

A Service of

ZBW

Leibniz-Informationszentrum Wirtschaft Leibniz Information Centre for Economics

Brunner, Sascha S.

## Working Paper Cost and quality differences in elective total hip arthroplasty between Swiss hospital departments

WWZ Working Paper, No. 2024/12

**Provided in Cooperation with:** Center of Business and Economics (WWZ), University of Basel

*Suggested Citation:* Brunner, Sascha S. (2024) : Cost and quality differences in elective total hip arthroplasty between Swiss hospital departments, WWZ Working Paper, No. 2024/12, University of Basel, Center of Business and Economics (WWZ), Basel, https://doi.org/10.5451/unibas-ep96743

This Version is available at: https://hdl.handle.net/10419/306527

#### Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

#### Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



# WWW.ECONSTOR.EU



Faculty of Business and Economics



October 2024

# Cost and Quality Differences in Elective Total Hip Arthroplasty between Swiss Hospital Departments

WWZ Working Paper 2024/12

Sascha Sebastian Brunner

A publication of the Center of Business and Economics (WWZ), University of Basel.

© WWZ 2024 and the authors. Reproduction for other purposes than the personal use needs the permission of the author

Universität Basel Peter Merian-Weg 6 4052 Basel, Switzerland wwz.unibas.ch **Corresponding Author:** Sascha Sebastian Brunner Tel. +41 61 207 33 59 Mail : <u>saschasebastian.brunner@unibas.ch</u>

# Cost and Quality Differences in Elective Total Hip Arthroplasty between Swiss Hospital Departments

Sascha S. Brunner

Wirtschaftswissenschaftliche Fakultät, Universität Basel, Peter Merian-Weg, 6, 4002 Basel, Schweiz

#### $02 \ 01 \ 2024$

#### WORKING PAPER

#### Abstract

**Objective:** Even in highly regulated and advanced health care markets, treatment costs and outcome quality substantially vary across facilities. Interhospital variation — particularly when it is not attributable to different patient compositions — might indicate inefficiencies in the health care system. This paper studies the sources and extent of these variations using novel data on total hip arthroplasty (THA) in Switzerland.

**Methodology:** The study sample comprises 20,918 patients from 110 hospital departments who underwent THA in 2016, 2017, and 2018. Two primary dependent variables are considered: quality of the outcome is measured by a binary variable that indicates a revision within two years of the surgery, and THA case costs represent the total inpatient costs of the initial THA as reported by the hospital. Interdepartmental cost and quality differences are explored using hierarchical two-level models. The study considers patient health factors and departmental factors, namely THA case volume, hospital type, and surgical techniques, as explanatory variables.

**Results:** Even after controlling for the patient mix, there is a significant variance in both costs and quality between hospital departments. Patient health factors account for 35% and 60% of the variance in interdepartmental differences for THA cost and quality, respectively. Several patient health factors, especially BMI and certain comorbidities, are important drivers of these differences in costs and quality. Additionally, both patient health and departmental factors together explain 50% and 80% of the significant

interdepartmental variance in costs and quality, respectively. Low case volume, the department's proportion of cemented implants, and department affiliation with university hospitals were significantly associated with higher inpatient costs.

**Discussion:** A relatively high share of interdepartmental variance in both quality and cost can be explained by different departmental patient mixes. The residual unexplained variance, albeit significant, is modest in international comparison. This suggests that there are no substantial unexplainable inefficiencies among hospital departments in Switzerland. The results of departmental level suggest that THA costs could be reduced by economies of scale and increasing the share of uncemented prostheses without compromising the quality of care. The high reduction of interdepartmental variance by controlling for patient mix and the fact that most but not all patient health risk factors are linked to both higher costs and lower quality underlines the importance of carefully-designed patient health risk adjustments in potential future pay for performance (P4P) programs in THA.

*Keywords:* inpatient costs, risk adjustment, total hip arthroplasty, hospital cost function, quality of care, multilevel model, hybrid model, between-within model, mixed model, fixed effects, random effects

JEL Classification: I11, C33, C50

## 1 Introduction

Total Hip Arthroplasty (THA) is a prevalent and well-established medical procedure in which the head of the femur and the acetabulum (socket located in the pelvis) are replaced with artificial implants. According to the OECD (2019), an average of 182 hip replacement surgeries were performed per 100,000 people across 32 OECD members<sup>1</sup> in 2017; among these 32 countries, Switzerland distinguished itself with a per capita rate of 307 operations per 100,000 people, ranking second only after Germany. The widespread application of THA can be attributed to its effectiveness as a comparatively safe and successful treatment option for patients suffering from persistent mobility limitations, pain or stiffness in the hip joint due to chronic conditions like osteoarthritis, arthritis, or osteonecrosis (Gandhi et al., 2010; Fordham et al., 2012). Given the aging population and the impracticability of outpatient hip replacements, the significance of THA is expected to increase in the future.

Despite THA's relatively high level of standardisation, heterogeneity in both treatment cost and quality has surfaced within and across hospitals. Cost (and quality) variations among health care providers are commonly perceived as a sign of inefficiency, suggesting that some facilities are not using their resources as efficiently as their best practice counterparts. However, variation might not always be connected to inefficiency but to the heterogeneity of patients, organizational structures, or treatment standards of each facility. The Swiss inpatient reimbursement system takes into account observable differences in patients and facilities, utilizing the Swiss Diagnosis-Related Group (DRG) catalog in tandem with hospital-specific base rates. In relation to THA, the 2023 DRG catalog introduces 11 unique DRG-weights, ranging from 1.514 to 4.238. These weights take into account different patient-specific treatments and health conditions, and thus represent the various patient mixes of the hospitals. DRG-weights are based on cost monitoring from previous years, with Swiss-wide data reported by network hospitals, which represent about half of all Swiss hospitals. In a parallel vein, hospital base rates are the result of hospital benchmarking and annual negotiations between hospitals and insurance entities, aiming to account for the distinct situation of each hospital. The resulting reimbursement for each patient is the product of the patient-specific DRG-weights multiplied by the hospital-wide base rate. This reimbursement is designed to equitably adjust for exogenous differences while encouraging cost-efficiency through its prospective character. Distinguishing between unavoidable differences and unexplainable variance that indicate inefficiency across providers raises the question of "What proportion of variance in cost is justifiable?" (Keeler, 1990). To address this question, it is essential to analyze the extent of observable and unobserved variations, not only in costs but also in quality

 $<sup>^1\</sup>mathrm{All}$  OECD members except the United States, Greece, Turkey and Japan

across departments. Understanding the alignment or lack thereof, between quality and cost is crucial, particularly in determining if variations in quality justify seeming cost-inefficiencies. Quality variance gains further importance in the Swiss context, as there is growing discourse around pay-for-performance (P4P) schemes, which would link quality indicators to hospital remuneration.

Prior studies have identified a similar set of patient health factors relevant for cost and quality variability in THA. However, their relative importance remains contentious. For instance, while Stargardt (2008) and Robinson et al. (2012) reported that patient mix accounts for approximately 5% of cost variance, Zegers et al. (2011) found that patient mix constitutes more than 20% of interdepartmental quality variation in THA. The literature also provides conflicting results regarding explanatory factors on a hospital level, especially concerning surgical procedures and THA case volume (Hooper et al., 2009; Castelli et al., 2015; Paxton et al., 2015; Vries et al., 2011). Some of the discrepancies in previous research may be attributed to differing research scopes and datasets. However, as Ali and Gibbons (2017) and Laudicella et al. (2010) point out, many THA studies lack an appropriate methodology to account for the hierarchical data structure, in which patients are nested in hospitals. Furthermore, apart from Ramos et al. (2014), most THA studies focus merely on either cost or quality, which results in an incomplete picture and limits the practical utility of the findings from a policy perspective. In light of the incomplete previous research and a lack of studies in Switzerland, this paper aims to investigate whether there are relevant differences in quality and costs of THA across Swiss hospital departments and to determine how much of this variation can be explained by the same set of observable patient health and organizational factors.

