPOR conjoint analysis good research practices task force. *Value in Health* 19(4), 300–315.
- Helter, T. M. and C. E. H. Boehler (2016). Developing attributes for discrete choice experiments in health: A systematic literature review and case study of alcohol misuse interventions. *Journal of Substance Use* 21(6), 662–668.
- Hensher, D., J. Louviere, and J. Swait (1998). Combining sources of preference data. *Journal of Econometrics* 89(1-2), 197–221.
- Hensher, D. A., J. M. Rose, and W. H. Greene (2015). *Applied Choice Analysis* (2nd ed.). Cambridge University Press.
- Hess, S. and J. M. Rose (2012). Can scale and coefficient heterogeneity be separated in random coefficients models? *Transportation* 39(6), 1225–1239.
- Huber, J. and K. Zwerina (1996). The importance of utility balance in efficient choice designs. *Journal of Marketing Research* 33(3), 307–317.

- Keane, M. and N. Wasi (2013). Comparing alternative models of heterogeneity in consumer choice behavior. *Journal of Applied Econometrics* 28(6), 1018–1045.
- Kessels, R., B. Jones, and P. Goos (2011). Bayesian optimal designs for discrete choice experiments with partial profiles. *Journal of Choice Modelling* 24(3), 52–74.
- Lancsar, E. and J. Louviere (2008). Conducting discrete choice experiments to inform healthcare decision making: A user’s guide. *PharmacoEconomics* 26, 661–677.
- Louviere, J. J., D. Street, L. Burgess, N. Wasi, T. Islam, and A. A. Marley (2008). Modeling the choices of individual decision-makers by combining efficient choice experiment designs with extra preference information. *Journal of Choice Modelling* 1(1), 128–164.
- McFadden, D. (1973). Conditional logit analysis of qualitative choice behaviour. In P. Zarembka (Ed.), *Frontiers in Econometrics*, pp. 105–142. New York, NY, USA: Academic Press New York.
- McFadden, D. (1974). The measurement of urban travel demand. *Journal of Public Economics* 3(4), 303–328.
- McFadden, D. (1981). Econometric models of probabilistic choice. *Structural Analysis of Discrete Data with Econometric Applications* 198272, 198–272.
- Mühlbacher, A. C. and S. Bethge (2015). Patients’ preferences: A discrete-choice experiment for treatment of non-small-cell lung cancer. *The European Journal of Health Economics* 16, 657–670.
- Mühlbacher, A. C. and S. Bethge (2016). What matters in type 2 diabetes mellitus oral treatment? A discrete choice experiment to evaluate patient preferences. *The European Journal of Health Economics* 17(9), 1125–1140.
- Mühlbacher, A. C., J. F. Bridges, S. Bethge, C.-M. Dintsios, A. Schwalm, A. Gerber-Grote, and M. Nübling (2017). Preferences for antiviral therapy of chronic hepatitis C: A discrete choice experiment. *The European Journal of Health Economics* 18, 155–165.

- Mühlbacher, A. C., U. Junker, C. Juhnke, E. Stemmler, T. Kohlmann, F. Leverkus, and M. Nübling (2015). Chronic pain patients’ treatment preferences: A discrete-choice experiment. *The European Journal of Health Economics* 16, 613–628.
- Mühlbacher, A. C., A. Sadler, F.-W. Dippel, and C. Juhnke (2018). Treatment preferences in Germany differ among apheresis patients with severe hypercholesterolemia. *Pharmacoeconomics* 36(4), 477–493.
- Nouwens, S. P. H., S. M. Marceta, M. Bui, D. M. A. H. van Dijk, C. G. M. Groothuis-Oudshoorn, J. Veldwijk, J. A. van Til, and E. W. de Bekker-Grob (2025). The evolving landscape of discrete choice experiments in health economics: A systematic review. *Pharmacoeconomics* 43, 879–936.
- Nylund, K. L., T. Asparouhov, and B. O. Muthén (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A Monte Carlo simulation study. *Structural Equation Modeling: A Multidisciplinary Journal* 14(4), 535–569.
- Orme, B. (2002). Formulating attributes and levels in conjoint analysis. *Sawtooth Software Research Paper*, 1–4.
- Rao, V. R. (2014). *Applied Conjoint Analysis*. Heidelberg: Springer.
- Sarrias, M. and R. Daziano (2017). Multinomial logit models with continuous and discrete individual heterogeneity in R: The gmnL package. *Journal of Statistical Software* 79(2), 1–46.
- Shen, J. (2009). Latent class model or mixed logit model? A comparison by transport mode choice data. *Applied Economics* 41(22), 2915–2924.
- Soekhai, V., E. W. de Bekker-Grob, A. R. Ellis, and C. M. Vass (2019). Discrete choice experiments in health economics: Past, present and future. *Pharmacoeconomics* 37, 201–226.
- Teichert, T. (2001). *Nutzenschätzung in Conjoint-Analysen: Theoretische Fundierung und empirische Aussagekraft*, Volume 282. Heidelberg: Springer.

- Train, K. E. (2009). *Discrete Choice Methods with Simulation*. Cambridge University Press.
- Walker, J. and M. Ben-Akiva (2002). Generalized random utility model. *Mathematical Social Sciences* 43(3), 303–343.
- Zeithammer, R. and P. Lenk (2009). Statistical benefits of choices from subsets. *Journal of Marketing Research* 46(6), 816–831.

3 Preferences of People Living with HIW for Long-Acting Antiretroviral Treatment in Germany: Evidence from a Discrete Choice Experiment¹⁰

¹⁰This chapter is based on an article published in *The Patient – Patient-Centered Outcomes Research* (DOI: <https://doi.org/10.1007/s40271-023-00641-y>); this was joint work with Martin Emmert, Jennifer Jahn, Katharina Fernando, and Michael Lauerer.



Preferences of People Living with HIV for Long-Acting Antiretroviral Treatment in Germany: Evidence from a Discrete Choice Experiment

Martin Emmert¹ · Stefan Rohrbacher¹ · Jennifer Jahn² · Katharina Fernando² · Michael Lauerer^{1,2}

Accepted: 25 June 2023 / Published online: 12 July 2023
© The Author(s) 2023

Abstract

Objective This study aimed to elicit preferences for attributes of current and novel long-acting antiretroviral therapy for human immunodeficiency virus treatment.

Methods Primary survey data were collected (July–October 2022) on a sample of 333 people living with human immunodeficiency virus in Germany from a patient recruitment agency. Respondents were invited by e-mail to respond to a web-based questionnaire. After performing a systematic literature review, we conducted qualitative semi-structured interviews to identify and select the key attributes of drug therapy for patients' preferences for human immunodeficiency virus treatment. Based on this, a discrete choice experiment survey elicited preferences for long-acting antiretroviral therapy characteristics, including the type of medication, frequency of dosing, the location of treatment, the risk of both short-term and long-term side effects, as well as possible interactions with other medications or (party) drugs. A statistical data analysis was performed using multinomial logit models. An additional latent class multinomial logit was performed to evaluate subgroup differences.

Results Overall, 226 respondents (86% male, mean age 46.1 years) were included in the analysis. The frequency of dosing (36.1%) and the risk of long-term side effects (28.2%) had the greatest influence on preferences. The latent class analysis identified two patient groups. While the first class ($n = 135$; 87% male, mean age 44.4 years) found the frequency of dosing (44.1%) to be most important, the second class ($n = 91$; 85% male, mean age 48.6 years) focused on the risk of long-term side effects (50.3%). The evaluation of structural variables showed that male respondents, those living in small cities or villages, and those with better health status results were significantly more likely to be assigned to the second class ($p < 0.05$ each).

Conclusions All attributes included in our survey were important to participants when choosing an antiretroviral therapy. We found evidence that the frequency of dosing as well as the risk of long-term side effects have a particular impact on the acceptance of novel therapy regimens and should be considered in order to optimize adherence and satisfaction.

1 Introduction

The human immunodeficiency virus (HIV) is a pathogen that damages the immune system, which weakens the ability to fight everyday infections and diseases [1]. As the virus destroys and impairs the function of immune cells, people with HIV gradually become immunodeficient. The most advanced stage of HIV infection is the acquired immunodeficiency syndrome. Acquired immunodeficiency syndrome is defined by the development of certain cancers, infections, or other severe long-term clinical manifestations [2]. To date, despite intensive studies, neither curative drugs nor preventive vaccines are available [2–4]. However, with increasing

access to effective antiretroviral therapies (ARTs) enabling a longer life expectancy for people living with HIV, the infection has become a manageable chronic health condition [2]. Current ARTs suppress viral replication and allow an individual's immune system recovery to strengthen and regain the capacity to fight off opportunistic infections and some cancers [2].

Human immunodeficiency virus continues to be a major global public health issue; for example, in 2021, 1.5 million people became newly infected with HIV and 650,000 people died from HIV-related causes [2, 3]. In this regard, the United Nations Program for HIV/Acquired Immunodeficiency Syndrome (called UNAIDS) had set itself three treatment goals ("90-90-90 target") to be reached by 2020: (1) 90% of all people living with HIV will know their HIV status; (2) 90% of all people with diagnosed HIV

Key Points for Decision Makers

Targeted antiretroviral therapy for human immunodeficiency virus treatment is important. Therefore, this study provides evidence on patient preferences for current and novel long-acting antiretroviral therapy attributes.

Based on a discrete choice experiment, the frequency of dosing (36.1%) and the risk of long-term side effects (28.2%) were most important for patients. The latent class analysis demonstrated significant differences between two patient groups; while the first class found the frequency of dosing (44.1%) to be most important, the second group focused on the risk of long-term side effects (50.3%).

The results may inform policymakers, physicians, research institutions, and other stakeholders about patients' preferences of novel long-acting antiretroviral therapies for human immunodeficiency virus treatment.

infection will receive sustained ART; and (3) 90% of all people receiving ART will have viral suppression [5]. In Germany, the total number of people living with HIV was estimated to be approximately 90,800 in 2021, with 1800 people who became newly infected with HIV [6]. Currently, approximately 96% of all people both living and diagnosed with HIV in Germany are accessing ART (79,100 out of 82,100) [6]. Thereby, 96% of all people receiving ART will have viral suppression (i.e., having a viral load of <200 copies/mm³) [7]. In 2020, all three treatment values were reached in Germany for the first time, with the second and third target values even being far exceeded (90%, 97%, and 96%, respectively) [7].

At the same time, a widening of the spectrum of available drug therapies regarding both the type of medication and the frequency of dosing can be observed. Innovative and novel therapeutic regimens are accompanied by new active principles (e.g., an attachment inhibitor [fostemsavir/Rukobia®] or a monoclonal antibody [ibalizumab/Trogarzo®], additional forms of doses (e.g., depot injection [cabotegravir/Vocabria® in conjunction with Rilpivirine /Rekambys®]), and longer application intervals (e.g., monthly or bimonthly). These research advances specifically comprise the availability of single-tablet daily regimens for HIV [8], injection-based therapies (e.g., under the skin, into muscle) [9], longer acting formulations of pill-based or injection-based regimens [9], and others. Achievements might have the potential to

modify current treatment routines for certain patient groups; for example, for those with suboptimal adherence, confidentiality/privacy concerns, as well as the emotional burden and other barriers of daily dosing [8, 9]. Thereby, the emergence of novel long-acting therapy regimens requires both patients and providers to navigate a complex array of ART characteristics (e.g., pill burden, side effects, drug interactions, long-term toxicities, out-of-pocket costs) moving away from a “one-size-fits-all model” [10, 10]. This may have an impact on patient satisfaction, ART initiation, adherence, treatment effectiveness, and the quality of life of people living with HIV [8, 9, 12, 13]. Therefore, it seems necessary to elicit patient preferences for both current and novel long-acting ART regimens to evaluate their acceptance. The results from this study could thus improve our understanding of patient preferences for innovations in HIV therapy regimens [10] as well as enable better matching of treatments to patient expectations [11].

Previous studies have already explored patient preferences for HIV treatment [1, 9–11, 13–18]. For example, Mühlbacher and colleagues have conducted a discrete choice experiment (DCE) in a German population in 2009/2010, and demonstrated a high impact of quality-of-life-related attributes on patient preferences for the selection of treatments. In particular, the emotional quality of life (i.e., disease not obvious for others) was most important followed by the physical quality of life (i.e., diarrhea, nausea less frequent) and the social quality of life (i.e., participation in social life possible) [17]. Another study showed that people living with HIV in the USA valued minimizing side effects and long-term toxicities over dosing and administration characteristics [13]. Another preference-revealing study is being carried out in the USA to learn more about preferences for both current and novel HIV treatment regimens; with the results still to be published [8]. So far, to the best of our knowledge, DCEs have not yet been conducted to elicit patient preferences for both current and novel HIV treatment regimens in Germany.

Therefore, the present study aims to elicit patient preferences for both current and novel HIV treatment regimens in Germany and to explore in greater detail the heterogeneity in preferences of people living with HIV. The latter might increase the understanding of how preferences might diverge depending on characteristics (e.g., age, gender) [10, 13, 18].

2 Methods

This study used a mixed-methods approach. After performing a systematic literature review, we conducted qualitative semi-structured interviews to identify and select the most important attributes for patients' preferences for

HIV treatment in Germany. Based on this, a DCE was developed and performed to elicit patients' preferences for HIV treatment relevant attributes and to quantitatively determine their relative value. Instead of ranking or rating different attributes, as is done in traditional importance elicitation formats, DCEs perform a pairwise comparison of hypothetical alternatives (i.e., differently configured HIV treatments) and ask the participants to choose between them [19]. Based on the decisions, DCEs can help understand which characteristics (termed attributes) are preferred by consumers and determine the relative value of each attribute [20, 21]. The design and analysis of the DCE were based on standardized research practices for undertaking a conjoint analysis of the ISPOR Conjoint Analysis Good Research Practices Task Force [22–24].

2.1 Systematic Search Procedure

In a first step, we conducted a systematic search procedure on MEDLINE (via PubMed), the Cochrane Library, and PsycINFO to identify studies assessing patients' preferences for HIV treatment. The search was carried out in March 2022 and aimed at identifying English and German language literature published since 2010 (Electronic Supplementary Material [ESM]). In addition, we screened reference lists of identified research articles for further articles. The review complied with the Guideline from the Cochrane Collaboration [25]. Our search strategy was segmented into three components. The first component referred to patient preferences (e.g., priorities, expectations, perceptions, attitude), the second component to drug therapy respectively anti-HIV agents (e.g., anti-HIV agents, drug therapy, ART), and the third component to HIV infections (e.g., HIV, acquired immunodeficiency syndrome) [see ESM for a detailed information of the search procedure]. As a result, 8608 potentially relevant papers were identified in the electronic databases. After eliminating duplicates and judging titles and abstracts in a first step as well as full papers in a second step, 43 studies were considered relevant. Those studies investigated patients' preferences for HIV treatment using qualitative research ($n = 11$), quantitative research ($n = 16$), mixed-methods studies ($n = 3$), systematic reviews/meta-analysis ($n = 3$), as well as DCEs ($n = 10$) [see ESM 3 for an overview of the included studies]. In sum, we derived 34 different attributes from the 43 studies (ESM). We added two further attributes (e.g., suitability for daily use, ease of use) based on German HIV expert consultations. To reduce the number of attributes, we discussed the relevance of each attribute regarding the context within the German healthcare system (e.g., the cost of services [26, 27] does not play an important role in the German setting from the patient's perspective), discussed the attributes with the involvement of HIV experts, and qualitatively ranked the importance of

each attribute. Following this, nine attributes remained for the next qualitative step.

2.2 Qualitative Steps

2.2.1 Semi-Structured Interviews

In total, we qualitatively surveyed (May to June 2022) 15 randomly selected people with confirmed HIV status from a patient recruitment agency (Liberating Research, LR) living in Germany (40% female, mean age 42.9 years). Written informed consent was obtained from all participants. Respondents were mailed a short survey before conducting the semi-structured interviews. This enabled us to learn more about the HIV history of the interviewees and their initial assessments of the relevance of the nine attributes. For example, each respondent was asked to rate each attribute on a 1–5 scale (1 = not all important, 5 = extremely important). The aim of the semi-structured interviews was to discuss the previously identified attributes and their relevance from the perspective of people living with HIV in the German healthcare setting, ask for possibly missing attributes, clarify the wording, specify both level characteristics as well as short descriptions, and to evaluate the comprehensibility of hypothetical choice tasks for the DCE (see below). The interviews were conducted online via MS Teams, recorded, transcribed, and analyzed verbatim by using MAXQDA (Release 2020.4.2). Individuals who completed the qualitative study received 30 Euros. We did not collect personal data such as addresses, names, or phone numbers. Based on all steps, six attributes were derived that were of major importance for HIV treatment from the patients' perspective, which are as follows: the type of medication, the frequency of dosing, the location of treatment, the risk of short-term side effects, the risk of long-term side effects, as well as possible interactions with other medication or (party) drugs (see Table 1 for an overview of all attributes and corresponding levels). In contrast, the following three attributes were not included in the DCE because of a lower relevance from the patients' perspective: the suitability for daily use, the ease of use, and the risk of HIV-related stigma (privacy concerns).

2.3 Quantitative Steps

We used Sawtooth Software (Lighthouse Studio Version 9.14.0) to both program and conduct the survey. The questionnaire began with demographic and self-reported health status questions, and ended with HIV-related questions after the DCE choice sets. The experiment was designed as a full profile design and was generated using the balanced overlap method to permit estimation of the main effects and interactions, as well as D-optimal procedures to maximize statistical efficiency [28]. [The strength of design for our research

Table 1 List of attributes and levels used in the DCE

Attributes	Description and levels given in the DCE questionnaire
Type of medication	Pill-based treatment Subcutaneous injection (into the subcutaneous fatty tissue) Intramuscular injection (into the muscle)
Frequency of dosing	Daily Weekly Every 2 months Every 3 months Every 6 months
Location of treatment	Physicians' practice/hospital (medical staff administers treatment) Home-based (self-administered treatment)
Risk of short-term side effects	Low risk (1 of 1000 patients will experience short-term side effects) Moderate risk (10 of 1000 patients will experience short-term side effects) Higher risk (100 of 1000 patients will experience short-term side effects)
Risk of long-term side effects	Low risk (1 of 1000 patients will experience long-term side effects) Moderate risk (10 of 1000 patients will experience long-term side effects) Higher risk (100 of 1000 patients will experience long-term side effects)
Interaction with other medication or (party) drugs	No medication/drug interactions occur Increased risk of reduced effect due to (party) drugs Increased risk of reduced effect due to other medication

DCE discrete choice experiment

model is 940.53503.] In this context, a full profile design refers to a design that ensures the inclusion of all possible attributes across the choice sets presented to survey respondents. We set forced-choice tasks (i.e., choice sets had no option to opt out, as opting for no treatment is not a rational option [1, 29]) and two hypothetical HIV treatments. Thus, respondents were forced to make trade-offs between attributes and their levels. The method offers practical advantages such as closeness to reality as trade-off decisions are part of everyday life [19]. This can provide valuable insights into how patients prioritize and trade-off different attributes or levels when making HIV-treatment related decisions. As stated above, HIV treatments differed in six attributes with between two and five levels (see Table 1). In DCE research, four to eight attributes per choice set are seen as appropriate [30, 31]. Figure 1 provides one hypothetical choice task as an example. The questionnaire was pilot tested for clarity and comprehensibility with 15 people living with HIV in Germany and slightly modified accordingly (e.g., we slightly modified the wording of two questions to improve understanding). The latter included a comprehensive qualitative discussion with the respondents about the final version of the survey as well as a quantitative analysis based on the 15 responses to detect implausible responses. [Please note that the final wording of all attributes, levels, and corresponding descriptions had already been examined during the qualitative part of this study.]

Orme's often-used rule-of-thumb calculates a sample of 104 participants for a DCE having our design specifications (i.e., 12 choice tasks per respondent, two alternatives, two to five levels per attribute) [32]. This sample

size was discussed as being the lowest limit for a main effects estimation. However, we aimed at doubling this number (i.e., to include at least 208 participants) following more advanced recommendations for statistical robustness [33].

2.4 Study Sample

To be included in either the qualitative or quantitative part of our study, patients had to be at least 18 years of age, to be a current resident in Germany, to have a confirmed HIV status, to be taking ART, and to have appropriate language skills in German. Final questionnaires were sent by e-mail to 333 randomly selected respondents with confirmed HIV status from a patient recruitment agency (LR) between July and November 2022. On registering, all patients provide LR with both their e-mail address and phone number to ensure direct contact for survey participation purposes. LR, which maintains the panel, invited eligible members and referred them to the online survey. The registered panel of people living with HIV in Germany at LR is representative in terms of gender and age for all patients in Germany and was also surveyed in other HIV-related research [9]. As an incentive, patients received between 20 Euros and 40 Euros, with at least 10% of the incentive being donated to a German HIV charity.

Fig. 1 Example choice task.
HIV human immunodeficiency virus

Which of the two treatments would you choose?

	HIV-Therapy A	HIV-Therapy B
Type of medication	Intramuscular injection (into the muscle)	Pill
Frequency of dosing	Daily	Weekly
Location of treatment	Home-based (self-administered treatment)	Physicians' practice/hospital (medical staff administers treatment)
Risk of short-term side effects	Higher risk (100 of 1000 patients will experience short term side-effects)	Low risk (1 of 1000 patients will experience short term side-effects)
Risk of long-term side effects	Higher risk (100 of 1000 patients will experience short term side-effects)	Moderate risk (10 of 1000 patients will experience short term side-effects)
Interaction with other medication or (party) drugs	No medication/drug interactions occur	Increased risk of reduced effect due to other medication
	Selection	Selection

2.5 Survey Administration

In total, the survey consisted of four parts (ESM). First, we collected general sociodemographic information on participants (e.g., age, gender) before asking for HIV-treatment related issues (e.g., the year of both diagnosis and first treatment, current HIV treatment) in the second part. Afterwards, respondents were presented with information on all six attributes as well as corresponding levels and were asked to rate each attribute on a 1–5 scale (1 = not all important; 5 = extremely important) as well as to select the single most important attribute for the therapy choice. In the following, participants had to answer 12 DCE choice tasks, in which they were asked to choose between two treatment options. In the fourth part, respondents were asked to respond to further preference-related items regarding the different types of therapy regimens.

2.6 Statistical Methods

Analyses of general survey questions were performed using SPSS (released 2019, IBM SPSS Statistics for Windows, Version 26.0; IBM Corp. Armonk, NY, USA). Descriptive statistics were used to examine demographic and experience-related variables. Discrete choice experiment analyses were performed using R Statistical Software (Version 4.2.2; R Foundation for Statistical Computing, Vienna, Austria) and

the corresponding mlogit-packages by Croissant (2020) [34] and “gmnl” by Sarrias and Daziano [35].

In our analysis, we tested different MNL and random parameter logit (RPL) models to identify the best-fitting model (ESM). As a benchmark, we used a MNL model with attributes as explanatory variables and assumed homoscedastic errors. In alternative models, we included heteroscedastic errors and additional explanatory variables. [We consider it relevant to test for heteroscedasticity, as unnoticed heteroscedasticity may bias the estimates of nonlinear models (e.g., logit models) [36]]. Overall, we did not find improvements in using the variations. Both, Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) point at the benchmark model to be the best-fitting model. Despite significant scaling factors, heteroscedastic errors were rejected by the Wald test, likelihood ratio test, and Lagrange multiplier test. Including an intercept or sociodemographic variables (i.e., age, gender, education, place of living, sexual orientation, and health state) as additional explanatory variables did not detect statistically significant changes. In order to test for unobserved heterogeneity, we analyzed the data by using an RPL model. We assumed that random parameters were normally distributed and uncorrelated with each other. We chose $R = 1000$ random draws, which is sufficiently large for convergence. However, including random parameters on all attributes did not lead to statistically significant additional

findings apart from those related to the attribute “Interaction with other medication or (party) drugs” (we also applied uniform distribution without any meaningful changes). Additionally, no information criteria supported the use of the RPL model compared with the benchmark estimation (see also the likelihood ratio test and Wald test). [Please note: When we restricted random parameters to the attribute “Interaction with other medication or (party) drugs” as the only attribute with significant random parameters, the RPL model performs slightly better, and the Wald test reports that the remaining random parameters were relevant. Yet, the benchmark MNL model remains superior and we conclude that including random parameters did not provide additional benefits.] Based on this, we analyzed the DCE by performing the initial MNL model (see above).

We applied effect coding to all attributes so that the levels were expressed as (1, 0, − 1); thereby, (− 1) indicating the reference category. We chose effect coding over dummy coding because with dummy coding the parameter estimate for the (omitted) baseline category is equal to zero and cannot be recovered. Hence, the estimates of the other levels are expressed relative to the benchmark level. Instead, with effect coding, we can calculate the parameter of the baseline category from the negative sum of the included categories and the standard error from the covariances of the included categories. This enables us to compare all levels of the attributes against the corresponding mean value indicating a positive or negative impact compared to the mean of the attribute and the magnitude specifying the size of the effect [1, 24]. For all analyses, *p* values less than 0.05 were considered statistically significant [28].

2.7 Latent Class Analysis

With a latent class analysis, groups of respondents with similar preferences may be identified. In contrast to a cluster analysis, respondents are not assigned to different segments in a discrete (all-or-nothing) manner, but have probabilities of membership in each group [37]. A latent class analysis assumes that the underlying distribution for the parameters is discrete. This includes the assumption that there are different decision-making strategies used by groups (labeled classes) of respondents in a sample but that within each of these classes, responses are homogeneous [38]. More specifically, respondents are assigned to a class based on their highest posterior probability of class membership [38, 39]. In order to identify differences between potential subgroups, we conducted a latent class multinomial logit model. Using the *gmnl* package by Sarrias and Daziano [35], we compared models with two to five classes. We included sociodemographic variables as regressors (see above) in order to capture their impact on class membership. Again, we applied effect coding on all variables except for age.

As recommended for analyses with a small sample size (< 300) as well with the intention to provide a probable allocation of the participants per group, we examined AIC, BIC, and sample-size adjusted BIC to select the best-fitting model [40–42]. The information criteria led to heterogeneous recommendations (ESM). However, if incorrect, BIC and adjusted BIC are likely to underestimate the number of classes while AIC could be shown to be more likely to overestimate the number of classes, especially for studies with a small sample size [46, 47]. Based on this, we chose a model with two classes for the following analysis and provide the results of the model with three classes in the ESM for transparency reasons.

3 Results

3.1 Participant Characteristics

Of 333 eligible people living with HIV in Germany who were contacted, a total of 237 respondents returned the survey (71.2% response rate). We report on the *n* = 226 participants who fully completed the DCE component of the questionnaire (Table 2). The mean age of respondents was 46.09 years (standard deviation 11.20 years). Most respondents (85.8%) were male and most respondents (38.0%) stated intermediate secondary school or less as the highest educational level. Overall, 179 respondents (79.2%) rated their current health as good or better. The year of both diagnosis and first treatment ranged between 1984 and 2020. Regarding HIV treatment-related issues, most participants report a pill-based therapy (96.5%), take one HIV-related pill daily (75.2%), have been 12 months or more under the current HIV-therapy regimen (88.0%), and report to be (very) satisfied with the current HIV therapy (91.2%). Because each respondent made 12 choices, the final samples consisted of 2712 (226 times 12) cases to be analyzed.

3.2 Patients' Preferences: Descriptive Findings

Respondents' assessments of the importance of different attributes for the HIV-related treatment choice are presented in the ESM. First, each treatment attribute was rated on a 1–5 scale (1 = not all important; 5 = extremely important). As shown, participants rated the risk of long-term side effects (4.70 ± 0.68) as most important. In addition, the risk of short-term side effects (4.25 ± 0.95), the location of treatment (4.24 ± 0.98), and the frequency of dosing (4.23 ± 0.97) was rated slightly more important than the type of medication (4.14 ± 0.99). In contrast, possible interactions with other medication or (party) drugs seemed to be less important (3.64 ± 1.40). Second, the survey results for the single most important information item for the treatment

Table 2 Participant characteristics: key characteristics ($n = 226$)

	Mean (SD) or n	Range or %
Demographic characteristics		
Age		
Mean (SD)	46.09 (11.20)	23–70
Gender		
Male	194	85.8
Female	31	13.7
Diverse	1	0.4
Educational attainment		
Secondary general school or less	19	8.4
Intermediate secondary school	69	30.5
(Technical) University entrance qualification	135	59.7
Others	3	1.3
Place of living		
Large city/suburbs (100,000 inhabitants or more)	178	78.7
Middle-size city (20,000–100,000 inhabitants)	19	8.4
Small city, village (less than 20,000 inhabitants)	29	12.8
Sexual orientation		
Heterosexual	29	12.8
Homosexual	184	81.4
Other	13	5.8
Health status		
Good or better	179	79.2
Satisfactory	42	18.6
Bad or worse	5	2.2
Year of HIV diagnosis and first treatment		
Year of HIV diagnosis	2006	1984–2020
Year of first treatment	2008	1984–2020
HIV-treatment related issues		
Current HIV treatment		
Pill-based	218	96.5
Injection-based	8	3.5
Number of daily HIV-related pills		
1	164	75.2
2	25	11.5
3+	29	13.3
Duration of current HIV therapy		
< 6 months	9	4.0
6–12 months	18	8.0
12 months +	199	88.0
Location of medical care		
Inpatient-based (e.g., hospital)	32	14.2
Outpatient-based (e.g., specialized practice)	194	85.8
Travel time to the location of medical care (minutes)		
< 30	150	66.4
30–60	45	19.9
60+	31	13.7
Satisfaction with current HIV therapy		
(Very) satisfied	206	91.2
More or less	14	6.2
(Very) dissatisfied	6	2.6

HIV human immunodeficiency virus, SD standard deviation

Table 3 Estimated parameters of the multinomial logit model (discrete choice experiment, patients)

Attributes and levels	Participants (<i>n</i> = 226)					
	Coeff.	OR	SE coeff.	95% CI	<i>p</i> value	Difference (highest to lowest level)
Type of medication						0.514
Pill-based treatment	0.289	1.335	0.041	0.209	0.369	0.000
Intramuscular injection (into the muscle)	− 0.225	0.798	0.041	− 0.306	− 0.145	0.000
Subcutaneous injection (into the subcutaneous fatty tissue)	− 0.064	0.938	0.040	− 0.142	0.014	0.107
Frequency of dosing						1.664
Daily	− 0.906	0.404	0.065	− 1.034	− 0.778	0.000
Weekly	− 0.456	0.634	0.060	− 0.575	− 0.337	0.000
Every 2 months	0.146	1.157	0.059	0.029	0.262	0.014
Every 3 months	0.459	1.583	0.059	0.342	0.576	0.000
Every 6 months	0.758	2.133	0.062	0.635	0.881	0.000
Location of treatment						0.340
Physicians' practice/hospital (medical staff administers treatment)	− 0.170	0.844	0.017	− 0.204	− 0.136	0.000
Home-based (self-administered treatment)	0.170	1.185	0.017	0.136	0.204	0.000
Risk of short-term side effects						0.494
Low risk (1 of 1000 patients will experience short-term side effects)	0.207	1.230	0.040	0.128	0.286	0.000
Moderate risk (10 of 1000 patients will experience short-term side effects)	0.080	1.083	0.040	0.001	0.159	0.046
Higher risk (100 of 1000 patients will experience short-term side effects)	− 0.287	0.751	0.041	− 0.367	− 0.207	0.000
Risk of long-term side effects						1.298
Low risk (1 of 1000 patients will experience short-term side effects)	0.537	1.711	0.041	0.456	0.619	0.000
Moderate risk (10 of 1000 patients will experience short-term side effects)	0.224	1.251	0.040	0.146	0.302	0.000
Higher risk (100 of 1000 patients will experience short-term side effects)	− 0.761	0.467	0.044	− 0.847	− 0.675	0.000
Interaction with other medication or (party) drugs						0.294
No medication/drug interactions occur	0.168	1.183	0.040	0.089	0.248	0.000
Increased risk of reduced effect due to other medication	− 0.125	0.882	0.040	− 0.204	− 0.047	0.002
Increased risk of reduced effect due to (party) drugs	− 0.043	0.958	0.040	− 0.122	0.036	0.286

CI confidence interval, *coeff.* coefficient, *OR* odds ratio, *SE* standard error

choice revealed the risk of long-term side effects (35.4%) and the frequency of dosing (31.4%) as most relevant.

