Qual Life Res
DOI 10.1007/s11136-014-0782-9

Comparison of pre- and post-operative health-related quality of life and length of stay after primary total hip replacement in matched English and German patient cohorts

Matthias Vogl · Reiner Leidl · Werner Plötz · Nils Gutacker

Accepted: 8 August 2014

Abstract

Purpose We compare pre- and post-operative health-related quality of life (HRQoL) and length of stay after total hip replacement (THR) in matched German and English patient cohorts to test for differences in admission thresholds, clinical effectiveness and resource utilisation between the healthcare systems.

Methods German data ($n = 271$) were collected in a large orthopaedic hospital in Munich, Germany; English data ($n = 26{,}254$) were collected as part of the national patient-reported outcome measures programme. HRQoL was measured using the EuroQoL-5D instrument. Propensity score matching was used to construct two patient cohorts that are comparable in terms of preoperative patient characteristics.

Results Before matching, patients in England showed lower preoperative EQ-5D scores (0.35 vs 0.52, $p < 0.001$) and experienced a larger improvement in HRQoL (0.43 vs 0.33, $p < 0.001$) than German patients. Patients in the German cohort were more likely to report no or only moderate problems with mobility and pain preoperatively than their English counterparts. After matching, improvements in HRQoL were comparable (0.32 vs 0.33, $p = 0.638$); post-operative scores were slightly higher in the German cohort (0.82 vs 0.85, $p = 0.585$). Length of stay was substantially lower in England than in Germany (4.5 vs 9.0 days, $p < 0.001$).

Conclusions Our results highlight differences in preoperative health status between countries, which may arise due to different admission thresholds and access to surgery. In terms of quality of life, THR surgery is equally effective in both countries when performed on similar patients, but hospital stay is shorter in England.

Keywords Health-related quality of life · EQ-5D · Total hip replacement · England · Germany

Introduction

Total hip replacement (THR) is a highly effective and cost-effective surgical procedure that alleviates pain, improves mobility and increases health-related quality of life (HRQoL) in patients with osteoarthritis and other indications [1]. While the effectiveness of THR as a procedure is generally undisputed, little is known about the relative performance of different healthcare systems in terms of improvements in HRQoL after surgery and admission

M. Vogl (✉) · R. Leidl
Institute of Health Economics and Health Care Management (IGM), Helmholtz Zentrum München, German Research Center for Environmental Health, P.O. Box 1129, 85758 Neuherberg, Munich, Germany
e-mail: matthias.vogl@helmholtz-muenchen.de

M. Vogl · R. Leidl
Institute of Health Economics and Health Care Management, Munich Centre of Health Sciences, Munich School of Management, Ludwig-Maximilians-Universität München, Munich, Germany

W. Plötz
Department of Orthopaedics and Trauma Surgery, Krankenhaus Barmherzige Brüder München, Romanstraße 93, 80639 Munich, Germany

W. Plötz
Klinikum Rechts der Isar, Technical University Munich, Munich, Germany

N. Gutacker
Centre for Health Economics, University of York, Heslington, York YO10 5DD, UK

Published online: 15 August 2014

thresholds. An international comparison dating back to 1998 reported lower levels of limitations in daily activities after THR in England than in Japan [2]. A more recent study documented higher 1-year HRQoL outcomes after THR in Denmark than Sweden but outcomes were not adjusted for differences in baseline HRQoL [3]. By comparing patient-reported outcome measures (PROM) data for patients in England and Germany before and after surgery, this paper seeks to test for differences in admission thresholds, effectiveness, i.e. improvements in quality of life, and resource utilisation between the two healthcare systems.

The English and German healthcare systems are likely to differ in the way healthcare is organised, and patients are selected for treatment. With 295 THRs per 100,000 inhabitants, Germany currently has the highest hip arthroplasty rates within the European Union [4]. In comparison, the incidence of THR in the UK is approximately 181 THRs per 100,000 inhabitants [4]. This difference in incidence may suggest that Germany implements lower selection thresholds with, on average, healthier patients undergoing surgery [5]. Also, patients in the English National Health Service (NHS) generally have to wait longer before they can undergo surgery; a form of non-price rationing. While the British government has made efforts to reduce waiting time for elective procedures, patients still wait, on average, 12 weeks before they are admitted to hospital for THR [6]. In contrast, patients in Germany generally face no or very short waiting times. Longer waiting times have been associated with poorer preoperative HRQoL and poorer post-operative outcomes [7, 8], although this finding has been contested [9]. Lower HRQoL at admission has been linked to poorer post-operative HRQoL and longer length of stay [10, 11].