This study considers two dependent variables: the total costs of the hospital stay associated with the initial THA recorded by the Swiss hip registry and the incidence of two-year revision<sup>2</sup>, which serves as the quality measure. The relationship between quality and cost has been explored in other disease areas than THA: either quality has been included in the hospital cost function or cost has been included in a quality model. However, these methodologies have often led to contradictory and inconclusive results (Schreyögg and Stargardt, 2010; Carey and Burgess, 1999; McKay and Deily, 2008; Street et al., 2010). As Schreyögg and Stargardt (2010) points out, the apparent effect of cost on quality, or the reverse, might be more due to the degree of control for patient mix in the study rather than a direct causal relationship. Essentially, the patient mix can confound the cost-quality relationship, making causal conclusions challenging, especially when it is un-

 $<sup>^{2}</sup>$ Revisions involve the removal and replacement of the artificial implant with a new one. Generally, implants are expected to hold more than 10 years; revision rates within two years after surgery is a common quality indicator that indicates a short useful life of the implant (SIRIS, 2022)

certain whether the patient mix has been thoroughly considered. As opposed to previous studies, this study uses two separate models for quality and cost but with the same explanatory variables. This has the advantage that the effect of the same explanatory variables on both outcomes can be interpreted without the confounding influence of the cost quality relationship. Consequently, the results can shed light on potential simultaneous impacts or trade-offs between quality and cost determinants. Given the nested structure of the data, results are estimated with hierarchical two-level models, which have become a popular tool in empirical health economics to analyse variance between health care providers (Rice and Jones, 1997). In this model, patients (representing the first level) are nested in hospital departments (the second level). Evaluating two prevalent types of hierarchical models—the random effects model, known as the mixed model in a multilevel context, and the hybrid model—adds an extra layer of robustness to the findings.

The remainder of the paper is organised as follows. The first section covers the econometric strategy and data sample in-depth. The second section presents evidence on cost and quality variances, highlighting relevant drivers. In the third section, results are interpreted in the context of prior literature and potential policy implications are discussed.

## 2 Methods and Data

#### 2.1 Sample

The study sample is derived from three Swiss datasets. The main dataset, the Swiss implant registry – SIRIS, enables the identification of all patients with a primary THA and potential revisions as well as the hospital department. The SIRIS registry has been matched to two datasets managed by the Swiss federal statistical office (SFSO)<sup>3</sup>. The Medical Statistics of Hospitals includes variables related to patient demographics, treatments, diagnoses, and hospital characteristics. The Diagnosis Related Case Cost Statistics includes comprehensive inpatient cost data.

Figure 1: THA Patients Nested in Hospital Departments



Note: own display of the simplified sample structure

The sample is at the patient level: each patient who underwent THA in 2016, 2017, or 2018 represents one observation. Data for 2019 and 2020 are available and used to observe all potential revision in the two-years post surgery. As depicted in Figure 1, the sample includes only patients without an additional private or half-private hospital insurance that resided in a network hospital since the *Diagnosis Related Case Cost Statistics* only comprises cost data for these patients. THAs resulting from hip fractures, developmental

 $<sup>^{3}</sup>$ SIRIS data has not been linked to the SFSO datasets before, but about 80% of the THAs can be matched based on the anonymized patient identifiers, treatments, and length of stay.

dysplasia, and perthes conditions were excluded, as they represent distinct pathologies and care. All patients who died or emigrated from Switzerland within two years of surgery were excluded. Departments with fewer than 30 THA cases were excluded to limit the sensitivity of the analysis to data-entry errors, i.e., assignment to the wrong department, and outliers. In two-level studies, hospital departments are generally favored over entire hospitals: departments represent units of work, while hospitals are politically and strategically motivated organizations (Laudicella et al., 2010). However, only eight hospitals have more than one department treating THA patients. The final sample comprises 20,918 patients are nested within 110 hospital departments.

## 2.2 Dependent Variables: Inpatient Costs and two-year Revisions

The costs of THA encompass the total inpatient cost of the hospital stay associated with the initial THA. Generally, hospitals bear these costs themselves<sup>4</sup> but report them in detail through the standardized REKOLE system. REKOLE-reported cost data is collected and published annually in the *Diagnosis Related Case Cost Statistics*. The correct use of REKOLE is controlled by an external inspection body and compulsory for network hospitals (Salzmann, 2021). Therefore, inpatient cost data, such as the costs of THA, is likely to be recorded uniformly and comparably across patients and hospitals. Table 1 shows the mean and standard deviation of THA costs. With an average of about 17,500 CHF, costs seem relatively stable over the three years. Costs are moderately right-skewed, with a median of about 16,000 CHF.

	2016	2017	2018	Mean
Frequency of initial THA	6,394	7,121	7,403	6,973
Total Inpatient Costs (CHF)				
Mean	$17,\!379$	$17,\!441$	$17,\!150$	$17,\!319$
Standard deviation	$6,\!951$	$8,\!358$	$5,\!976$	$7,\!155$
Two-year Revision (%)				
Mean	3.19	2.75	2.99	2.97

Table 1: Case Costs and Revision Incidence per THA (2016-2018)

In Figure 2, one can easily see the differences between hospital departments. The average costs in most hospitals are between 15,000 and 20,000 CHF. However, some departments

<sup>&</sup>lt;sup>4</sup>For each case, hospitals receive payment that is calculated as the product of the patient-specific DRGweight and the base rate. If a patient's hospital stay is unusually long or short for the corresponding DRG, the payment is slightly adjusted. DRG-weights are calculated based on the *Diagnosis Related Case Cost Statistics* from past years. Consequently, there is a strong correlation of 0.62 between case payments and case costs

have average costs of about 25,000 CHF. Visually, these fluctuation across departments persist, albeit on a lower level, even in a subsample of "standard" patients, see APPENDIX A. A standard patient is defined as THA patient without special health risks that was assigned to the most common DRG regarding THA, I46C.

Related CaseDiagnosis Cost *Statistics* distinguishes between direct costs, overhead costs, and facility fees<sup>5</sup>. The overhead costs and facility fees are proportionally allocated to individual patient cases, while direct costs are already noted at the patient level. The direct and overhead costs are further divided into and twelve four categories, respectively. This dissection of the total inpatient costs based on the **REKOLE** catalogue is illustrated in Table 2. The mean and standard deviation for each cost

Figure 2: Differences in Departmental Costs



Note: perpendicular lines represent the averages in the corresponding categories

category are calculated across all departments, as well as separately for the departments in the top 10% for most expensive THA cases and the bottom 10% for least expensive THA cases. In terms of average total costs, THA in one of the top 10% most expensive departments costs about 4,000 CHF more than the overall average and 7,500 CHF more than in one of the 10% least expensive departments. Departments that are less expensive maintain a lower expenditure in all categories except for the "Physician Salaries" category, which is directly tied to the efforts physicians invest in attending to the patients. The 10% least expensive departments report higher mean costs in this area compared to the average across all departments. Cost categories exhibit a high degree of heterogeneity, encompassing categories with both high and low standard deviation. It seems that the total variation in costs is mainly composed of the categories of direct costs

<sup>&</sup>lt;sup>5</sup>all costs related to the usage of the hospital facilities