3.3 Patients' Preferences: Findings from the DCE Analysis

Table 3 presents the results of multinomial logit models of 226 participants' choices, totaling 2712 comparisons (see above). Overall, patients preferred a pill-based treatment with less frequent dosing regimens, which can be administered at home and further characterized by lower risks of both short-term and long-term side effects as well as by no interactions with other medications or (party) drugs. We showed statistically significant coefficients for all levels except for "Subcutaneous injection" and "Increased risk of reduced effect due to

(party) drugs". The coefficient values within an attribute add up to zero because of the effect coding (Table 3). The range of coefficient values within an attribute shows the importance of an attribute on a choice decision. We calculated the relative importance of each attribute by each attribute's coefficient range (i.e., the difference between the coefficients of the highest and lowest level of each attribute) expressed as a share of the total range across attributes. As shown in Fig. 2, the frequency of dosing (36.1%; level range of 1.664) as well as the risk of long-term side effects (28.2%; level range of 1.298) had the strongest influence on preferences. In contrast, the type of medication (11.2%; level range of 0.514), the risk of short-term side effects (10.7%; level range of 0.494), the location of treatment (7.4%; level range of 0.340), and possible

interactions with other medications or (party) drugs (6.4%; level range of 0.294) were less important.

3.4 Latent Class Analysis

Figure 3 displays the level estimates and the mean relative importance of attributes for both subgroups (see ESM for a detailed overview). Thereby, slightly more respondents were assigned to class 1 ($n = 135$; 59.7%) than to class 2 ($n = 91$; 40.3%). Regarding the preference pattern of class 1, the frequency of dosing (44.1%; level range of 1.990) is weighted highest. Furthermore, the type of medication (16.8%; level range of 0.759) was ranked second in this class. The remaining attributes risk of long-term side effects (10.8%; level range of 0.486), location of treatment (10.4%; level range of 0.470), possible interactions with other medications or (party) drugs (9.0%; level range of 0.408), and the risk of short-term side effects (8.8%; level range of 0.399) were rated less important in this class. Regarding the preference pattern of class 2, the risk of long-term side effects (50.3%; level range of 3.645) was weighted highest, followed by the frequency of dosing (27.4%; level range of 1.987). The attribute risk of short-term side effects (14.1%; level range of 1.022) was ranked third in this class. The remaining three attributes, type of medication (3.9%; level range of 0.280), location of treatment (2.9%; level range of 0.212), and possible interactions with other medications or (party) drugs (1.4%; level range of 0.105) were less important.

The evaluation of structural variables for both classes shows that age, educational attainment, and sexual orientation did not show any significant impact on the estimated class probabilities ($p > 0.05$ each) [ESM]. For class 1, further sociodemographic variables (i.e., gender, place of living, and health status) did not show any significant impact either. In contrast, male respondents, those living in small cities or villages, and those with higher health status results were significantly more likely to be assigned to class 2. Table 4 presents the participant characteristics of both subgroups. As shown, we found significant differences between both groups for all characteristics. For example, members of class 1 were more likely to be younger ($p < 0.001$), have a lower educational level ($p < 0.001$), live in small cities or villages ($p < 0.001$), and are less satisfied with the current HIV therapy ($p = 0.013$) compared with members of class 2.

4 Discussion

This study provides new insights into preference patterns that people living with HIV in Germany place on certain attributes of current and novel long-acting ART regimens. Thereby, the study's DCE-based findings appear broadly consistent with the results from both the rating-based and

ranking-based survey results and may allow conclusions to be drawn for targeting therapy regimens: based on our findings, the frequency of dosing and the risk of long-term side effects had the greatest influence on preferences. The corresponding attributes were also the most important to the study participants when asked about the most significant information items for choosing a certain HIV treatment. In addition, consistent with these findings, study participants rated the risk of long-term side effects as most important when rating all attributes on a scale of 1–5.

As mentioned above, previous studies have also explored patient preferences for HIV treatment in different parts of the world [1, 9–11, 13–18]. For example, Mühlbacher and colleagues have surveyed 218 people living with HIV in Germany in 2009/2010 and have shown that the attribute “Emotional quality of life (disease not obvious for others)” was most important, followed by “Physical quality of life (diarrhoea, nausea less frequent)” and “Social quality of life (participation in social life possible)”. Less important but still statistically significant were the attributes “Life expectancy,” “Flexibility of dosing,” and “Long term side effects improbable” [17]. The study is comparable to ours in that it included two similar attributes and one attribute that refers to the flexibility of dosing. It identified similar preference structures for people living with HIV in Germany. For example, the possibility of taking medication at home is also a significant attribute in our study, although it was only ranked the fifth most important attribute. However, the “Frequency of dosing” and “Risk of long-term side effects” were the two most significant factors in our study, whereas they were the two least significant factors in the mentioned study.

Another study including respondents from Germany (along with those from France, Italy, and the UK) was conducted by Akinwunmi and colleagues addressing in detail the importance of the attribute “Frequency of dosing” [9]. Here, the authors assessed whether people living with HIV as well as physicians would be interested in trying and offering a novel long-acting regimen (LAR) for HIV treatment that requires a dosing every 2 months instead of daily. As a result, it could be shown that two thirds of all respondents were very strongly interested in trying LAR. As one main reason, people living with HIV viewed the described LAR as addressing several unmet needs, such as suboptimal adherence, confidentiality/privacy concerns, and the emotional burden from taking pills every day. Further regression results demonstrated that the odds of being interested in trying LAR were 4.21 times higher among those reporting versus not reporting that taking their HIV treatment less often “would reduce the shame or stigma I feel for having HIV” [9]. Even though the study has applied a different design it does confirm, at least to some extent, the absolute importance of the

Fig. 2 Graphic display of level estimates with the 95% confidence interval (discrete choice experiment) and mean relative importance of attributes (based on each attribute's coefficient range, i.e., the difference between the coefficients of the highest and lowest level of each attribute)

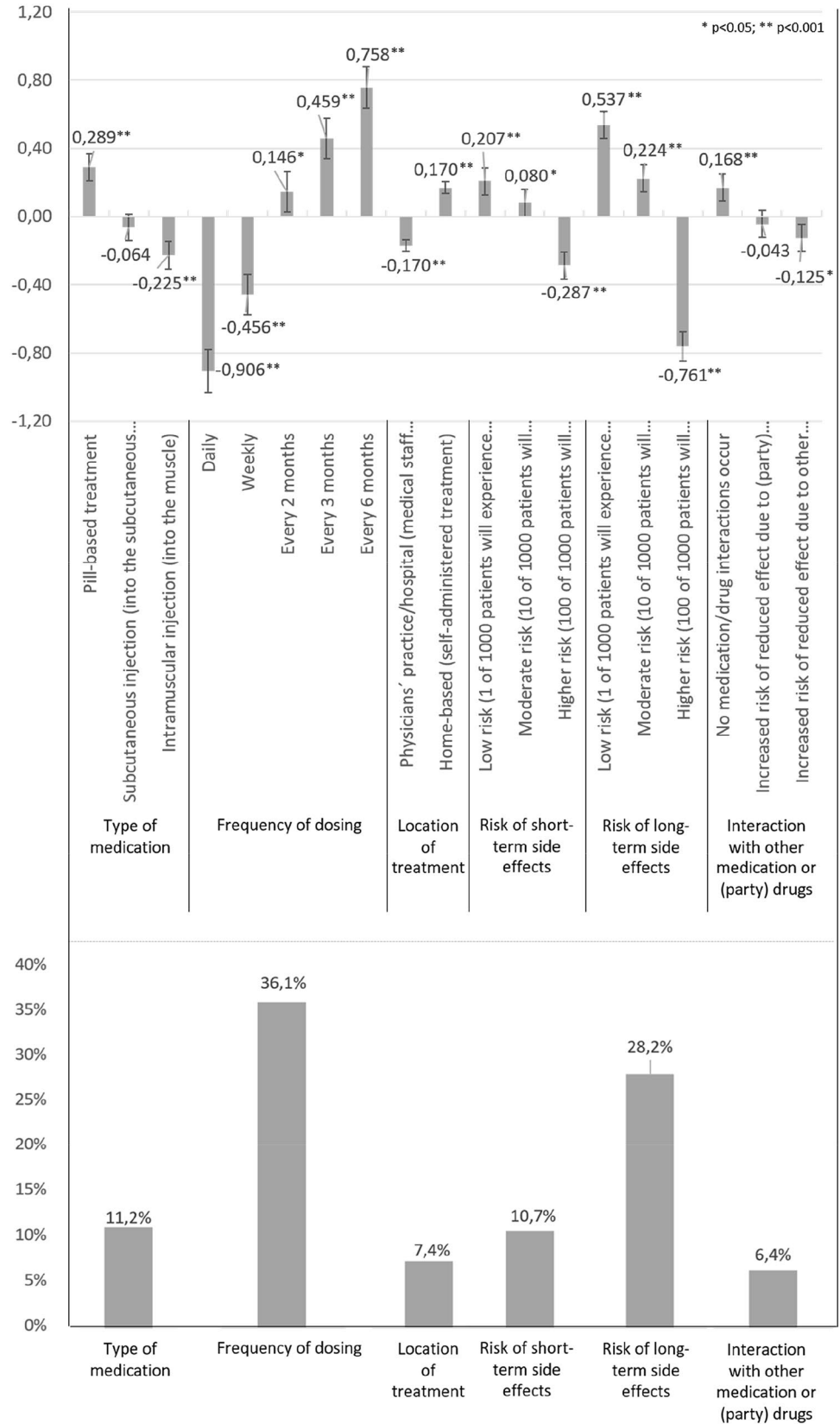


Fig. 3 Graphic display of level estimates with the 95% confidence interval (discrete choice experiment) and mean relative importance of attributes for both subgroups (based on each attribute's coefficient range, i.e., the difference between the coefficients of the highest and lowest level of each attribute)

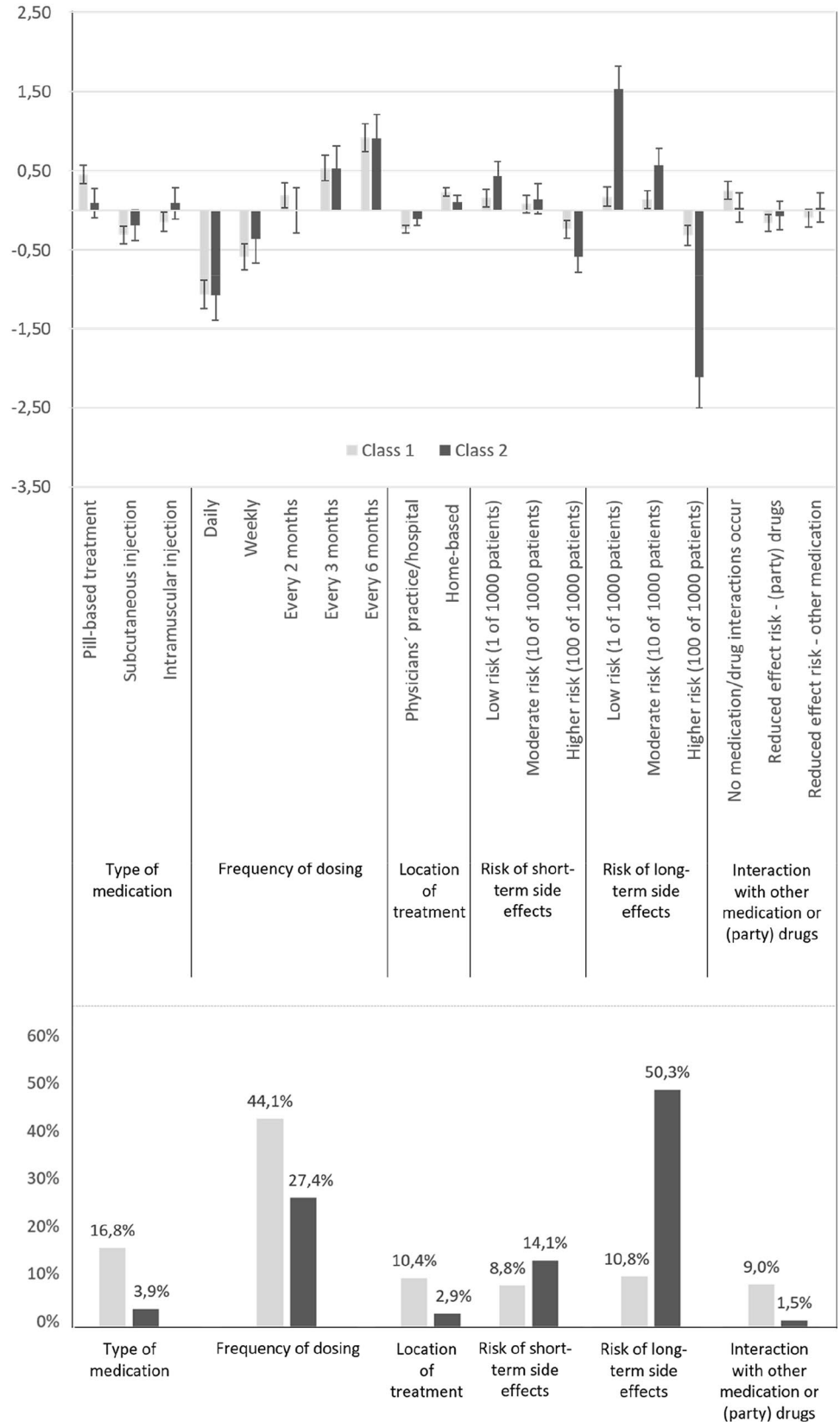


Table 4 Participant characteristics of both subgroups within the latent class analysis (*p* value was calculated based on the Wilcoxon test)

	Class 1 (<i>n</i> = 135; 59.7%) Mean (SD) or <i>n</i> (%)	Class 2 (<i>n</i> = 91; 40.3%) Mean (SD) or <i>n</i> (%)	<i>p</i> value
Demographic characteristics			
Age			
Mean (SD)	44.41 (10.58)	48.58 (11.57)	0.000
Gender			
Male	117 (86.7)	77 (84.6)	0.045
Female	17 (12.6)	14 (15.4)	
Diverse	1 (0.7)	0 (0.0)	
Educational attainment			
Secondary general school or less	15 (11.1)	4 (4.4)	0.000
Intermediate secondary school	43 (31.9)	26 (28.6)	
(Technical) University entrance qualification	75 (55.6)	60 (65.9)	
Others	2 (1.5)	1 (1.1)	
Place of living			
Large city/suburbs (100.000 inhabitants or more)	101 (74.8)	77 (84.6)	0.000
Middle-size city (20.000–100.000 inhabitants)	12 (8.9)	7 (7.7)	
Small city, village (<20,000 inhabitants)	22 (16.3)	7 (7.7)	
Sexual orientation			
Heterosexual	17 (12.6)	12 (13.2)	0.00
Homosexual	113 (83.7)	71 (78.0)	
Other	5 (3.7)	8 (8.8)	
Health status			
Good or better	110 (81.5)	69 (75.8)	0.000
Satisfactory	21 (15.6)	21 (23.1)	
Bad or worse	4 (3.0)	1 (1.1)	
Year of HIV diagnosis and first treatment			
Year of HIV diagnosis (range)	2007 (1985–2020)	2005 (1984–2019)	0.000
Year of first treatment (range)	2009 (1989–2020)	2007 (1984–2019)	0.000
HIV treatment-related issues			
Current HIV treatment			
Pill-based	129 (95.6)	89 (97.8)	0.000
Injection-based	6 (4.4)	2 (2.2)	
Number of daily HIV-related pills (<i>n</i> = 218)			
1	103 (79.8)	61 (68.5)	0.000
2	13 (10.1)	12 (13.5)	
3+	12 (9.3)	10 (9.6)	
Duration of current HIV therapy (months)			
<6	5 (3.7)	4 (4.4)	0.000
6–12	13 (9.6)	5 (5.5)	
12+	117 (86.7)	82 (90.1)	
Location of medical care			
Inpatient-based (e.g., hospital)	22 (16.3)	10 (11.0)	0.000
Outpatient-based (e.g., specialized practice)	113 (83.7)	81 (89.0)	
Travel time to the location of medical care (minutes)			
<30	87 (64.4)	63 (69.2)	0.000
30–60	27 (20.0)	18 (19.8)	
60+	21 (15.6)	10 (11.0)	
Satisfaction with current HIV therapy			
(Very) satisfied	122 (90.4)	84 (92.3)	0.013
More or less	9 (6.7)	5 (5.5)	

Table 4 (continued)

	Class 1 (<i>n</i> = 135; 59.7%) Mean (SD) or <i>n</i> (%)	Class 2 (<i>n</i> = 91; 40.3%) Mean (SD) or <i>n</i> (%)	<i>p</i> value
(Very) dissatisfied	4 (3.0)	2 (2.2)	

HIV human immunodeficiency virus, SD standard deviation

attribute “Frequency of dosing” as one major attribute when choosing HIV therapy regimens.

In another European study, Bréigéon-Ronot and colleagues surveyed 101 people living with HIV in France in 2014 to elicit relative preferences for attributes of ARTs [11]. The sample is comparable to ours regarding age, gender, as well as therapy-based characteristics (e.g., educational level, current health status, year of both diagnosis, and first treatment). As a result from the DCE analysis, it appears that patients placed significant value on the avoidance of both short-term side effects (i.e., diarrhea) as well as long-term health problems, such as cardiovascular disease and kidney disease [11]. Also similar to our findings, patients were concerned about the potential risk of drug–drug interactions and demonstrated a strong preference for a treatment with limited drug–drug interactions. Patients also preferred to avoid problems associated with treatment failure or therapy that left them with a higher viral load after the first weeks of treatment. Even though some attributes evaluated in this study were different from those in ours, in both studies, we detected limited drug–drug interactions and both short-term and long-term health problems to be significant attributes for patients’ preferences in ART regimens.

In addition, we identified two DCE studies from the USA with the intention to identify patient preferences for novel ART regimens. First, Ostermann and colleagues surveyed 403 people living with HIV in the USA in 2017/2018. They included four attributes in their DCE but those were similar to ours, namely dosing (i.e., the number of pills per dose and the dosing frequency), administration (i.e., meal requirement and pill size), short-term side effects (e.g., diarrhea, dizziness), and long-term side effects (e.g., risk of heart attack, risk of new or worse kidney problems) [13]. As a result, they could show that people living with HIV valued minimizing short-term side effects (relative importance: 44%) and long-term risks (32%) over dosing (17%) and administration characteristics (8%) [13]. Those findings are partly similar, except that in our study, the long-term side effects and the frequency of dosing were most important. In contrast, the short-term side effects were less relevant and the administration was not among the most relevant attributes in our study. Second, another study is currently being carried out in the USA to learn more about preferences for both current and novel HIV treatment regimens; with the results still to be published [8]. The findings of this study will

enable interesting comparisons, as the attributes identified as relevant in this study are partly very similar to ours (e.g., location, frequency of dosing, side effects), but partly also address alternative aspects (e.g., late dose leeway).

In another interesting DCE study, Tieosapjaroen and colleagues surveyed 335 people living with HIV in Melbourne, Australia in 2021 to examine the relative importance of antiretroviral medication side effects so as to improve patient satisfaction, adherence, and treatment. In particular, the authors focused on weight gain as a possible side effect of HIV treatment. As a result, it could be shown that weight gain was the second most important short-term side effect following depression and the third most important undesirable long-term side effect, following risks of heart attack and kidney problems. Interestingly, the authors could show that respondents who weighed more than 85 kg preferred medications not causing weight gain while respondents with a lower weight preferred medications causing weight gain [43].

We have identified further studies that have been conducted in developing countries (e.g., Colombia [1, 16], Zambia [15], Kenya [10], urban Zimbabwe [18]). However, these studies seem to be less comparable to ours compared with studies from other European countries (e.g., France) and the USA, as they are likely to focus on other aspects of care [3, 5, 44]. For example, Dommaraju and colleagues have surveyed 104 persons living with HIV in Kenya to elicit preferences for differentiated care models. Here, the authors concluded that clinically stable patients expressed a preference for facility-based care provided clinical visits were extended to biannually. Here, the most important attributes were the location of clinical review (24.1%), location of ART refills (21.6%), the frequency of clinical visits (15.2%), and the frequency of ART refills (13.5%), followed by the person providing ART (9.7%), adherence support (8.2%), and refill pick-up/delivery time (7.8%) [10].

As mentioned above, patients preferred a pill-based treatment with less frequent dosing regimens, which can be administered at home and further characterized by lower risks of both short-term and long-term side effects as well as by no interactions with other medications or (party) drugs. In line with the current HIV healthcare delivery in the USA [8], most people living with HIV in Germany take one or more pills daily as well. Based on our results, more than 86% take one or two HIV-related pills daily combining two

or more antiretroviral medications for treatment (see above). Using the example of frequency of dosing, it appears evident that the reality of care and the preferences of people living with HIV diverge, at least to some extent.

Finally, our latent class analysis-based findings have demonstrated meaningful differences between both subgroups. Interestingly, we detected one attribute to be of major importance in both groups. While the frequency of dosing was most important in class 1, the risk of long-term side effects was most important in class 2. The relative weight of those attributes accounted almost for 50% of all attributes. This means that these criteria were almost as important as the other five criteria combined. In both groups, we found one additional attribute to be of higher importance; while the type of medication was relevant in class 1, the frequency of dosing was weighted important in class 2. In contrast, the three attributes location of treatment, risk of short-term side effects, and the interaction with other medications or (party) drugs did not play an important role in either class.

With this in mind, our findings on the therapy preferences of people living with HIV have several implications: first, the results are directed to pharmaceutical researchers, who should consider them when developing innovative and novel therapy regimens. It should be noted that the desideratum of drug research and development corresponding to preferences is currently already being met, at least to some extent. Present research advances including the availability of single-tablet daily regimens for HIV [8], injection-based therapies (e.g., under the skin, into muscle), longer acting formulations of pill-based regimens, and others might have the potential to modify current treatment routine, at least for some patients; for example, for patients with suboptimal adherence, confidentiality/privacy concerns, as well as the emotional burden and other barriers of daily dosing [8, 9]. Thus, the emerging treatment options already correspond, in part, to patient preferences emerging from the present study. This applies, for example, to longer application intervals. As mentioned above, new long-acting therapy regimens might then have the potential to increase ART initiation and adherence [8, 12].

Second, our findings are relevant to professional societies, which should consider them when revising therapeutic guidelines or therapy regimens. Third, the results should also be considered by actors and institutions entrusted with regulatory tasks and responsible for priority setting. This applies, for example, to benefit assessments or health economic evaluations and thus also to pricing as well as decisions on reimbursement or inclusion in the benefit catalog of health insurance funds. Fourth, the results are directed at the medical profession, as they can contribute to joint decision making with patients when discussing

different therapy options so as to increase therapy adherence. This, in turn, would also have a positive impact on both the outcomes and economic aspects of HIV treatment. For example, the frequency of treatment administration could be shown to be a key predictor of patient adherence [45].

In summary, taking into account the preferences of people living with HIV could further improve decisions and medical care. In addition, a strengthened role for preferences should be regarded as a form of participation that helps increase trust in health systems as a whole that may also strengthen the legitimacy of decisions within that system.

4.1 Limitations

As with any study, there are several limitations that have to be considered when interpreting the results of our research. First, our surveyed study respondents were recruited by a patient recruitment agency (LR). The registered panel of people living with HIV in Germany at LR is representative in terms of gender and age for all patients in Germany and was also surveyed in other HIV-related research [9]. In our study in particular, the proportion of male participants was slightly above the average in Germany (85.8% vs 80.1% [6]). Overall, it must be stated that there is little sociodemographic information available of people living with HIV in Germany [46]. Therefore, we cannot ensure that our sample is representative of the German population with respect to other characteristics as well (e.g., education, marital status, place of living). In addition, the recruitment via LR could have influenced the study population with respect to individual parameters, which could not be examined in total [47]. However, based on the sociodemographic data, it seems that our study population is similar to those of other studies (see above), at least to a certain extent. For example, the sociodemographic data showed a similar pattern to other studies in HIV-related preference research [9, 11, 17]. Second, it must be noted that the selected attributes for the DCE referred to both currently already available as well as novel, innovative HIV treatment regimens. Therefore, we cannot exclude that the participants in the experiment were more familiar with the already available therapy regimens and therefore chose those primarily. Third, the selection of the levels for the attribute “Frequency of Dosing” was associated with some challenges as related novel therapy regimens are currently not yet available. In the qualitative steps, however, we aimed to assure not to integrate any unrealistic characteristics (i.e., levels) but those that are actually currently being investigated in research. Therefore, the pool of experts from HIV research provided support in this regard. Fourth, the participants surveyed in this study had a mean year of first diagnosis of 2006, with the earliest diagnosis in 1984 and the latest diagnosis in 2020. Because of this relatively broad

range, a different amount of experience of the participants can be assumed. Preferences of patients who have received one or more treatments and preferences of those who are newly diagnosed might vary as preferences might change over time [48]. Therefore, it should be taken into account that preferences of individual patients might change in the future. However, our results shall be interpreted as a current snapshot of preferences of people living with HIV in Germany [48]. Fifth, the generalizability (external validity) of our results to other settings should also be mentioned. As described above, studies from other European countries (Germany [17], France, Italy, and the UK [9], France [11]) or the USA [8, 13] tend to have included more similar attributes and the samples surveyed appear more comparable to our study. Therefore, our results are more likely to be transferable to these countries. In contrast, studies that have been conducted in developing countries (e.g., Colombia [1, 16], Zambia [15], Kenya [10], Urban Zimbabwe [18]) are likely to focus on other aspects of care as well [3, 5, 44]. Here, the transferability is likely to be lower. Sixth, patients received a financial incentive for participating in the quantitative part of our study. It should be mentioned that the incentive has been increased over time from 20 Euros to 40 Euros to improve the response rate. Please note that the average paid incentive was calculated to be 28.32 Euros. Finally, it should be noted that we did not randomly allocate the order of the presentation of attributes between participants. We refrained from doing so, as previous research has shown that the ordering of the attributes was found to have no statistically significant effect on the importance attached to them by respondents [49, 50]. Nevertheless, we follow Norman and colleagues by recognizing that in some situations (e.g., relatively difficult choice tasks, unfamiliar choice task), respondents may have the tendency to rely on simplifying heuristics such as focusing on the earlier or later dimensions [49]. However, in our study, we felt that the choice tasks were neither very difficult nor unfamiliar to the respondents because we surveyed only people with confirmed HIV status.

5 Conclusions

This study provides new insights into preference patterns that people living with HIV in Germany place on certain attributes of current and novel ART regimens. Thereby, the study's DCE-based findings appear broadly consistent with the results from both the rating-based and ranking-based survey results. One of the major results is the importance placed by people living with HIV in Germany on the characteristics of HIV-related therapy regimens related to the frequency of dosing and the risk of long-term side effects. The latter had the greatest influence on

peoples' preferences. This should be considered when further developing HIV-related therapy regimens in order to increase treatment adherence and the satisfaction of people living with HIV.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s40271-023-00641-y>.

Acknowledgments The authors thank all survey participants.

Funding Open Access funding enabled and organized by Projekt DEAL.

Declarations

Funding This research was financed by Gilead Sciences GmbH, Planegg, Germany.

Conflicts of interest/competing interests Martin Emmert, Stefan Rohrbacher, Jennifer Jahn, Katharina Fernando, and Michael Lauerer have no conflicts of interest that are directly relevant to the content of this article.

Ethics approval Ethics approval for this study was not necessary according to the Ethics Committee from the University of Bayreuth, Germany.

Consent to participate All respondents were informed about the study and its potential risks and benefits prior to participation. Respondents had to sign an informed consent to participate in the study, and to have their data used to develop the results contained in this paper. In addition, participation was voluntarily and the participant could stop at any time.

Consent for publication Not applicable.

Availability of data and material The datasets generated and/or analyzed during the current study are available from the corresponding author on reasonable request.

Code availability Not applicable.

Authors' contributions ME designed and directed this study as the principal investigator. ME and ML conceived and planned the experimental design of the study. ME and SR performed the analysis on all samples. All authors contributed to the interpretation of the results. JJ and KF both conducted and analyzed the qualitative interviews. ME took the lead in preparing the manuscript. All authors provided critical feedback and commented on the manuscript. Finally, all authors approved the latest version of the manuscript to be published.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License, which permits any non-commercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc/4.0/>.