Since 2009, all providers in the English NHS have been required to collect PROM data from THR patients before and 6 months after surgery [12]. Patients are asked to describe their health status using the generic EuroQoL-5D (EQ-5D) instrument [13]. The Department of Health in England explicitly intended PROMs data to be used for comparative effectiveness analyses [12]. Similar data for Germany have also been collected in a single-centre prospective study in a large German centre of orthopaedic surgery.

The objective of the present study was to compare baseline HRQoL scores, as measured by the EQ-5D, and improvements in HRQoL scores, and test for differences in length of stay after THR in Germany and England using patient-level data. We hypothesized (1) that preoperative HRQoL would differ between countries and (2) that this would result in differences in post-operative HRQoL scores, change scores, and length of stay between countries.

Table 1 Comparison of age and gender profile of responders and non-responders in English and German patient cohort

Country	Variable	Responders	Non-responders	p value
England	Age (years) (mean, SD)	68.4 (10.1)	68.6 (11.3)	0.178
	Gender (n, %)			
	Male	11,437 (41)	11,012 (38)	<0.001
	Female	16,783 (59)	17,797 (62)	
Germany	Age (years) (mean, SD)	68.7 (9.4)	68.6 (10.3)	0.237
	Gender (n, %)			
	Male	117 (43)	22 (29)	<0.001
	Female	154 (57)	55 (71)	

For continuous data: independent sample t test adjusted for unequal variances

For categorical data: chi-squared test

Patients and methods

Study cohort

Patients were eligible to participate in this study if they were aged 18 or over, had a primary diagnosis of osteoarthritis, underwent elective primary unilateral THR surgery, and were operated in the admitting hospital (i.e. not transferred). Revisions of previous THR were excluded. Only patients with complete pre- and post-operative HRQoL responses were included. The German sample consisted of 348 patients that underwent surgery between January and June 2012 and fulfilled the eligibility criteria. Of those, 342 (98.3 %) participated and provided assessments of their preoperative HRQoL. A total of 271 (77.8 %) patients provided complete follow-up responses at approximately 6 months after surgery. The English sample consisted of 57,108 patients who were operated between April 2011 and March 2012. Of these, 34,867 (61.1 %) provided preoperative assessments of their HRQoL and 28,261 (49.5 %) provided follow-up responses at approximately 6 months after surgery. There was no difference in mean age between responders and non-responders, but the latter were more likely to be female (Table 1). A further 2,016 (3.5 %) patients were excluded because of missing data on other relevant characteristics, resulting in a final sample of 26,245 (46.0 %) patients.

Measurement instruments

HRQoL data were collected using the EQ-5D instrument. The EQ-5D is a generic instrument to measure HRQoL in five dimensions: mobility, self-care, usual activity, pain/discomfort, and anxiety/depression [13]. For each dimension, patients are asked to indicate whether they have no/some/severe problems. This allows for 243 different health states. Each health state is then assigned a utility value to create an index score, where the utility values reflect the preferences of the general population for given health states. This study applied the well-acknowledged UK value set to both German and English data [13]. The EQ-5D instrument is one of the most frequently used HRQoL instruments with demonstrated construct validity, reliability, and useful psychometric characteristics in THR [14].

Statistical analysis

The outcomes of interest were post-operative EQ-5D scores after 6 months, change in EQ-5D scores, and length of inpatient stay. We conducted a retrospective analysis of observational data. Because patients could not be randomly assigned to the English and German 'treatment', observed and unobserved preoperative patient characteristics were likely to differ across the two countries analysed, leading to potentially biased estimates of differences in outcomes. Propensity score matching (PSM) techniques were used to address this issue. PSM simulates the random allocation of patients to the samples based on observed preoperative patient characteristics [15]. To match patients, we pooled data and then predicted for each patient the probability ('propensity score') of being treated in Germany using logistic regression. Several preoperative patient characteristics that were found to be associated with post-operative HRQoL in previous clinical studies [16–18] were controlled for, including age, gender, preoperative EQ-5D score, living arrangements (alone, co-habitant), Charlson comorbidity score [19], and the number of additionally coded comorbidities not already covered. Post-operative EQ-5D score, change in EQ-5D score, and length of stay were excluded from the set of matching variables.