	Full sample			Subsamples (each 10% of sample)				
	All departments			Most	expensive	Least	expensive	
			% of total	depa	rtments	depa	artments	
	Mean	Std.Dev.	$cost variance^{a}$	Mean	Std.Dev.	Mean	Std.Dev.	
Direct Costs								
Medication	240	1338	14.32	331	2161	264	3377	
Physician Salaries	1289	1647	14.92	788	1516	1396	1522	
Material	5057	1469	16.79	5222	1749	4519	1255	
Undefined	119	368	0.91	117	256	73	254	
<b>Overhead</b> Costs								
Diagnostics	17	118	0.09	72	287	6	75	
Dialysis	2	107	0.09	1	15	0	19	
Laboratory	136	242	0.22	212	334	74	114	
Emergency	17	92	0.04	27	113	19	109	
Nursing	3075	2378	13.53	4088	2930	2208	1955	
Picturing	121	156	0.13	166	252	109	108	
Anesthesia &	2842	1387	10.37	4139	1704	2216	916	
Operation Room								
Physician Staff	1359	2020	19.44	2077	1589	997	1209	
Intensive Care	160	915	5.25	237	875	32	384	
Physical &	346	339	0.52	534	550	202	215	
Occupational Therapy								
Hotel Service	892	563	0.95	1148	726	623	396	
Undefined	256	163	0.17	299	168	162	88	
Facility Fee								
Facility Fee	1374	792	1.91	2049	987	1064	547	
Total Costs	17319	7155	100	21525	7955	14011	5968	
Observations		2091	.8	6 2	2174	-	1994	

Table 2: Inpatient	Costs (	Íin	CHF)	$\operatorname{per}$	Patient	for	each	Category
--------------------	---------	-----	------	----------------------	---------	-----	------	----------

<sup>a</sup> ANOVA Variance decomposition

— medication, physician salaries, and material — as well as some categories of overhead costs, namely nursing, anesthesia & operation room, and physician staff. The physician staff category alone accounts for almost 20% of the variation in total costs. Figure 3 summarised the relative size of the most relevant categories. The direct material costs, encompassing the implant, represent the largest share of costs, amounting to almost 30% of the total costs.

In this study, revisions serve as the sole quality indicator. Revisions are recorded as a binary variable using the SIRIS registry. The registry defines revision clearly, leaving little room for different reporting between hospitals. Each THA that led to a revision within two years post-surgery is flagged. Aggregated, this represents a sample two-year revision probability or revision rate, as shown in Table 1. In contrast to some other THA studies, hospital readmissions or follow-up complications related to the initial THA are not considered, as they had not yet been registered in SIRIS<sup>6</sup>.

<sup>&</sup>lt;sup>6</sup>It is possible to identify complications and readmissions using a set of different treatment codes in the SFSO data. However, this identification process is complex and susceptible to errors, making them unsuitable as a quality indicator alongside standardised revisions



Figure 3: Average Shares of Relevant Cost Categories per THA

Note: full pie is 100% of inpatient case costs

As seen in Figure 4, there are substantial differences between departments regarding the average revision rates. While some departments reported no revision at all, a substantial number of departments had more than twice as many revisions as the average. Visually, these difference persist in the graph with only standard patients; see APPENDIX А.



Figure 4: Differences in Departmental Revision Rates

Note: perpendicular lines represent the averages in the corresponding categories

### 2.3 Patient Health Risks Factors (level-one variables)

	Mean	Std. dev.
Demographics		
Male $(in \%)$	0.51	0.50
Patient's age	67.80	11.57
General Health		
Patient's BMI	27.83	5.50
ASA classification score	2.47	1.25
Pathology		
Osteoarthritis	0.94	0.25
Inflammatory arthritis	0.01	0.07
Osteonecrosis	0.06	0.24
$Charlson\ comorbidity\ indicators$		
AMI (Acute Myocardial)	0.01	0.12
CHF (Congestive Heart)	0.02	0.13
PVD (Peripheral Vascular)	0.02	0.15
CEVD (Cerebrovascular)	0.01	0.09
Dementia	0.01	0.08
COPD (Chronic Obstructive Pulmonary)	0.05	0.21
Rheumatoid Disease	0.02	0.14
PUD (Peptic Ulcer)	0.00	0.04
Mild LD (Liver Disease)	0.00	0.06
Diabetes	0.09	0.29
Diabetes with Complications	0.01	0.07
HP/PAPL (Hemiplegia or Paraplegia)	0.00	0.06
RD (Renal Disease)	0.05	0.22
Cancer	0.01	0.09
Moderate/Severe LD (Liver Disease)	0.00	0.02
Metastatic Cancer	0.00	0.03
AIDS	0.00	0.01
Observations	20918	

Table 3: Patient-level Explanatory Variables

A set of patient-level explanatory variables, as shown in Table 3, represents the patient's health risk prior to their hospital stay. These variables include standard patient characteristics for THA studies, such as gender, age and body mass index (BMI) (Lenguerrand et al., 2018; Robinson et al., 2012; Prosser et al., 2010; Robinson et al., 2012). Age is included as a categorical variable in groups of 10 years. Based on the BMI, patients are assigned to WHO weight categories: underweight (BMI $\leq$ 18.5), normal range (BMI 18.6-24.9), overweight (BM: 25-29.9), obese (BMI 30-34.9), severely obese (BMI $\geq$ 35) (WHO, 2005). Besides patient characteristics, the patient's health status is represented by dummies for the primary diagnosis, the Charlson comorbidity indicators (CCI), as well as a linear variable for the American society of anesthesiologists (ASA) score for general

health. The CCI is a well-established categorization for comorbidities, and there is strong evidence for its association with THA costs and quality (Rosas et al., 2017; Lenguerrand et al., 2018). CCI is based on diagnoses from the international classification of diseases (ICD) and comprises 17 comorbidities (Lix et al., 2016). ASA is an ordinal scale ranging from 1 (healthy patient) to 5 (moribund patient). ASA and primary diagnosis are common explanatory variables (Robinson et al., 2012; Prosser et al., 2010; Street et al., 2010). Laudicella et al. (2010) also include DRG dummies as an explanatory variable for patient health risk, arguing that they already reflect some risk adjustment. However, DRGs depends also on departmental procedures and assessments, which makes them susceptible to measurement error and manipulation. In contrast, the health risk factors should not be dependent on departmental decisions and are thus exogenous to the departmental system.

#### 2.4 Departmental Factors (level-two variables)

Departmental factors that are sometimes also called environmental or system variables to underline that they are exogenous to the decisions of the hospital departments (Street et al., 2010). They reflect the organisational differences between departments. Descriptive statistics of departmental factors are provided in Table 4.

Controlling for specialized hospitals<sup>7</sup>, university hospitals and THA case volume provide a benchmark for comparison (Street et al., 2010; Castelli et al., 2015; Ramkumar et al., 2018). For example, departments in university hospitals are expected to report higher costs due to their teaching duties, even though direct teaching costs are not part of the case costs. Departments in specialized and university hospitals are identified using the hospital typology of the SFO (BFS, 2022). THA case volume is represented by dummies: departments were categorized into low-, medium-, and high-volume groups based on their total THA case numbers in the sample. In order to distribute patients equally among these three volume groups, departments with fewer than 330 cases are considered low-volume departments, departments with 330 to 750 cases are considered medium-volume departments.

As proposed by (Street et al., 2010), two medical procedures as departmental shares are considered to capture the state of "technological equipment" of a department. Unlike patient-level medical procedures that involve decisions of the treating physicians, department shares of decisive procedures might capture the modernity or technological orientation of a department. Department's share of partly/fully cemented THA, indicates the

<sup>&</sup>lt;sup>7</sup>Specialized hospitals are hospitals that only provide care in one or two disease areas.