References

1. Sijstermans E, Cheung KL, Goossens AJM, et al. A discrete choice experiment to assess patients' preferences for HIV treatment in the urban population in Colombia. *J Med Econ.* 2020;23(8):812–8. <https://doi.org/10.1080/13696998.2020.1735399>.
2. World Health Organization. HIV: key facts 2022. <https://www.who.int/news-room/fact-sheets/detail/hiv-aids>. Accessed 4 Jul 2023.
3. Joint United Nations Programme on HIV/AIDS (UNAIDS). Global HIV & AIDS statistics: fact sheet 2022. <https://www.unaids.org/en/resources/fact-sheet>. Accessed 4 Jul 2023.
4. German Center for Infection Research. HIV: HIV scientists concentrate their research on both remission and cure. 2022. <https://www.dzif.de/en/hiv>. Accessed 4 Jul 2023.
5. Joint United Nations Programme on HIV/AIDS (UNAIDS). 90-90-90: an ambitious treatment target to help end the AIDS epidemic 2014. https://www.unaids.org/sites/default/files/media_asset/90-90-90_en.pdf. Accessed 4 Jul 2023.
6. Robert Koch-Institut (RKI). HIV/AIDS in Deutschland: Eckdaten der Schätzung. Epidemiologische Kurzinformation des Robert Koch-Instituts, Stand: Ende 2021. Robert Koch-Institut (RKI): Berlin; 2022.
7. Robert Koch-Institut (RKI). Epidemiologisches Bulletin: HIV in Deutschland 2020. Erfassung der SARS-CoV-2-PCR-Testzahlen. Robert Koch-Institut (RKI): Berlin; 2021.
8. Barthold D, Brah AT, Graham SM, et al. Improvements to survey design from pilot testing a discrete-choice experiment of the preferences of persons living with HIV for long-acting antiretroviral therapies. *Patient.* 2022;15(5):513–20. <https://doi.org/10.1007/s40271-022-00581-z>.
9. Akinwunmi B, Buchenberger D, Scherzer J, et al. Factors associated with interest in a long-acting HIV regimen: perspectives of people living with HIV and healthcare providers in four European countries. *Sex Transm Infect.* 2021;97(8):566–73. <https://doi.org/10.1136/sextrans-2020-054648>.
10. Dommaraju S, Hagey J, Odeny TA, et al. Preferences of people living with HIV for differentiated care models in Kenya: a discrete choice experiment. *PLoS ONE.* 2021;16(8):e0255650. <https://doi.org/10.1371/journal.pone.0255650>.
11. Bréigeon-Ronot S, Cheret A, Cabié A, et al. Evaluating patient preference and satisfaction for human immunodeficiency virus therapy in France. *Patient Prefer Adherence.* 2017;11:1159–69. <https://doi.org/10.2147/PPA.S130276>.
12. Cohen J, Beaubrun A, Bashyal R, et al. Real-world adherence and persistence for newly-prescribed HIV treatment: single versus multiple tablet regimen comparison among US medicaid beneficiaries. *AIDS Res Ther.* 2020;17(1):12. <https://doi.org/10.1186/s12981-020-00268-1>.
13. Ostermann J, Mühlbacher A, Brown DS, et al. Heterogeneous patient preferences for modern antiretroviral therapy: results of a discrete choice experiment. *Value Health.* 2020;23(7):851–61. <https://doi.org/10.1016/j.jval.2020.03.007>.
14. Eshun-Wilson I, Kim H-Y, Schwartz S, et al. Exploring relative preferences for HIV service features using discrete choice experiments: a synthetic review. *Curr HIV/AIDS Rep.* 2020;17(5):467–77.
15. Eshun-Wilson I, Mukumbwa-Mwenechanya M, Kim H-Y, et al. Differentiated care preferences of stable patients on antiretroviral therapy in Zambia: a discrete choice experiment. *J Acquir Immune Defic Syndr.* 2019;81(5):540–6.
16. Goossens AJM, Cheung KL, Sijstermans E, et al. A discrete choice experiment to assess patients' preferences for HIV treatment in the rural population in Colombia. *J Med Econ.* 2020;23(8):803–11. <https://doi.org/10.1080/13696998.2020.1735398>.
17. Mühlbacher AC, Stoll M, Mahlich J, et al. Patient preferences for HIV/AIDS therapy: a discrete choice experiment. *Health Econ Rev.* 2013;3(1):14. <https://doi.org/10.1186/2191-1991-3-14>.
18. Strauss M, George G, Mantell JE, et al. Optimizing differentiated HIV treatment models in urban Zimbabwe: assessing patient preferences using a discrete choice experiment. *AIDS Behav.* 2021;25(2):397–413.
19. Mühlbacher AC, Bethge S. Patients' preferences: a discrete-choice experiment for treatment of non-small-cell lung cancer. *Eur J Health Econ.* 2015;16(6):657–70. <https://doi.org/10.1007/s10198-014-0622-4>.
20. Vass C, Rigby D, Payne K. The role of qualitative research methods in discrete choice experiments. *Med Decis Making.* 2017;37(3):298–313. <https://doi.org/10.1177/0272989X16683934>.
21. Naik Panvelkar P, Armour C, Saini B. Community pharmacy-based asthma services: what do patients prefer? *J Asthma.* 2010;47(10):1085–93.
22. Reed Johnson F, Lancsar E, Marshall D, et al. Constructing experimental designs for discrete-choice experiments: report of the ISPOR Conjoint Analysis Experimental Design Good Research Practices Task Force. *Value Health.* 2013;16(1):3–13.
23. Bridges JFP, Hauber AB, Marshall D, et al. Conjoint analysis applications in health—a checklist: a report of the ISPOR Good Research Practices for Conjoint Analysis Task Force. *Value Health.* 2011;14(4):403–13. <https://doi.org/10.1016/j.jval.2010.11.013>.
24. Hauber AB, González JM, Groothuis-Oudshoorn CGM, et al. Statistical methods for the analysis of discrete choice experiments: a report of the ISPOR Conjoint Analysis Good Research Practices Task Force. *Value Health.* 2016;19(4):300–15. <https://doi.org/10.1016/j.jval.2016.04.004>.
25. Higgins J, Thomas J, Chandler J, et al editors. *Cochrane handbook for systematic reviews of interventions.* 2nd ed. Chichester: Wiley; 2019.
26. Opuni M, Bishai D, Gray GE, et al. Preferences for characteristics of antiretroviral therapy provision in Johannesburg, South Africa: results of a conjoint analysis. *AIDS Behav.* 2010;14(4):807–15. <https://doi.org/10.1007/s10461-009-9584-4>.
27. Kruk ME, Riley PL, Palma AM, et al. How can the health system retain women in HIV treatment for a lifetime? A discrete choice experiment in Ethiopia and Mozambique. *PLoS ONE.* 2016;11(8):e0160764. <https://doi.org/10.1371/journal.pone.0160764>.
28. Marshall DA, Deal K, Conner-Spady B, et al. How do patients trade-off surgeon choice and waiting times for total joint replacement: a discrete choice experiment. *Osteoarthritis Cartil.* 2018;26(4):522–30.
29. Veldwijk J, Lambooy MS, de Bekker-Grob EW, et al. The effect of including an opt-out option in discrete choice experiments. *PLoS ONE.* 2014;9(11):e111805. <https://doi.org/10.1371/journal.pone.0111805>.
30. Ryan M, Gerard K. Using discrete choice experiments to value health care programmes: current practice and future research reflections. *Appl Health Econ Health Policy.* 2003;2(1):55–64.
31. Sawtooth Software. The CBC system for choice-based conjoint analysis. In: Sawtooth Software, Inc., editor. CBC: technical paper. Technical Paper Series. Sawtooth Software, Inc.: Washington DC; 2008.
32. Orme BK. Getting started with conjoint analysis: strategies for product design and pricing research. Research Publishers LLC: USA; 2020.

33. Orme BK, Chrzan K. Becoming an expert in conjoint analysis: choice modeling for pros. Sawtooth Software, Inc.: North Orem (UT); 2017.
34. Croissant Y. Estimation of random utility models in R The mlogit Package. *J Stat Soft.* 2020;95(11):1–41.
35. Sarrias M, Daziano R. Multinomial logit models with continuous and discrete individual heterogeneity in R The gmn Package. *J Stat Soft.* 2017;79(2):1–46.
36. White H. Estimation, inference, and specification analysis. Cambridge: Cambridge University Press; 1994.
37. Sawtooth Software, Inc. Technical paper series: the latent class technical paper V4.8. Utah; 2021.
38. Street AE, Street DJ, Flynn GM. Who gets the last bed? A discrete-choice experiment examining general population preferences for intensive care bed prioritization in a pandemic. *Med Decis Making.* 2021;41(4):408–18. <https://doi.org/10.1177/0272989X21996615>.
39. Greene WH, Hensher DA. A latent class model for discrete choice analysis: contrasts with mixed logit. *Transp Res B-Meth.* 2003;37(8):681–98.
40. Mühlbacher AC, Sadler A, Dippel F-W, et al. Treatment preferences in Germany differ among apheresis patients with severe hypercholesterolemia. *Pharmacoeconomics.* 2018;36(4):477–93.
41. Mühlbacher AC, Bethge S. Reduce mortality risk above all else: a discrete-choice experiment in acute coronary syndrome patients. *Pharmacoeconomics.* 2015;33(1):71–81.
42. Sinha P, Calfee CS, Delucchi KL. Practitioner's guide to latent class analysis: methodological considerations and common pitfalls. *Crit Care Med.* 2021;49(1):e63–79.
43. Tieosapjaroen W, Fairley CK, Chow EPF, et al. Preferences for weight gain compared with other antiretroviral therapy side effects in people living with HIV: a discrete choice experiment. *J Acquir Immune Defic Syndr.* 2022;91(3):305–11.
44. Byanyima W. In danger: UNAIDS global AIDS update 2022: foreword. UNAIDS: Geneva; 2022.
45. Cohen CJ, Meyers JL, Davis KL. Association between daily antiretroviral pill burden and treatment adherence, hospitalisation risk, and other healthcare utilisation and costs in a US medicaid population with HIV. *BMJ Open.* 2013. <https://doi.org/10.1136/bmjopen-2013-003028>.
46. Robert Koch-Institut (RKI). Epidemiologisches Bulletin: HIV in Deutschland 2021. Robert Koch-Institut (RKI): Berlin; 2022.
47. Mühlbacher A, Bethge S. What matters in type 2 diabetes mellitus oral treatment? A discrete choice experiment to evaluate patient preferences. *Eur J Health Econ.* 2016;17(9):1125–40. <https://doi.org/10.1007/s10198-015-0750-5>.
48. Mühlbacher AC, Junker U, Juhnke C, et al. Chronic pain patients' treatment preferences: a discrete-choice experiment. *Eur J Health Econ.* 2015;16(6):613–28. <https://doi.org/10.1007/s10198-014-0614-4>.
49. Norman R, Kemmler G, Viney R, et al. Order of presentation of dimensions does not systematically bias utility weights from a discrete choice experiment. *Value Health.* 2016;19(8):1033–8. <https://doi.org/10.1016/j.jval.2016.07.003>.
50. Farrar S, Ryan M. Response-ordering effects: a methodological issue in conjoint analysis. *Health Econ.* 1999;8(1):75–9.

Authors and Affiliations

Martin Emmert¹  · Stefan Rohrbacher¹ · Jennifer Jahn² · Katharina Fernando² · Michael Lauerer^{1,2}

✉ Martin Emmert
Martin.Emmert@uni-bayreuth.de

² GWS-Gesundheit Wissenschaft Strategie GmbH
(Health-Science-Strategy Ltd.), Bayreuth, Germany

¹ Faculty of Law, Business and Economics, Institute for Healthcare Management and Health Sciences, University of Bayreuth, Prieserstraße 2, 95444 Bayreuth, Germany

4 The Elicitation of Patient and Physician Preferences for Calculating Consumer-Based Composite Measures on Hospital Report Cards: Results of Two Discrete Choice Experiments¹¹

¹¹This chapter is based on an article published in *The European Journal of Health Economics* (DOI: <https://doi.org/10.1007/s10198-023-01650-2>); this was joint work with Martin Emmert, Florian Meier, Laura Heppe, Cordula Drach, Anja Schindler, Uwe Sander, Christiane Patzelt, Cornelia Frömke, Oliver Schöffski, and Michael Lauerer.



The elicitation of patient and physician preferences for calculating consumer-based composite measures on hospital report cards: results of two discrete choice experiments

Martin Emmert¹ · Stefan Rohrbacher¹ · Florian Meier² · Laura Heppe³ · Cordula Drach³ · Anja Schindler⁴ · Uwe Sander⁴ · Christiane Patzelt⁴ · Cornelia Frömke⁴ · Oliver Schöffski³ · Michael Lauerer¹

Received: 26 May 2023 / Accepted: 14 November 2023
© The Author(s) 2023

Abstract

Purpose The calculation of aggregated composite measures is a widely used strategy to reduce the amount of data on hospital report cards. Therefore, this study aims to elicit and compare preferences of both patients as well as referring physicians regarding publicly available hospital quality information

Methods Based on systematic literature reviews as well as qualitative analysis, two discrete choice experiments (DCEs) were applied to elicit patients' and referring physicians' preferences. The DCEs were conducted using a fractional factorial design. Statistical data analysis was performed using multinomial logit models

Results Apart from five identical attributes, one specific attribute was identified for each study group, respectively. Overall, 322 patients (mean age 68.99) and 187 referring physicians (mean age 53.60) were included. Our models displayed significant coefficients for all attributes ($p < 0.001$ each). Among patients, "Postoperative complication rate" (20.6%; level range of 1.164) was rated highest, followed by "Mobility at hospital discharge" (19.9%; level range of 1.127), and "The number of cases treated" (18.5%; level range of 1.045). In contrast, referring physicians valued most the "One-year revision surgery rate" (30.4%; level range of 1.989), followed by "The number of cases treated" (21.0%; level range of 1.372), and "Postoperative complication rate" (17.2%; level range of 1.123)

Conclusion We determined considerable differences between both study groups when calculating the relative value of publicly available hospital quality information. This may have an impact when calculating aggregated composite measures based on consumer-based weighting.

Keywords Public reporting · Hospital choice · Discrete choice experiment · Composite measures · Hospital report cards

JEL Classification I12

✉ Martin Emmert
Martin.Emmert@uni-bayreuth.de

¹ Faculty of Law, Business and Economics, Institute for Healthcare Management and Health Sciences, University of Bayreuth, Prieserstraße 2, 95444 Bayreuth, Germany

² Department of Management and Economics, SRH Wilhelm Löhe University of Applied Sciences, 90763 Fürth, Germany

³ School of Business and Economics, Chair of Health Care Management, Friedrich-Alexander-University of Erlangen-Nuremberg, Lange Gasse 20, 90403 Nuremberg, Germany

⁴ Department of Information and Communication, Faculty for Media, Information and Design, University of Applied Sciences and Arts, Hannover, Germany

Introduction

Public reporting of performance information encompasses data, publicly available or available to a broad audience free of charge or at a nominal cost, about a health-care structure, process, or outcome at any provider level (e.g., hospitals) [1]. The overall aim of public reporting is to improve health-care quality by both stimulating quality improvement on the provider level ("Improvement Through Changes in Care") and also by helping patients, referring physicians, and other consumers select the "right" provider ("Improvement Through Selection") [2, 3]. Therefore, hospital report cards (HRCs) publicly display quality-related information about hospitals and enable hospital comparisons [4].

However, studies have shown heterogeneous results regarding the impact of HRCs on the choice behavior of consumers [5–10]. To increase the uptake of HRCs of consumers, HRCs should present the information that consumers value most when making hospital choices [11]. Besides, it should be attempted to limit the richness of consumers' choice sets, i.e., the number of hospitals as well as the variety of quality measures [12].

In this context, one promising strategy is to provide customized information that both reflects the preferences of consumers and reduces the amount of data by aggregating individual performance measures into summary scores (i.e., composite measures) [11, 13–17]. For example, a recent study has shown that the introduction of summary measures on Nursing Home Compare—a web-based guide detailing quality of care at over 17,000 Medicare- or Medicaid-certified nursing homes in the United States—was associated with a significant change in consumer demand for low- and high-scoring facilities [18]. So far, composite measures have been published in several countries such as the United States, the UK, and Germany (e.g., Hospital Compare, Nursing Home Compare, NHS Choices, US News Best Hospital, AOK Hospital Navigator) [17–19].

There are different approaches to calculate aggregated composite measures such as opportunity-based weights, numerator-based weights, all-or-none measures, expert panels, and consumer-based weighting [16]. As shown, consumer-based weighting aligns mostly closely with patient-centered care and should, therefore, be prioritized [16]; for this purpose, it is crucial to learn more about consumers' preferences when choosing a hospital. So far, current composite measure approaches do not consider consumer preferences in more detail when calculating summary scores. For example, the overall star rating system for hospitals for the Centers for Medicare and Medicaid Services (CMS) in the United States—which is publicly displayed on Hospital Compare—is calculated by taking the weighted average of the hospital's scores across five areas of quality (i.e., mortality, safety of care, readmission, patient experience, and timely and effective care). Therefore, the relative weight for mortality, safety, readmission, and patient experience is worth 22% each, and timely and effective care is worth 12% of the overall score [20].

In addition, most research has been conducted to investigate whether patients—as the main target group of HRCs—use publicly reported quality information to search for and select hospitals [21–23]. However, less information is available regarding whether publicly available quality information plays a role from the referring physicians' perspective [24, 25]. This seems to be surprising since most patients trust their referring physicians' recommendation regarding what hospital to choose [25, 26]. Therefore, referring physicians should be regarded as another major target group of HRCs

so as to direct patients to well-performing hospitals and to increase the impact of public reporting [24, 27].

In this context, the present study aims to identify the most important publicly reported hospital quality information for hospital choice for elective hip replacement surgery and to determine their relative importance among both patients as well as referring physicians. We focused on elective hip replacement surgery for the following reasons: first, we aimed to address one procedure that is both standard for public reporting initiatives as well as included in German hospital quality assurance initiatives [28, 29]. Second, non-acute procedures should be prioritized since public reporting aims to support patients, referring physicians, and other consumers to select the “right” provider. [2, 3] Third, elective hip replacement surgery is one of the most frequently conducted procedures in Germany; for example, 160,910 surgeries were performed in 1,254 hospitals in 2020 [30]. The results might help us learn how to weight different quality measures when calculating composite measures. In more detail, we sought to address the following three questions: (1) what is the essential publicly available hospital quality information for choosing a hospital for elective hip replacement surgery from the perspective of patients and referring physicians? (2) How do patients and referring physicians' rate different publicly available hospital quality information? (3) What relative importance do both patients and referring physicians assign to different quality information?

Materials and methods

This study used a mixed methods approach. After performing a systematic literature review, we conducted qualitative research methods (i.e., semi-structured interviews) to identify and select the most important quality measures for choosing a hospital for hip replacement surgery from the perspective of both patients and referring physicians. Based on this, two separate discrete choice experiments (DCEs) were developed and performed to elicit and compare patients' and referring physicians' preferences for all relevant hospital choice characteristics and to determine the relative value of each quality measure. DCEs are increasingly used in the health-care context to inform on consumer preferences for health-care services such as hospital choice [31]. DCEs are a stated preference method that use (survey) data to systematically elicit individuals' preferences based on a series of hypothetical choice scenarios (termed choice sets) [32]. Two key economic theories (i.e., Random Utility Theory and Lancaster's Theory [33, 34]) suggest that respondents choose the option from each choice set which provides them with the most satisfaction or “utility” [32]. Based on the decisions, DCEs can help understand which characteristics (termed attributes) are preferred by

consumers and determine the relative value of each attribute. [32, 35]

Study design

The design and analysis of both DCEs were based on standardized research practices for undertaking conjoint analysis of the ISPOR Conjoint Analysis Good Research Practices Task Force [36–38]. Ethics approval for both studies was obtained from the Friedrich-Alexander-University Erlangen-Nuremberg Ethics Board (196_19 B). Informed consent was obtained from each study participant.

Systematic search procedure

First, we conducted two systematic search procedures on Medline (via PubMed) and the Cochrane Library to identify studies which aimed to identify relevant criteria for hospital choice for hip replacement surgery from the perspective of both patients and referring physicians. The searches were carried out in September 2020 and aimed to identify English and German language literature published since 2010. In addition, reference lists of identified research articles were screened for further articles. The reviews were complied with the Guideline from the Cochrane Collaboration. [39]

The search strategy addressing the *perspective of patients* was segmented into three components. The first component referred to hospitals (e.g., hospital, clinic), the second component referred to choice (e.g., choice, selection), and the third component addressed hip replacement surgery (e.g., hip, coxarthrosis). As a result, 5,510 potentially relevant papers were identified. After eliminating duplicates and judging titles and abstracts in a first step as well as full papers in a second step, ten studies were considered relevant [40–49]. In sum, 73 individual criteria were derived from the 10 studies. We added four further quality measures from the German hospital quality report administered by the Institute for Quality Assurance and Transparency in Healthcare (IQTIG). This led to a final sample of 77 quality measures which were then merged into 23 more general measures (e.g., the following three individual criteria “28-day mortality rate (%)”, “28-day mortality rate following discharge”, and “Postoperative mortality” were merged into the category “Mortality rate”). Afterward, we excluded ten criteria which are not available in the German hospital sector for public reporting purposes (e.g., waiting time, 28-day emergency readmission rate), so that 13 criteria remained for the next step (See Supplemental Material 1).

The search strategy addressing the *perspective of referring physicians* was based on previously published literature and slightly modified [50]. As a result, 2,246 potentially relevant papers were identified. After eliminating duplicates and judging titles and abstracts in a first step as well as full

papers in a second step, 16 studies were considered relevant [25, 51–65]. In sum, 39 criteria were derived from those studies. Then we excluded 20 criteria that are not publicly available (see above) (e.g., waiting times, MRSA events), so that 19 criteria remained.

Qualitative steps (semi-structured interviews)

In total, we qualitatively surveyed (January 2021 to February 2021) 20 randomly selected hip replacement surgery patients (mean age 63.9 years, 55% female) from a German statutory health insurance who had undergone elective hip arthroplasty surgery within 3 years prior to answering the survey as well as 15 referring physicians (mean age 56.3 years, 20% female). Written informed consent was obtained from all participants. Respondents were sent a short survey through postal mail before conducting the semi-structured interviews to learn more about (the) past hospital choice(s) and the importance of the 13 resp. 19 criteria for choosing a hospital. After receiving the short survey, semi-structured interviews were conducted to explore the stated preferences, to determine the most relevant criteria as well as corresponding levels, to clarify the wording, and to evaluate the comprehensibility of hypothetical choice tasks for the DCE (see below). In particular, we evaluated the comprehensibility of the labels used for the attributes and their corresponding descriptions to assure the intended interpretation of the explanations. Based on all steps, five attributes were derived which were of major importance for choosing a hospital from the perspective of both patients and referring physicians which were as follows: the rate of confirmed diagnosis prior the surgery, the certification as an Endoprosthetics Center, the number of cases treated, the rate of postoperative complications, as well as the rate of mobility at hospital discharge. This was supplemented with the assessment of actions to prevent falls of patients for patients and the 1-year revision surgery rate for referring physicians, respectively (Table 1). Individuals who completed the qualitative study received 50 Euro.

Quantitative steps

The survey instrument

Both survey instruments were similar and consisted of four parts. First, we asked for general sociodemographic information (e.g., age, gender) before collecting information about the experience with HRCs when selecting hospitals in the second part. Second, respondents were presented with publicly available hospital quality information items (see above) as well as short descriptions and were asked to rate each item on a 1–5 scale (1 = not all important; 5 = very important). Furthermore, we asked the respondents to select the single

Table 1 Summary of attributes and levels used in the discrete choice experiments (DCE)

Attribute	Description and levels given in DCE questionnaire
Confirmed diagnosis rate ^{§,#}	<i>Hip replacement primary implantation with fulfilled indication criteria</i> Quality targets reached Quality targets not reached No assessment intended resp. results not (yet) available
EndoCert certificate ^{§,#}	<i>Hospitals which meet requirements of the EndoCert certification system</i> No certificate Certified EndoProstheticsCenter (EPZ) Certified EndoProstheticsCenter of Maximum Care (EPZmax)
The number of cases treated ^{§,#}	<i>The number of total elective hip replacement surgeries in a hospital</i> 84 patients (below average) 148 patients (average) 230 patients (above average)
Postoperative complication rate ^{§,#}	<i>Postoperative specific complications after hip replacement surgery (e.g., infections, bleeding, thromboembolism)</i> Quality targets reached Quality targets not reached No assessment intended resp. results not (yet) available
Mobility at hospital discharge ^{§,#}	<i>Functional mobility status on discharge after hip replacement surgery</i> Quality targets reached Quality targets not reached No assessment intended resp. results not (yet) available
Prevention of falls measures [§]	<i>Hospitals should be active in minimizing fall risk among patients who have undergone hip replacement surgeries</i> Quality targets reached Quality targets not reached No assessment intended resp. results not (yet) available
1-year revision surgery rate [#]	<i>Percentage of patients who underwent revision surgery within one year after initial surgery</i> 1.41% (better than average) 2.48% (average) 3.89% (lower than average)

[§]This attribute applies for patients

[#]This attribute applies for referring physicians

most important information item for the hospital choice. Afterward, all respondents were asked to respond to the DCE survey block. Both DCE experiments were designed using Sawtooth Software Lighthouse Studio 9.11.0.

Instead of ranking or rating different quality measures, as is done in traditional importance elicitation formats, DCEs perform a pairwise comparison of hypothetical alternatives (i.e., differently configured hospitals) and ask the participants to choose between them [66]. Both experiments were designed as a full profile design (i.e., each choice set included all six attributes) and were generated using the balanced overlap method so as to achieve standard errors below 0.05 for main effect utilities and 0.10 or smaller for interaction effects and the highest D-efficiency score [67, 68]. We created 10 versions (termed blocks) of the DCE, each containing 10 choice tasks, thus

generating 100 unique choice tasks (fractional factorial design). Choice tasks were generated that maximized both balance (i.e., meaning each level appears with the same frequency) and orthogonality (i.e., meaning each pair of levels appears with the same frequency across each pair of attributes) [69]. The choice sets in both surveys were designed as forced-choice tasks; this means that respondents had to choose one of two hypothetical hospitals by making trade-offs between attributes and their levels. This method offers practical advantages such as closeness to reality as trade-off decisions are part of everyday life [66]. As stated above, hospitals differed in six attributes with three levels each. In DCE research, four to eight attributes per choice set are seen as appropriate [70, 71]. Figure 1 provides one hypothetical choice task as an example. The questionnaire was pilot tested for clarity and understanding

Which of the following two hospitals would you choose?

	Hospital 1	Hospital 2
Confirmed diagnosis rate	Quality targets reached	Quality targets not reached
Endocert Certificate	Certified EndoProstheticsCenter of Maximum Care (EPZmax)	No certificate
The number of cases treated	230 Patients (above average)	84 Patients (below average)
Prevention of falls measures	Quality targets not reached	Quality targets reached
Postoperative complication rate	Quality targets reached	Quality targets not reached
Mobility at hospital discharge	Quality targets reached	Quality targets not reached
	Hospital	Hospital

Fig. 1 Example choice task for patients

with 15 hip replacement patients as well as 15 referring physicians and slightly modified accordingly.

The survey sample

Final questionnaires were sent through postal mail to 1000 randomly selected people from a German statutory health insurance (AOK Lower Saxony) who had undergone elective hip arthroplasty surgery within 3 years prior to answering the survey. One reminder was sent out 2 weeks after the initial invitation. Furthermore, we purchased a database containing contact information (e.g., postal address, email) for orthopedists in the German outpatient sector from a commercial provider (ArztData AG). This database covers about 88% of all orthopedists in the German outpatient sector [72]. We used a sequential mixed-mode strategy to achieve high response rates. In a first step, 1,650 referring orthopedists were contacted via email, which contained a link to participate online (web-based survey). After 1 week, a first reminder was sent out. In a second step, the remaining orthopedists were contacted via postal mail and received a printed version of the survey; after 2 weeks, we sent out a second reminder. The surveys were conducted between

March and June 2021 in the German language. As an incentive, respondents received a payment of approximately €50.

Sample size

Johnson’s often-used rule-of-thumb calculates a sample of 75 participants for a DCE having our design specifications (i.e., ten choice tasks per respondent, two alternatives, three levels per attribute as maximum) [73]. This sample size was discussed as being the lowest limit for main effects estimation. However, we aimed at doubling this number (i.e., to include at least 150 participants) following more advanced recommendations for statistical robustness [74, 75].

Statistical methods

Analyses of general survey questions were performed using SPSS (IBM Corp. Released 2019. IBM SPSS Statistics for Windows, Version 26.0. Armonk, NY: IBM Corp). Descriptive statistics (i.e., means for continuous variables, percentages for categorical variables) were used to examine the demographic and experience variables.

For analyzing the DCEs, we performed multinomial logit models using R (Version 4.2.2) and the corresponding

mlogit-package by Croissant (2020) [76]. We constructed the data sets in long format. Thus, the data set about patients included 6,440 observations (322 patients $\times$ 10 choice tasks per subject $\times$ 2 alternatives per choice set). Similarly, the data set for physicians comprised 3,740 observations (187 physicians $\times$ 10 choice tasks per subject $\times$ 2 alternatives per choice set). As for data encoding, we started from assigning each variable equidistant numerical values of scale {0; 1; 2} as described by Mühlbacher et al. [77]. For example, regarding the attribute “Confirmed diagnosis rate”, we designated the value {0} to “Quality targets not reached”, value {1} to “No assessment intended resp. results not (yet) available” (please note: due to the implicit uncertainty of this alternative, we assumed that the absence of results lay between the certain answers “targets reached” and “targets not reached”), and value {2} to “Quality targets reached”. Afterward, we repeated the estimations treating the attributes as categorical variables. Both approaches provided qualitatively very similar results, whereby the estimates from categorical attributes could be shown to be more precise.

We employed effect coding by following standard guidelines [38, 78]. Therefore, we expressed every attribute as $n-1$ variables where n is the number of levels. Each variable corresponds to one level, while the omitted level serves as the reference. If the variable matches the level presented in the corresponding profile, we assigned value {1}. If the level in the corresponding profile equals another non-omitted level, we assigned value {0}. If the level in the corresponding profile is the omitted level, we assign value {-1}. For example, the attribute “Confirmed diagnosis rate” was operationalized by three possible levels: “Quality targets reached”, “Quality targets not reached”, “No assessment intended resp. results not (yet) available”. As reference level, we chose “No assessment intended resp. results not (yet) available”. Hence, we constructed two variables; the first covered the level “Quality targets reached”, the other one the level “Quality targets not reached”. If the given occurrence equals the reference, we filled both variables with {-1}, otherwise we assigned values {1} or {0}. As reference categories, we assigned “Certified Endoprosthetics Center of Maximum Care (EPZmax)” for the attribute “Endocert Certificate”, middle positions for the attributes “Number of cases treated” and “One-year revision surgery rate”, and finally “No assessment intended resp. results not (yet) available” for all remaining attributes. We chose effect coding over dummy coding since with dummy coding, the parameter estimate for the (omitted) baseline category is equal to zero and cannot be recovered. Thus, the estimates of the other levels are only relative to the benchmark level. Instead, with effect coding, we can extract the parameter of the reference category from the negative sum of the included categories and the standard error from the covariances of the included categories. This enables us to compare all levels of the attributes against the corresponding

mean value with the sign of the coefficients indicating a positive or negative impact compared to the mean of the attribute and the magnitude specifying the size of the effect [38, 78, 79]. In all approaches, we applied Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and log-likelihood value (LL) to identify the best fitting model. The relative importance of each attribute was determined based on each attribute’s coefficient range. This means that we calculated the difference between the coefficients of the highest and lowest level of each attribute and expressed this as a share of the total range across all attributes.

Results

Sample characteristics

Overall, 378 patients (37.8% response rate) and 267 referring orthopedists (16.2% response rate) participated in our study and returned the survey. The following analysis reports on the 322 patients and 187 orthopedists who fully completed the DCE component of the questionnaire and provided consistent responses (see Table 2 for the demographic characteristics of both study groups). The mean age of patients was 68.99 (SD 10.44) years. Slightly more than half of patients (51.6%) were female, and most patients (53.7%) stated secondary general school or less as the highest educational level. Approximately one in four patients was aware of HRCs when answering the survey (26.1%). The mean age of participating orthopedists was 53.60 (SD 8.14) years, a large majority of respondents were male (90.4%), and 39 respondents (21.2%) indicated prior HRC experience in searching for a hospital. Almost three out of four surveyed referring orthopedists (72.7%) stated that they perceived substantial differences in the quality of care between hospitals.