Nearest neighbour (3:1) matching with replacement was performed, where three English patients were matched to each German patient based on the similarity in propensity scores. 'One to many' matching increases precision and confidence intervals compared to 1:1 matching, although matches other than the closest match may introduce bias [20]. To mitigate potential bias, possible matches were rejected if the difference in propensity scores between the German patient and the English matched patient exceeded 0.25 of the standard deviation of the logit of the propensity score [21, 22]. Post-matching balance in covariates was assessed on the basis of standardised differences in means [15], i.e. the difference in means divided by the pooled standard deviation. Improvements in overall balance were assessed by re-estimating the logistic regression on the matched samples and assessing changes in the goodness of fit using Pseudo-R^2 and chi-squared statistics.

We tested for differences between English and German patients in individual preoperative and post-operative patient characteristics and outcomes using two-sample *t* tests for continuous variables and chi-squared tests for categorical variables. A critical value of alpha = 0.05 was chosen for all analyses. The null hypotheses for all statistical tests were that there were no differences between the two patient samples.

Results

Before matching

Patient demographics before and after matching are presented in Table 2. Prior to matching, patient cohorts in both countries were comparable in terms of age and gender, Charlson score, and percentage living alone. The number of additionally coded diagnoses was higher in English patients (mean = 3.51) than in German ones (mean = 2.56) ($p < 0.001$). Pre-treatment EQ-5D index scores were significantly lower (poorer) in the English cohort (mean = 0.35) compared to the German cohort (mean = 0.52) ($p < 0.001$) (Fig. 1). After THR, the post-operative EQ-5D scores were lower in English patients (mean = 0.78) than in German ones (mean = 0.85) ($p < 0.001$). There was an average improvement in HRQoL of 0.43 points for the English cohort and 0.33 points for the German cohort ($p < 0.001$). A comparison of responses on each of the five EQ-5D dimensions revealed that German patients were generally more likely to report having no problems before surgery than English patients (Table 3). This was particularly evident for the dimensions mobility, self-care, and usual activities. Post-operative levels were similar with respect to pain and anxiety, but statistically significant difference for the other three dimensions remained (Table 3). Average length of stay was 4.7 days in England and 9.0 days in Germany ($p < 0.001$).

After matching

After matching, differences in preoperative patient characteristics between countries were reduced for most variables, but actually increased in some (age, number of Charlson comorbidities; see Table 4). The standardised difference in preoperative EQ-5D index scores reduced from 0.51 to 0.06 and the absolute difference reduced from 0.17 to 0.02 (Table 2), indicating a much better balanced sample in terms of preoperative HRQoL. The logistic regression model was predictive of country of treatment before matching (χ^2: 104.2; $p < 0.001$) but was jointly

Table 2 Patient characteristics and outcomes in the German and English cohort before and after matching

Variable	Before matching			After matching	
	German cohort ($n = 271$)	English cohort ($n = 26{,}254$)	*p* value	English cohort ($n = 813$)	*p* value
Age (years) (mean, SD)	68.7 (9.4)	68.3 (10.1)	0.454	69.7 (9.0)	0.121
Gender (*n*, %)					
Male	117 (43)	10,696 (41)	0.420	353 (43)	0.944
Female	154 (57)	15,549 (59)		460 (57)	
Living arrangement (*n*, %)					
Alone	77 (28)	6,757 (26)	0.318	247 (30)	0.540
With partner/care home	194 (72)	19,488 (74)		566 (70)	
Charlson comorbidities (mean, SD)	0.4 (0.8)	0.3 (0.7)	0.184	0.6 (1.1)	<0.001
Additional comorbidities (mean, SD)	2.6 (2.2)	3.5 (2.2)	<0.001	2.9 (1.7)	0.019
Preoperative EQ-5D score (mean, SD)	0.52 (0.30)	0.35 (0.32)	<0.001	0.50 (0.30)	0.370
Post-operative EQ-5D score (mean, SD)	0.85 (0.18)	0.78 (0.25)	<0.001	0.82 (0.22)	0.585
Change in EQ-5D score (mean, SD)	0.33 (0.31)	0.43 (0.34)	<0.001	0.32 (0.31)	0.638
Length of stay (mean, SD)	9.0 (2.8)	4.7 (3.4)	<0.001	4.5 (2.3)	<0.001