	Mean	Std. dev.
Medical Procedures		
Department's share of partly/fully cemented THA	0.16	0.16
Department's share of THA with anterior approach	0.44	0.39
Hospital type		
Share of Departments in University Hospitals	0.05	0.21
Share of Departments in General Hospitals	0.82	0.39
Share of Departments Specialised Hospitals	0.14	0.34
Case Volumen category		
Share of low-volume Departments	0.64	0.48
Share of mid-volume Departments	0.32	0.47
Share of high-volume Departments	0.10	0.30
Observations	110	

Table 4: Department-level Explanatory Variables

department's proportion of patients who receive at least a partly cemented implant. Conversely, all other patients receive a cementless prosthesis, which is generally considered a more modern approach. *Department's share of THA with anterior approach*, indicates the proportion of anterior THA surgeries in a department. Compared to the posterior and lateral approaches, the anterior approach requires longer surgery but is less intrusive for the patient's musculature. Unlike

#### 2.5 Econometric Model

The cost model considers patients who have undergone THA, with each patient (i) nested in a department (d). The model is represented by the equation:

$$ln(c_{di}) = x'_{di}\beta + z'_d\gamma + \alpha_d + \varepsilon_{di} \tag{1}$$

As costs are right-skewed, costs were transformed logarithmically into a log-level model, thus  $ln(c_{di})$  represents the natural logarithm of costs for each patient.  $x_{di}$  is a vector of level-one variables, the patient health factors.  $z_d$  is a vector of level-two variables that are invariant within a department. The corresponding model parameters,  $\beta$  and  $\gamma$ , are estimated. Some multilevel studies include cluster-specific interactions of  $\beta$  to account for the fact that explanatory variables can have different impacts on different departments, so-called random slopes. While random<sup>8</sup> slopes could improve the overall fit of the model, they make little sense in the theoretical structure as patient health risk factors should be viewed as something general and predetermined. In line with Castelli et al. (2015) and Lenguerrand et al. (2018) random slopes are not used. The classical error term  $u_{di}$ 

<sup>&</sup>lt;sup>8</sup>In this context, random means "unique" for each department as opposed to "fixed" across departments. This should not be confused with random and fixed in the estimation context. Without random slopes, the estimated coefficients represent a weighted average of all "hypothetical" departmental slopes

is decomposed into two parts:  $\alpha_d$ , which represents the department-specific, invariant unobservables, i.e., the level-two error term, and  $\varepsilon_{di}$ , which represents the patient-specific unobservables.

To avoid the bias of a pooled or aggregated hospital-level estimation of the nested structure of THA data, two common hierarchical models are used: the hybrid model and the mixed model. If non-random slopes are assumed, the mixed model is equivalent to Random Effect (RE) estimation<sup>9</sup>. RE offers a compromise between pooled and within estimation, and is able to estimate  $\gamma$ . RE models account for intradepartmental correlation by including the covariance matrix without completely ignoring between-departmental variation<sup>10</sup>. Nevertheless,  $\beta_{RE}$  can be biased. Whether the reduction of bias in  $\beta_{FE/WI}$ compared to  $\beta_{RE}$  justifies the efficiency loss of  $\beta_{FE/WI}$  can be tested – most commonly with a Hausman test, but also the Akaike information criterion (AIC) values can be compared. To overcome the shortcoming of RE and FE, the so-called hybrid was invented<sup>11</sup>. The hybrid model estimates  $\beta$  and  $\gamma$  in a single likelihood maximization. The hybrid model applies within-transformation only to  $x_{di}$  and includes  $z_d$  unaltered. To the best of my understanding, this approach has not yet been applied in the context of THA; however, its superiority to standard multilevel RE models has been discussed in econometric theory and applied studies in other disease areas (Park et al., 2011; Dieleman and Templin, 2014; Allison, 2009; Sjölander et al., 2013; Bell et al., 2019). Assuming  $E(\varepsilon_{di} \mid x_{di}, z_d) = 0$ , the hybrid model in the linear case is given by

$$ln(c_{di}) = \beta_{FE/WI} \left( x_{ij} - \bar{x}_d \right) + \beta_{BE} \bar{x}_i + \gamma z_i + a_d \tag{2}$$

 $\beta_{FE/WI}$  is the same as in the standard FE model since they do not depend on betweendepartmental variation (level two).  $\beta_{BE}$  is the between-department effect of the level-one variables. This represents the effect of the aggregated department's risk factors on costs. Isolated  $\beta_{BE}$  values have limited interpretative value on their own, since they might be strongly biased by  $\alpha_d$ , however,  $\beta_{BE}$  captures the contextual relationship to the outcomes. The combined within and between effects capture the complete observed impact of patient health factors, representing the observable patient mix.

<sup>&</sup>lt;sup>9</sup>In this context, random refers to the estimation. The definition of random varies depending on the field and author. RE estimation is also called variance component model, Feasible Generalized Least Square (FGLS), random intercept model. The mixed in multilevel contetext refers to "random" and "fixed" slopes and not to FE and RE estimation. This notation can be misleading because, given non-random slopes, the random effect estimate is equivalent to the mixed model estimates. In an attempt to reduce confusion, Hayashi (2011) advertises the rather technical name 3SLS estimator with common coefficients for  $\beta_{RE}$ .

<sup>&</sup>lt;sup>10</sup>In fact, it can be shown that  $\beta_{RE}$  has to be between  $\beta_{pooled}$  and  $\beta_{FE/WI}$  as  $\beta_{RE}$  estimation is the result of a quasi-demeaning of the regressors. It represents a matrix-weighted average of  $\beta_{FE/WI}$  and  $\beta_{pooled}$ .

<sup>&</sup>lt;sup>11</sup>The idea was originally proposed by Mundlak (1978) and adapted and branded by Allison (2009) as a hybrid model and by Sjölander et al. (2013) as the between-within model.

Clearly,  $E(\alpha_d | \bar{x}d, zd) = 0$  must hold to estimate  $\gamma$  unbiasedly. In contrast to the RE and pooled, however,  $\alpha_d$  can be dependent on  $z_d$  through the departmental means  $\bar{x}_d$ . Only a correlation between  $\alpha_d$  and  $z_d$  that is not accounted for by  $\bar{x}_d$  biases the results. Without quantifying this bias reduction, it is evident that the inclusion of cluster averages,  $\bar{x}_d$ , as controls should pick up some relevant omitted variable bias (OVB).

In the non-linear case, the revision model with a binary response, slight adaptions should be considered. The hybrid model with logit link is is given by:

$$P(R_{di} = 1 \mid x, z) = \frac{1}{1 + e^{-(\beta_{FE/WI}(x_{ij} - \bar{x}_d) + \beta_{BE}\bar{x}_i + \gamma z_i + a_d)}}$$
(3)

 $P(R_{di} = 1 \mid x, z)$  is the probability — given the covariates x, z — that a patient had a revision during two years after discharge from the initial THA hospitalization.  $x_{di}$  and  $z_d$  represent the same explanatory variables as in the non-linear case, except for the IMR, which is not needed since the full sample is used.

With regard to the estimation, there are slight differences compared to the linear model. FE for non-linear models are only possible in special cases. For instance, conditional FE is feasible for logit and poison models. In line with Bilger et al. (2022), logit models are used for the revision models. Lenguerrand et al. (2018), on the other hand, opt for a poison regression because revisions are a rare event, and poison models tend to perform well for rare outputs. However, comparing AICs showed that logit models fit the data better than Poisson (and probit) models, while estimation results were similar. The hybrid approach considers only a linear relationship between  $\bar{x}$  (level-one variables) and  $\alpha_d$ . In the non-linear case, unlike in the linear case, hybrid and conditional FE are not completely equivalent due to possible non-linear dependencies<sup>12</sup>.

For all non-hybrid estimations results are displayed as average partial effects, specifically for any arbitrary variable  $x_j$   $(x_1, ..., x_j, ..., x_j$  or  $z_j)$ , the APE is given by:

$$APE_{j}(x_{di}, z_{d}) = E\left[PE_{j}(x_{di}, z_{d})\right] = E\left[\frac{\partial P\left(y_{di} = 1 \mid x_{di}, z_{d}\right)}{\partial x_{dij}}\right]$$
(4)

In the case of hybrid, estimation results only provide the direction and relative magnitude. Without further assumption it is not possible to estimate useful APE (Wooldridge, 2019). All quality hybrid results are therefore displayed as model coefficients. For patient health risks, additional odds ratios as a robustness check are provided in APPENDIX B.