Descriptive rating and ranking results

The results regarding the importance of different information items for the hospital choice—on a 1–5 scale (1 = not all important; 5 = extremely important)—showed that patients rated confirmed diagnosis (hip surgery) rate (4.65 ± 0.80), mobility at hospital discharge (4.60 ± 0.73), and complication rate (4.54 ± 0.81) as most important (see Supplemental material 2). In contrast, referring physicians rated complication rate (4.19 ± 1.06), the number of cases treated (4.09 ± 0.98), and the 1-year revision surgery rate (4.00 ± 1.21) as most important. Furthermore, the survey results for the single most important information item for the hospital choice revealed confirmed diagnosis (hip surgery) rate (44.5%) as well as the number of cases treated (35.9%) as most relevant from the perspective of patients and referring physicians, respectively.

Table 2 Key characteristics for both patients ($n=322$) and referring physicians ($n=187$)

	Patients ($n=322$)		Referring physicians ($n=187$)	
	Mean (SD) or n	Range or %	Mean (SD) or n	Range or %
Age				
Mean (SD)	68.99 (10.44)	35–89	53.60 (8.14)	33–73
Gender				
Male	151	46.9%	169	90.4%
Female	166	51.6%	18	9.6%
Missing	5	1.6%	0	0.0%
Educational attainment				
Secondary general school or less	173	53.7%	n.a	n.a
Intermediate secondary school	93	28.9%	n.a	n.a
(Technical) University entrance qualification	38	11.8%	n.a	n.a
Missing	18	5.6%	n.a	n.a
Chronic conditions				
Any chronic condition	141	43.8%	n.a	n.a
No chronic condition	159	49.4%	n.a	n.a
Missing	22	6.8%	n.a	n.a
Health Status				
Good or better	168	52.2%	124	66.3%
Satisfactory	122	37.9%	51	27.3%
Bad or worse	26	8.1%	12	6.4%
Missing	6	1.9%	0	0.0%
Perceived differences in the hospital quality				
Big differences	81	25.2%	136	72.7%
Small differences	89	27.6%	48	25.7%
No differences	23	7.1%	2	1.1%
I don't know	122	37.9%	1	0.5%
Missing	7	2.2%	0	0.0%
Knowledge of hospital report cards				
Yes	84	26.1%	39	21.2%
No	188	56.4%	145	78.8%
Missing	50	15.5%	0	0.0%
Practice type				
Single physician practice	n.a	n.a	65	34.8%
Group physician practice	n.a	n.a	76	40.6%
Outpatient-based health-care center	n.a	n.a	40	21.4%
Other	n.a	n.a	6	3.2%
Attending doctor (i.e., with hospital affiliation)				
Yes and I perform the surgery myself	n.a	n.a	40	21.4%
Yes but I do not perform the surgery myself	n.a	n.a	21	11.2%
No	n.a	n.a	124	66.3%
Missing	n.a	n.a	2	1.1%

Results of the discrete choice experiments

The applied models displayed significant coefficients for all attributes included in both analyses ($p < 0.001$ each). This means that, independent of the placement, all attributes were relevant to the decision of both patients and referring physicians (Supplemental material 3). [Please also note that some

mean levels could not be shown to be statistically significant; for example, the level “No assessment intended/results not (yet) available” did not reach statistical significance in case of the “Confirmed diagnosis rate” and the “Mobility of hospital discharge”.] Among patients, the highest preference weights for hospital choice were calculated for “Postoperative complication rate” (coef.: 0.560; SE: 0.037), “Mobility

at hospital discharge” (coef.: 0.551; SE: 0.036), and “The number of cases treated” (coef.: 0.547; SE: 0.036). The fourth position was occupied by “Confirmed diagnosis (hip surgery) rate” (Coef.: 0.437; SE: 0.033), followed by “Prevention of falls measures” (Coef. 0.431; SE: 0.035), and “Endocert Certificate” (Coef. 0.341; SE: 0.033). In contrast, referring physicians showed a clear preference for a decrease in “One-year revision surgery rate” (coef.: 1.014; SE: 0.053). The following positions were occupied by “The number of cases treated” (coef.: 0.696; SE: 0.051) and “Postoperative complication rate” (coef.: 0.546; SE: 0.050). Here, the “Confirmed diagnosis (hip surgery) rate” (coef.: 0.390; SE: 0.052), “Endocert Certificate” (coef.: 0.310; SE: 0.047), and the “Mobility at hospital discharge” (coef.: 0.236; SE: 0.046) were less important for referring physicians. The same order can be derived from the calculation of the relative importance of all attributes for both study groups. The relative importance was calculated based on each attribute’s coefficient range (i.e., the difference between the coefficients of the highest and lowest level of each attribute) expressed as a share of the total range across attributes (Table 3). Table 3 also presents the odds ratio (OR) for each level to analyze its impact on the hospital choice. For example, the likelihood of voting a hospital with a successful result regarding the “Confirmed diagnosis rate” (i.e., quality targets reached) is 1.535 times higher than the probability of choosing a hospital with the mean of “confirmed diagnosis rate”.

As shown in Fig. 2, the preference pattern of patients showed a relatively small range of results. The following three attributes, “Postoperative complication rate” (20.6%; level range of 1.164), “Mobility at hospital discharge” (19.9%; level range of 1.127), and “The number of cases treated” (18.5%; level range of 1.045) were weighted highest. The lowest weight was detected for “Endocert Certificate” (11.4%; level range of 0.646). In contrast, the preference pattern of referring orthopedists showed larger differences in the relative importance of presented quality information measures. Here, the “One-year revision surgery rate” was identified as the main dominant attribute (30.4%; level range of 1.989). Furthermore, “The number of cases treated” (21.0%; level range of 1.372) and “Postoperative complication rate” (17.2%; level range of 1.123) were relatively highly weighted. Here, the lowest weight was detected for “Mobility at hospital discharge” (8.2%; level range of 0.536).

Discussion

The objective of the present study was to identify the most important publicly reported hospital quality information for hospital choice and to determine their relative importance among both former patients and referring physicians

in Germany. The results might assist us in learning how to weight different quality measures when calculating aggregated patient-centered composite measures.

Based on our results, several conclusions can be drawn. First, patients and referring physicians selected similar publicly available quality information for the hospital choice. However, it should be mentioned that we focused on publicly available quality information and did not include information which is not available for public reporting purposes or which is not routinely collected in the health-care system in our experiments. Therefore, we cannot exclude that the integration of such information might have led to different findings. For example, Geraedts and colleagues showed that many surgical patients decided in favor of a hospital because of the trust they had built up through previous treatment in the hospital, which they experienced as satisfactory [80].

Second, HRCs should present the information that consumers value most when making hospital choices. At least five of the seven most relevant publicly reported quality information items are displayed on most German HRCs (e.g., Weisse Liste, AOK Hospital Navigator) [28]. This is mainly due to the fact that those information items are part of the German external hospital quality assurance system whose results are made available for public reporting purposes [81]. In contrast, the certification as a Certified Endoprosthetics Center is only presented on a few HRCs so far (e.g., Weisse Liste). Besides, the 1-year revision rate is only presented on the hospital navigator of the Allgemeine Ortskrankenkasse (AOK) [81]. The AOK is a major provider of statutory health insurance in the German health-care system and calculates further quality information based on claims data [29]. Following this, providers of HRCs might compare whether the most relevant quality information is already included on its own HRC. If not, providers might consider adjusting the content as a first step toward more consumer orientation.

Third, the study’s DCE-based findings appear broadly consistent with the results from the rating-based survey. For example, the preference pattern of patients showed a relatively small range of results. While the rating-based findings (on a scale of 1–5) varied between 4.09 ± 1.03 (Endocert Certificate) and 4.65 ± 0.80 (confirmed diagnosis hip surgery rate), the DCE-based relative importance values ranged between 11.4% (“Endocert Certificate”) and 20.6% (Postoperative complication rate), respectively. In both cases, “Endocert Certificate” was the least important attribute for the hospital choice while “Complication rate” and “Mobility at hospital discharge” were among the most important attributes for the hospital choice. However, the ranking-based results showed slightly different findings. While the confirmed diagnosis (hip surgery) rate (44.5%) was shown to be the single most important information item for the hospital choice, the DCE-based

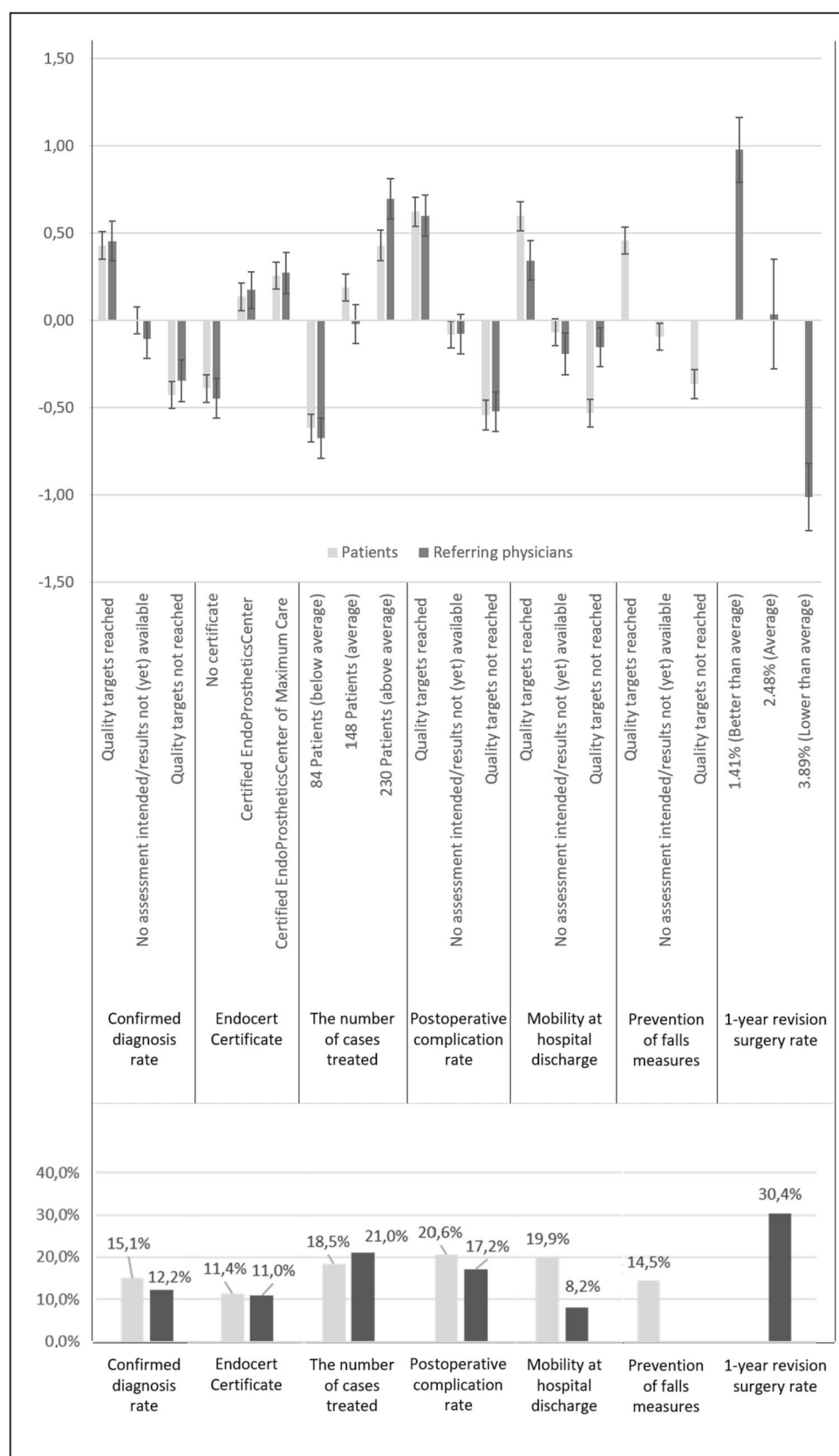
Table 3 Estimated parameters of multinomial logit (MNL) models

Attributes and levels	Patients (<i>n</i> = 322)					Referring physicians (<i>n</i> = 187)								
	Coeff	OR	SE coeff	95% CI	p	Difference (high-est—lowest level)	Coeff	OR	SE coeff	95% CI	p	Difference (high-est—lowest level)		
Confirmed diagnosis rate ^{§, #}														
Quality targets reached	0.429	1.535	0.040	0.350	0.507	<.001	0.856	0.454	1.575	0.058	0.340	0.568	<.001	0.800
No assessment intended/results not (yet) available	− 0.001	0.999	0.039	− 0.078	0.076	0.977		− 0.108	0.897	0.056	− 0.218	0.001	0.051	
Quality targets not reached	− 0.427	0.652	0.038	− 0.503	− 0.352	<.001		− 0.346	0.708	0.062	− 0.467	− 0.224	<.001	
[#] Endocert Certificate [§]														
No certificate	− 0.390	0.677	0.040	− 0.469	− 0.311	<.001	0.646	− 0.446	0.640	0.058	− 0.561	− 0.331	<.001	0.718
Certified EndoProstheticsCenter	0.135	1.144	0.040	0.055	0.214	0.001		0.174	1.191	0.054	0.068	0.281	0.001	
Certified EndoProstheticsCenter of Maximum Care	0.256	1.291	0.039	0.179	0.332	<.001		0.272	1.312	0.059	0.155	0.388	<.001	
The number of cases treated ^{§, #}														
84 patients (below average)	− 0.616	0.540	0.040	− 0.695	− 0.537	<.001	1.045	− 0.676	0.509	0.058	− 0.791	− 0.560	<.001	1.372
148 patients (average)	0.188	1.207	0.039	0.112	0.264	<.001		− 0.020	0.980	0.056	− 0.131	0.090	0.716	
230 patients (above average)	0.428	1.535	0.044	0.341	0.516	<.001		0.696	2.006	0.058	0.581	0.811	<.001	
Postoperative complication rate ^{§, #}														
Quality targets reached	0.622	1.863	0.042	0.540	0.704	<.001	1.164	0.600	1.823	0.061	0.481	0.720	<.001	1.123
No assessment intended/results not (yet) available	− 0.080	0.923	0.039	− 0.157	− 0.003	0.040		− 0.078	0.925	0.057	− 0.190	0.034	0.170	
Quality targets not reached	− 0.542	0.582	0.044	− 0.628	− 0.455	<.001		− 0.522	0.593	0.057	− 0.636	− 0.409	<.001	
Mobility at hospital discharge ^{§, #}														
Quality targets reached	0.597	1.817	0.043	0.513	0.682	<.001	1.127	0.344	1.410	0.057	0.231	0.456	<.001	0.536
No assessment intended/results not (yet) available	− 0.068	0.935	0.039	− 0.145	0.010	0.085		− 0.192	0.826	0.061	− 0.313	− 0.071	0.002	
Quality targets not reached	− 0.530	0.589	0.041	− 0.610	− 0.449	<.001		− 0.152	0.859	0.056	− 0.263	− 0.042	0.007	
Prevention of falls measures [§]														
Quality targets reached	0.458	1.581	0.039	0.381	0.535	<.001	0.822	n.a	n.a	n.a	n.a	n.a	n.a	
No assessment intended/results not (yet) available	− 0.094	0.911	0.040	− 0.172	− 0.016	0.018		n.a	n.a	n.a	n.a	n.a	n.a	
Quality targets not reached	− 0.364	0.695	0.042	− 0.447	− 0.281	<.001		n.a	n.a	n.a	n.a	n.a	n.a	
1-year revision surgery rate [#]														
1.41% (better than average)	n.a	n.a	n.a	n.a	n.a	n.a	n.a	0.977	2.657	0.094	0.791	1.163	<.001	1.989
2.48% (average)	n.a	n.a	n.a	n.a	n.a	n.a		0.035	1.035	0.159	− 0.279	0.348	0.827	
3.89% (lower than average)	n.a	n.a	n.a	n.a	n.a	n.a		− 1.012	0.363	0.097	− 1.203	− 0.821	<.001	
Model constant (Intercept)	− 0.309	0.734	0.047	− 0.402	− 0.216	<.001		n.a	n.a	n.a	n.a	n.a	n.a	

OR odds ratio (OR = exp(coefficient)); SE standard errors, CI confidence interval

[§]This attribute applies for patients[#]This attribute applies for referring physicians

Fig. 2 Graphic display of level estimates with 95% CI (discrete choice experiment) and mean relative importance of attributes for both subgroups for the hospital choice



findings revealed a relative importance value of 15.1%. In addition, the overall results regarding the preference pattern of referring physicians were shown to be very consistent, regardless of the methodology. For example, the three attributes “The number of cases treated”, “Complication rate”, and “1-year revision surgery rate” showed the strongest results. In contrast, the “Confirmed diagnosis (hip surgery) rate”, “Endocert Certificate”, and “Mobility at hospital discharge” were shown to be less important.

Fourth, the results of this study could be used for the calculation of composite measures [11, 13–17]. Therefore, we applied a consumer-based weighting approach as recommended and determined the relative weight of relevant information [16]. For example, the preference pattern of referring orthopedists showed the “One-year revision surgery rate” to be most important with a relative weight of 30.4%. The challenge now is to convert the hospital quality results into (numeric) scores for calculating the composite measure. As stated by van Til and colleagues [82], the overall preference for alternatives (i.e., hospitals) might be estimated based on the sum of the part-worth utilities for the selected level of all attributes. For example, a clinic certified as an Endoprosthetics Center of Maximum Care (coef.: 0.272), with a below-average 1-year revision surgery rate (coef.: 0.977), an above-average number of cases (coef.: 0.696), that also achieves the statutory quality targets regarding confirmed diagnosis rate (coef.: 0.454), postoperative complication rates (coef.: 0.600), and mobility at hospital discharge (coef.: 0.344), would have an overall score of 3.343 from the perspective of referring physicians. In contrast, an overall score of -3.194 would be calculated for a hospital that achieves the most negative level expressions for each attribute. [Please consider this example to be simplified. A more advanced approach would consider the corresponding utility function to estimate the expected overall utility of each alternative (i.e., hospital).] Based on this, hospitals could be ranked and then grouped into several performance groups. Therefore, our study has shown that consumer preferences might differ substantially between target groups and each preference structure should be considered individually when developing rating systems based on composite measures. However, it should also be noted that our approach has certain limitations that must be considered carefully. For example, the attribute levels are used to operationalize the alternatives included in the choice sets. Therefore, it is important to select attributes and corresponding levels that properly describe the health-care product or service (i.e., the hospitals). If the levels are not defined in the appropriate range, the estimated coefficients could be biased [83]. Other methodological issues might also have an impact on the results derived by DCE studies and should be considered carefully [83, 84].

Finally, the results might serve as the basis for improved physician–patient communication. The findings might increase the mutual understanding of hospital choice preferences. In particular, referring physicians might better understand what patients actually want and which hospital they would prefer. They might have an important “agent” function in the hospital choice especially for patients with little time prior to admission and those who do not decide themselves [85]. Therefore, by understanding factors that influence patients’ hospital choice decisions, referring physicians may become more sensitive to patients’ preferences, which may have a positive effect on the process and outcomes of shared decision making [86]. This would increase the likelihood that patients may be satisfied with both the hospital choice process and the hospital choice itself [87]. Our study makes some contributions to such understanding, since it shows that, in contrast to referring physicians, the 1-year revision rate did not influence patients’ stated preferences, whereas the prevention of falls measures did.

As with any study, there are several limitations that have to be considered. First, it is important to mention that this study was conducted in Germany and might be of limited relevance for other countries. Nevertheless, the results presented in this paper are of interest for all countries with public reporting initiatives, such as the United States and others. For example, Hospital Compare in the United States already provides an overall composite measure for hospitals but does not consider preferences of patients and referring physicians in detail [20]. Second, it must be noted that the selected attributes for the DCEs referred to publicly available quality measures. Therefore, we cannot exclude that other quality measures are more relevant from the perspective of both study groups. Nevertheless, it seems reasonable for our study purpose to follow the approach used. In comparable studies, very similar approaches have been applied [47]. Third, our study analyses stated that preferences of the respondents to hypothetical scenarios and their actual responses may differ [47]. Fourth, the sample of surveyed referring physicians ($n = 187$) was smaller than that of patients surveyed ($n = 322$). In this regard, it should be noted that we aimed to reach a minimum number of 150 participants in both studies to meet Johnson’s often-used rule-of-thumb mentioned above [73], taking into account more advanced recommendations for statistical robustness [74, 75]. Following this, we were able to achieve our set target regarding the recruitment in both studies. Finally, the comprehensibility of the study attributes and corresponding levels is crucial for conducting DCEs and analyzing the corresponding data. During our study, we aimed to address this topic by means of different approaches. First, we qualitatively interviewed 20 patients who had undergone elective hip arthroplasty surgery within 3 years

prior to answering the survey. Thereby, we evaluated the comprehensibility of the labels used for the attributes and their corresponding descriptions in particular to assure the intended interpretation of the explanations. Based on those interviews, slight modifications were made to eliminate remaining uncertainties. Second, the respondents had to work with the presented study attributes and corresponding levels before conducting the choice-based part of the DCE. Both approaches might have positively influenced the comprehensibility of the study attributes and levels. In addition, the surveyed patient sample consisted of patients who had undergone elective hip arthroplasty surgery within 3 years prior to answering the survey. This might indicate—at least to a certain amount—some experience with hip-surgery-related issues.

In sum, this study adds to the literature by presenting results of 2 surveys among 322 patients and 187 referring physicians from the German outpatient sector to identify the most important publicly reported hospital quality information for hospital choice and to determine their relative importance. The results might support us in learning how to weight different quality measures when calculating aggregated composite measures based on consumer-based weighting. It could be shown that patients and referring physicians have similar interests with regard to the selection of publicly available quality information for the choice of hospitals. Nevertheless, we saw meaningful differences between patients and referring physicians in the relative weight of relevant information. This might have a significant impact on the hospital ranking results for each study group. Future research should address in detail how to convert hospital quality results into (numeric) scores for calculating composite measures. Besides, by understanding factors that influence patients' hospital choice decisions, referring physicians may become more sensitive to patients' preferences, which may have a positive effect on the process and outcomes of shared decision making.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10198-023-01650-2>.

Author contributions All authors contributed to the study conception. ME, SR, and LH developed the study design and preference elicitation framework, ME and SR derived the analytical results. The first draft of the manuscript was written by ME and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding Open Access funding enabled and organized by Projekt DEAL. Financial support for this study was provided entirely by The German health care Innovation Fund (01VSF18034). The funding agreement ensured the authors' independence in designing the study, interpreting the data, writing, and publishing the report.

Availability of data and materials Data are available on request from the corresponding author (martin.emmert@uni-bayreuth.de).

Declarations

Conflict of interest The authors declare that they have no competing interests.

Role of the funder/sponsor The funder had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Ethics approval and consent Ethics approval for this study was obtained from the Friedrich-Alexander-University Erlangen-Nuremberg Ethics Board (196_19 B).

Consent to participate Informed consent was obtained from each participant. All participants participated voluntarily and could end their participation at any time. All participants were informed about the study prior to participation and gave consent to use the data for publication.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Totten, A.M., Wagner, J., Tiwari, A., O'Haire, C., Griffin, J., Walker, M.: Public reporting as a quality improvement strategy: closing the quality gap: revisiting the state of the science. In: AHRQ Publication No. 12-E011-EF. Evidence Report/Technology Assessment; Number 208. (2012)
2. Berwick, D.M., James, B., Coye, M.J.: Connections between quality measurement and improvement. *Med. Care* **41**(1), 130–138 (2003)
3. Contandriopoulos, D., Champagne, F., Denis, J.-L.: The multiple causal pathways between performance measures' use and effects. *Med. Care Res. Rev.* **71**(1), 3–20 (2014). <https://doi.org/10.1177/1077558713496320>
4. Emmert, M., Schindler, A., Drach, C., et al.: The use intention of hospital report cards among patients in the presence or absence of patient-reported outcomes. *Health Policy* **126**(6), 541–548 (2022). <https://doi.org/10.1016/j.healthpol.2022.03.015>
5. Hussey, P.S., Luft, H.S., McNamara, P.: Public reporting of provider performance at a crossroads in the United States: summary of current barriers and recommendations on how to move forward. *Med. Care Res. Re. MCRR.* **71**(5 Suppl), 5S-16S (2014). <https://doi.org/10.1177/1077558714535980>
6. Metcalfe, D., Rios Diaz, A.J., Olufajo, O.A., et al.: Impact of public release of performance data on the behaviour of health-care consumers and providers. *Cochrane Database Syst. Rev.* **9**, CD004538 (2018). <https://doi.org/10.1002/14651858.CD004538.pub3>
7. Haviland, A.M., Damberg, C.L., Mathews, M., Paddock, S.M., Elliott, M.N.: Shifting from passive quality reporting to active

- nudging to influence consumer choice of health plan. *Med. Care Res. Rev.* (2018). <https://doi.org/10.1177/1077558718798534>
8. de Groot, I.B., Otten, W., Dijs-Elsinga, J., Smeets, H.J., Kievit, J., Marang-van de Mheen, P.J.: Choosing between hospitals: the influence of the experiences of other patients. *Med. Decis. Making* **32**(6), 764–778 (2012). <https://doi.org/10.1177/0272989X12443416>
 9. Pross, C., Busse, R., Geissler, A.: Hospital quality variation matters—a time-trend and cross-section analysis of outcomes in German hospitals from 2006 to 2014. *Health Policy* **121**(8), 842–852 (2017). <https://doi.org/10.1016/j.healthpol.2017.06.009>
 10. Strumann, C., Geissler, A., Busse, R., Pross, C.: Can competition improve hospital quality of care? A difference-in-differences approach to evaluate the effect of increasing quality transparency on hospital quality. *Eur. J. Health Econ.* **23**(7), 1229–1242 (2022). <https://doi.org/10.1007/s10198-021-01423-9>
 11. Sandmeyer, B., Fraser, I.: New evidence on what works in effective public reporting. *Health Serv. Res.* **51**(Suppl 2), 1159–1166 (2016). <https://doi.org/10.1111/1475-6773.12502>
 12. Schlesinger, M., Kanouse, D.E., Martino, S.C., Shaller, D., Rybowski, L.: Complexity, public reporting, and choice of doctors: a look inside the blackest box of consumer behavior. *Med. Care Res. Rev.* **71**(5 Suppl), 38S–64S (2014). <https://doi.org/10.1177/1077558713496321>
 13. Smith, M.A., Bednarz, L., Nordby, P.A., et al.: Increasing consumer engagement by tailoring a public reporting website on the quality of diabetes care: a qualitative Study. *JMIR.* **18**(12), e332 (2016). <https://doi.org/10.2196/jmir.6555>
 14. Huckman, R.S., Kelley, M.A.: Public reporting, consumerism, and patient empowerment. *N. Engl. J. Med.* **369**(20), 1875–1877 (2013). <https://doi.org/10.1056/NEJMp1310419>
 15. Sinaiko, A.D., Eastman, D., Rosenthal, M.B.: How report cards on physicians, physician groups, and hospitals can have greater impact on consumer choices. *Health Affairs (Project Hope)*. **31**(3), 602–611 (2012). <https://doi.org/10.1377/hlthaff.2011.1197>
 16. Schwartz, M., Restuccia, J.D., Rosen, A.K.: Composite measures of health care provider performance: a description of approaches. *Milbank Q.* **93**(4), 788–825 (2015). <https://doi.org/10.1111/1468-0009.12165>
 17. Schwartz, M., Rosen, A.K., Burgess, J.F.: Can composite measures provide a different perspective on provider performance than individual measures? *Med. Care* **55**(12), e99–e103 (2017). <https://doi.org/10.1097/MLR.0000000000000407>
 18. Werner, R.M., Konetzka, R.T., Polsky, D.: Changes in consumer demand following public reporting of summary quality ratings: an evaluation in nursing homes. *Health Serv. Res.* **51**(Suppl 2), 1291–1309 (2016). <https://doi.org/10.1111/1475-6773.12459>
 19. Bhandari, N., Scanlon, D.P., Shi, Y., Smith, R.A.: Why do so few consumers use health care quality report cards? A framework for understanding the limited consumer impact of comparative quality information. *Med. Care Res. Rev.* **76**(5), 515–537 (2019). <https://doi.org/10.1177/1077558718774945>
 20. The Centers for Medicare and Medicaid Services: Overall hospital quality star rating. <https://data.cms.gov/provider-data/topics/hospitals/overall-hospital-quality-star-rating/#measure-included-by-categories>
 21. Hibbard, J.H., Greene, J., Daniel, D.: What is quality anyway? Performance reports that clearly communicate to consumers the meaning of quality of care. *Med. Care Res. Rev.* **67**(3), 275–293 (2010). <https://doi.org/10.1177/1077558709356300>
 22. Emmert, M., Meier, F., Pisch, F., Sander, U.: Physician choice making and characteristics associated with using physician-rating websites: cross-sectional study. *J. Med. Internet Res.* **15**(8), e187 (2013). <https://doi.org/10.2196/jmir.2702>
 23. Mukamel, D.B., Weimer, D.L., Zwanziger, J., Gorthy, S.-F.H., Mushlin, A.I.: Quality report cards, selection of cardiac surgeons, and racial disparities: a study of the publication of the New York State Cardiac Surgery Reports. *Inquiry* **41**(4):435–446 (2004–2005)
 24. Emmert, M., Meszmer, N., Jablonski, L., Zinth, L., Schöffski, O., Taheri-Zadeh, F.: Public release of hospital quality data for referral practices in Germany: results from a cluster-randomised controlled trial. *Health Econ Rev.* **7**(33), I30 (2017). <https://doi.org/10.1186/s13561-017-0171-5>
 25. Geraedts, M., Hermeling, P., Ortwein, A., de Cruppé, W.: Public reporting of hospital quality data: what do referring physicians want to know? *Health Policy* **122**(11), 1177–1182 (2018). <https://doi.org/10.1016/j.healthpol.2018.09.010>
 26. Salampessy, B.H., Ikkersheim, D., Portrait, F.R.M., Koolman, X.: Do patients' preferences prevail in hospital selection? A comparison between discrete choice experiments and revealed hospital choice. *BMC Health Serv. Res.* **22**(1), 1136 (2022). <https://doi.org/10.1186/s12913-022-08403-6>
 27. Berwick, D.M., James, B., Coye, M.J.: Connections between quality measurement and improvement. *Med. Care* **41**(1 Suppl), I30–I38 (2003)
 28. Emmert, M., Meszmer, N., Simon, A., Sander, U.: Internetportale für die Krankenhauswahl in Deutschland: Eine leistungsbereichsspezifische Betrachtung. *Gesundheitswesen* (2015). <https://doi.org/10.1055/s-0035-1549968>
 29. Wissenschaftliches Instituts der AOK: Qualitätssicherung mit Routinedaten. <https://www.qualitaetssicherung-mit-routinedat.de/>
 30. Institut für Qualitätssicherung und Transparenz im Gesundheitswesen: Hüftendoprothesenversorgung - Bundesauswertung zum Erfassungsjahr 2020: Qualitätsindikatoren und Kennzahlen. Berlin (2021). https://iqtig.org/downloads/auswertung/2020/hep/QSKH_HEP_2020_BUAW_V01_2021-08-10.pdf
 31. Determann, D., Gyrd-Hansen, D., de Wit, G.A., et al.: Designing unforced choice experiments to inform health care decision making: implications of using opt-out, neither, or status quo alternatives in discrete choice experiments. *Med. Decis. Making* **39**(6), 681–692 (2019). <https://doi.org/10.1177/0272989X19862275>
 32. Vass, C., Rigby, D., Payne, K.: The role of qualitative research methods in discrete choice experiments. *Med. Decis. Making* **37**(3), 298–313 (2017). <https://doi.org/10.1177/0272989X16683934>
 33. McFadden, D.: Conditional logit analysis of qualitative choice behavior. In: *Frontiers in econometrics*, pp 105–142. Academic Press, New York (1974)
 34. Lancaster, K.J.: A new approach to consumer theory. *J. Polit. Econ.* **LXXIV**, 2 (1966)
 35. Naik Panvelkar, P., Armour, C., Saini, B.: Community pharmacy-based asthma services—what do patients prefer? *J. Asthma* **47**(10), 1085–1093 (2010). <https://doi.org/10.3109/02770903.2010.514638>
 36. Reed Johnson, F., Lancsar, E., Marshall, D., et al.: Constructing experimental designs for discrete-choice experiments: report of the ISPOR conjoint analysis experimental design good research practices task force. *Value Health.* **16**(1), 3–13 (2013). <https://doi.org/10.1016/j.jval.2012.08.2223>
 37. Bridges, J.F.P., Hauber, A.B., Marshall, D., et al.: Conjoint analysis applications in health—a checklist: a report of the ISPOR good research practices for conjoint analysis task force. *Value Health.* **14**(4), 403–413 (2011). <https://doi.org/10.1016/j.jval.2010.11.013>
 38. Hauber, A.B., González, J.M., Groothuis-Oudshoorn, C.G.M., et al.: Statistical methods for the analysis of discrete choice experiments: a report of the ISPOR conjoint analysis good research