For continuous data: two-sample *t* test adjusted for unequal variances

For categorical data: chi-squared test

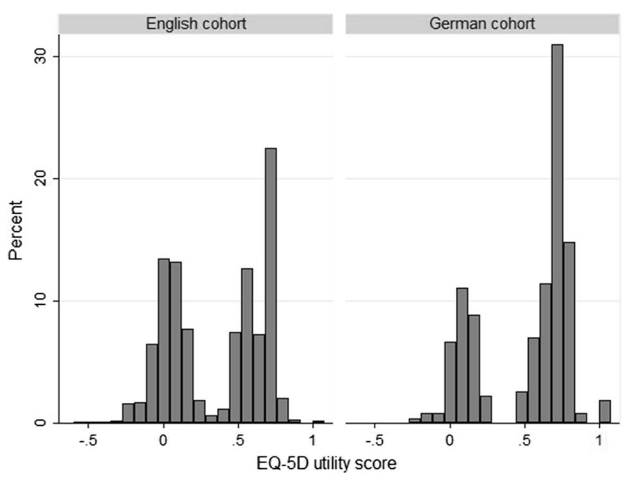

Fig. 1 Distribution of preoperative EQ-5D index scores across countries

insignificant after matching (χ^2: 11.8; $p = 0.066$). The Pseudo-R^2 decreased from 5.5 to 1.3 %.

The matched post-operative EQ-5D scores were lower in English than in German patients, but now by only 0.03 points ($p = 0.585$) (Table 2). Changes in EQ-5D scores were comparable in size between patients from the two countries, with German patients improving 0.33 points and English patients improving 0.32 points ($p = 0.638$). Length of stay differences between the two patient cohorts were largely unaffected by matching (4.5 vs 9.0 days) ($p < 0.001$).

We also conducted a sensitivity analysis in which we restricted the patient cohorts to those aged 55–85, which can be expected to be more homogeneous with respect to diagnosis and expected outcome. The results were qualitatively similar and are available from the authors on request.

Discussion

Previous international comparisons have demonstrated considerable differences in health outcomes and resource use for otherwise similar patients treated for a wide range of conditions [23–25]. These differences in effectiveness and costs have been linked to institutional features, such as care pathways, training of medical professionals, access to services, or payment systems. However, there is a lack of international studies comparing health outcomes on the basis of validated PROMs. Given differences in incidence of THR and waiting times for treatment between the English and German healthcare systems, we hypothesised that HRQoL outcomes for THR patients might also differ across systems. On the basis of previous evidence, we also expected hospital resource utilisation to differ between countries [24].

Our results suggest that the German THR patients in our sample have substantially higher preoperative HRQoL levels but only marginally higher post-operative HRQoL levels than the English patients, so that the overall improvement in HRQoL is smaller. Matching analysis suggests that these observed differences in outcomes are mostly attributable to differences in HRQoL levels at admission and that THR surgery is approximately equally effective in both countries when applied to similar patients. This is encouraging, as it does not indicate any systematic

Table 3 Comparison of pre- and post-operative responses on five EQ-5D dimensions (1 = no problems, 2 = some problems, 3 = severe problems)

EQ-5D dimension	Preoperative			Post-operative		
	German cohort	English cohort	*p* value*	German cohort	English cohort	*p* value*
Mobility (*n*, %)						
Level 1	80 (30)	1,660 (6)	<0.001	215 (79)	15,323 (58)	<0.001
Level 2	191 (70)	24,522 (93)		56 (21)	10,908 (42)	
Level 3	0 (0)	63 (0)		0 (0)	14 (0)	
Self-care (*n*, %)						
Level 1	206 (76)	11,947 (46)	<0.001	252 (93)	20,789 (79)	<0.001
Level 2	64 (24)	14,060 (54)		19 (7)	5,326 (20)	
Level 3	1 (0)	238 (1)		0 (0)	130 (0)	
Usual activities (*n*, %)						
Level 1	91 (34)	1,503 (6)	<0.001	211 (78)	14,108 (54)	<0.001
Level 2	175 (65)	19,888 (76)		59 (22)	11,233 (43)	
Level 3	5 (2)	4,854 (18)		1 (0)	904 (3)	
Pain/discomfort (*n*, %)						
Level 1	9 (3)	167 (1)	<0.001	146 (54)	14,112 (54)	0.184
Level 2	180 (66)	14,972 (57)		121 (45)	11,215 (43)	
Level 3	82 (30)	11,106 (42)		4 (1)	918 (3)	
Anxiety/depression (*n*, %)						
Level 1	183 (68)	15,441 (59)	0.002	228 (84)	21,538 (82)	0.583
Level 2	85 (31)	9,612 (37)		40 (15)	4,255 (16)	
Level 3	3 (1)	1,192 (5)		3 (1)	452 (2)	