To break down  $\alpha_d$ , models are estimated three times: null model (no explanatory vari-

<sup>&</sup>lt;sup>12</sup>The differences between hybrid (or Mundlak) and conditional FE logit in the revision model are, nevertheless, negligibly small; see Allison (2014) for an in-depth comment on that issue.

ables), level-one model with patient health factors only, and full model with both patient health and departmental factors. Assessing the reduction of  $\sigma_{\alpha}^2$  compared to the null model is a common approach to getting a measurement comparable to  $R^2$  approaches in the multilevel context. Finally, the extent of the cluster structure in the sample is assessed by the intra-cluster correlation (ICC) given by the formula

$$ICC = \frac{\sigma_{\alpha}^2}{\sigma_{\alpha}^2 + \sigma_{\epsilon}^2} \tag{5}$$

ICC is the proportion of the unexplained between-department variance compared to the total unexplained variance.

In order to discuss the robustness of coefficients, models are estimated with the hybrid, RE, and pooled if appropriate approach. Although the linear models could be fitted using ordinary least squares (OLS), all models are estimated as generalized linear model using the maximum likelihood to ease comparison. Additionally, it is the preferred approach for multilevel models (Luo et al., 2021; StataCorp, 2013).

## **3** Results

#### 3.1 Interdepartmental Variance in Cost and Quality

ICC,  $\sigma_{\alpha}^2$ , and AICs for the null models, the level-one models, and the full hybrid models can be found in Table 5. Comparable results of the RE model are provided in ADDENDIX B.

		Cost Model			Quality Mod	el
	Null <sup>a</sup>	$Level-one^{b}$	Full <sup>c</sup>	$Null^{a}$	$Level-one^{b}$	Full <sup>c</sup>
$\sigma_{\alpha}^2$	$0.0151^{***}$	0.00994***	0.00779***	0.106**	$0.0434^{*}$	0.0210
ICC(rho)	20.6%			3.4%		
Observations	20918	20918	20918	20918	20918	20918
AIC	343.2	-1946.0	-2050.8	5575.1	5545.5	5545.8

Table 5: Null, Level-one and Full Model Comparison

\* p < 0.05, \*\* p < 0.01, \*\*\* p < 0.001

<sup>a</sup> Model with:  $\varepsilon_{di}$ ,  $\alpha_d$  — no within nor between variance is explained

<sup>b</sup> Model with:  $\varepsilon_{di}$ ,  $\alpha_d$ ,  $(x_{ij} - \bar{x}_d)$ ,  $\bar{x}_i$  — patient factors included

<sup>c</sup> Model with:  $\varepsilon_{di}$ ,  $\alpha_d$ ,  $(x_{ij} - \bar{x}_d)$ ,  $\bar{x}_i$ ,  $\gamma z_i$  — patient and departmental factors included

The ICC shows pronounced cluster structure in the cost sample. 21% of the total variation in costs is due to interdepartmental variance. The ICC for the quality model only amount to about 3.4%, revealing rather modest<sup>13</sup> cluster structure. The patient health characteristics together with their departmental means as representation of the department's patient mix explain about 35% and 60% of  $\sigma_{\alpha}^2$  in cost and quality respectively. While these reductions in  $\sigma_{\alpha}^2$  is substantial especially in the quality model, significant  $\sigma_{\alpha}^2$ remains in both quality and cost variation.

<sup>&</sup>lt;sup>13</sup>There is no agreed minimum threshold for the ICC, but even relatively small ICCs of about 5% has been shown to lead to significant heteroscedasticity problems (Hayes, 2006)



Compared to the null model, the full model reduces  $\sigma_{\alpha}^2$  in cost and quality by 50% and 80%, respectively. These reductions in interdepartmental variance are graphically summarized by Figure 5. In the revision case,  $\sigma_{\alpha(full)}^2$  is not significantly different from zero, suggesting that most differences between departments can be explained by observed patient mix, procedures, case volume and hospital type. In the cost model,  $\sigma_{\alpha(full)}^2$  remains highly significant significant, i.e. 99.9 % confidence level. As expected, the hybrid model, compared to the RE model in APPENDIX C, reduces more of the interdepartmental variance because of the inclusion of departmental health risk means,  $\bar{x}_d$ .

The break down of total costs to cost categories, in Table 6, shows that even after accounting for patient mixes, there are significant interdepartmental differences regarding all cost categories. However, the extent of differences varies greatly between different cost categories. While only 20% of the variation in *Nursing Costs* and *Material* is attributable to interdepartmental differences, 64% of the variation in *Other Direct Costs*, which encompass physician salaries, is due to differences between departments.

Additionally, predicted versus observed plots that allow direct comparison to the study of Bilger et al., 2022<sup>14</sup> are provided in APPENDIX D. Many hospitals have more than twice the expected revision rate based on their patient mix than expected. Similarly, regarding costs, there are salient differences between expected and observed total costs: a substantial number of low-volume and high-volume departments have a more than 1.3 times higher average costs than expected

 $<sup>^{14}\</sup>mathrm{The}$  only Swiss THA quality study that analysis THA revisions

Cost Category	ICC	$\sigma_{lpha}^2$
Materials	20%	0.0246***
Other Direct	64%	$1.391^{***}$
Nursing	20%	$0.0436^{***}$
Anesthesia & Operation Room	61%	$0.291^{***}$
Physician Stuff	61%	$0.291^{***}$
Other Overhead	38%	$0.0431^{***}$
Facility fee	29%	0.0823***
* $p < 0.05$ , ** $p < 0.01$ , *** $p < 0.001$		

Table 6: Patienen mix adjusted ICC and  $\sigma_{\alpha}^2$  of Cost Categories

#### 3.2 Association of Patient Health Factors to Cost and Quality

The patient level estimates for the cost model are presented in Table 7. In both models, hybrid and RE, age groups 81-90 years and 91+, severely obese, ASA score, and the comorbidities COPD (Chronic Obstructive Pulmonary Disease) and CHF (Congestive Heart Failure), AMI (Acute Myocardial Infarction), PVD (Peripheral Vascular Disease), CEVD (Cerebrovascular Disease), Dementia, PUD (Peptic Ulcer Disease), HP/PAPL (Hemiplegia/Paraplegia), cancer, and diabetes with complications are significantly associated with higher THA costs. The magnitudes of some comorbidities are substantial: THA patients with CHF, dementia, PUD, hypertension, or pulmonary disease are, ceteris paribus, associated with higher inpatient costs of at least 15%. Additionally, age is a strong predictor of costs. In both the hybrid and RE model, patients in the age groups 81-90 and 91+ have higher costs of  $11\%((=e^{\beta}-1)*100)$ ) and 17%, respectively. Compared to osteoarthritis, THAs due osteonecrosis are significantly linked to higher inpatient costs. The estimates for  $\beta_{FE/within}$  and  $\beta_{RE}$  are very similar, indicating robust associations. It seems that  $\beta_{RE}$ is not overly biased. The results of a Hausman test support this observation; the difference between FE and RE is just short of being significant. It seems that in the adjusted sample with the covariates, consistent estimation is possible even without  $FE^{15}$ . However, according to the AIC  $\beta_{FE}$  fits the data best. While the difference in AIC between RE and FE is small, the difference between hierarchical models and pooled are more pronounced. This is in line with the Hausmann test of RE (or FE) vs. pooled that shows a significant bias problem in the pooled model.

<sup>&</sup>lt;sup>15</sup>When using the original 'not-cleaned' sample (i.e., including very small departments, THAs of fractures, deceased and emigrated patients), the differences between  $\beta_{FE}$ ,  $\beta_{RE}$  and  $\beta_{pooled}$  are much higher and significant according to the Hausman test.