- practices task force. *Value Health*. **19**(4), 300–315 (2016). <https://doi.org/10.1016/j.jval.2016.04.004>
39. Higgins, J., Thomas, J., Chandler, J., et al. (eds): *Cochrane Handbook for Systematic Reviews of Interventions*. 2nd Edition. In: *Cochrane Book Series*. John, Chichester (2019)
 40. Ketelaar, N.A.B.M., Faber, M.J., Braspenning, J.C., Westert, G.P.: Patients' expectations of variation in quality of care relates to their search for comparative performance information. *BMC Health Serv. Res.* **14**, 617 (2014). <https://doi.org/10.1186/s12913-014-0617-y>
 41. Zwijnenberg, N., Damman, O., Spreeuwenberg, P., Hendriks, M., Rademakers, J.: Different patient subgroup, different ranking? Which quality indicators do patients find important when choosing a hospital for hip- or knee arthroplasty? *BMC Health Serv. Res.* **11**, 299 (2011). <https://doi.org/10.1186/1472-6963-11-299>
 42. Blake, R.S., Clarke, H.D.: Hospital compare and hospital choice: public reporting and hospital choice by hip replacement patients in texas. *Med. Care Res. Rev.* **76**(2), 184–207 (2019). <https://doi.org/10.1177/1077558717699311>
 43. Beukers Puck, D.C., Kemp Ron, G.M., Varkevisser, M.: Patient hospital choice for hip replacement: empirical evidence from the Netherlands. *Eur. J. Health Econ. HEPAC Health Econ. Prevent. Care* **15**(9), 927–936 (2014). <https://doi.org/10.1007/s10198-013-0535-7>
 44. Gutacker, N., Siciliani, L., Moscelli, G., Gravelle, H.: Choice of hospital: which type of quality matters? *J. Health Econ.* **50**, 230–246 (2016). <https://doi.org/10.1016/j.jhealeco.2016.08.001>
 45. Moscelli, G., Siciliani, L., Gutacker, N., Gravelle, H.: Location, quality and choice of hospital: evidence from England 2002–2013. *Reg. Sci. Urban Econ.* **60**, 112–124 (2016). <https://doi.org/10.1016/j.regsciurbeco.2016.07.001>
 46. Damman, O.C., Spreeuwenberg, P., Rademakers, J., Hendriks, M.: Creating compact comparative health care information: what are the key quality attributes to present for cataract and total hip or knee replacement surgery? *Med. Decis. Making* **32**(2), 287–300 (2012). <https://doi.org/10.1177/0272989X11415115>
 47. O'Hara, L.M., Caturegli, I., O'Hara, N.N., et al.: What publicly available quality metrics do hip and knee arthroplasty patients care about most when selecting a hospital in Maryland: a discrete choice experiment. *BMJ Open* **9**(5), e028202 (2019). <https://doi.org/10.1136/bmjopen-2018-028202>
 48. Moser, A., Korstjens, I., van der Weijden, T., Tange, H.: Patient's decision making in selecting a hospital for elective orthopaedic surgery. *J. Eval. Clin. Pract.* **16**(6), 1262–1268 (2010). <https://doi.org/10.1111/j.1365-2753.2009.01311.x>
 49. Kandrack, R., Mehrotra, A., DeVries, A., Wu, S.-J., SooHoo, N.F., Martsof, G.R.: Patient use of cost and quality data when choosing a joint replacement provider in the context of reference pricing. *Health Serv. Res. Manag. Epidemiol.* (2015). <https://doi.org/10.1177/2333392815598310>
 50. Kolb, B., Emmert, M., Sander, U., Patzelt, C., Schöffski, O.: Do German public reporting websites provide information that office-based physicians consider before referring patients to hospital? A four-step analysis. *Z. Evid. Fortbild. Qual. Gesundheitswes.* (2018). <https://doi.org/10.1016/j.zefq.2018.07.010>
 51. Ikkersheim, D., Koolman, X.: The use of quality information by general practitioners: does it alter choices? A randomized clustered study. *BMC Fam. Pract.* **14**, 95 (2013). <https://doi.org/10.1186/1471-2296-14-95>
 52. Ringard, A.: Why do general practitioners abandon the local hospital? An analysis of referral decisions related to elective treatment. *Scand. J. Public Health*. **38**(6), 597–604 (2010). <https://doi.org/10.1177/1403494810371019>
 53. Prang, K.-H., Canaway, R., Bismark, M., Dunt, D., Kelaher, M.: The use of public performance reporting by general practitioners: a study of perceptions and referral behaviours. *BMC Fam. Pract.* **19**(1), 29 (2018). <https://doi.org/10.1186/s12875-018-0719-4>
 54. Rotar, A.M., van den Berg, M.J., Schäfer, W., Kringos, D.S., Klazinga, N.S.: Shared decision making between patient and GP about referrals from primary care: Does gatekeeping make a difference? *PLoS ONE* **13**(6), e0198729 (2018). <https://doi.org/10.1371/journal.pone.0198729>
 55. Potappel, A.J.C., Meijers, M.C., Kloek, C., et al.: To what degree do patients actively choose their healthcare provider at the point of referral by their GP? A video observation study. *BMC Fam. Pract.* **20**(1), 166 (2019). <https://doi.org/10.1186/s12875-019-1060-2>
 56. Beckert, W.: Choice in the presence of experts: the role of general practitioners in patients' hospital choice. *J. Health Econ.* **60**, 98–117 (2018). <https://doi.org/10.1016/j.jhealeco.2018.06.003>
 57. Birk, H.O., Henriksen, L.O.: Which factors decided general practitioners' choice of hospital on behalf of their patients in an area with free choice of public hospital? A questionnaire study. *BMC Health Serv. Res.* **12**, 126 (2012). <https://doi.org/10.1186/1472-6963-12-126>
 58. Doering, N., Maarse, H.: The use of publicly available quality information when choosing a hospital or health-care provider: the role of the GP. *Health Expect.* **18**(6), 2174–2182 (2015). <https://doi.org/10.1111/hex.12187>
 59. Geraedts, M., Schwartz, D., Molzahn, T.: Hospital quality reports in Germany: patient and physician opinion of the reported quality indicators. *BMC Health Serv. Res.* **7**, 157 (2007). <https://doi.org/10.1186/1472-6963-7-157>
 60. Frezza, E.E., Wachtel, M.S.: The quality of a surgeon defined by internal medicine and family practice physicians: a closed-ended survey with importance scale. *Am. Surg.* **73**(5), 481–483 (2007)
 61. Ferrua, M., Sicotte, C., Lalloue, B., Minvielle, E.: Comparative quality indicators for hospital choice: do general practitioners care? *PLoS ONE* **11**(2), e0147296 (2016). <https://doi.org/10.1371/journal.pone.0147296>
 62. Ketelaar, N.A., Faber, M.J., Elwyn, G., Westert, G.P., Braspenning, J.C.: Comparative performance information plays no role in the referral behaviour of GPs. *BMC Fam. Pract.* **15**, 146 (2014). <https://doi.org/10.1186/1471-2296-15-146>
 63. Hermeling, P., de Cruppé, W., Geraedts, M.: Informationsbedarf niedergelassener Ärzte bei Ein- und Überweisungen. *Gesundheitswesen* **75**(7), 448–455 (2013). <https://doi.org/10.1055/s-0032-1327745>
 64. Hesse, A., Schreyögg, J.: Determinanten eines erfolgreichen Einweisermarketings für Krankenhäuser - eine explorative Analyse. *Gesundheitsökonomie und Qualitätsmanagement*. **12**(5), 310–314 (2007). <https://doi.org/10.1055/s-2007-962846>
 65. Konrad, S., Raab, A.: Welche Faktoren beeinflussen die Einweisungsentscheidung von niedergelassenen Ärzten in der Praxis? *Gesundheitsökonomie und Qualitätsmanagement*. **19**(03), 137–143 (2014). <https://doi.org/10.1055/s-0034-1365968>
 66. Mühlbacher, A.C., Bethge, S.: Patients' preferences: a discrete-choice experiment for treatment of non-small-cell lung cancer. *Eur. J. Health Econ.* **16**(6), 657–670 (2015). <https://doi.org/10.1007/s10198-014-0622-4>
 67. Haaijer, R., Wedel, M.: Conjoint choice experiments: general characteristics and alternative model specifications. In: Gustafsson, A. (ed.) *Conjoint measurement: Methods and applications*; with 91 tables. 3. ed. Springer, Berlin, Heidelberg, New York (2003). http://digitale-objekte.hbz-nrw.de/webclient/DeliveryManager?pid=1487557&custom_att_2=simple_viewer
 68. Cunningham, C.E., Deal, K., Chen, Y.: Adaptive choice-based conjoint analysis: a new patient-centered approach to the assessment of health service preferences. *Patient*. **3**(4), 257–273 (2010). <https://doi.org/10.2165/11537870-000000000-00000>
 69. Dommaraju, S., Hagey, J., Odeny, T.A., et al.: Preferences of people living with HIV for differentiated care models in Kenya: a

- discrete choice experiment. *PLoS ONE* **16**(8), e0255650 (2021). <https://doi.org/10.1371/journal.pone.0255650>
70. Ryan, M., Gerard, K.: Using discrete choice experiments to value health care programmes: current practice and future research reflections. *Appl. Health Econ. Health Policy* **2**(1), 55–64 (2003)
71. Sawtooth Software: The CBC system for choice-based conjoint analysis. In: Sawtooth Software, Inc., ed. *CBC—Technical Paper [Technical Paper Series]*. Washington D.C (2008)
72. The German Medical Association: Ärzttestatistik zum 31. Dezember 2020. Berlin (2021). https://www.bundesaerztekammer.de/fileadmin/user_upload/downloads/pdf-Ordner/Statistik_2020/2020-Statistik.pdf.
73. Orme, B.K.: *Getting Started with Conjoint Analysis: Strategies for Product Design and Pricing Research*. Fourth Edition (2020)
74. Orme, B.K., Chrzan, K.: *Becoming an Expert in Conjoint Analysis: Choice Modeling for Pros*. North Orem, UT (2017)
75. Mühlbacher, A., Bethge, S.: What matters in type 2 diabetes mellitus oral treatment? A discrete choice experiment to evaluate patient preferences. *Eur. J. Health Econ.* **17**(9), 1125–1140 (2016). <https://doi.org/10.1007/s10198-015-0750-5>
76. Croissant, Y.: Estimation of random utility models in R : the mlogit package. *J. Stat. Soft.* (2020). <https://doi.org/10.18637/jss.v095.i11>
77. Mühlbacher, A.C., Bridges, J.F.P., Bethge, S., et al.: Preferences for antiviral therapy of chronic hepatitis C: a discrete choice experiment. *Eur. J. Health Econ.* **18**(2), 155–165 (2017). <https://doi.org/10.1007/s10198-016-0763-8>
78. Sijstermans, E., Cheung, K.L., Goossens, A.J.M., Conde, R., Gonzalez, J.G.R., Hilgsmann, M.: A discrete choice experiment to assess patients' preferences for HIV treatment in the urban population in Colombia. *J. Med. Econ.* **23**(8), 812–818 (2020). <https://doi.org/10.1080/13696998.2020.1735399>
79. Eggers, F., Sattler, H., Teichert, T., Völckner, F.: Choice-based conjoint analysis. In: Homburg, C., Klarmann, M., Vomberg, A. (eds.) *Handbook of market research*, pp. 781–819. Springer, Cham (2022)
80. de Cruppé, W., Kummer, M.-C., Geraedts, M.: Wie wählen chirurgische patienten ihr krankenhaus? Eine querschnittstudie how do surgical patients choose their hospital? A cross-sectional study. *Zentralbl. Chir.* (2021). <https://doi.org/10.1055/a-1555-3907>
81. Emmert, M., Meszmer, N., Jablonski, L., Zinth, L., Schöffski, O., Taheri-Zadeh, F.: Public release of hospital quality data for referral practices in Germany: results from a cluster-randomised controlled trial. *Health Econ. Rev.* (2017). <https://doi.org/10.1186/s13561-017-0171-5>
82. van Til, J.A., Stiggelbout, A.M., Ijzerman, M.J.: The effect of information on preferences stated in a choice-based conjoint analysis. *Patient Educ. Couns.* **74**(2), 264–271 (2009). <https://doi.org/10.1016/j.pec.2008.08.025>
83. van den Broek-Altenburg, E., Atherly, A.: Using discrete choice experiments to measure preferences for hard to observe choice attributes to inform health policy decisions. *Health Econ. Rev.* (2020). <https://doi.org/10.1186/s13561-020-00276-x>
84. van der Pol, M., Currie, G., Kromm, S., Ryan, M.: Specification of the utility function in discrete choice experiments. *Value Health.* **17**(2), 297–301 (2014). <https://doi.org/10.1016/j.jval.2013.11.009>
85. de Cruppé, W., Geraedts, M.: Hospital choice in Germany from the patient's perspective: a cross-sectional study. *BMC Health Serv. Res.* **17**(1), 720 (2017). <https://doi.org/10.1186/s12913-017-2712-3>
86. de Bekker-Grob, E.W., Bliemer, M.C.J., Donkers, B., et al.: Patients' and urologists' preferences for prostate cancer treatment: a discrete choice experiment. *Br. J. Cancer* **109**(3), 633–640 (2013). <https://doi.org/10.1038/bjc.2013.370>
87. de Cruppé, W., Geraedts, M.: Wie wählen Patienten ein Krankenhaus für elektive operative Eingriffe? Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz **54**(8), 951–957 (2011). <https://doi.org/10.1007/s00103-011-1320-3>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

5 The Elicitation of Patient Preferences for Hip Replacement Surgery: A Discrete Choice Experiment¹²

¹²This chapter is based on an article published in *BMC Health Services Research* (DOI: <https://doi.org/10.1186/s12913-025-12393-6>); this was joint work with Martin Emmert.

RESEARCH

Open Access



The elicitation of patient preferences for hip replacement surgery: a discrete choice experiment

Stefan Rohrbacher^{1,2*} and Martin Emmert^{1,2}

Abstract

Background The calculation of aggregated composite measures is a widely used approach to reduce the amount of quality-related data on hospital report cards (HRCs). This study aims to elicit patient preferences for hospital choice concerning publicly available hospital quality information for total hip replacement surgery. The results may assist in the development of weighted composite measures for elective hip replacement, which facilitates a conscious selection of the hospital.

Methods We collect primary survey data on a sample of 364 randomly selected users of the German HRC “Weisse Liste” (WL) (4/5 2023). The key attributes for hospital choice are based on the information provided in WL. We run various model specifications to identify patient preferences, allowing the analysis of unobserved preference heterogeneity.

Results Our sample consists of 177 respondents (mean age 56.46; 43.5% female). All attributes used are statistically significant for hospital choice ($p < 0.01$). Patients consider the “Quality of treatment” (26.95%; level range 1.734) and “Number of cases treated” (24.78%; level range 1.594) to be the most important. In contrast, “EndoCert Certificate” (17.50%; level range 1.126), “Equipment and qualification” (15.83%; level range 1.018), and “Recommendation from other patients” (14.94%; level range 0.960) remain less important. We find no evidence for unobserved heterogeneity regarding the preferences for hospital choice.

Conclusion Based on our findings, HRC users value publicly available hospital quality information for elective hip replacement differently. These differences should be taken into account when calculating aggregated composite measures. Our results may allow the calculation of a weighted aggregate composite measure from the perspective of HRC users.

Keywords Public reporting, Hospital choice, Patient preferences, Discrete choice experiment, Composite measures, Hospital report cards

Introduction

Total hip replacement (THR) is one of the most frequent surgeries worldwide; approximately one million patients undergo THR surgery annually [1]. In Germany, THR is one of the most common inpatient surgeries, with nearly 250,000 procedures per year. More than 160,000 patients with hip arthrosis undergo primary elective THR annually in approximately 1,250 hospitals [2]. Thereby, the frequency of THR surgeries in Germany differs widely

*Correspondence:

Stefan Rohrbacher

stefan.rohrbacher@uni-bayreuth.de

¹ Institute for Health Care Management and Health Sciences, University of Bayreuth, Prieserstrasse 2, Bayreuth 95444, Germany

² Bavarian Cancer Registry, Bavarian Health and Food Safety Authority, Meistersingerstrasse 2, Bayreuth 95444, Germany



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

from region to region; Schäfer et al. [3] report a variation factor of 2.8 between federal states, with higher rates in the south and the north-west. They claim that this may be caused – at least to some extent – by “the absence of standardized decision criteria as basis for the indication of THR in a transparent and consistent way”.

The numbers shown above demonstrate that patients in Germany have the option to choose from a large number of hospitals to undergo THR surgery. However, data from the German external hospital quality assurance system indicate that the effectiveness of treatment (i.e. the success of the surgery) varies between hospitals. Therein, for each patient, hospitals must document certain interventions as THR surgery based on a set of in-hospital related quality indicators [2]. The latest German hospital quality report from 2024 lists the share of hospitals with quality deficits ranging between 4.15%, regarding “Prevention of falls measures” (47 out of 1,132 hospitals), and 24.46%, regarding “Indication for hip endoprosthesis or component replacement” (251 out of 1,026 hospitals) [4]. These quality differences are not surprising since other studies also demonstrate hospital-related variations in the quality of THR surgery. For example, a recently published meta-analysis of 44 studies focusing on the impact of hospital volume on the outcomes of THR indicates that low-volume hospitals are associated with higher rates of surgical site infections, 90-day complications, costs, and mortality [1]. Previous evidence from Germany shows lower risk-adjusted in-hospital mortality rates for THR in high-volume hospitals compared to low-volume hospitals (0.10% vs. 0.23%) [5].

Against this backdrop, it seems important for patients to be able to make an informed choice; i.e. to select the best-performing or “right” hospital. Therefore, public reporting aims to support patients and other consumers by providing quality information about health care providers such as hospitals, nursing homes, or practitioners [6]. For this purpose, publicly available Internet rating websites have been developed and implemented in many high-income countries like the United Kingdom, the United States, and Germany [7–9]. For example, the leading German public reporting portal “Weisse Liste” (in English, white list, WL) was jointly initiated by the non-profit foundation “Bertelsmann Stiftung” and the main patients’ and consumers’ associations in 2008 [9]. More recently, the operation of WL ceased in March 2024; however, it will serve as an important element of its successor “Klinik-Atlas” in terms of, e.g. user guidance, selected quality measures, and other aspects [10]. The latter is the first German government-run public reporting website with the intention to provide quality-related information on the hospital level for the public as of May 2024 [11]. So far, there is limited quality information

regarding the detailed types of information that will be provided on “Klinik-Atlas” but it is likely to be extended shortly and become similar to WL (e.g. number of cases treated, clinical measures, staff-related information), apart from novel types of information [12].

Public reporting websites should fulfil a number of requirements in order to have an impact on health care delivery [13–16]. They should present the information that consumers value most while keeping the amount of quality information low in order to minimise complexity [17, 18]. As shown in previous literature, calculating composite measures based on consumer preferences is one way to achieve this [18–26].

The objective of our study is to learn more about consumers’ preferences when choosing a hospital for THR surgery on the German public reporting website WL. We focus exclusively on the publicly reported quality information about hospitals regarding elective THR surgery and exclude parameters related to patients as individuals. This may help in the future to calculate transparent weighted composite measures, which may ease patients making a conscious choice. Hence, we address the following three questions:

1. How do patients rate different publicly available hospital quality information regarding THR surgery?
2. What relative importance do patients assign to different types of quality information?
3. Can we identify groups of respondents with similar preferences?

Methods

Methodology of the discrete choice experiment

The objective of this study is to elicit patients’ preferences regarding hospital choice for THR surgery through the utilisation of a discrete choice experiment (DCE). Building upon the theoretical work of McFadden [27] and Lancaster [28], a DCE is a method that uses (survey) data on stated preferences based on a set of hypothetical choice scenarios (e.g. hospital choice) to systematically investigate the structure of individuals’ preferences. DCEs allow for reduced complexity for respondents due to the pairwise comparison of alternatives, which is an advantage compared with traditional formats for the elicitation of stated preferences, like rating- and ranking-based methods [29]. In addition, they also facilitate realistic trade-off decisions [30–32]. Therefore, DCEs appear increasingly often in the field of preference research in health care [33].

The design and analysis of the DCE are based on standardised research practices for conducting conjoint analysis developed by the ISPOR Conjoint Analysis Good Research Practices Task Force [34–36]. However,

in contrast to previous studies focusing on the relative importance of quality information for hospital choice in general (e.g. [25]), we aim to investigate how patients rate the hospital quality information with respect to THR surgery publicly available on WL and the relative importance patients assign to this information. Based on this, our results might help determine relative weights for publicly reported quality information on WL for the calculation of a composite performance measure. Such weighted composite measures may be integrated in health report services in the future to support patients making a conscious hospital choice. Thus, instead of conducting both a systematic literature review and qualitative research (e.g. semi-structured interviews, focus groups) to identify and choose the most important attributes for patients' preferences (see, e.g. [25, 29, 37]), we select our attributes in line with the publicly reported quality information provided on WL, one of the major health report cards in Germany at the time of the experiment. We focus on the information items primarily presented to users in the search interface of WL: (1) "Quality of treatment", which indicates whether a hospital fulfils obligatory quality targets such as, e.g. certain treatment outcomes; (2) "Recommendation from other patients", which provides information on the experiences of previous patients; (3) "Annual number of cases treated", which is supposed to reflect the experience of a hospital, based on the assumption that high numbers of cases indicate more in-depth

experience with the special treatment; (4) "Equipment and qualification", which covers the (non-)satisfaction of compulsory quality targets in the provision of medical equipment and the appropriate qualification of medical staff; and (5) "EndoCert Certificate", which indicates possible certifications that a hospital may have received from the "EndoCert Certification System", an independent initiative for quality auditing for endoprosthetic care founded by the German Society for Orthopaedics and Orthopaedic Surgery [38]. We focus exclusively on criteria on hospital quality and refrain from regarding individual patient-related features as e.g. travelling distance which depend on each patient's situation and willingness to share such information. Moreover, due to transparency and simplicity the composite measure should only build upon primary information in order to be useful for each patient. Each attribute contains three levels (see Table 1). We explain and illustrate each attribute to participants using simple examples and basic facts (for a detailed list of underlying information for each attribute as used on WL, please refer to Table A.1 of the Supplementary Material). We employ effect coding on all attributes (e.g. [36, 39]). Effect coding enables us to construct estimates and standard errors of all levels against their deviation from each attribute's mean. Hence, both estimates and standard errors are stable and independent of an arbitrary benchmark level.

Table 1 Summary of attributes and levels used in the discrete choice experiment

Attribute	Description and levels given in the questionnaire
Quality of treatment	Total hip replacements with fulfilled treatment criteria Quality targets reached No assessment available / results not (yet) available Quality targets not reached
Recommendation from other patients	Share of positive recommendations of other patients Recommended by 85% (above average) Recommended by 80% (average) Recommended by 76% (below average)
Number of cases treated	The annual number of total elective hip replacement surgeries in a hospital 230 patients (above average) 148 patients (average) 84 patients (below average)
Equipment and qualification	Hospitals with fulfilled criteria on medical equipment and staff qualification Quality targets reached No assessment available / results not (yet) available Quality targets not reached
EndoCert Certificate	Hospitals which meet requirements of the EndoCert Certification System Certified EndoProstheticsCentre of Maximum Care (EPCmax) Certified EndoProstheticsCentre (EPC) No certificate

Survey design

There are four parts to the survey. In the first part, we collect information about the participants' motivation to use WL, their expectations, and their previous experience with HRCs. In the second part, respondents are presented with information on all five attributes and their levels and are asked to rate the attributes on a scale of 1–5 (1 = not at all important; 5 = very important) and to rank them against each other. The third section contains ten DCE choice tasks in which respondents are asked to choose between two hypothetical hospitals. The final section collects socio-demographic and health-related information on participants for subgroup analyses. Participants were also given the opportunity to provide feedback on the survey and to enter a prize drawing for 10 online vouchers worth EUR 50 each. Before the survey was launched, the questionnaire was anonymously piloted for clarity and comprehensibility by 20 people and modified accordingly.

Experimental design

We designed the survey using Sawtooth Software (Lighthouse Studio Version 9.14.0) as a full profile design, i.e. each choice set includes all five attributes. We generated the final set using the balanced overlap method, which permits the estimation of both main and interaction effects with standard errors below 0.05 and 0.1 and the highest D-efficiency score [40–42]. The choice tasks are forced-choice tasks, i.e. respondents have to choose one of two hypothetical hospitals by making trade-offs between attributes and levels [25]. With this approach, the experiment provides a setting that is close to reality, as comparable trade-off decisions are part of daily life [32]. We administered the survey as an onsite-based survey on the German HRC Weisse Liste (WL). Over a 2-month period (April and May 2023), we invited all users of WL who either searched for information on THR surgery or performed hospital comparisons on THR surgery to participate in our study. Participants were free to take part and could terminate the survey at any time. Please refer to the supplementary material for an English translation of the questionnaire.

Sample size

Our study design consists of ten choice tasks per respondent, including two alternatives (i.e. hospitals) per choice task and a maximum number of three levels across all attributes. As suggested by Orme [43], the design specifications require a minimum of participants according to $N \geq \frac{500l}{ta}$, where N is the number of respondents, t the number of tasks, a the number of alternatives and l the maximum of levels. In our setting with $\{t, a, l\} = \{10, 2, 3\}$, we derive a minimum of 75

participants. However, since this number indicates only the lower bound limit for main effect estimation, we require at least 150 respondents, as recommended for achieving statistical robustness [44, 45].

Data analyses

For data analyses, we use R Statistical Software (Version 4.2.2; R Foundation for Statistical Computing, Vienna, Austria) and the corresponding packages “mlogit” [46] and “gmnl” [47]. As a starting point, we use the standard multinomial logit model (MNL), as initially developed by McFadden [27]. MNL is attractive because of its simplicity in terms of estimation and interpretation. However, it relies on the rather restrictive assumptions that preferences are homogeneous across individuals and that error terms are independently and identically distributed, both of which may conceal unobserved heterogeneity. We relax both assumptions with various alternative model specifications to test for unobserved heterogeneity. We consider this important because undetected heterogeneity may affect the estimation results and thus the appropriate weights of the included attributes, which may then be used for the calculation of a composite measure. Especially in the case of undetected group heterogeneity, we would need to adjust the composite measure for appropriate subgroups.

We use the random parameter model (RPL), as proposed by McFadden and Train [48], in order to relax the assumption of taste homogeneity (while maintaining independent and identically distributed errors). This model extends MNL by using a continuously distributed random parameter for each individual. In addition, we retreat from independently and identically distributed errors (while maintaining homogeneous preferences) and assume that idiosyncratic errors are not identical but individually scaled, as suggested by Bhat [49] and Fiebig et al. [50] (S-MNL). Similar to RPL, S-MNL also captures unobserved heterogeneity. However, its advantage over RPL is its more parsimonious description, which allows for more efficient estimation. Next, we use the general MNL model (G-MNL) introduced by Fiebig et al. [50], which combines the characteristics of RPL and S-MNL. As pointed out by Fiebig et al. [50], G-MNL puts more weight on randomness in the tails compared to RPL due to the inclusion of scaled error terms. Therefore, G-MNL captures unobserved heterogeneity better than RPL, which focuses more on the centre of the distribution. Besides assuming a continuous distribution of preference heterogeneity, as in the models above, we also estimate standard latent class models (LC) [51] and latent class models enhanced by random parameters (MM-MNL) [52]. These model types assume a discrete distribution of heterogeneity. In consequence, the

unobserved heterogeneity clusters in groups rather than at the individual level, as in RPL or G-MNL. In the case of MM-MNL, both types of heterogeneity are combined so that individual unobserved heterogeneity can occur within clustered groups. In order to identify the best-fitting model, the Bayesian information criterion (BIC) and conditional Akaike information criterion (CAIC) are used. We refrain from using the Akaike information criterion (AIC) since it is known to be insufficiently restrictive, especially in multi-class models, favouring too many groups, whereas the BIC and CAIC perform well (see, e.g. [53]).

Results

Sample characteristics

In total, 5,042 consumers of WL opened the survey link provided on WL. Of these, 4,678 stopped answering the survey directly after the short introduction page. Thus, 364 respondents returned the questionnaire. Clearly, the drop-out rate of 96.5% is rather high. However, our main target is to derive possibly accurate weights for a composite measure and therefore not necessarily to draw a representative sample which is likely to contain observations from users who do not have clear preferences or are unwilling to participate, as these we consider the primary source of noisy data. In the light of our objectives, we assume the drop-out rate not problematic. The following analysis reports only on the 177 respondents who fully completed the DCE part of the survey and provided consistent responses (48.63% completion rate). Table 2 summarises the key characteristics of the sample. The median age of all respondents is 58 years, slightly more than half of all respondents are female (54.80%), and 68.93% stated (technical) university entrance qualification as their highest educational level. In addition, 53.56% of all participants report their health status to be good or better, and 54.80% claim that they suffer from a chronic condition. Finally, 58.76% of respondents state that they have used hospital report cards during the last 12 months, and more than eight out of ten surveyed respondents (84.75%) state that they perceived big differences in the quality of care between hospitals. Please note that all respondents acknowledge differences in hospital quality which is in line with the studies and reports mentioned in the introduction [1, 4, 5].

