

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	<i>Total hip replacements with fulfilled treatment criteria</i> Quality targets reached No assessment available / results not (yet) available Quality targets not reached
Recommendation from other patients	<i>Share of positive recommendations of other patients</i> Recommended by 85% (above average) Recommended by 80% (average) Recommended by 76% (below average)
Number of cases treated	<i>The annual number of total elective hip replacement surgeries in a hospital</i> 230 patients (above average) 148 patients (average) 84 patients (below average)
Equipment and qualification	<i>Hospitals with fulfilled criteria on medical equipment and staff qualification</i> Quality targets reached No assessment available / results not (yet) available Quality targets not reached
EndoCert Certificate	<i>Hospitals which meet requirements of the EndoCert Certification System</i> 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 “gmn1” [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.

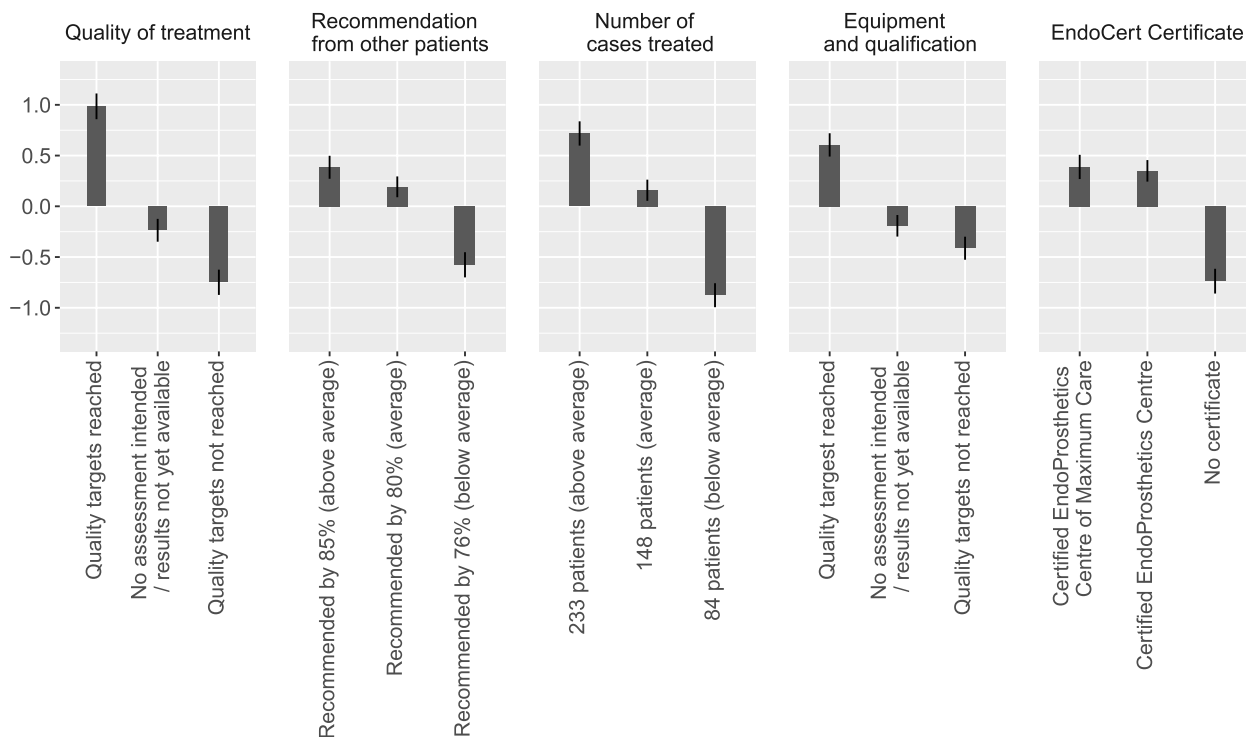


Fig. 1 Graphical display of level estimates with 95%CI of attributes for hospital choice

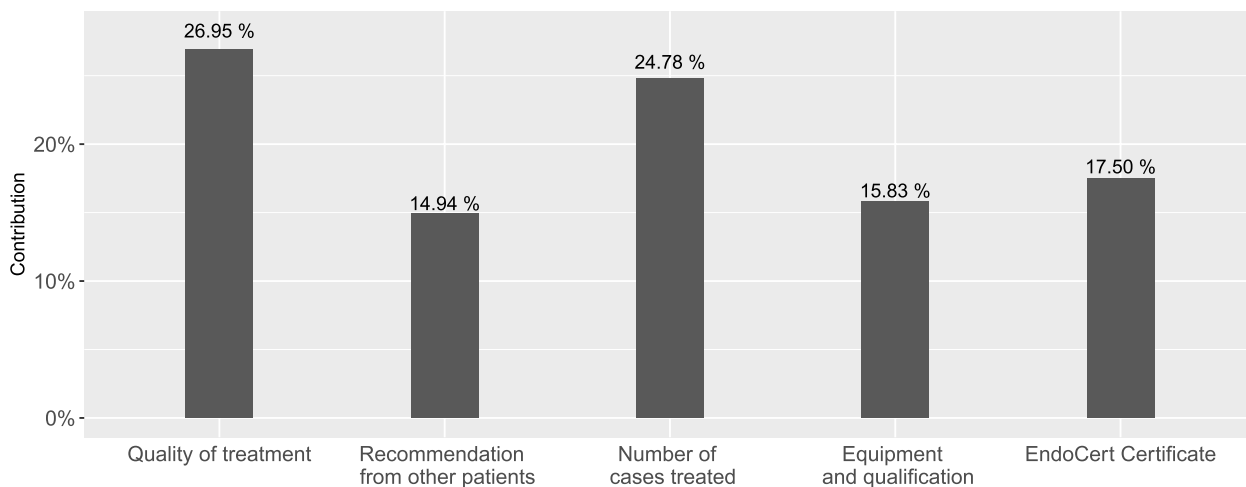


Fig. 2 Graphical display of mean relative importance of attributes for hospital choice

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. *Osteoarthr 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_Bundesqualitaetsbericht-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, Coye 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 blackest 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 New York; 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. *Osteoarthr 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 gmln 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. Nylund 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, 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*. 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. *Osteoarthr 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 (Berufspolitische 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, Lütznert 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.