	$\beta_{hybrid/FE}$		$\beta_R$	EE	$\beta_{pooled}$	
Gender (female)	ref.		ref.		ref.	
Male	-0.0322***	(-9.35)	-0.0322***	(-9.34)	$-0.0317^{***}$	(-7.44)
Age (31-40 years)	ref.	. ,	ref.	· · ·	ref.	. ,
0-30 years old	0.0435	(1.34)	0.0438	(1.34)	0.0649	(1.69)
41-50 years old	0.0141	(0.55)	0.0142	(0.55)	0.0179	(0.67)
51-60 years old	0.00822	(0.32)	0.00825	(0.32)	0.00354	(0.13)
61-70 years old	0.0236	(0.89)	0.0237	(0.89)	0.0217	(0.76)
71-80 years old	0.0496	(1.83)	0.0497	(1.84)	0.0496	(1.64)
81-90 years old	$0.108^{***}$	(3.85)	$0.108^{***}$	(3.86)	$0.107^{***}$	(3.44)
91+ years old	$0.156^{***}$	(3.94)	$0.156^{***}$	(3.93)	$0.158^{**}$	(3.38)
BMI (norm. range)	ref.	. ,	ref.	. ,	ref.	. ,
Underweight	0.0177	(0.91)	0.0176	(0.90)	0.0144	(0.61)
Overweight	0.00494	(1.09)	0.00468	(1.03)	-0.00857	(-0.86)
Obese	$0.0362^{***}$	(6.13)	$0.0361^{***}$	(6.07)	0.0232	(1.98)
Obese (severe)	$0.133^{***}$	(8.06)	$0.133^{***}$	(8.04)	$0.121^{***}$	(6.14)
$ASA \ classification$	$0.00997^{*}$	(2.50)	$0.00985^{*}$	(2.44)	0.000705	(0.07)
Pathology (Osteo.)	ref.		ref.		ref.	
Inflammatory arthritis	0.00732	(0.52)	0.00748	(0.53)	0.0255	(0.90)
Osteonecrosis	$0.0687^{***}$	(6.18)	$0.0687^{***}$	(6.19)	$0.0672^{***}$	(4.29)
Charlson indicators						
AMI	$0.0630^{***}$	(4.00)	$0.0627^{***}$	(3.99)	$0.0437^{*}$	(2.24)
CHF	$0.190^{***}$	(6.68)	$0.190^{***}$	(6.69)	$0.192^{***}$	(6.19)
PVD (Peripheral Vascular)	$0.0364^{**}$	(2.99)	$0.0366^{**}$	(3.00)	$0.0450^{***}$	(3.42)
CEVD	$0.0937^{***}$	(3.44)	$0.0941^{***}$	(3.45)	$0.118^{***}$	(3.73)
Dementia	$0.209^{***}$	(6.44)	$0.210^{***}$	(6.45)	$0.217^{***}$	(6.12)
COPD	$0.0420^{***}$	(5.10)	$0.0421^{***}$	(5.12)	$0.0514^{***}$	(5.23)
Rheumatoid Disease	$0.0358^{**}$	(2.95)	$0.0358^{**}$	(2.95)	$0.0368^{*}$	(2.56)
PUD	$0.325^{***}$	(3.43)	$0.324^{***}$	(3.43)	$0.305^{**}$	(3.28)
Mild LD	0.0774	(0.93)	0.0776	(0.94)	0.104	(1.23)
Diabetes	$0.0337^{***}$	(5.61)	$0.0339^{***}$	(5.65)	$0.0427^{***}$	(6.55)
Diabetes + Complications	$0.139^{***}$	(3.83)	$0.139^{***}$	(3.84)	$0.138^{**}$	(3.31)
HP/PAPL	$0.285^{***}$	(5.45)	$0.285^{***}$	(5.46)	$0.286^{***}$	(5.30)
RD	$0.0950^{***}$	(7.32)	$0.0955^{***}$	(7.38)	$0.115^{***}$	(8.00)
Cancer	$0.0816^{***}$	(3.63)	$0.0817^{***}$	(3.65)	$0.0795^{**}$	(3.18)
Moderate/Severe LD	0.107	(0.73)	0.107	(0.73)	0.143	(1.05)
Metastatic Cancer	0.0864	(0.79)	0.0866	(0.80)	0.157	(1.50)
AIDS	0.0545	(0.34)	0.0597	(0.37)	$0.232^{*}$	(2.49)
Constant	$9.635^{***}$	(350.05)	$9.657^{***}$	(317.45)	$9.663^{***}$	(271.97)
AIC	-2585.1		-2061.2		2311.4	
Observations	20918		20918		20918	

Table 7: Link of Patient Health Factors to Total Inpatient Costs

t statistics in parentheses

\* p < 0.05, \*\* p < 0.01, \*\*\* p < 0.001

The results of the hybrid, RE, and pooled estimation for the quality model are displayed in Table 8. Results are less robust across the models than for the cost model; nevertheless, all models identify, at the 95% confidence interval level, almost the same risk factors for higher revision probability: severe obesity, obesity, and the comorbidities CHF, COPD, Mild Liver Disease, HP/PAPL, and Renal Disease. Compared to the baseline of patients aged 31 to 40, younger and older patients are negatively associated, but this association is not significant for any age group. This is one of the few salient differences in the revision model to the cost as most patient risk factors have the same direction. In terms of magnitude, in particular severe obesity, PUD, Mild Liver Disease, and HP/PAPL stand out, all of which are associated with at least 3 percentage points higher revision probability in the RE model. Translated to odds ratios of conditional FE – provided in APPENDIX B – this translates to 3.5, 4.4, and 3.5 times higher revision probabilities for severe obesity, mild liver disease, and HP/PAPL, respectively.

Modifying the revision probability model by including costs alongside patient factors as an additional regressor shows the strong correlation between higher costs and lower quality despite controlling for the observational patient mix, see Table 13 and 14 in APPENDIX E. The strongly significantly positive link persists for all cost categories except for *Direct Physician Costs*.