Descriptive rating and ranking results

The results regarding the importance of the five presented quality information items for the hospital choice for THR surgery (on a scale of 1–5, with 1 = not at all important and 5 = extremely important) show that the “Quality of treatment” is rated as the most relevant (4.76 ± 0.56), followed by “Equipment

Table 2 Key characteristics of participants

Participants (N = 177)	n	%
Age		
39 years and younger	22	12.43%
40 to 49 years	24	13.56%
50 to 59 years	46	25.99%
60 to 69 years	58	32.77%
70 years and older	24	15.25%
missing	3	1.69%
Gender		
Female	97	54.80%
Male	77	43.50%
Diverse	3	1.69%
Educational Attainment		
(Technical) University entrance qualification	122	68.93%
Intermediate secondary school	42	23.73%
Secondary general school or less	12	6.78%
Missing	1	0.56%
Health Status		
Good or better	95	53.67%
Satisfactory	57	32.20%
Bad or worse	20	11.30%
Missing	5	2.82%
Chronic conditions		
Any chronic condition	97	54.80%
No chronic condition	75	42.37%
Missing	5	2.82%
Health insurance		
Public health insurance	128	72.32%
Private health insurance	48	27.12%
Missing	1	0.56%
Perceived differences in hospital quality		
Big differences	150	84.75%
Small differences	27	15.25%
No differences	0	0%
Experience with hospital report cards		
Yes	104	58.76%
No	73	41.24%

and qualification” (4.73 ± 0.49) and “Number of cases treated” (4.62 ± 0.70) (see Table 3). In contrast, holding an “EndoCert Certificate” (4.15 ± 0.85) and “Recommendation from other patients” (3.96 ± 0.84) seem to be less important. The ranking results for the single most relevant information item reveal “Quality of treatment” (47.46%) and “Number of cases treated” (35.03%) as the most relevant, while “EndoCert Certificate” (9.60%), “Equipment and qualification” (6.21%), and “Recommendation from other patients” (1.69%) are stated less frequently.

Table 3 Importance of different information items for the treatment choice

Attributes (N = 177)	Rated importance of different information items		Ranked single most important information items	
	Mean	SD	n	%
Quality of treatment	4.76	0.56	84	47.46%
Recommendation from other patients	3.96	0.84	3	1.69%
Number of cases treated	4.62	0.70	62	35.03%
Equipment and qualification	4.73	0.49	11	6.21%
EndoCert Certificate	4.15	0.85	17	9.60%

Model choice

For single-class models, we find that the MNL model represents the data best with respect to all consulted information criteria. In addition, relative estimation results vary only marginally across single-class models (cf. Table A.3 of Supplementary Material). The same picture emerges with respect to multi-class models. Regarding the number of classes for the multi-class models, the BIC and CAIC point at using a single class only (cf. Table A.4 of Supplementary Material). Moreover, we find only a

little variation of relative estimations even for multi-class models. For both LC and MM-MNL with two groups, for example, one group evaluates the number of cases treated higher and the other prefers recommendations, while all other attributes remain roughly as before (cf. Table A.5 of Supplementary Material). However, using two groups does not improve the data match sufficiently compared to the (single-class) MNL model such that the BIC and CAIC reject these models. Expanding to even more classes provides no further insights. This indicates that unobserved heterogeneity, either as idiosyncratic continuous error or in the form of clustered groups, is negligible in our sample. We conclude that additional model assumptions provide no meaningful information. Therefore, we hereinafter restrict our discussion to the MNL model.

Findings from the DCE

Table 4 summarises the estimation results of the MNL model. Since we employ effect coding, all estimates are with respect to the (hypothetical) mean of the corresponding attribute, implying that the coefficient values within an attribute add up to zero. As shown, all estimates are highly significant. Consequently, all attributes are highly relevant for the hospital choice of the respondents. Overall, consumers prefer hospitals that

Table 4 Estimated parameters of the multinomial logit model (MNL)

Attributes and levels	Coeff.	SE	95% CI		p-value	Difference highest to lowest level
Quality of treatment						1.734
Quality targets reached	0.985	0.064	0.858	1.112	<0.001	
No assessment intended	-0.236	0.057	-0.349	-0.124	<0.001	
Quality targets not reached	-0.749	0.063	-0.873	-0.624	<0.001	
Recommendation from other patients						0.960
Recommended by 85% (above average)	0.384	0.057	0.271	0.498	<0.001	
Recommended by 80% (average)	0.192	0.052	0.089	0.294	<0.001	
Recommended by 76% (below average)	-0.576	0.063	-0.700	-0.452	<0.001	
Number of cases treated						1.594
230 patients (above average)	0.718	0.061	0.598	0.837	<0.001	
148 patients (average)	0.158	0.053	0.054	0.263	0.003	
84 patients (below average)	-0.876	0.060	-0.994	-0.758	<0.001	
Equipment and qualification						1.018
Quality targets reached	0.605	0.058	0.490	0.720	<0.001	
No assessment intended	-0.192	0.054	-0.298	-0.086	<0.001	
Quality targets not reached	-0.413	0.058	-0.527	-0.300	<0.001	
EndoCert Certificate						1.126
Certified EPCmax	0.388	0.060	0.269	0.507	<0.001	
Certified EPC	0.350	0.054	0.243	0.456	<0.001	
No certificate	-0.738	0.062	-0.860	-0.616	<0.001	

SE standard error, CI confidence interval

achieve the quality targets for the treatment quality, as well as for equipment and qualification, and have above-average numbers for the cases treated and patient recommendations, as well as being certified as an EndoProstheticsCentre of Maximum Care (EPC-max). This pattern also emerges from Fig. 1, which depicts the corresponding level estimates.

Figure 2 illustrates the mean relative importance of attributes for hospital choice. Thereby, the range of coefficient values within an attribute reflects the importance of an attribute for the hospital choice. We compute the relative importance of each attribute using each attribute's coefficient range (i.e. the difference between the coefficients of the highest and lowest level of each attribute), expressed as a share of the total range across attributes.

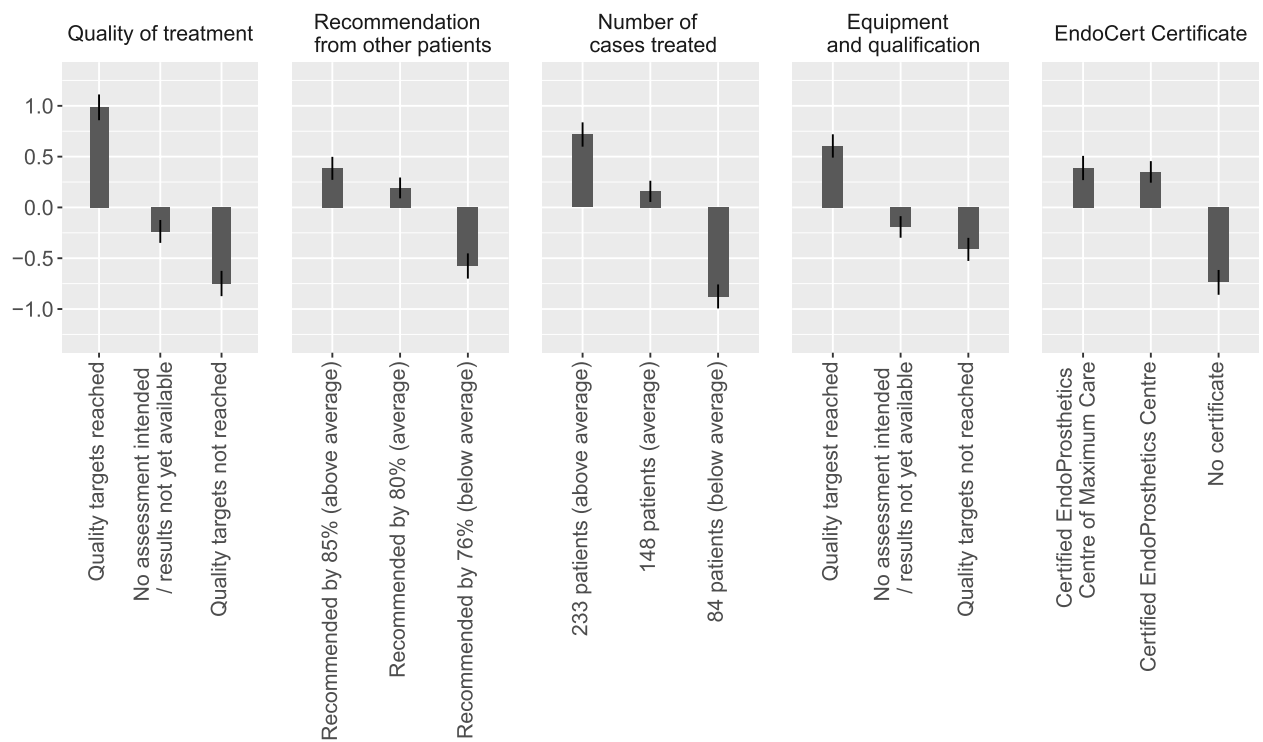


Fig. 1 Graphical display of level estimates with 95%CI of attributes for hospital choice

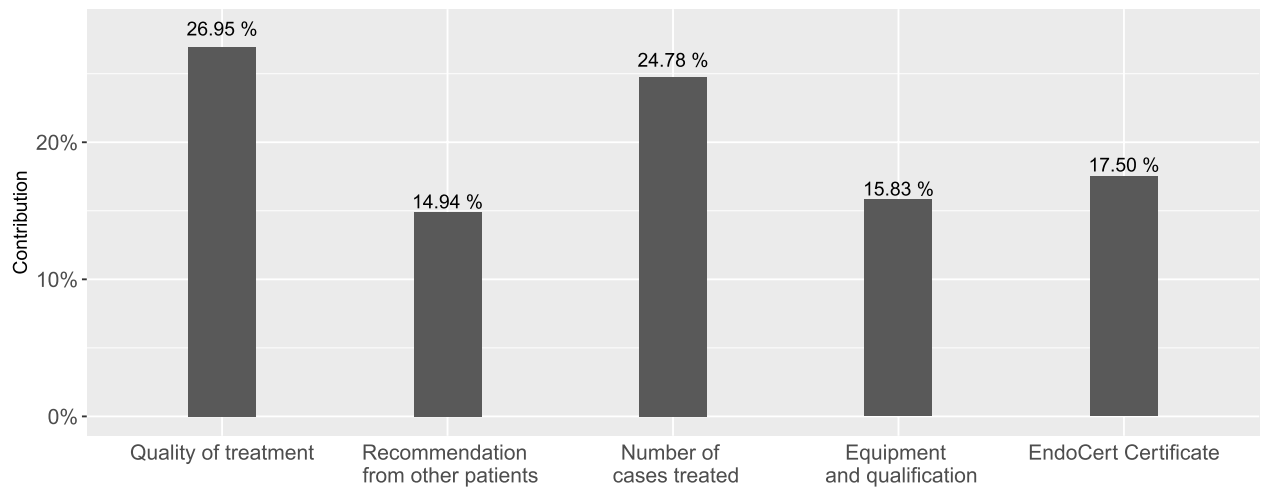


Fig. 2 Graphical display of mean relative importance of attributes for hospital choice

As shown, patients consider the quality of treatment as the most important (26.96%; level range of 1.734), followed by the annual number of cases treated (24.78%; level range of 1.594). In contrast, holding an EndoCert Certificate (17.51%; level range of 1.126), reaching quality targets with respect to equipment and qualification (15.83%; level range of 1.018), and recommendations from other patients (14.93%; level range of 0.960) are less important.

Discussion

The objective of this study is to elicit patient preferences concerning publicly available hospital quality information for elective hip replacement surgery by conducting an onsite-survey on the German HRC Weisse Liste (WL). The findings may support us in developing weighted composite measures from the consumers' perspective. We can draw several conclusions from our results. First, we can see that the DCE-based findings appear broadly consistent with the results of the rating- and ranking-based findings. This applies to both the order and relative importance of all quality information items. As shown, "Quality of treatment" is the most important information in all three analyses. For example, the DCE shows a relative importance of 26.95%, whereas 47.46% of the respondents ranked it in first place; it was also rated highest in terms of the mean importance, with a score of 4.76 (on a 1–5 scale with 1 = not at all important and 5 = extremely important). In addition, "Recommendations from other patients" appears to be the least important information in all three approaches. Here, the DCE shows a relative importance of 14.94%, whereas 1.69% of the respondents ranked it in first place; it was also rated lowest in terms of the mean importance, with a score of 3.96. We find slightly different results for the remaining three attributes, which, however, are not contradictory overall. For example, "Number of cases treated" is placed second with respect to the relative importance of all five attributes. Here, we compute a relative weight of 24.78% based on the DCE, and 35.03% of the respondents ranked it in first place as the single most important information item; however, the mean importance is rated 4.62 and thus is slightly lower than the mean importance of "Equipment and qualification". Nevertheless, considering the standard deviations and the small differences of the means, the higher mean of "Equipment and qualification" from the rating-based approach may not be meaningful compared with the results from the ranking-based approach and the DCE-based findings.

Second, the attributes for hospital choice for THR employed in this study – or joint replacement in general – have been utilised in other DCE studies, and thus appear to be of significant importance in general. For example,

several studies include "Quality of treatment" [54–56]. The "Number of cases treated" is also identified as a factor influencing hospital choice [25, 55–57]. In other studies, "Equipment and qualification" [58], "EndoCert Certificate" [25, 57], and "Recommendations from other patients" [57, 59] are applied in a hypothetical hospital choice scenario. Subsequently, our study may incorporate the most important attributes for hospital choice, despite the absence of qualitative research to identify and select the most pertinent attributes for patient preferences in general [25, 29, 37]. However, we select our attributes in accordance with the publicly reported quality information provided on WL (see above).

Furthermore, we find evidence for the importance of the attributes of this study from more general approaches. Quality improvements of hospitals are associated with high case volume [1, 60, 61], certification [62, 63] or better equipment and experience of medical staff [64]. Additionally, all but "Recommendation from other patients" are among the top ten relevant information items for hospital choice among THR patients in Germany [65]. However, we would like to emphasise that all the studies mentioned above include a variety of attributes that may not be present in our setting or may have been used differently. For example, "Quality of treatment" might also contain adverse medical outcomes such as complication rates or readmission rates. In addition to this, "Equipment and qualification" might also cover the conduct of medical staff.

Third, despite the differences in the general setting and combination of attributes, which do not allow for the direct comparison of findings, we may compare the relative importance of our attributes with findings from other research. In general, our results seem to be similar to those from other studies. "Quality of treatment" shows high importance in other studies [25, 54–56, 59]. For example, Emmert et al. [25] compare the preferences of both patients and referring physicians regarding hospital quality information for THR surgery. They demonstrate that traditional quality measures are rated most important (e.g. postoperative complication rates, one-year revision surgery). In the case of "Number of cases treated", we find more heterogeneous results. While Emmert et al. [25] and Kuklinski et al. [57] report "Number of cases treated" to be a highly important attribute, Groenewoud et al. [55] and Damman et al. [56] determine a medium relevance yet still find it to be more important than patient recommendations or the qualifications of staff, respectively. With respect to the remaining attributes, our results demonstrate their lower importance for hospital choice. For example, "Equipment and qualification" [56, 58], "EndoCert Certificate" [25], and "Recommendation from other patients" [59] appear to be less important.

Moreover, Kuklinski et al. [57] find for knee replacement that patient recommendation is more important than certification with case volume being the most important quality indicator. Finally, it should be mentioned that our study is similar to the research of Emmert et al. [25], who analyse the preferences of hip replacement patients after surgery. Given that we recruited participants for the survey during their hospital search on WL, we expect that our results are more likely to reflect the preference picture before THR treatment. Both studies demonstrate that patients evaluate “Number of cases treated” as relatively important and holding an “EndoCert Certificate” as less relevant.

Fourth, we find that the order of levels is consistent and as expected for all attributes. However, we find that the preference for EPCmax certificates is only 10.9% higher than that for EndoprostheticsCentre certificates (EPC). Similarly, Emmert et al. [25] report relatively small differences between preferences for EPCmax and EPC certificates among both patients and physicians. We suspect that this may be due to patients not being aware of the differences between the two certificates. The certificates are issued by EndoCert Limited, which is a subsidiary of the German Society for Orthopaedics and Orthopaedic Surgery [66]. The certification process requires the fulfilment of comprehensive quality objectives regarding structure, processes, and outcomes. Depending on which criteria are met, a hospital will receive one or the other certificate. For example, an EPC must treat at least 100 patients per year, whereas an EPCmax must treat at least 200 patients per year [38]. Alternatively, the added value of higher certificate types may be of limited relevance. This would imply that patients may know the differences but may not care about the type of certificate a hospital holds as long as it provides one. However, at this point we can only speculate about the reasons for this finding. Since our findings are consistent with Emmert et al. [25], we recommend further research to explore the reason for this result in more detail.

Fifth, we cannot detect various respondent groups with similar preferences. All considered multi-class models are inferior to the MNL model as the best model with respect to BIC and CAIC. The differences seem minor or implausible and do not make a case for dividing into sub-groups (cf. Tables A.4 and A.5 of Supplementary Material).

Finally, as suggested by previous research, the results of our study on patient preferences can be used to develop weighted composite measures from the consumer perspective [22, 25]. Such a composite measure would aggregate existing information into one summary score, thereby improving the usability of a health report card like WL by reducing the complexity of the information

provided. In this context, Schlesinger et al. [17] or Emmert et al. [67] show that reducing the complexity of report cards increases the quality of users’ hospital choices. Based on our findings, it may be possible to calculate the weights for a patient-centred composite measure that is publicly reported on WL. The next step would be to convert real hospital quality results into hospital-related composite measures. The preference for each hospital can then be estimated based on the sum of part-worth utilities for the selected level of all attributes [68]. For example, a hospital that achieves quality targets regarding the treatment quality (coef. 0.985) and equipment and qualification (coef. 0.605), with above-average recommendations from other patients (coef. 0.384), above-average case numbers (coef. 0.718), and certification as an Endoprosthetics Centre of Maximum Care (coef. 0.388), would have an overall score of 3.080. Comparing this value for each hospital against the overall score for all hospitals, we can group hospitals into several performance groups [25]. In this context, it seems highly relevant that our different models do not detect unobserved preference heterogeneity in our sample. This implies that it would be sufficient to compute one single weighted composite measure that corresponds to the preferences of all users. Otherwise, it would be appropriate to take into account specific individual or group characteristics.

Limitations

Our findings should be considered in light of some limitations. First, since we focus on German data from a German population subgroup (i.e. users of the German hospital report card WL), our findings and conclusions might be of limited importance for other countries. However, we think that our methodological approach and results may be of interest for countries with public reporting initiatives (e.g. the United States or United Kingdom). Second, the preferences of respondents derived from hypothetical hospital scenarios might differ from the actual search and selection behaviour in real life when people are confronted with similar decisions [59]. Third, the choice of attributes is based on a pre-existing selection of variables provided on WL during the time of the experiment and not on qualitative pre-study research as recommended (e.g. [29, 37]). Therefore, it is important to mention that other potentially relevant attributes are not considered in our study. We cannot exclude the possibility that integrating other information may lead to different findings. However, the selection of quality information on WL is based on publicly available quality measures from the patient’s perspective; a similar approach was published recently by other authors [59]. Moreover, the chosen attributes are similar to those

utilised by Emmert et al. [25], who derive the attributes by means of comprehensive literature research and qualitative research. In both studies, “Number of cases treated” and “EndoCert Certificate” appear, while our “Quality of treatment” and “Equipment and qualification” comprise their “Postoperative complication rate”, “Confirmed diagnosis rate”, “Mobility at hospital discharge”, and “Prevention of fall measures”. Only our attribute “Recommendation from other patients” is entirely novel. Fourth, for “Quality of treatment” and “Equipment and qualification” we model the level types as quality targets reached/not reached. This may lead to overestimating the relative importance of the two attributes since respondents do not need to wager for relative differences as e.g. with level types modelled as mean share of quality targets reached and relative deviations above and below. Yet, even if we overestimate these two, our results for relative importance are very likely to remain qualitatively robust with “Quality of treatment” and “Number of cases treated” being notably more important than the other attributes. Fifth, our results refer to patients for elective hip replacement, i.e. to treatments that focus on improving the quality of life. As shown by Kuklinski et al. [57], patients value quality information differently depending whether the treatment is life-saving or improves quality of life. Hence, our results may not be transferable to other indications as e.g. cancer or stroke where the primary goal of treatment is saving life. Furthermore, they may only add to a weighted composite measure for elective THR or similar indications where treatment is meant to improve the quality of life. Sixth, we need to emphasise that the external validity of our study is limited. Compared to patients of elective THR in 2023, our sample contains slightly too many of age below 50 years, slightly too few of age above 70 years and slightly too many male [69]. In comparison to the general population of Germany, our sample contains higher educational attainment [70]. These differences may be explained partially by regular web-users in Germany, who are younger, more male and better educated [67] as well as by the intertemporal correlation between year of birth and education attainment [70]. Compared to WL users in general, our sample is slightly older, less female and better educated [67]. Taking into account that the major activities of WL next to THR are total knee replacement where patient structure is similar to THR [69] and breast cancer with patients being more female and younger, differences regarding age and sex may narrow down with respect to THR patients on WL. However, our main target is to derive weights for a composite measure. This implies a trade-off between including a representative sample versus only those who see a benefit in participating, i.e., those who are aware of their preferences and eagerly return to use the composite

measure. Therefore, the remaining deviations in representativeness seem tolerable. Finally, we cannot entirely exclude sample biases. However, we tried to mitigate the potential risk of biased data by collecting a sufficiently representative sample of adequate size, by providing anonymity for respondents to increase honesty and response quality as well as clear and neutral questions for easy understanding [71].

Conclusion

This study provides new insights into the preference patterns of HRC users for hospital characteristics for patients undergoing elective THR surgery in Germany. Our results show that patients consider “Quality of treatment” and “Number of cases treated” as highly important. In contrast, “EndoCert Certificate”, “Equipment and qualification”, and “Recommendations from other patients” seem to be less important. We do not detect any meaningful heterogeneity in the preferences of all respondents. Based on our findings, the computation of a weighted composite measure may be the next step.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12913-025-12393-6>.

Supplementary Material 1.

Acknowledgements

The authors thank all survey participants and two anonymous reviewers for valuable and constructive comments.

Authors' contributions

All authors contributed to the study conception. SR designed and directed this study as the principal investigator. ME conceived and planned the experimental design of the study. SR performed the analysis on all samples. All authors contributed to the interpretation of the results. SR took the lead in preparing the manuscript. All authors provided critical feedback and commented on the manuscript. Finally, all authors approved the latest version of the manuscript to be published.

Funding

Open Access funding enabled and organized by Projekt DEAL.

Data availability

Data are available on request from the corresponding author (stefan.rohrbacher@uni-bayreuth.de).

Declarations

Ethics approval and consent to participate

Ethics approval for this study was obtained from the University of Bayreuth Ethics Board (24-052). The research was conducted in adherence to the Declaration of Helsinki.

All participants participated voluntarily and could end their participation at any time. All participants were informed about the study prior to participation and gave consent to use the data for publication.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 12 September 2024 Accepted: 7 February 2025

Published online: 18 February 2025

References

- Mufarrih SH, Ghani MOA, Martins RS, Qureshi NQ, Mufarrih SA, Malik AT, Noordin S. Effect of hospital volume on outcomes of total hip arthroplasty: a systematic review and meta-analysis. *J Orthop Surg Res*. 2019;14:1–13.
- Institute for Quality Assurance and Transparency in Health Care. Hüftendoprothesenversorgung – Bundesauswertung zum Erfassungsjahr 2020: Qualitätsindikatoren und Kennzahlen. 2021. https://iqtig.org/downloads/auswertung/2020/hep/QSKH_HEP_2020_BUAW_V01_2021-08-10.pdf. Accessed 08 May 2024.
- Schäfer T, Pritzkeleit R, Jeszenszky C, Malzahn J, Maier W, Günther KP, Niethard F. Trends and geographical variation of primary hip and knee joint replacement in Germany. *Osteoarthritis Cartil*. 2013;21(2):279–88. <https://doi.org/10.1016/j.joca.2012.11.006>.
- Institute for Quality Assurance and Transparency in Health Care. Bundesqualitätsbericht 2024. 2024. https://iqtig.org/downloads/berichte/2024/IQTIG_Bundesqualitätsbericht-2024_2024-10-22.pdf. Accessed 12 Dec 2024.
- Nimptsch U, Mansky T. Hospital volume and mortality for 25 types of inpatient treatment in German hospitals: observational study using complete national data from 2009 to 2014. *BMJ Open*. 2017;7(9). <https://doi.org/10.1136/bmjopen-2017-016184>. <https://bmjopen.bmj.com/content/7/9/e016184.full.pdf>.
- Berwick DM, James B, Coyne MJ. Connections between quality measurement and improvement. *Med Care*. 2003;41(1):30.
- Emmert M, Schlesinger M. Hospital quality reporting in the United States: does report card design and incorporation of patient narrative comments affect hospital choice? *Health Serv Res*. 2017;52(3):933–58.
- Greaves F, Millett C, Nuki P. England's experience incorporating "anecdotal" reports from consumers into their national reporting system: lessons for the United States of what to do or not to do? *Med Care Res Rev*. 2014;71(5):65–80.
- Pross C, Busse R, Geissler A. Hospital quality variation matters – a time-trend and cross-section analysis of outcomes in German hospitals from 2006 to 2014. *Health Policy*. 2017;121(8):842–52. <https://doi.org/10.1016/j.healthpol.2017.06.009>.
- Koch MC. Mehr Transparenz: Klinik-Atlas startet bald mit Erkenntnissen aus Weißer Liste. 2024. <https://www.heise.de/news/Klinik-Atlas-Gesundheitsministerium-setzt-auf-Wissen-der-BertelsmannStiftung-9697434.html>. Accessed 08 May 2024.
- German Federal Government. Interaktiver Krankenhaus-Atlas für Patientinnen und Patienten. 2024. <https://www.bundesregierung.de/breg-de/service/gesetzesvorhaben/krankenhaustransparenz-gesetz-2223136>. Accessed 08 May 2024.
- Federal Ministry of Health. Bundes-Klinik-Atlas – Datengrundlage. 2024. <https://bundes-klinik-atlas.de/datengrundlage/>. Accessed 22 Jul 2024.
- Hussey PS, Luft HS, McNamara P. Public reporting of provider performance at a crossroads in the United States: summary of current barriers and recommendations on how to move forward. *Med Care Res Rev*. 2014;71(5_suppl):5–16. <https://doi.org/10.1177/1077558714535980>. PMID: 24871273
- Metcalfe D, Diaz AJR, Olufajo OA, Massa MS, Ketelaar NA, Flottorp SA, Perry DC. Impact of public release of performance data on the behaviour of healthcare consumers and providers. *Cochrane Database Syst Rev*. 2018;9. <https://doi.org/10.1002/14651858.CD004538.pub3>.
- Haviland AM, Damberg CL, Mathews M, Paddock SM, Elliott MN. Shifting from passive quality reporting to active nudging to influence consumer choice of health plan. *Med Care Res Rev*. 2020;77(4):345–356. <https://doi.org/10.1177/1077558718798534>. PMID: 30255721
- Groot IB, Otten W, Dijks-Elsinga J, Smeets HJ, Kievit J, Mheen PJMV, CHOICE-2 Study Group. Choosing between hospitals: the influence of the experiences of other patients. *Med Decis Making*. 2012;32(6):764–778. <https://doi.org/10.1177/0272989X12443416>. PMID: 22546750
- Schlesinger M, Kanouse DE, Martino SC, Shaller D, Rybowski L. Complexity, public reporting, and choice of doctors: a look inside the black box of consumer behavior. *Med Care Res Rev*. 2014;71(5_suppl):38–64. <https://doi.org/10.1177/1077558713496321>. PMID: 23999489
- Sandmeyer B, Fraser I. New evidence on what works in effective public reporting. *Health Serv Res*. 2016;51:1159–66.
- Smith MA, Bednarz L, Nordby PA, Fink J, Greenlee RT, Bolt D, Magnan EM. Increasing consumer engagement by tailoring a public reporting website on the quality of diabetes care: a qualitative study. *J Med Internet Res*. 2016;18(12):332. <https://doi.org/10.2196/jmir.6555>.
- Huckman RS, Kelley MA. Public reporting, consumerism, and patient empowerment. *N Engl J Med*. 2013;369(20):1875–7.
- Sinaiko AD, Eastman D, Rosenthal MB. How report cards on physicians, physician groups, and hospitals can have greater impact on consumer choices. *Health Aff*. 2012;31(3):602–11.
- Shwartz M, Restuccia JD, Rosen AK. Composite measures of health care provider performance: a description of approaches. *Milbank Q*. 2015;93(4):788–825.
- Shwartz M, Rosen AK, Burgess JF Jr. Can composite measures provide a different perspective on provider performance than individual measures? *Med Care*. 2017;55(12):99–103.
- Werner RM, Konetzka RT, Polsky D. Changes in consumer demand following public reporting of summary quality ratings: an evaluation in nursing homes. *Health Serv Res*. 2016;51:1291–309.
- Emmert M, Rohrbacher S, Meier F, Heppel L, Drach C, Schindler A, Sander U, Patzelt C, Frömke C, Schöffski O, Lauerer M. The elicitation of patient and physician preferences for calculating consumer-based composite measures on hospital report cards: results of two discrete choice experiments. *Eur J Health Econ*. 2024;25(6):1071–85. <https://doi.org/10.1007/s10198-023-01650-2>.
- Chen J, Ou L, Hollis SJ. A systematic review of the impact of routine collection of patient reported outcome measures on patients, providers and health organisations in an oncologic setting. *BMC Health Serv Res*. 2013;13:1–24.
- McFadden D. Conditional logit analysis of qualitative choice behaviour. In: Zarembka P, editor. *Frontiers in Econometrics*. New York: Academic Press; 1973. pp. 105–42.
- Lancaster KJ. A new approach to consumer theory. *J Polit Econ*. 1966;74(2):132–57.
- Emmert M, Rohrbacher S, Jahn J, Fernando K, Lauerer M. Preferences of people living with HIV for long-acting antiretroviral treatment in Germany: evidence from a discrete choice experiment. *Patient Patient-Centered Outcome Res*. 2023;16(5):537–53. <https://doi.org/10.1007/s40271-023-00641-y>.
- Lancsar E, Louviere J. Conducting discrete choice experiments to inform healthcare decision making: a user's guide. *Pharmacoeconomics*. 2008;26:661–77.
- Johnson FR. Why not ask? Measuring patient preferences for healthcare decision making. *Patient Patient-Centered Outcome Res*. 2008;1:245–8.
- Mühlbacher AC, Bethge S. Patients' preferences: a discrete-choice experiment for treatment of non-small-cell lung cancer. *Eur J Health Econ*. 2015;16:657–70.
- Bekker-Grob EW, Ryan M, Gerard K. Discrete choice experiments in health economics: a review of the literature. *Health Econ*. 2012;21(2):145–72.
- Bridges JFP, Hauber AB, Marshall D, Lloyd A, Prosser LA, Regier DA, Johnson FR, Mauskopf J. Conjoint analysis applications in health—a checklist: a report of the ISPOR Good Research Practices for Conjoint Analysis Task Force. *Value Health*. 2011;14(4):403–13. <https://doi.org/10.1016/j.jval.2010.11.013>.
- Johnson FR, Lancsar E, Marshall D, Kilambi V, Mühlbacher A, Regier DA, Bresnahan BW, Kanninen B, Bridges JFP. Constructing experimental designs for discrete-choice experiments: report of the ISPOR Conjoint Analysis Experimental Design Good Research Practices Task Force. *Value Health*. 2013;16(1):3–13. <https://doi.org/10.1016/j.jval.2012.08.2223>.
- Hauber AB, González JM, Groothuis-Oudshoorn CGM, Prior T, Marshall DA, Cunningham C, IJzerman MJ, Bridges JFP. Statistical methods for the analysis of discrete choice experiments: a report of the ISPOR Conjoint Analysis Good Research Practices Task Force. *Value Health*. 2016;19(4):300–15. <https://doi.org/10.1016/j.jval.2016.04.004>.