* Chi-squared test

Table 4 Standardised differences between cohorts and fit statistics before and after matching

Variable	Unmatched cohorts	Matched cohorts
Age (years)	0.04	−0.11
Gender—male	0.05	−0.00
Living arrangement—living alone	0.06	−0.04
Charlson comorbidities	0.10	−0.24
Additional comorbidities	−0.44	−0.18
Preoperative EQ-5D score	0.51	0.06
Fit statistics		
Pseudo-R^2	5.5 %	1.3 %
Chi-squared test (statistic, *p* value)	104.2 (<0.001)	11.8 (0.066)

differences in the quality of care provided in both healthcare systems. But it also raises the question why English patients had lower preoperative HRQoL? Do patients in the English NHS face difficulties accessing healthcare service and thus have to endure poorer health over a longer period before undergoing surgery, or do patients in Germany—as hypothesised by some German health insurances [5]—often receive operations too early. One potential explanation for differences in preoperative HRQoL is the considerably longer waiting time in the English NHS, although evidence on the effects of this mechanism is mixed [7, 8]. Other explanations are higher referral and treatment thresholds in England compared to Germany. This argument is supported by the higher age-sex standardised incidence rates in Germany and the observation that patients in the German cohort reported more frequently to have no problems in terms of mobility or usual activities preoperatively than their English counterparts. This corroborates the threshold hypothesis and raises questions about the optimal time point for referral for primary THR.

A recent EU-wide research project documented variation in treatment costs for THR of up to 300 % across nine European countries (adjusted for purchasing power differences) [24]. Unadjusted average resource utilisation for THR surgery was lower in England (5,700 EUR) than in Germany (6,400 EUR). Our study confirms large differences in resource use between the two countries for patients with similar baseline characteristics. Matched THR patients in England are expected to be discharged after 4.5 days, which is 4.5 days earlier than similar patients in Germany. This finding is both clinically and statistically important. Given our finding of comparable health outcomes at 6 month after surgery, this indicates that further improvements in discharge management and coordination of aftercare might be possible in Germany. This hypothesis is further corroborated by the observation that the particular hospital from which the German patient cohort was recruited already operates a well-functioning clinical pathway with comparably effective discharge management and short LOS (9 days, compared to the German average of 12 days [26]). An alternative explanation is that English hospitals rely more on external rehabilitation providers, whereas German hospitals begin post-operative rehabilitation measures while the patient is still in hospital. Unfortunately, our data do not permit exploring these hypotheses further.

The relationship between resource use and health outcomes is one of the most poorly understood issues in health policy. Our findings suggest that shorter length of stay is not associated with worse health outcomes after 6 months. This is consistent with previous studies that report no or negative association between the costs or length of stay and health outcomes in the English NHS [10, 27]. Hence, given these results, one would not expect reductions in post-operative length of stay in Germany to have adverse effects on health outcomes.

Our study contributes to the literature in several ways. First, to our knowledge, this is the first study to use a broadly accepted generic HRQoL measure as an endpoint and to pool patient-level data from two countries to compare pre- and post-operative levels of HRQoL for THR patients in different countries [2]. Other studies only had access to post-operative HRQoL data and could therefore not explore the importance of differences in baseline HRQoL levels [3]. Second, random allocation of patients to healthcare systems is generally not feasible, thereby impeding international comparisons of the effectiveness of orthopaedic surgery. To address this issue, we employed matching methods and simulated random allocation based on observed preoperative characteristics. Third, length of stay is measured in a natural metric (bed days) and can therefore be compared across countries. In contrast, costs are generally not directly comparable because of different accounting rules or purchasing power differences.