	$\beta_{hyb}$	orid	$APE_{RE}$		$APE_{pooled}$	
Gender (ref. female)	ref.		ref.	-	ref.	
Male	0.00951	(0.11)	-0.0000165	(-0.01)	-0.0000487	(-0.02)
Age (ref. 31-40 years)	ref.		ref.		ref.	
0-30 years old	-1.386	(-1.32)	-0.0378	(-1.34)	-0.0372	(-1.33)
41-50 years old	-0.323	(-0.96)	-0.00996	(-1.00)	-0.0100	(-1.02)
51-60 years old	-0.405	(-1.28)	-0.0127	(-1.45)	-0.0128	(-1.46)
61-70 years old	-0.553	(-1.77)	-0.0166	(-1.71)	-0.0164	(-1.71)
71-80 years old	-0.550	(-1.75)	-0.0171	(-1.81)	-0.0168	(-1.82)
81-90 years old	-0.595	(-1.80)	-0.0178	(-1.79)	-0.0171	(-1.76)
91+ years old	-0.771	(-1.15)	-0.0235	(-1.43)	-0.0218	(-1.32)
BMI (ref. norm. range)	ref.	. ,	ref.	. ,	ref.	
Underweight	-0.475	(-0.92)	-0.0134	(-0.90)	-0.0130	(-0.89)
Overweight	0.0210	(0.20)	0.0000773	(0.03)	-0.000215	(-0.08)
Obese	$0.407^{***}$	(3.67)	$0.0108^{***}$	(4.03)	$0.0104^{***}$	(3.94)
Obese (severe)	$1.201^{***}$	(5.94)	$0.0338^{***}$	(6.93)	$0.0336^{***}$	(6.96)
ASA classification	0.00415	(0.09)	0.000794	(0.69)	0.00106	(1.02)
Pathology (ref. Osteo.)	ref.	. ,	ref.	. ,	ref.	. ,
Inflammatory arthritis	0.702	(1.72)	0.0196	(1.64)	0.0195	(1.52)
Osteonecrosis	-0.133	(-0.74)	-0.00365	(-0.73)	-0.00333	(-0.67)
Charlson indicators						
AMI	-0.106	(-0.30)	-0.00392	(-0.45)	-0.00526	(-0.60)
CHF	$0.549^{*}$	(2.36)	$0.0160^{*}$	(2.27)	$0.0161^{*}$	(2.28)
PVD	0.267	(1.14)	0.00847	(1.43)	0.00856	(1.45)
CEVD	-0.981	(-1.74)	-0.0274	(-1.93)	-0.0259	(-1.83)
Dementia	-0.108	(-0.21)	-0.00412	(-0.29)	-0.00465	(-0.33)
COPD	$0.490^{**}$	(3.18)	$0.0143^{***}$	(3.94)	$0.0144^{***}$	(4.09)
Rheumatoid Disease	0.383	(1.54)	0.0107	(1.76)	0.00991	(1.60)
PUD	$1.186^{*}$	(2.12)	$0.0356^{**}$	(2.75)	$0.0373^{**}$	(2.89)
Mild LD	$1.017^{*}$	(2.56)	$0.0311^{**}$	(2.78)	$0.0317^{**}$	(2.75)
Diabetes	-0.0359	(-0.26)	-0.00110	(-0.30)	-0.00118	(-0.32)
Diabetes + Complications	-0.114	(-0.24)	-0.00254	(-0.21)	-0.000946	(-0.07)
HP/PAPL	$1.467^{**}$	(3.10)	$0.0413^{**}$	(3.17)	$0.0412^{**}$	(3.23)
RD	$0.669^{***}$	(4.36)	$0.0186^{***}$	(4.08)	$0.0181^{***}$	(4.01)
Cancer	0.0407	(0.09)	0.000874	(0.07)	0.00251	(0.21)
Moderate/Severe LD	0.858	(0.74)	0.0248	(0.80)	0.0267	(0.89)
Metastatic Cancer	0.268	(0.23)	0.0130	(0.43)	0.0132	(0.45)
AIDS	-14.35	(-0.00)	0	(.)	0	(.)
AIC	5545.5		5514.2		5532.9	
Observations	20918		20916		20916	

Table 8: Link of Patient Health Factors to Revision Probability

t statistics in parentheses

\* p < 0.05, \*\* p < 0.01, \*\*\* p < 0.001

#### 3.3 Association of Departmental Factors to Cost and Quality

Results of the cost model in Table 9 suggest that, ceteris paribus, departments in university hospitals are more expensive. The cemented approach is associated with significantly higher costs, while the anterior approach appears to be linked – albeit not significantly – to lower costs. When examining case volume, there appears to be a relatively linear relationship with costs; higher THA case volumes correlate with reduced costs<sup>16</sup>. Across all models, high-volume departments maintains a significant association with lower costs at the 95% confidence level. Moreover, in the RE model, the medium-volume retains significance at the 95% confidence level, but exhibits a weaker effect at a 90% confidence level in the pooled and hybrid models. Based on the hybrid model, THA in a mediumor high-volume department. There are noticeable variations in results across different models, which lead to divergent significance levels. In considering the AIC, a small discrepancy is observed between RE and hybrid models, with a clearly better fit demonstrated by the hierarchical models compared to the pooled estimation.

	$\gamma_{hybrid}$		$\gamma_{RE}$		$\gamma_{po}$	oled
Category(ref. General hospital)	ref.		ref.		ref.	
University hospitals	$0.164^{*}$	(2.37)	$0.118^{*}$	(2.28)	$0.160^{***}$	(3.75)
Specialised hospitals	0.0291	(0.89)	0.0576	(1.84)	$0.108^{**}$	(2.71)
Processes						
% partly cemented	$0.217^{**}$	(3.27)	$0.217^{***}$	(3.31)	$0.168^{*}$	(2.25)
% anterior approach	-0.0440	(-1.58)	-0.0478	(-1.79)	-0.0365	(-1.15)
Case Volume (ref. low volume)	ref.		ref.		ref.	
Medium volume	-0.0294	(-1.30)	$-0.0501^{*}$	(-2.16)	-0.0321	(-1.40)
High volume	$-0.0940^{*}$	(-2.32)	$-0.106^{**}$	(-2.85)	$-0.107^{**}$	(-2.93)
Constant	$10.31^{***}$	(9.15)	$9.659^{***}$	(364.68)	$9.671^{***}$	(255.31)
Observations	20918		20918		20918	
AIC	-2050.8		-2076.0		1126.5	

Table 9: Link of Departmental Factors to Total Inpatient Costs

t statistics in parentheses

Patient health factors are part of the model but their coefficients are omitted for better readability

\* p < 0.05, \*\* p < 0.01, \*\*\* p < 0.001

Results from the quality model, shown in Table 10, offer few clear associations: the sole significant link is between the share of the anterior approach and lower revision rates in the hybrid model. In all models, medium-volume departments are associated with lower revision rates, while high-volume departments are correlated with higher revision rates; however, these links are not significant in either case<sup>17</sup>. While the RE model still provides

<sup>&</sup>lt;sup>16</sup>Replacing the volume dummies with a linear volume variable and its squared and cubed terms confirms this linear relationship, with only the linear term being significantly negative

<sup>&</sup>lt;sup>17</sup>By increasing the share of medium-volume departments, it is possible to get a significant link between medium-volume departments and lower revision rates

the best fit for the data according to the AIC, the differences are less pronounced when compared to the cost models.

	$\gamma_{hy}$	brid	APE	$E_{RE}$	$APE_{p}$	pooled
Category(ref. General Hosp.)	ref.		ref.		ref.	
University hospitals	0.414	(1.40)	0.0115	(1.68)	0.0107	(1.74)
Specialised hospitals	0.0659	(0.38)	-0.00579	(-1.29)	-0.00587	(-1.26)
Processes						
% partly cemented	-0.0124	(-0.03)	-0.00774	(-0.68)	-0.00211	(-0.18)
% anterior approach	$-0.322^{*}$	(-2.03)	-0.00163	(-0.40)	-0.00121	(-0.27)
Case Volume (ref. low vol.)	ref.		ref.		ref.	
Medium volume	-0.152	(-1.27)	-0.00169	(-0.48)	-0.00148	(-0.43)
High volume	0.343	(1.90)	0.00204	(0.44)	0.00166	(0.33)
Observations	20918		20918		20918	
AIC	5541.6		5518.5		5571.6	

Table 10: Link of Departmental Factors to Revision Probability

 $t\ {\rm statistics}$  in parentheses

Patient health factors are part of the model but their coefficients are omitted for better readability

\* p < 0.05, \*\* p < 0.01, \*\*\* p < 0.001

### 4 Discussion

This study contributes to the understanding of interdepartmental heterogeneity of quality and costs in the case of THA, allowing us to identify what part of interdepartmental differences might be problematic for the health care system. Results show notable differences between Swiss departments as regards quality and costs for THA. However, the relative importance of inter-departmental variance, as indicated by the ICC, is moderate when contrasted with international THA studies (Gómez et al., 2021; Robinson et al., 2012; Pablo et al., 2004). Specifically, Robinson et al. (2012) found an ICC of 40% for cost data, nearly double than in this study. Moreover, a relatively high share of interdepartmental variance is observable in patient mix and organisational structure(Stargardt, 2008; Robinson et al., 2012; Zegers et al., 2011; Bilger et al., 2022). For example, Robinson et al. (2012) found that patient health factors only explain about 4.5% of the total cost variation; a methodically similar two-level study from the Netherlands by Zegers et al. (2011) found that the patient mix and department-level characteristics together account for only 23% of the interdepartmental variance. Compared to Switzerland, THA provision seems to be more varied in other countries, leading to greater observable and especially unobservable differences between healthcare departments. Yet, even compared to the Swiss study by Bilger et al. (2022), this study found fewer departmental variations when accounting for patient mix. Bilger et al. (2022) only provides a predicted vs. observed plot with respect to revision but the differences are even graphically apparent. For example, while this study reports only seven departments with a revision rate at least twice the expected one, Bilger et al. (2022) reports 10 hospitals with that ratio. As sample and methodology are comparable, the discrepancies can only be due to additional patient health factors from SIRIS used in this study. In particular, BMI weight groups, as a strong predictor of revisions, might be responsible for the difference. While the interdepartmental variations found in this study are moderately compared to prior research, significant variance beyond patient mix remains in both cost and quality. Relatively high inter-departmental cost variance persist for all cost categories, suggesting that the heterogeneity in THA costs cannot be pinned down to a specific category like implant costs. Consequently, focusing on certain cost categories, as Robinson et al., 2012, does not allow to understand the complexity of THA costs. Finally, significant variations in costs, but not quality, continue after accounting for benchmark factors, pointing out at least some degree of different cost-efficiency among departments with similar organizational structures and patient mixes also in Switzerland.