37. Mühlbacher AC, Sadler A, Dippel F-W, Juhnke C. Treatment preferences in Germany differ among apheresis patients with severe hypercholesterolemia. *Pharmacoeconomics*. 2018;36(4):477–93. <https://doi.org/10.1007/s40273-018-0614-9>.
38. Haas H, Grifka J, Günther KP, Heller KD, Niethard FU, Windhagen H, Ebner M, Mittelmeier W. EndoCert – Zertifizierung von Endoprothetischen Versorgungszentren in Deutschland. 2013. https://www.endocert.de/images/pdf/EndoCert_thieme_1_Auflage_2013.pdf. Accessed 23 Jul 2024.
39. Sijstermans E, Cheung KL, Goossens AJ, Conde R, Gonzalez JG, Hilgsmann M. A discrete choice experiment to assess patients' preferences for HIV treatment in the urban population in Colombia. *J Med Econ*. 2020;23(8):812–8.
40. Haaijer R, Wedel M. Conjoint choice experiments: general characteristics and alternative model specifications. In: *Conjoint Measurement: Methods and Applications*. Heidelberg: Springer Berlin, Heidelberg; 2003. p 371–412.
41. Cunningham CE, Deal K, Chen Y. Adaptive choice-based conjoint analysis: a new patient-centered approach to the assessment of health service preferences. *Patient Patient-Centered Outcome Res*. 2010;3:257–73.
42. Marshall DA, Deal K, Conner-Spady B, Bohm E, Hawker G, Loucks L, MacDonald KV, Noseworthy T. How do patients trade-off surgeon choice and waiting times for total joint replacement: a discrete choice experiment. *Osteoarthritis Cartil*. 2018;26(4):522–30. <https://doi.org/10.1016/j.joca.2018.01.008>.
43. Orme BK. *Getting Started with Conjoint Analysis: Strategies for Product Design and Pricing Research*. Sawtooth Software; 2020.
44. Orme BK, Chrzan K. *Becoming an Expert in Conjoint Analysis: Choice Modeling for Pros*. Sawtooth Software; 2017.
45. Mühlbacher AC, Bethge S. What matters in type 2 diabetes mellitus oral treatment? A discrete choice experiment to evaluate patient preferences. *Eur J Health Econ*. 2016;17:1125–40. <https://doi.org/10.1007/s10198-015-0750-5>.
46. Croissant Y. Estimation of random utility models in R: the mlogit package. *J Stat Softw*. 2020;95:1–41. <https://doi.org/10.18637/jss.v095.i11>.
47. Sarrias M, Daziano R. Multinomial logit models with continuous and discrete individual heterogeneity in R: the gnm1 package. *J Stat Softw*. 2017;79(2):1–46. <https://doi.org/10.18637/jss.v079.i02>.
48. McFadden D, Train K. Mixed MNL models for discrete response. *J Appl Econ*. 2000;15(5):447–70.
49. Bhat CR. A heteroscedastic extreme value model of intercity travel mode choice. *Transp Res B Methodol*. 1995;29(6):471–83. [https://doi.org/10.1016/0191-2615\(95\)00015-6](https://doi.org/10.1016/0191-2615(95)00015-6).
50. Fiebig D, Keane M, Louviere J, Wasi N. The generalized multinomial logit model: accounting for scale and coefficient heterogeneity. *Mark Sci*. 2010;29(3):393–421.
51. Greene WH, Hensher DA. A latent class model for discrete choice analysis: contrasts with mixed logit. *Transp Res B Methodol*. 2003;37(8):681–98. [https://doi.org/10.1016/S0191-2615\(02\)00046-2](https://doi.org/10.1016/S0191-2615(02)00046-2).
52. Greene WH, Hensher DA. Revealing additional dimensions of preference heterogeneity in a latent class mixed multinomial logit model. *Appl Econ*. 2013;45(14):1897–902. <https://doi.org/10.1080/00036846.2011.650325>.
53. Nyland KL, Asparouhov T, Muthén BO. Deciding on the number of classes in latent class analysis and growth mixture modeling: a Monte Carlo simulation study. *Struct Equ Model Multidiscip J*. 2007;14(4):535–69. <https://doi.org/10.1080/10705510701575396>.
54. Gutacker N, Siciliani L, Moscelli G, Gravelle H. Choice of hospital: which type of quality matters? *J Health Econ*. 2016;50:230–46.
55. Groenewoud S, Van Exel NJA, Bobinac A, Berg M, Huijsman R, Stolk EA. What influences patients' decisions when choosing a health care provider? Measuring preferences of patients with knee arthrosis, chronic depression, or Alzheimer's disease, using discrete choice experiments. *Health Serv Res*. 2015;50(6):1941–72.
56. Damman OC, Spreuunenberg P, Rademakers J, Hendriks M. Creating compact comparative health care information: what are the key quality attributes to present for cataract and total hip or knee replacement surgery? *Med Decis Making*. 2012;32(2):287–300.
57. Kuklinski D, Vogel J, Geissler A. The impact of quality on hospital choice. Which information affects patients' behavior for colorectal resection or knee replacement? *Health Care Manag Sci*. 2021;24:185–202.
58. Arslan I, Huls S, Bekker-Grob E, Rozendaal R, Persoons M, Spruijt-van Hell M, Bindels P, Bierma-Zeinstra S, Schiphof D. Patients', healthcare providers', and insurance company employees' preferences for knee and hip osteoarthritis care: a discrete choice experiment. *Osteoarthritis Cartil*. 2020;28(10):1316–24.
59. O'Hara LM, Caturegli I, O'Hara NN, O'Toole RV, Dalury DF, Harris AD, Manson TT. What publicly available quality metrics do hip and knee arthroplasty patients care about most when selecting a hospital in Maryland: a discrete choice experiment. *BMJ Open*. 2019;9(5):028202.
60. Jeschke E, Gehrke T, Günster C, Heller K-D, Leicht H, Malzahn J, Niethard FU, Schröder P, Zacher J, Halder AM. Low hospital volume increases revision rate and mortality following revision total hip arthroplasty: an analysis of 17,773 cases. *J Arthroplast*. 2019;34(9):2045–50.
61. Mäkelä KT, Häkkinen U, Peltola M, Linna M, Kröger H, Remes V. The effect of hospital volume on length of stay, re-admissions, and complications of total hip arthroplasty: a population-based register analysis of 72 hospitals and 30,266 replacements. *Acta Orthop*. 2011;82(1):20–6.
62. Shaw CD, Groene O, Botje D, Sunol R, Kutryba B, Bruneau C, Hammer A, Wang A, Arah OA, et al. The effect of certification and accreditation on quality management in 4 clinical services in 73 European hospitals. *Int J Qual Health Care*. 2014;26(suppl_1):100–7.
63. Bergschmidt P, Maruniewicz JP, Westphal T, Klinder A, Mittelmeier W. Retrospective comparative study of the influence on quality of primary total hip arthroplasty by the EndoCert-Certification System in a communal hospital. *Z Orthop Unfallchirurgie*. 2021;159(04):397–405.
64. Cram P, Vaughan-Sarrazin MS, Wolf B, Katz JN, Rosenthal GE. A comparison of total hip and knee replacement in specialty and general hospitals. *J Bone Joint Surg*. 2007;89(8):1675–84.
65. Simon A. Patienteninvolvement und Informationspräferenzen zur Krankenhausqualität: Ergebnisse einer empirischen Analyse (Berufspolitisches Forum). *Unfallchirurg*. 2011;114(1):73–8.
66. Osmanski-Zenk K, Klinder A, Pingsmann A, Lohmann CH, Bail HJ, Kladny B, Mittelmeier W. Institutional surgical setting and volume effects of certified arthroplasty centers in Germany: evaluation of the quality of care in a 5-year comparison. *Healthcare*. 2024; 12(9):904. <https://doi.org/10.3390/healthcare12090904>.
67. Emmert M, Kast K, Sander U. Characteristics and decision making of hospital report card consumers: lessons from an onsite-based cross-sectional study. *Health Policy*. 2019;123(11):1061–7. <https://doi.org/10.1016/j.healthpol.2019.07.013>.
68. Til JA, Stiggelbout A, IJzerman MJ. The effect of information on preferences stated in a choice-based conjoint analysis. *Patient Educ Couns*. 2009;74(2):264–71.
69. Grimberg A, Kirschner S, Lütznier J, Melsheimer O, Morlock M, Steinbrück A. Endoprothesenregister Deutschland (EPRD): Jahresbericht 2024. 2024. https://www.eprd.de/fileadmin/user_upload/Dateien/Publikationen/Berichte/Jahresbericht2024-Status5_2024-10-22_F.pdf. Accessed 13 Feb 2025.
70. Federal Statistical Office of Germany. Bildungsstand der Bevölkerung - Ergebnisse des Mikrozensus 2019. 2020. https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bildung-Forschung-Kultur/Bildungsstand/Publikationen/Downloads-Bildungsstand/bildungsstand-bevoelkerung-5210002197004.pdf?__blob=publicationFile. Accessed 27 Dec 2024.
71. Menon V, Muraleedharan A. Internet-based surveys: Relevance, methodological considerations and troubleshooting strategies. *Gen Psychiatry*. 2020;33(5):e100264.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

6 Discussion

The following section synthesises the findings and methodological approaches of the studies reviewed in the preceding chapters. The analysis commences with a comparative assessment of the experimental designs implemented across the selected studies. Subsequently, the discussion evaluates the choice of model specifications and critically reviews the associated estimation results. The section concludes with an examination of how the relative importance of attributes differs according to the adopted model specifications.

6.1 Comparison of Experimental Designs

The three studies conduct four DCEs that are in line with general guidelines of experimental design and apply appropriate estimation techniques. All discussed studies satisfy the requirements of the ISPOR Conjoint Analysis Good Research Practices Task Force (Bridges et al., 2011; Johnson et al., 2013; Hauber et al., 2016). Similar to most health-related DCEs (de Bekker-Grob et al., 2012; Soekhai et al., 2019; Nouwens et al., 2025), each of the mentioned DCEs compares two alternatives. Also, with respect to the number of attributes, each DCE is in line with the majority of comparable studies, with either five (Rohrbacher and Emmert, 2025) or six attributes (Emmert et al., 2023, 2024). The number of levels is predominantly three, but in Emmert et al. (2023) only, it ranges from two to five. The number of tasks is ten (Emmert et al., 2024; Rohrbacher and Emmert, 2025) and twelve (Emmert et al., 2023), respectively, providing sufficient tasks for reliable data while avoiding the risk of fatigue among respondents. All studies follow a full profile design, i.e. each choice set includes all attributes. None of the studies incorporate no-choice options. This is reasonable, as all decisions assume unavoidable treatment, meaning that there is no option to not take any alternative. Instead, all studies make use of forced-choice tasks, highlighting the need to make a decision on treatment. All studies perform pilot testing for clarity and comprehensibility prior to each survey. This way, all studies keep the tasks sufficiently realistic and manageable for respondents while generating reliable data.

Concerning the selection of attributes, there are two approaches. On the one hand, Emmert et al. (2023) and Emmert et al. (2024) conduct, for all three DCEs, first, a systematic literature review, followed by semi-structured interviews in order to identify the most important attributes. Rohrbacher and Emmert (2025) on the other hand use the pre-existing selection of attributes reported on WL. As a result, their selection of attributes may not contain the attributes considered most important by patients. However, the authors assume that the selection of attributes by an established health report card such as WL is sufficiently appropriate. This way, Rohrbacher and Emmert exploit the groundwork of WL to circumvent time- and cost-intensive interviews. In the end, the chosen attributes remain relatively similar across both THR studies, supporting the simplified approach of Rohrbacher and Emmert (2025).

All mentioned DCEs apply fractional factorial design. Therefore, the experimental design is not balanced and orthogonal by definition. To achieve statistical efficiency, all studies choose the most efficient design by maximising D-efficiency measures. Recently, the majority of DCEs in health economics have applied efficient rather than orthogonal designs (Soekhai et al., 2019; Nouwens et al., 2025). In order to optimise the choice design, all utilise balance overlap. Each discussed DCE employs effect coding and evaluates utility using a partworth utility function model correspondingly. Additionally, Emmert et al. (2024) report estimates using numerical variables combined with a linear model. However, the usage of numerical variables is limited to the quick identification of significant attributes, as this approach is less precise and more restrictive. Thus, their final results refer only to the model based on effect-coded variables, which allows for non-linear relations between levels and preference.

According to Orme (2020), the number of respondents N should be at least $N \geq \frac{500l}{ta}$, where l is the maximum number of levels, t is the number of tasks, and a is the number of alternatives. To ensure statistical robustness, doubling this number should be considered the lower limit for the sample size (Orme and Chrzan, 2017; Orme, 2020). The sample sizes in all discussed DCEs satisfy this requirement, indicating robust samples with reliable results. All in all, all studies comply with the standard guidelines and techniques of conducting discrete choice experiments.

6.2 Estimation Results

In the following, I estimate all models mentioned in Section 2.1 for all discussed experiments.¹³ For the sake of brevity, I numerate all DCEs according to appearance: Emmert et al. (2023) contains DCE 1, Emmert et al. (2024) contains DCEs 2 and 3, which refer to the surveys of patients and medical staff, respectively, and Rohrbacher and Emmert (2025) contains DCE 4. Since not all models were part of the original studies, I provide the computation of the additional models if it was not reported before. Tables 6.1–6.4 introduce the estimation results of all single-class models for all DCEs: Table 6.1 presents the model estimations for DCE 1; Table 6.2 presents the model estimations for DCE 2; Table 6.3 presents the model estimations for DCE 3; Table 6.4 presents the model estimations for DCE 4. Similarly, Tables 6.5–6.8 describe the estimation results of a latent-class model with two groups: Table 6.5 illustrates the estimation results of the LC2 model for DCE 1; Table 6.6 illustrates the estimation results of the LC2 model for DCE 2; Table 6.7 illustrates the estimation results of the LC2 model for DCE 3; Table 6.8 illustrates the estimation results of the LC2 model for DCE 4. For all DCEs, I select the final models using the BIC and CAIC reported in the aforementioned tables. For the SMNL, GMNL, and RPL models, I use 2000 repetitions of random draws, which is sufficient for convergence (Hensher et al., 2015). As before, the best-fitting model for DCE 1 is a latent-class model with two groups, followed now by the GMNL model and MNL model for single-class types. In all other DCEs, the standard single-class multinomial logit model turns out to be the most favourable model.

Focusing first on the single-class models (see Table 6.1), the GMNL model appears to be most favourable for DCE 1, a model not considered in Emmert et al. (2023). Nevertheless, the results do not vary qualitatively from any other single-class model and all models provide the same conclusions: All attributes appear to be important, with most levels being significant with at least $p < 0.05$. Only ‘Subcutaneous injection’

¹³I refrain from estimating probit models due to their methodological proximity to logit models and the mentioned issues of identification and estimation.

for the attribute ‘Type of medication’ and the level ‘Increased risk of reduced effect due to (party) drugs’ for the attribute ‘Interaction with other medication or (party) drugs’ remain insignificant. For the significant levels, the results align consistently with expectations and still draw the following picture: Patients with HIV prefer pill-based treatment with low frequency, self-administered at home, with a low risk of short- and long-run side effects and no interaction with medication or (party) drugs.

Within the GMNL model, just as in the RPL model, unobserved heterogeneity enters through the assumption that preference parameters have some random component. Frequently, these random parameters are assumed to be normally distributed, although other forms are also possible (Sarrias and Daziano, 2017). Consequently, the estimation comprises both the mean value and the standard deviation for each attribute level. Significant estimates for the standard deviation of single attribute levels indicate sources of individual unobserved heterogeneity. In the case of DCE 1, I find meaningful variation for most of the attribute levels, implying unobserved heterogeneity across all attributes. However, this affects the qualitative results of the estimates of level means only marginally.

Considering now the latent-class model with two groups, Table 6.5 reports that the groups differ notably. While group LC1 ($n = 135$ of 226) mimics the results of the single-class models relatively closely except for the evaluation of the low risk of short-term side effects, group LC2 ($n = 91$ of 226) deviates in several aspects. For this group, neither ‘Type of medication’ nor ‘Interaction with other medication or (party) drugs’ plays a significant role, whereas both the risk of short-term side effects and the risk of long-term side effects are more pronounced. Despite these differences, the expected order of the attribute levels persists for both groups among significant levels.

$N = 226$									
	MNL	HEV	SMNL	GMNL	RPL				
	Coefficients	Coefficients	Coefficients	Coefficients	Coefficients	SD	SD	Coefficients	SD
Type of medication									
Pill-based treatment	0.289*** (0.041)	0.282*** (0.041)	0.322*** (0.053)	0.534*** (0.124)	1.476*** (0.266)	0.412** (0.130)	0.450 (1.574)		
Intramuscular injection	-0.225*** (0.041)	-0.217*** (0.041)	-0.265*** (0.055)	-0.416*** (0.099)	0.690*** (0.149)	-0.306** (0.099)	0.097 (1.414)		
Subcutaneous injection	-0.064 (0.040)	-0.065 (0.038)	-0.057 (0.046)	-0.118 (0.087)	0.785*** (0.152)	-0.105 (0.065)	0.353 (0.439)		
Frequency of dosing									
Daily	-0.906*** (0.065)	-0.864*** (0.071)	-1.062*** (0.130)	-1.662*** (0.265)	0.454* (0.210)	-1.259*** (0.365)	0.064 (2.987)		
Weekly	-0.456*** (0.060)	-0.438*** (0.062)	-0.536*** (0.088)	-0.803*** (0.159)	0.507* (0.216)	-0.630*** (0.186)	0.115 (1.735)		
Every 2 months	0.146* (0.059)	0.140* (0.058)	0.170* (0.070)	0.302* (0.122)	1.450** (0.507)	0.198 (0.106)	0.344 (8.577)		
Every 3 months	0.459*** (0.059)	0.443*** (0.062)	0.539*** (0.086)	0.776*** (0.147)	0.051 (0.356)	0.640** (0.194)	0.002 (7.716)		
Every 6 months	0.758*** (0.062)	0.719*** (0.068)	0.889*** (0.114)	1.388*** (0.234)	0.437* (0.212)	1.052*** (0.298)	0.525 (0.472)		
Location of treatment									
Practice / hospital	-0.170*** (0.017)	-0.165*** (0.018)	-0.200*** (0.028)	-0.306*** (0.059)	0.370*** (0.076)	-0.235*** (0.069)	0.227 (0.286)		
Home	0.170*** (0.017)	0.165*** (0.018)	0.200*** (0.028)	0.306*** (0.059)	0.370*** (0.076)	0.235*** (0.069)	0.227 (0.286)		
Short-term side effects									
Low risk	0.207*** (0.040)	0.197*** (0.039)	0.246*** (0.053)	0.353*** (0.079)	0.011 (0.145)	0.284** (0.090)	0.053 (1.811)		
Moderate risk	0.080* (0.040)	0.076 (0.040)	0.102* (0.048)	0.189* (0.084)	0.538** (0.198)	0.101 (0.065)	0.083 (4.050)		
Higher risk	-0.287*** (0.041)	-0.273*** (0.040)	-0.348*** (0.061)	-0.542*** (0.111)	0.526*** (0.132)	-0.386*** (0.115)	0.030 (3.602)		
Long-term side effects									
Low risk	0.537*** (0.041)	0.513*** (0.045)	0.635*** (0.822)	0.995*** (0.173)	0.487*** (0.142)	0.742*** (0.211)	0.003 (2.547)		
Moderate risk	0.224*** (0.040)	-0.215*** (0.045)	0.257*** (0.051)	0.396*** (0.105)	1.291*** (0.273)	0.300** (0.098)	0.015 (5.102)		
Higher risk	-0.761*** (0.044)	-0.728*** (0.052)	-0.892*** (0.102)	-1.390*** (0.225)	0.804*** (0.158)	-1.043*** (0.290)	0.018 (4.317)		
Interaction risks									
No interactions	0.168*** (0.040)	0.158*** (0.038)	0.200*** (0.051)	0.245*** (0.086)	0.745*** (0.203)	0.236** (0.090)	1.364* (0.594)		
Interaction with other medication	-0.125*** (0.040)	-0.115** (0.039)	-0.160*** (0.051)	-0.176* (0.078)	0.404*** (0.122)	-0.167* (0.074)	0.732* (0.342)		
Interaction with (party) drugs	-0.043 (0.040)	-0.042 (0.040)	-0.040 (0.047)	-0.068 (0.071)	0.342* (0.157)	-0.069** (0.060)	0.632 (0.326)		
θ	-	0.942*** (0.090)	-	-	-	-	-		
τ	-	-	0.569** (0.186)	0.378 (0.214)	-	-	-		
γ	-	-	-	-11.691 (290.187)	-	-	-		
Log likelihood	-1434.04	-1434.81	-1432.22	-1361.10	-1431.41				
BIC	2940.93	2948.06	2942.88	2879.08	3008.50				
CAIC	2953.93	2962.06	2956.88	2907.08	3034.50				
Random draws	-	-	2000	2000	2000				

Note: *p<0.05; **p<0.01; ***p<0.001

Table 6.1: DCE 1 – Results of single-class regression models

	MNL		HEV		SMNL		GMNL		RPL	
	Coefficients		Coefficients		Coefficients		Coefficients		Coefficients	SD
Confirmed diagnosis rate										
Quality targets reached	0.429*** (0.040)		0.440*** (0.063)		0.515*** (0.069)		0.838*** (0.174)		0.517*** (0.101)	0.449 (0.379)
No assessment intended	-0.001 (0.039)		-0.004 (0.041)		-0.005 (0.048)		0.017 (0.080)		-0.022 (0.055)	1.023 (0.662)
Quality targets not reached	-0.427*** (0.038)		-0.437*** (0.067)		-0.520*** (0.069)		-0.855*** (0.171)		-0.539*** (0.117)	0.575 (0.357)
EndoCert Certificate										
Certified EPCmax	0.256*** (0.039)		0.268*** (0.051)		0.288*** (0.052)		0.442*** (0.109)		0.325*** (0.080)	0.067 (0.804)
Certified EPC	0.135*** (0.040)		0.135*** (0.049)		0.168*** (0.050)		0.268** (0.089)		0.173*** (0.066)	0.064 (0.775)
No certificate	-0.390*** (0.040)		-0.403*** (0.069)		-0.456*** (0.060)		-0.710*** (0.150)		-0.498*** (0.114)	0.003 (0.194)
Number of cases treated										
230 patients (above average)	0.428*** (0.044)		0.446*** (0.073)		0.519*** (0.074)		0.886*** (0.191)		0.538*** (0.117)	0.008 (0.449)
148 patients (average)	0.188*** (0.039)		0.189*** (0.045)		0.221*** (0.050)		0.355*** (0.101)		0.255*** (0.080)	0.030 (0.584)
84 patients (below average)	-0.616*** (0.040)		-0.635*** (0.085)		-0.740*** (0.084)		-1.242*** (0.247)		-0.793*** (0.172)	0.022 (0.382)
Postoperative complication rate										
Quality targets reached	0.622*** (0.042)		0.637*** (0.088)		0.752*** (0.088)		1.167*** (0.225)		0.800*** (0.179)	0.004 (0.357)
No assessment intended	-0.080* (0.039)		-0.089* (0.043)		-0.093 (0.048)		-0.112 (0.078)		-0.090 (0.052)	0.008 (0.446)
Quality targets not reached	-0.542*** (0.044)		-0.549*** (0.081)		-0.659*** (0.085)		-1.056*** (0.208)		-0.710*** (0.168)	0.005 (0.272)
Mobility at hospital discharge										
Quality targets reached	0.597*** (0.043)		0.611*** (0.082)		0.718*** (0.084)		1.140*** (0.223)		0.758*** (0.166)	0.004 (0.221)
No assessment intended	-0.068 (0.039)		-0.059 (0.042)		-0.091 (0.048)		-0.156 (0.083)		-0.107 (0.065)	0.015 (0.391)
Quality targets not reached	-0.530*** (0.041)		-0.552*** (0.076)		-0.627*** (0.074)		-0.984*** (0.190)		-0.651*** (0.127)	0.011 (0.328)
Prevention of fall measures										
Quality targets reached	0.458*** (0.039)		0.468*** (0.068)		0.542*** (0.068)		0.807*** (0.154)		0.583*** (0.128)	0.658 (0.357)
No assessment intended	-0.094* (0.040)		-0.098* (0.048)		-0.098* (0.048)		-0.144 (0.077)		-0.140* (0.067)	0.659 (0.499)
Quality targets not reached	-0.364*** (0.042)		-0.370*** (0.055)		-0.444*** (0.068)		-0.663*** (0.139)		-0.442*** (0.094)	0.001 (0.353)
Intercept	-0.309*** (0.047)		-0.366* (0.145)		-0.352*** (0.062)		-0.554*** (0.134)		-0.423*** (0.116)	
θ	-		1.113*** (0.221)		-		-		-	
τ	-		-		0.581*** (0.155)		0.868*** (0.152)		-	
γ	-		-		-		11.934 (202.104)		-	
Log likelihood										
Log likelihood	-1627.71		-1630.00		1624.63		-1595.73		-1626.59	
BIC	3360.42		3373.09		3362.34		3409.55		3455.12	
CAIC	3360.42		3373.09		3376.34		3436.55		3480.12	
Random draws	-		-		2000		2000		2000	

Note: * p<0.05; ** p<0.01; *** p<0.001

Table 6.2: DCE 2 – Results of single-class regression models

$N = 187$									
	MNL	HEV	SMNL	GMNL		RPL			
	Coefficients	Coefficients	Coefficients	Coefficients	SD	Coefficients	SD		
Confirmed diagnosis rate									
Quality targets reached	0.454*** (0.058)	0.441*** (0.061)	0.463*** (0.066)	0.645*** (0.099)	0.076 (0.180)	0.620*** (0.083)	0.003 (0.189)		
No assessment intended	-0.108 (0.056)	-0.106 (0.056)	-0.110 (0.057)	-0.105 (0.076)	0.101 (0.232)	-0.102 (0.072)	0.017 (0.233)		
Quality targets not reached	-0.346*** (0.062)	-0.335*** (0.061)	-0.352*** (0.068)	-0.540*** (0.100)	0.026 (0.152)	-0.517*** (0.083)	0.014 (0.146)		
EndoCert Certificate									
Certified EPCmax	0.272*** (0.059)	0.258*** (0.067)	0.274*** (0.066)	0.425*** (0.100)	0.833*** (0.201)	0.411*** (0.091)	0.837*** (0.195)		
Certified EPC	0.174** (0.054)	0.175** (0.054)	0.177** (0.056)	0.223** (0.078)	0.277*** (0.164)	0.226*** (0.073)	0.273 (0.147)		
No certificate	-0.446*** (0.058)	-0.433*** (0.067)	-0.451*** (0.063)	-0.659*** (0.109)	0.555*** (0.126)	-0.638*** (0.094)	0.563*** (0.110)		
Number of cases treated									
230 patients (above average)	0.696*** (0.058)	0.667*** (0.060)	0.713*** (0.087)	0.991*** (0.132)	0.033 (0.339)	0.935*** (0.095)	0.101 (0.289)		
148 patients (average)	-0.020 (0.056)	-0.023 (0.057)	-0.020 (0.057)	-0.031 (0.075)	0.068 (0.424)	-0.027 (0.072)	0.107 (0.350)		
84 patients (below average)	-0.676*** (0.058)	-0.644*** (0.061)	-0.693*** (0.086)	-0.960*** (0.131)	0.035 (0.268)	-0.908*** (0.094)	0.006 (0.256)		
Postop. complication rate									
Quality targets reached	0.600*** (0.061)	0.573*** (0.060)	0.616*** (0.084)	0.873*** (0.122)	0.120 (0.233)	0.828*** (0.080)	0.028 (0.373)		
No assessment intended	-0.078 (0.057)	-0.072 (0.057)	-0.084 (0.063)	-0.293 (0.081)	0.295 (0.352)	-0.075 (0.075)	0.180 (0.479)		
Quality targets not reached	-0.522*** (0.057)	-0.501*** (0.064)	-0.532*** (0.068)	-0.781*** (0.113)	0.174 (0.277)	-0.753*** (0.090)	0.152 (0.282)		
Mobility at hosp. discharge									
Quality targets reached	0.344*** (0.057)	0.333*** (0.058)	0.354*** (0.068)	0.488*** (0.091)	0.270 (0.171)	0.466*** (0.080)	0.264 (0.155)		
No assessment intended	-0.192** (0.061)	-0.184** (0.057)	-0.197** (0.066)	-0.275** (0.090)	0.040 (0.318)	-0.268** (0.083)	0.331 (0.456)		
Quality targets not reached	-0.152** (0.041)	-0.149** (0.053)	-0.156** (0.059)	-0.213** (0.077)	0.127 (0.268)	-0.198*** (0.071)	0.067 (0.426)		
Revision surgery rate									
1.41% (better than average)	0.977*** (0.094)	0.942*** (0.102)	0.997*** (0.120)	1.446*** (0.212)	1.071*** (0.229)	1.398*** (0.171)	1.042*** (0.190)		
2.48% (average)	0.035 (0.159)	0.032 (0.161)	0.039 (0.164)	0.030 (0.254)	2.412*** (0.381)	0.011 (0.248)	2.394*** (0.292)		
3.89% (worse than average)	-1.012*** (0.097)	-0.974*** (0.101)	-1.036*** (0.133)	-1.476*** (0.224)	1.340*** (0.239)	-1.410*** (0.183)	1.352*** (0.204)		
θ	-	0.965*** (0.114)	-	-	-	-	-		
τ	-	-	0.183 (0.326)	0.230 (0.223)	-	-	-		
γ	-	-	-	2.391 (19.965)	-	-	-		
Log likelihood	-819.61	-820.59	819.58	-776.37	-776.95	-776.95	-776.95		
BIC	1729.62	1739.12	1737.11	1734.71	1792.41	1792.41	1792.41		
CAIC	1741.62	1752.12	1750.11	1774.61	1758.71	1758.71	1758.71		
Random draws	-	-	2000	2000	2000	2000	2000		

Note: * p<0.05; ** p<0.01; *** p<0.001

Table 6.3: DCE 3 – Results of single-class regression models

For all remaining DCEs, the standard single-class multinomial logit model turns out to be the most favourable, as described in the original papers. As displayed in Tables 6.2–6.4, estimates of other model classes show only moderate numerical variations, while qualitative results remain stable. Most of the levels are significant with at least $p < 0.05$, indicating significance for all attributes. In the case of DCE 2, only the middle positions of ‘No assessment intended / results not (yet) available’ for the attributes ‘Confirmed diagnosis rate’ and ‘Mobility at hospital discharge’ are not significant. For DCE 3, the middle positions of the attributes ‘Confirmed diagnosis rate’, ‘Number of cases treated’, ‘Postoperative complication rate’, and ‘1-year revision surgery rate’ remain insignificant, whereas all attribute levels of DCE 4 are significant. For DCE 2 and DCE 4, the expected order of levels is fulfilled for all attributes. In the case of DCE 3, the levels of the attribute ‘Mobility at hospital discharge’ do not show a strict order, with ‘Quality targets not reached’ and ‘No assessment intended / results not (yet) available’ reversed with close numerical values.