This study also has several limitations. First, German data were only available for patients treated in one hospital. This raises concerns about the representativeness of our sample. The Krankenhaus Barmherzige Brüder München is one of the largest orthopaedic units in Germany, performing approximately 2,000 joint replacements in patients with osteoarthritis each year. Patients treated in this hospital may be substantially different from those treated elsewhere in Germany; similarly the quality of care provided by this hospital and its discharge management (see above) may not be representative. Second, for the German sample, no comparable data were available on patients' socio-economic background [28], their preoperative expectations of surgery [29] or hip-specific health status measured by, e.g. the Oxford Hip Score (OHS). Third, coding differences between countries, especially with respect to the number of recorded comorbidities, may have affected our analysis. Fourth, patients in different countries may report their underlying health status differently, thereby introducing a cultural or language bias to our comparison [30]. For example, Murray and colleagues compared how participants in the WHO Multi-country Survey Study on Health and Responsiveness described a series of patient health profiles, so-called 'anchoring vignettes' and found substantial cross-country variation [31]; see also [32–34]. Unfortunately, our data do not include suitable anchoring vignettes to adjust for reporting heterogeneity in the descriptive part of the EQ-5D. The corresponding issue on the valuation side is the choice of weights used to aggregate health dimensions into index scores. Knies and colleagues found that the UK value set produces larger decrements than a comparable German value set [35]. To investigate sensitivity to the valuation approach, we also conducted analysis using a German value set which is not based on hypothetical but on experienced health states from the general population [36]. We found results to be broadly similar, although the difference in post-operative scores after matching is now statistically significant [0.79 (SD = 0.15) in England compared to 0.77 (SD = 0.15) in Germany, $p = 0.031$]; full results are available from the authors on request. The bimodal distribution of preoperative EQ-5D values found in Fig. 1 has also been reported by other studies [37]. This distribution appears much more similar to the normal distribution when using the German experience-based value set. The preoperative OHS is also distributed approximately normal in the English patient cohort (not shown), which points to the need to further consider the valuation approach when making international comparisons. Finally, while post-matching balance was greatly increased, we note that remaining imbalances in unobserved characteristics may have biased the results in unknown directions.

There are several avenues for further research. First, this analysis could be repeated with more countries and more observations within each country to provide a more representative picture of country differences in treatment pathways and to facilitate a comparative effectiveness research. Second, future analysis should include a condition-specific PROM such as the OHS or the Western Ontario and McMaster Universities Osteoarthritis score (WOMAC) alongside a generic measure such as the EQ-5D. This could potentially reveal more specific clinical differences not currently captured by the EQ-5D. Third, potential differences in reporting behaviour across countries and patient groups should be explored further to ensure that observed differences in outcomes are due to differences in health, not health perception or reporting. Fourth, it may be instructive to compare the robustness of our findings when using health state valuations derived from THR patients instead of the general public [36, 38]. Fifth, we focused on primary replacements and there is scope to extend this research to encompass revisions. Sixth, future studies should aim to capture all relevant costs in order to allow for meaningful comparative cost-effectiveness analyses. Length of stay is only a partial measure of overall costs, i.e. it correlates well with some cost components (e.g., wages, estates) but not with others (e.g., implant cost, costs occurring outside the hospital). Finally, more research is needed to understand how care pathways and admission thresholds differ across countries and how this translates into different patients' health outcomes.

Conclusions

Compared to their German colleagues, English surgeons operate on patients with, on average, lower HRQoL. The difference in baseline HRQoL levels is reflected in larger improvements in HRQoL after surgery experienced by English patients, and larger post-operative scores reported by German patients. However, after controlling for these baseline differences, there does not seem to be any systematic differences in the effectiveness of THR surgery in England and Germany.

This study highlights the potential for collecting and comparing PROM data across countries to study differences in preoperative health levels, which are assumed to be reflective of different admission thresholds, and post-operative outcomes. However, with few exceptions, healthcare systems generally do not collect PROM data on a routine basis. We encourage researchers and policy makers to facilitate such data collection and collaborate across jurisdictions in order to improve care pathways and enhance the HRQoL of patients.

Acknowledgments The authors would like to thank Karen Bloor, Richard Cookson, Anne Mason, and Andrew Street as well as participants at the Health Economic seminar series at the Helmholtz Centre Munich for helpful comments and suggestions. Parts of this work were carried out while MV was visiting the Centre for Health Economics, York.

Conflict of interest The authors do not have financial or other conflicts of interest that might bias this work.

Ethical standards The German data were collected prospectively as part of a health state evaluation study carried out in a large orthopaedic hospital in Munich, Germany ('Krankenhaus Barmherzige Brüder München'). Ethics approval for this study was granted by the Ethics committee of Klinikum rechts der Isar, Technical University Munich. The observational study was in accordance with the Helsinki Declaration of 1975, as revised in 2000. All patients gave their informed consent prior to their inclusion in the study. English HRQoL data were collected retrospectively as part of the routine national PROM survey and linked to routine inpatient data from the Hospital Episode Statistics dataset. No further ethics approval was required for secondary data analysis, and no patient-identifiable data were available. Data collection protocols were similar in both countries and followed guidance set out by the Department of Health in England.

References

1. Jenkins, P. J., Clement, N. D., Hamilton, D. F., Gaston, P., Patton, J. T., & Howie, C. R. (2013). Predicting the cost-effectiveness of total hip and knee replacement: a health economic analysis. *Bone and Joint Journal, 95-B*(1), 115–121.
2. Imamura, K., & Black, N. (1998). Outcome of total hip replacement in Japan and England. Comparison of two retrospective cohorts. *International Journal of Technology Assessment in Health Care, 14*(4), 762–773.
3. Gordon, M., Paulsen, A., Overgaard, S., Garellick, G., Pedersen, A. B., & Rolfson, O. (2013). Factors influencing health-related quality of life after total hip replacement—a comparison of data from the Swedish and Danish hip arthroplasty registers. *BMC Musculoskeletal Disorders, 14*, 316.
4. OECD. (2012, 17 July 2013). Health at a glance Europe 2012. Retrieved January 17, 2014, from http://www.oecd.org/els/health-systems/HealthAtAGlanceEurope2012.pdf.
5. Klauber, J., Geraedts, M., Friedrich, J., & Wasem, J. (2013). *Krankenhaus-Report 2013 Mengendynamik: Mehr Menge, mehr Nutzen?* Stuttgart: Schattauer.
6. Siciliani, L., Moran, V., & Borowitz, M. (2013). Measuring and comparing health care waiting times in OECD Countries, OECD Health Working paper No. 67. Retrieved February 2, 2014, from http://search.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=DELSA/HEA/WD/HWP%282013%298&docLanguage=En.
7. Garbuz, D. S., Xu, M., Duncan, C. P., Masri, B. A., & Sobolev, B. (2006). Delays worsen quality of life outcome of primary total hip arthroplasty. *Clinical Orthopaedics and Related Research, 447*, 79–84.
8. Vergara, I., Bilbao, A., Gonzalez, N., Escobar, A., & Quintana, J. (2011). Factors and consequences of waiting times for total hip arthroplasty. *Clinical Orthopaedics and Related Research®, 469*(5), 1413–1420.
9. Hoogeboom, T. J., van den Ende, C. H. M., van der Sluis, G., Elings, J., Dronkers, J. J., Aiken, A. B., et al. (2009). The impact of waiting for total joint replacement on pain and functional status: A systematic review. *Osteoarthritis and Cartilage, 17*(11), 1420–1427.
10. Street, A., Gutacker, N., Bojke, C., Devlin, N., & Daidone, S. (2014). Variations in outcome and costs among NHS providers for common surgical procedures: econometric analyses of routinely collected data. *Health Serv Deliv Res, 2*(1).
11. Dall, G. F., Ohly, N. E., Ballantyne, J. A., & Brenkel, I. J. (2009). The influence of pre-operative factors on the length of in-patient stay following primary total hip replacement for osteoarthritis: A multivariate analysis of 2302 patients. *Journal of Bone and Joint Surgery British, 91*(4), 434–440.
12. Department of Health. (2010, 13 September 2013). Guidance on the routine collection of patient reported outcome measures (PROMs). Retrieved January 12, 2014, from http://webarchive.nationalarchives.gov.uk/20130107105354/http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/@dh/@en/documents/digitalasset/dh_092625.pdf.
13. Dolan, P. (1996). Modelling valuations for health states: The effect of duration. *Health Policy, 38*(3), 189–203.
14. Kind, P., Brooks, R., & Rabin, R. (2005). *EQ-5D concepts and methods: A developmental history*. Dordrecht: Springer.