Each DCE refers to the most reliable estimation method according to the BIC and CAIC. The derived results appear solid and robust. If attribute levels have a clear hierarchy a priori, the estimation results confirm this order in most cases. Only for three attributes is a clear hierarchy not given, namely, ‘Type of medication’ and ‘Interaction with other medication or (party) drugs’ for DCE 1 and ‘Mobility at hospital discharge’ for DCE 3. However, for the two attributes of DCE 1 there is no order that exists a priori, as these are true categorical variables. Therefore, only in the case of the attribute in DCE 3 do the results possibly appear conspicuous. Yet, physicians evaluate positively if the quality targets are reached for ‘Mobility at hospital discharge’ and disapprove if they are not met or if this information is not available. One may read this as indicating that physicians are risk averse and consider the uncertainty about the fulfilment of the quality target as highly negative: If the quality target is not definitively reached, this is taken negatively. Thus, the expected hierarchy is at least partially fulfilled. Moreover, the abnormality dissolves in the latent-class model, indicating that it is driven by group LC2, who pay little attention to this attribute, while for group LC1, the expected order is now given.

Hence, the results presented in Chapters 3–5 hold even after taking additional, more complex estimation models into account. Patients undergoing total hip replacement (DCE 2) prefer hospitals satisfying quality targets with respect to the confirmed diagnosis rate, postoperative complication rate, mobility at hospital discharge, and prevention of fall measures, with large numbers of cases treated and an EndoCert certificate of ‘Certified EndoProstheticsCentre of Maximum Care’. The preferences of physicians surveyed in DCE 3 favour hospitals reaching quality targets regarding the confirmed diagnosis rate, postoperative complication rate, and mobility at hospital discharge, combined with high numbers of cases treated, a low revision surgery rate, and an EndoCert certificate of ‘Certified EndoProstheticsCentre of Maximum Care’. Finally, users searching WL for information about THR surgery (DCE 4) prefer hospitals that reach quality targets concerning treatment quality, as well as equipment and qualifications, and have high numbers of cases treated, a high number of recommendations, and the certificate ‘Certified EndoProstheticsCentre of Maximum Care’.

6.3 Comparison of Relative Importance of Attributes

Unfortunately, estimates like those reported in Tables 6.1–6.8 are not comparable when standing alone. Normalising the estimates using the relative importance of attributes as a measure solves this issue. This way, the results across all estimated models become comparable. Figures 6.1 and 6.2 display the shares of the relative importance of attributes across all DCEs for single-class models and the latent-class model with two groups, respectively. For the sake of brevity, I declare the attributes in alphabetical order according to their appearance in Tables 6.1–6.8. Consider, for example, DCE 2: The attribute ‘Number of cases treated’ turns into ‘Attribute C’, while ‘Prevention of fall measures’ becomes ‘Attribute F’. Focusing first on single-class models, Tables 6.1–6.4 show that the relative importance of attributes varies only moderately between all single-class models despite the numerical differences in estimations. The largest deviations occur in DCE 2, where attributes C and F show variations of more than one percentage point. Thus, it seems that all variation captured in individual heterogeneity

<i>N</i> = 226		MNL		LC	
		Coefficients	LC 1 (n = 135)	LC 2 (n = 91)	
Type of medication					
Pill-based treatment		0.289*** (0.041)	0.446*** (0.064)	0.100 (0.093)	
Intramuscular injection		−0.225*** (0.041)	−0.301*** (0.059)	−0.175 (0.099)	
Subcutaneous injection		−0.064 (0.063)	−0.144* (0.063)	0.075 (0.101)	
Frequency of dosing					
Daily		−0.906*** (0.065)	−1.078*** (0.097)	−0.966*** (0.161)	
Weekly		−0.456*** (0.060)	−0.590*** (0.091)	−0.310* (0.152)	
Every 2 months		0.146* (0.059)	0.203* (0.081)	−0.003 (0.145)	
Every 3 months		0.459*** (0.059)	0.554*** (0.086)	0.448*** (0.141)	
Every 6 months		0.758*** (0.062)	0.912*** (0.092)	0.831*** (0.163)	
Location of treatment					
Physicians' practice / hospital (medical staff administers treatment)		−0.170*** (0.017)	−0.230*** (0.026)	−0.116* (0.046)	
Home-based (self-administered treatment)		0.170*** (0.017)	0.230*** (0.026)	0.116* (0.046)	
Risk of short-term side effects					
Low risk (1 of 1000 patients will experience short-term side effects)		0.207*** (0.040)	0.152* (0.059)	0.448*** (0.093)	
Moderate (10 of 1000 patients will experience short-term side effects)		0.080* (0.040)	0.077 (0.056)	0.132 (0.093)	
Higher risk (100 of 1000 patients will experience short-term side effects)		−0.287*** (0.041)	−0.227*** (0.064)	−0.580*** (0.099)	
Risk of long-term side effects					
Low risk (1 of 1000 patients will experience long-term side effects)		0.537*** (0.041)	0.172** (0.064)	1.455*** (0.180)	
Moderate (10 of 1000 patients will experience long-term side effects)		0.224*** (0.040)	0.147* (0.059)	0.545*** (0.103)	
Higher risk (100 of 1000 patients will experience long-term side effects)		−0.761*** (0.044)	−0.319*** (0.069)	−1.995*** (0.228)	
Interaction with other medication or (party) drugs					
No medication / drug interactions occur		0.168*** (0.040)	0.251*** (0.057)	0.069 (0.094)	
Increased risk of reduced effect due to other medication		−0.125** (0.040)	−0.156** (0.056)	−0.092 (0.091)	
Increased risk of reduced effect due to (party) drugs		−0.043 (0.040)	−0.095 (0.057)	0.023 (0.094)	
intercept class 2		−	−	−0.440*** (0.094)	
Log likelihood		−1434.04		−1357.26	
BIC		2940.93		2865.79	
CAIC		2953.93		2892.79	

Note: *p<0.05; **p<0.01; ***p<0.001

Table 6.5: DCE 1 – Results of latent-class model with two classes

<i>N</i> = 322	MNL	LC	
	Coefficients	LC 1 (n = 161)	LC 2 (n = 161)
Confirmed diagnosis rate			
Quality targets reached	0.429*** (0.040)	0.688*** (0.143)	0.369*** (0.074)
No assessment intended / results not (yet) available	−0.001 (0.039)	0.006 (0.120)	−0.001 (0.063)
Quality targets not reached	−0.427*** (0.038)	−0.694*** (0.132)	0.363*** (0.067)
EndoCert Certificate			
Certified EndoProstheticsCentre of Maximum Care	0.256*** (0.039)	0.405*** (0.112)	0.192*** (0.086)
Certified EndoProstheticsCentre	0.135*** (0.040)	0.181 (0.101)	0.146* (0.063)
No certificate	−0.390*** (0.040)	−0.586*** (0.114)	−0.337*** (0.065)
Number of cases treated			
230 patients (above average)	0.428*** (0.044)	1.340*** (0.245)	0.029 (0.089)
148 patients (average)	0.188*** (0.039)	0.237* (0.096)	0.150* (0.063)
84 patients (below average)	−0.616*** (0.040)	−1.578*** (0.234)	−0.179 (0.094)
Postoperative complication rate			
Quality targets reached	0.622*** (0.042)	0.922*** (0.120)	0.554*** (0.080)
No assessment intended / results not (yet) available	−0.080* (0.039)	−0.136 (0.097)	−0.032 (0.063)
Quality targets not reached	−0.542*** (0.044)	−0.786*** (0.113)	−0.521*** (0.074)
Mobility at hospital discharge			
Quality targets reached	0.597*** (0.043)	1.107*** (0.154)	0.446*** (0.081)
No assessment intended / results not (yet) available	−0.068 (0.039)	−0.173 (0.102)	−0.011 (0.063)
Quality targets not reached	−0.530*** (0.041)	−0.934*** (0.139)	−0.434*** (0.077)
Prevention of fall measures			
Quality targets reached	0.458*** (0.039)	0.578*** (0.104)	0.450*** (0.063)
No assessment intended / results not (yet) available	−0.094* (0.040)	−0.078 (0.104)	−0.097 (0.064)
Quality targets not reached	−0.364*** (0.042)	−0.500*** (0.106)	−0.323*** (0.065)
Intercept	−0.309*** (0.047)	−491** (0.157)	−0.285** (0.088)
Intercept class 2	−	−	0.150 (0.114)
Log likelihood	−1627.71	−1587.11	
BIC	3360.42	3392.30	
CAIC	3373.42	3419.30	

Note: *p<0.05; **p<0.01; ***p<0.001

Table 6.6: DCE 2 – Results of latent-class model with two classes

<i>N</i> = 187	MNL	LC	
	Coefficients	LC 1 (n = 131)	LC 2 (n = 56)
Confirmed diagnosis rate			
Quality targets reached	0.454*** (0.058)	0.535*** (0.090)	0.553*** (0.127)
No assessment intended / results not (yet) available	−0.108 (0.056)	−0.054 (0.087)	−0.168 (0.116)
Quality targets not reached	−0.346*** (0.062)	−0.480*** (0.090)	−0.384*** (0.122)
EndoCert Certificate			
Certified EndoProstheticsCentre of Maximum Care	0.272*** (0.059)	0.201* (0.101)	0.599*** (0.149)
Certified EndoProstheticsCentre	0.174** (0.054)	0.215* (0.091)	0.181 (0.134)
No certificate	−0.446*** (0.058)	−0.417*** (0.100)	−0.780** (0.184)
Number of cases treated			
230 patients (above average)	0.696*** (0.058)	0.608*** (0.099)	1.138*** (0.202)
148 patients (average)	−0.020 (0.056)	−0.070 (0.084)	0.002 (0.110)
84 patients (below average)	−0.676*** (0.058)	−0.539*** (0.096)	−1.139*** (0.214)
Postoperative complication rate			
Quality targets reached	0.600*** (0.058)	0.810*** (0.103)	0.516*** (0.148)
No assessment intended / results not (yet) available	−0.078 (0.057)	−0.144 (0.091)	−0.009 (0.113)
Quality targets not reached	−0.522*** (0.057)	−0.666*** (0.095)	−0.526** (0.161)
Mobility at hospital discharge			
Quality targets reached	0.344*** (0.057)	0.545*** (0.097)	0.153 (0.130)
No assessment intended / results not (yet) available	−0.192** (0.061)	−0.255* (0.101)	−0.190 (0.128)
Quality targets not reached	−0.152** (0.041)	−0.290** (0.092)	0.037 (0.123)
1-year revision surgery rate			
1.41% (better than average)	0.977*** (0.094)	1.389*** (0.183)	0.490* (0.206)
2.48% (average)	0.035 (0.159)	0.315 (0.251)	−0.408 (0.375)
3.89% (worse than average)	−1.012*** (0.097)	−1.704*** (0.225)	−0.082 (0.253)
Intercept class 2	−	−	−0.759*** (0.119)
Log likelihood	−819.61	−778.77	
BIC	1729.62	1745.87	
CAIC	1741.62	1779.87	

Note: *p<0.05; **p<0.01; ***p<0.001

Table 6.7: DCE 3 – Results of latent-class model with two classes

$N = 177$	MNL	LC	
	Coefficients	LC 1 ($n = 77$)	LC 2 ($n = 100$)
Quality of treatment			
Quality targets reached	0.985*** (0.064)	1.305*** (0.194)	1.078*** (0.098)
No assessment intended / results not (yet) available	-0.246*** (0.057)	-0.680*** (0.190)	-0.136 (0.089)
Quality targets not reached	-0.749*** (0.063)	-0.625*** (0.177)	-0.941*** (0.092)
Recommendation from other patients			
Recommended by 85% (above average)	0.384*** (0.057)	0.168 (0.149)	0.553*** (0.093)
Recommended by 80% (average)	0.192*** (0.052)	0.374** (0.125)	0.171* (0.082)
Recommended by 76% (below average)	-0.576*** (0.063)	-0.543*** (0.163)	-0.724*** (0.094)
Number of cases treated			
230 patients (above average)	0.718*** (0.061)	1.820*** (0.258)	0.221 (0.129)
148 patients (average)	0.158*** (0.053)	0.114 (0.122)	0.214*** (0.075)
84 patients (below average)	-0.876*** (0.060)	-1.934*** (0.259)	-0.435*** (0.125)
Equipment and qualification			
Quality targets reached	0.605*** (0.058)	0.761*** (0.131)	0.633*** (0.087)
No assessment intended / results not (yet) available	-0.192*** (0.054)	-0.250 (0.134)	-0.196** (0.078)
Quality targets not reached	-0.413*** (0.058)	-0.511*** (0.142)	-0.438*** (0.092)
EndoCert Certificate			
Certified EndoProstheticsCentre of Maximum Care	0.388*** (0.060)	0.663*** (0.155)	0.293* (0.115)
Certified EndoProstheticsCentre	0.350*** (0.054)	0.496*** (0.131)	0.346*** (0.076)
No certificate	-0.738*** (0.062)	-1.159*** (0.157)	-0.639** (0.115)
Intercept class 2	—	—	0.317** (0.119)
Log likelihood	-825.7	-787.1	
BIC	1726.18	1731.25	
CAIC	1736.18	1752.25	

Note: * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Table 6.8: DCE 4 – Results of latent-class model with two classes

plays no further role in the relative importance of attributes, regardless of the model choice.

As already reported previously, it appears for DCE 1 that the attributes ‘Frequency of dosing’ (B) and ‘Risk of long-term side effects’ (E) gain the largest share of importance. In DCE 2 on patients undergoing THR, the shares are more equally distributed, such that single attributes do not stand out, as in the case of DCE 1. In that case, the attributes ‘Number of cases treated’ (C) and ‘Mobility at hospital discharge’ (E) take the lead with an importance of about 20% each, followed closely by ‘Postoperative complication rate’ (D). In DCE 3, physicians demonstrate a clear preference for the attribute ‘1-year revision surgery rate’ (F; about 30%), followed by ‘Number of cases treated’ (C; about 20%) and ‘Postoperative complication rate’ (D; about 17%), with respect to THR. It appears that patients and physicians share a strong preference for ‘Number of cases treated’ and ‘Postoperative complication rate’, while physicians evaluate ‘Mobility at hospital discharge’ lower than patients (E; about 8%). The attributes ‘Confirmed diagnosis rate’ and ‘EndoCert Certificate’ receive minor importance for both groups. Finally, in DCE 4, on people searching for information on THR, ‘Quality of treatment’ (A; about 27%) and ‘Number of cases treated’ (C; about 24%) are the most dominant attributes. This aligns with the results of Emmert et al. (2024), as ‘Quality of treatment’ (DCE 4), ‘Postoperative complication rate’ (DCEs 2 and 3), and ‘1-year revision surgery rate’ (DCE 3) capture similar outcomes. As in the case of DCEs 2 and 3 (B in each), ‘EndoCert Certificate’ is considered less important in DCE 4 (E).

Contrarily, group-specific heterogeneity does affect the relative importance of attributes, as shown in Figure 6.2. Unsurprisingly, the most notable differences occur in DCE 1, for which the LC2 model is in general the most favourable model. The relative importance of attributes varies between the MNL model as the benchmark and the two groups LC1 and LC2 for all attributes. For example, LC1 especially considers attributes A (‘Type of medication’) and B (‘Frequency of dosing’) more than in the MNL model, whereas LC2 favours attribute E (‘Risk of long-term side effects’). In the remaining DCEs 2–4, the results vary to a lesser extent compared to DCE 1, not least due to the lack of unobserved taste heterogeneity and the resulting dominance of the single-class MNL

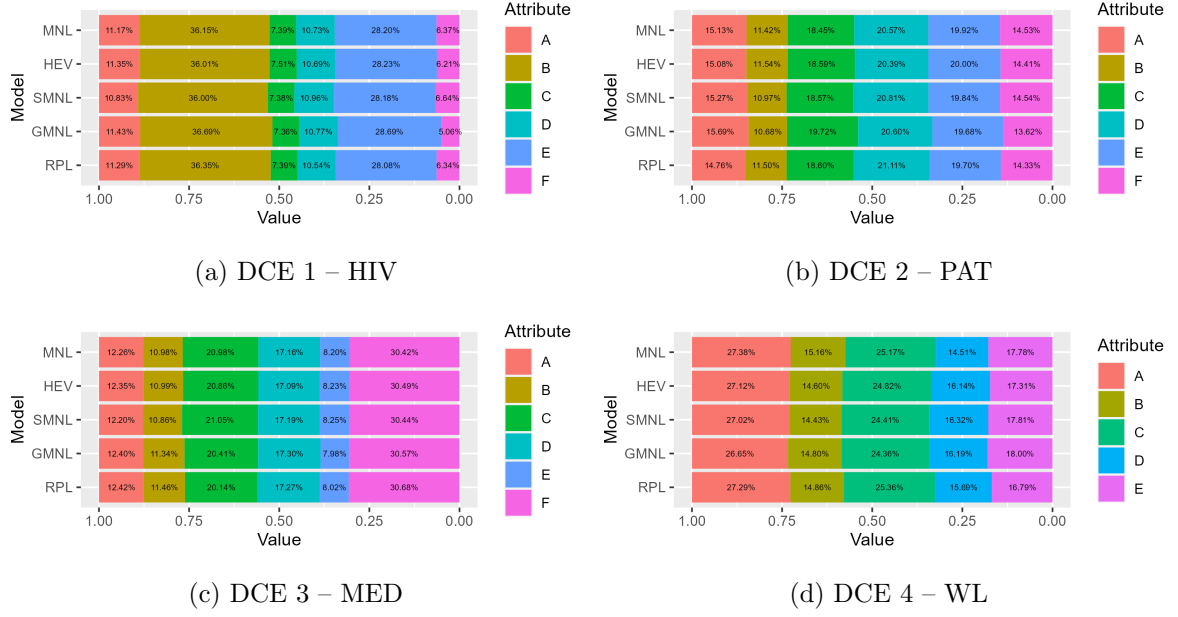


Figure 6.1: Relative importance of attributes for single-class models

model. However, even in these surveys, notable differences emerge between the groups of the latent-class model and the benchmark MNL. In DCE 2, LC1 focuses more on attribute C (‘Number of cases treated’) and LC2 focuses on D (‘Postoperative complication rate’). In DCE 3, LC1 evaluates F (‘1-year revision surgery rate’) even higher, whereas LC2 clearly considers C (‘Number of cases treated’) more important. In DCE 4, LC1 shows a clear preference for C (‘Number of cases treated’) compared to the MNL model and LC2 shows a preference for A (‘Quality of treatment’) and B (‘Recommendation from other patients’). Hence, group-specific heterogeneity seems to have a larger impact on the relative importance of attributes than individual unobserved heterogeneity. From this I conclude, regarding the relative importance of attributes, that in applied DCEs, it is vital to always bear in mind to test for latent-class models, whereas individual unobserved heterogeneity, as captured in, e.g. RPL models, may be considered as a secondary problem.

To summarise, the results of all DCEs vary very little across single-class models concerning the relative importance of attributes. Contrary to this, they differ notably in the case of the latent-class models with multiple classes, even if the BIC and CAIC

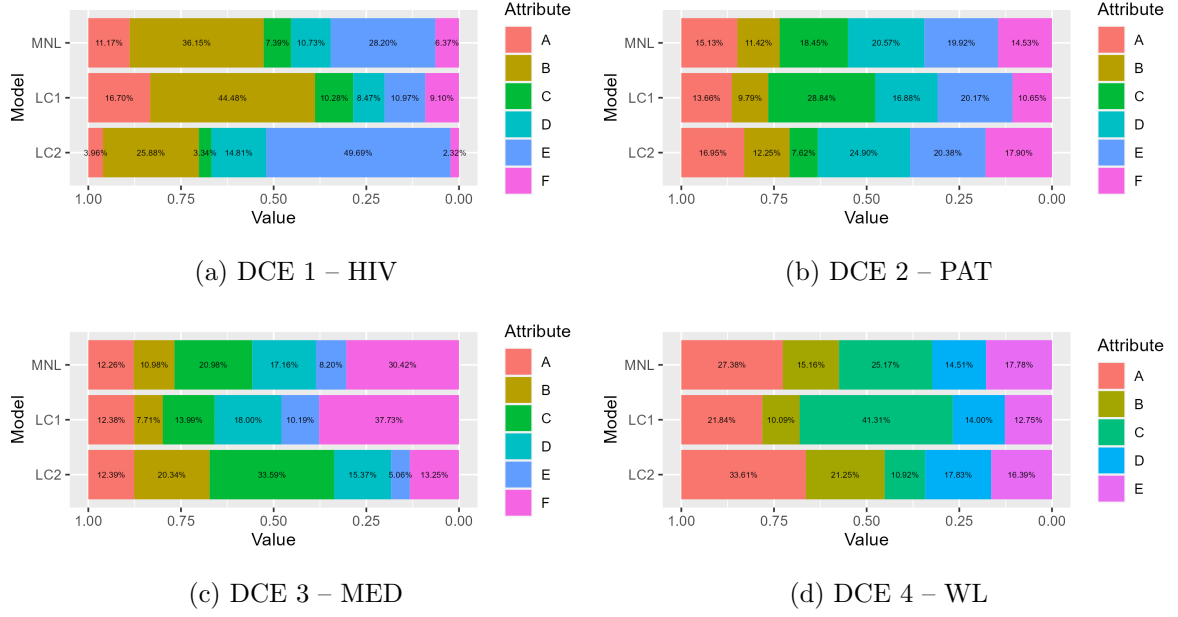


Figure 6.2: Relative importance of attributes for latent-class models with two groups

refer to a single-class model as best fitting. I conclude that concerning the relative importance of attributes, it may be sufficient to rely on simple models such as the MNL model for single classes and the latent-class model for capturing unobserved heterogeneity. This finds support in the results of Shen (2009), who reports that latent-class models are superior to RPL models. Moreover, it contradicts Keane and Wasi (2013), who discuss various models, favouring complex models such as, e.g. GMNL, for DCEs. I agree with Keane and Wasi that log-likelihoods improve with the degrees of freedom allowed by increasing the complexity of the models such that more unobserved heterogeneity can describe the data better, as reported in Tables 6.1–6.4. However, this gain is only marginal in the case of Emmert et al. (2023), Emmert et al. (2024), and Rohrbacher and Emmert (2025), where information criteria such as the BIC or CAIC reject the more complex single-class models. Regarding the relative importance of attributes, more complex structures do not affect the results and are in our view not necessary.

Given this, the continuing extensive utilisation of the RPL model in health economics may appear counterintuitive (de Bekker-Grob, 2009; de Bekker-Grob et al., 2012;

Soekhai et al., 2019; Nouwens et al., 2025). If both the more complex and computationally more demanding RPL model and the rather simple MNL model provide similar results, one may wonder about the need for the complex RPL model. Despite this, also from a theoretical point of view, other models allowing for unobserved heterogeneity may seem superior to the RPL model. The GMNL model allows for more degrees of freedom and captures more heterogeneity on the tails, whereas the SMNL model captures heterogeneity similarly to the RPL model while being much less complex. Therefore, one could utilise a less complex model with a similar output (SMNL) or introduce an even finer modelling approach (GMNL) (Louviere et al., 2002; Louviere and Eagle, 2006; Fiebig et al., 2010). Yet, these kind of models are still used relatively rarely compared to the RPL model (de Bekker-Grob et al., 2012; Soekhai et al., 2019; Nouwens et al., 2025).

Nevertheless, some arguments remain in favour of more complex models. All discussed DCEs comprise attributes with a categorical level structure. Hence, complex models may provide varying results in DCEs with continuous variables. Also, complex models still find support from a theoretical point of view. Recall that the MNL model contains the strong assumption of the IIA, which can be relaxed through other (and more complex) model structures. If the IIA may not hold in a DCE, more complex models may be the right approach. Additionally, the small amount of variation applies to the relative importance of attributes but does not hold for the estimated error terms. In order to estimate the appropriate covariance matrix and the corresponding error terms, one may consider more complex models. However, the model choice should rely on appropriate tools, such as information criteria and model testing. Furthermore, it seems in general advisable to be cautious with using advanced models, as researchers need to make additional decisions with respect to the distribution of random parameters (RPL, GMNL) or the number of latent classes (LC). Furthermore, the interpretation of complex models may rapidly become challenging (de Bekker-Grob, 2009).

References

- Bridges, J. F., A. B. Hauber, D. Marshall, A. Lloyd, L. A. Prosser, D. A. Regier, F. R. Johnson, and J. Mauskopf (2011). Conjoint analysis applications in health-a checklist: A report of the ISPOR good research practices for conjoint analysis task force. *Value in Health* 14(4), 403–413.
- de Bekker-Grob, E. (2009). *Discrete choice experiments in health care: Theory and applications*. Doctoral thesis, Erasmus University Rotterdam.
- de Bekker-Grob, E. W., M. Ryan, and K. Gerard (2012). Discrete choice experiments in health economics: A review of the literature. *Health Economics* 21(2), 145–172.
- Emmert, M., S. Rohrbacher, J. Jahn, K. Fernando, and M. Lauerer (2023). Preferences of people living with HIV for long-acting antiretroviral treatment in Germany: Evidence from a discrete choice experiment. *The Patient – Patient-Centered Outcomes Research* 16(5), 537–553.
- Emmert, M., S. Rohrbacher, F. Meier, L. Heppe, C. Drach, A. Schindler, U. Sander, C. Patzelt, C. Frömke, O. Schöffski, and M. Lauerer (2024). The elicitation of patient and physician preferences for calculating consumer-based composite measures on hospital report cards: Results of two discrete choice experiments. *The European Journal of Health Economics* 25(6), 1071–1085.
- Fiebig, D., M. Keane, J. Louviere, and N. Wasi (2010). The generalized multinomial logit model: Accounting for scale and coefficient heterogeneity. *Marketing Science* 29(3), 393–421.
- Hauber, A. B., J. M. González, C. G. Groothuis-Oudshoorn, T. Prior, D. A. Marshall, C. Cunningham, M. J. IJzerman, and J. F. Bridges (2016). Statistical methods for the analysis of discrete choice experiments: A report of the ISPOR conjoint analysis good research practices task force. *Value in Health* 19(4), 300–315.
- Hensher, D. A., J. M. Rose, and W. H. Greene (2015). *Applied Choice Analysis* (2nd ed.). Cambridge University Press.

- Johnson, F. R., E. Lancsar, D. Marshall, V. Kilambi, A. Mühlbacher, D. A. Regier, B. W. Bresnahan, B. Kanninen, and J. F. Bridges (2013). Constructing experimental designs for discrete-choice experiments: Report of the ISPOR conjoint analysis experimental design good research practices task force. *Value in Health* 16(1), 3–13.
- Keane, M. and N. Wasi (2013). Comparing alternative models of heterogeneity in consumer choice behavior. *Journal of Applied Econometrics* 28(6), 1018–1045.
- Louviere, J., D. Street, R. Carson, A. Ainslie, J. DeShazo, T. Cameron, D. Hensher, R. Kohn, and T. Marley (2002). Dissecting the random component of utility. *Marketing Letters* 13(3), 177–193.
- Louviere, J. J. and T. Eagle (2006). Confound it! That pesky little scale constant messes up our convenient assumptions. In *Sawtooth Software Conference*. Sawtooth Software Inc.
- Nouwens, S. P. H., S. M. Marceta, M. Bui, D. M. A. H. van Dijk, C. G. M. Groothuis-Oudshoorn, J. Veldwijk, J. A. van Til, and E. W. de Bekker-Grob (2025). The evolving landscape of discrete choice experiments in health economics: A systematic review. *Pharmacoeconomics* 43, 879–936.
- Orme, B. K. (2020). *Getting Started with Conjoint Analysis: Strategies for Product Design and Pricing Research*. Sawtooth Software.
- Orme, B. K. and K. Chrzan (2017). *Becoming an Expert in Conjoint Analysis: Choice Modeling for Pros*. Sawtooth Software.
- Rohrbacher, S. and M. Emmert (2025). The elicitation of patient preferences for hip replacement surgery: A discrete choice experiment. *BMC Health Services Research* 25(1), 268.
- Sarrias, M. and R. Daziano (2017). Multinomial logit models with continuous and discrete individual heterogeneity in R: The gmnL package. *Journal of Statistical Software* 79(2), 1–46.

- Shen, J. (2009). Latent class model or mixed logit model? A comparison by transport mode choice data. *Applied Economics* 41(22), 2915–2924.
- Soekhai, V., E. W. de Bekker-Grob, A. R. Ellis, and C. M. Vass (2019). Discrete choice experiments in health economics: Past, present and future. *PharmacoEconomics* 37, 201–226.

7 Conclusion

This dissertation addresses the theoretical aspects of discrete choice experiments and their application in health economics. DCEs are an appropriate and easy way to elicit preferences in line with random utility theory. I discuss the most popular estimation methods and present how to appropriately design a DCE. I highlight the importance of good practice for experimental design.

The DCEs presented in this thesis provide valuable insights regarding the preferences of patients and physicians concerning human immunodeficiency virus and total hip replacement. I identify the preferences and the relative importance of attributes based on a comprehensive analysis of various models. The derived results may help to improve therapies and enhance health services such as, e.g. health report cards.

Furthermore, I analyse the sensitivity of the relative importance of attributes regarding the model approach. Despite numerical differences in estimation results across model types, I find barely any differences between single-class models, whereas multi-class models show varying results. Hence, I conclude that simple models such as multinomial logit models and latent-class models may be sufficient to provide a quick but robust picture of the relative importance of attributes.

However, this practice needs to be handled with caution, as it neglects appropriate model choice based on valid information criteria. Moreover, complex models remain necessary to correctly estimate error terms. Finally, I derive the results from data based only on variables with a categorical level structure. Thus far it remains unknown if the results hold in the presence of other variable types. This question I leave for future research.