15. Austin, P. C. (2009). Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Statistics in Medicine, 28*(25), 3083–3107.
16. Ethgen, O., Bruyere, O., Richy, F., Dardennes, C., & Reginster, J. Y. (2004). Health-related quality of life in total hip and total knee arthroplasty: A qualitative and systematic review of the literature. *The Journal of Bone and Joint Surgery, 86-A*(5), 963–974.
17. Biring, G. S., Masri, B. A., Greidanus, N. V., Duncan, C. P., & Garbuz, D. S. (2007). Predictors of quality of life outcomes after revision total hip replacement. *Journal of Bone and Joint Surgery British, 89*(11), 1446–1451.
18. Husted, H., Holm, G., & Jacobsen, S. (2008). Predictors of length of stay and patient satisfaction after hip and knee replacement surgery: Fast-track experience in 712 patients. *Acta Orthopaedica, 79*(2), 168–173.
19. Bjorgul, K., Novicoff, W., & Saleh, K. (2010). Evaluating comorbidities in total hip and knee arthroplasty: available instruments. *Journal of Orthopaedics and Traumatology, 11*(4), 203–209.
20. Rassen, J. A., Shelat, A. A., Myers, J., Glynn, R. J., Rothman, K. J., & Schneeweiss, S. (2012). One-to-many propensity score matching in cohort studies. *Pharmacoepidemiology and Drug Safety, 21*(Suppl 2), 69–80.
21. Austin, P. C. (2011). Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies. *Pharmaceutical Statistics, 10*(2), 150–161.
22. Austin, P. C. (2009). Some methods of propensity-score matching had superior performance to others: Results of an empirical investigation and Monte Carlo simulations. *Biometrical Journal, 51*(1), 171–184.
23. Pearse, R. M., Moreno, R. P., Bauer, P., Pelosi, P., Metnitz, P., Spies, C., et al. (2012). Mortality after surgery in Europe: A 7 day cohort study. *Lancet, 380*(9847), 1059–1065.
24. Stargardt, T. (2008). Health service costs in Europe: Cost and reimbursement of primary hip replacement in nine countries. *Health Economics, 17*(1 Suppl), S9–20.
25. Coleman, M. P., Quaresma, M., Berrino, F., Lutz, J. M., De Angelis, R., Capocaccia, R., et al. (2008). Cancer survival in five continents: A worldwide population-based study (CONCORD). *The Lancet Oncology, 9*(8), 730–756.
26. Institut für das Entgeltsystem im Krankenhaus (InEK). (2013). G-DRG Browser 2013; Datenveröffentlichung gem. § 21 KHEntgG. Retrieved February 11, 2014, from http://www.g-drg.de/cms/Datenveroeffentlichung_gem._21_KHEntgG.
27. Gutacker, N., Bojke, C., Daidone, S., Devlin, N. J., Parkin, D., & Street, A. (2013). Truly inefficient or providing better quality of care? Analysing the relationship between risk-adjusted hospital costs and patients' health outcomes. *Health Economics, 22*(8), 931–947.
28. Laudicella, M., Siciliani, L., & Cookson, R. (2012). Waiting times and socioeconomic status: Evidence from England. *Social Science and Medicine, 74*(9), 1331–1341.
29. Haanstra, T. M., van den Berg, T., Ostelo, R. W., Poolman, R. W., Jansma, E. P., Cuijpers, P., et al. (2012). Systematic review: Do patient expectations influence treatment outcomes in total knee and total hip arthroplasty? *Health Quality Life Outcomes, 10*, 152.
30. Rice, N., Robone, S., & Smith, P. C. (2012). Vignettes and health systems responsiveness in cross-country comparative analyses. *Journal of the Royal Statistical Society: Series A (Statistics in Society), 175*(2), 337–369.
31. Murray, C. J. L., Özaltin, E., Tandon, A., Salomon, J. A., Sadana, R., & Chatterji, S. (2003). Empirical evaluation of the anchoring vignette approach in health surveys (Chapter 30). In C. J. L. Murray & D. B. Evans (Eds.), *Health systems performance assessment: Debates, methods and empiricism*. Geneva: Word Health Organisation.
32. Bago d'Uva, T., Van Doorslaer, E., Lindeboom, M., & O'Donnell, O. (2008). Does reporting heterogeneity bias the measurement of health disparities? *Health Economics, 17*(3), 351–375.
33. Salomon, J. A., Tandon, A., & Murray, C. J. (2004). Comparability of self rated health: Cross sectional multi-country survey using anchoring vignettes. *BMJ, 328*(7434), 258.
34. Jurges, H. (2007). True health vs response styles: Exploring cross-country differences in self-reported health. *Health Economics, 16*(2), 163–178.
35. Knies, S., Evers, S. M., Candel, M. J., Severens, J. L., & Ament, A. J. (2009). Utilities of the EQ-5D: transferable or not? *Pharmacoeconomics, 27*(9), 767–779.
36. Leidl, R., & Reitmeir, P. (2011). A value set for the EQ-5D based on experienced health states: Development and testing for the German population. *Pharmacoeconomics, 29*(6), 521–534.
37. Jansson, K. A., & Granath, F. (2011). Health-related quality of life (EQ-5D) before and after orthopedic surgery. *Acta Orthopaedica, 82*(1), 82–89.
38. Pinedo-Villanueva, R. A., Turner, D., Judge, A., Raftery, J. P., & Arden, N. K. (2013). Mapping the Oxford hip score onto the EQ-5D utility index. *Quality of Life Research, 22*(3), 665–675.