While patient health does not entirely explain interdepartmental differences in cost and quality, it is an important driver of both. This study emphasizes several health factors, such as renal disease, COPD, CHF, PUD, and HP/PAPL, as significant determinants of

higher costs, consistent with prior research (Rosas et al., 2017; Kiridly et al., 2014; Best et al., 2015). Additionally, it reaffirms the link between a higher ASA score, age, and BMI with rising costs in the Swiss context, echoing earlier findings (Castelli et al., 2015; Rosas et al., 2017). A novel discovery is the clear correlation of osteonecrosis to increased costs, in contrast to Robinson et al. (2012). As for quality, most results mirror prior research (Lenguerrand et al., 2018; Werner et al., 2017; Issa et al., 2016). In particular the association of comorbidities like CHF, COPD, and renal and liver disease with revision risks. However, the magnitude of these associations in this study tends to be more pronounced, as for example in Lenguerrand et al. (2018). The notable correlation between weight and revision risk, thoroughly addressed by Werner et al. (2017) with a US sample, is also confirmed for Switzerland, though their more narrow categorisation of the heaviest category, e.g. super obese with a BMI of  $50 \ge$ , resulted in a slightly higher magnitude, i.e. 4.5 times for that category. Most health factors demonstrate a consistent association with both higher cost and lower quality. However, age is an exception; it is linked to higher inpatient costs but has a weak negative association with revision rates. This is likely because older individuals lead less active lifestyles, placing fewer demands on their hips. This pattern resonates with prior observations (Lenguerrand et al., 2018; Prosser et al., 2010). The reported strong links between patient health factors costs and quality emphasizes the need for adequate patient health risk adjustments in payment models. While the DRG system accounts - albeit as Laudicella et al. (2010) point out possibly incompletely - for patient health, potential P4P for THA might not include such risk-adjustment. Without proper risk-adjustments, P4P could exacerbate small imprecision in the 11 DRG-weights for THA: patients with health risks typically require more expensive care from hospitals and, on average, experience lower quality outcomes.

The results support the economies of scale argument for THA in Switzerland, consistent with what has been suggested by international studies (Stargardt, 2008; Ramkumar et al., 2018). While a greater THA case volume is associated with reduced costs, hospital specializations in surgery do not seem to improve the department's cost-efficiency. In fact, there is even an insignificant positive correlation between higher costs and departments in specialized clinics. University hospitals are generally linked to increased expenses, a finding that is presumably attributed to their additional teaching responsibilities. However, this contradicts the findings of Castelli et al. (2015), who found that a department's teaching status was inversely correlated with costs. It is plausible that the discrepancy in findings resulted from an omitted variable bias (OVB) in their study, stemming from not controlling for the THA case volume. There is a weak, i.e. not significance at 95% levels, inverse-U relationship between THA case volume and revision, which has been suggested by Paxton et al. (2015). Results show clearly that limiting the number of departments

offering THA could potentially reduce the number of overly costly departments, without compromising quality. While no definitive link has been established between cementing techniques and the need for revisions, results suggest that the use of (partly) cemented prostheses is associated with higher costs. This observation aligns with a previous study by Castelli et al. (2015). It is important to note, however, that these increased costs must result from longer hospital stays and/or more post-surgery complications, since surgeries using cemented implants are less costly (Pennington et al., 2013; Maggs and Wilson, 2017). L'Hommedieu et al. (2016) and Joseph et al. (2017) noted that the anterior approach is related to lower costs, which is in line with the study results: increasing the proportion of cementless prostheses and anterior approaches in Switzerland could be a way to reduce costs without compromising THA quality.

The relationship between cost and quality in healthcare is complex. When cost and quality are combined in one model, results indicate a strong significant correlation of higher costs to lower quality. However, this is likely due to confounding effects of the patient health conditions: health risk factors correlate with both high costs and low quality. This notion is supported by the observation that most health risk factors align with cost and quality in the same direction. Even after controlling for observable factors, the strong association between costs and quality only slightly diminishes, suggesting that not all elements of the patient mix are captured. This resonates with the discussions of Schreyögg and Stargardt (2010), who highlighted the limited possibilities to mitigate the confounding role of patient mix in the cost-quality relationship. Consequently, it is difficult to derive definitive conclusions from a direct cost-quality model. The association between most cost categories to quality is negative and highly significant. However, a notable exception is the link between physician salaries and quality, which is insignificant. If you assume that the confounding influence of patient mix also biases this result, it would indicate that there might actually be a link between higher physician salaries and higher quality. Moreover, physician salaries tend to be higher, on average, in the bottom 10%of the least expensive departments compared to the top 10% of the most expensive ones, and compared to the average of all departments. These findings suggest that directing investments towards enabling physicians to spend more time with patients could lead to both lower overall costs and improved quality of care.

The use of hierarchical models<sup>18</sup> and the comprehensive sample in this study allow consistent results regarding quality and costs of THA in Switzerland. Results can therefore provide a base for further studies regarding specific quality and costs relationship in THA

<sup>&</sup>lt;sup>18</sup>Generally, the standard RE multilevel model seems to fit the data better than hybrid model. This is convenient for further research, as standard multilevel models are more accessible and flexible. In Stata for example, many hybrid model applications require manual computation

as well as help in identifying areas of the health care system that could be improved. However, a limitation of this study is the use of revisions as the only quality indicator. Revisions are relatively rare so estimation is not possible in departments with no variation. Moreover, revisions reflect only one aspect of quality. Nevertheless, apart from revisions, there are few suitable indicators in the case of THA. Consequently, it is not surprising that a relatively new quality indicator, the patient-reported outcome measures (PROMs), is increasingly being used in the context of THA<sup>19</sup>. PROMs are patient-filled questionnaires about their satisfaction regarding pain and functionality. PROMs can provide a more comprehensive assessment of the patient experience and help evaluate the impact of different surgical techniques and other factors on THA outcomes. The second limitation is the endogenous problem of studies without treatment control. As with all observational studies, estimations cannot be directly interpreted as causal effects but only as conditional associations. If there are structurally unobserved patient health factors, this would bias the results regarding regression coefficients and variance components. Especially at the departmental level, where results seem less robust, further causal research is needed to confirm proposed relationships.

## 5 Conclusion

In Switzerland, there are notable — albeit modest when compared internationally — variances between hospital departments regarding inpatient costs and quality for THA. The impact of patient health factors on both cost and quality is significant, emphasizing the need for risk adjustment in payment schemes. This is especially relevant as Switzerland considers P4P. Even after adjusting for patient health, considerable disparities persist in both costs and quality across departments. This indicates that the cost and quality of a THA procedure for the same patient might differ based on the department conducting the surgery. Economies of scale and the promotion of the use of uncemented implants might provide cost advantages without sacrificing quality. While this study enhances the current body of knowledge by offering detailed insights into the factors influencing THA costs and quality in Swiss hospitals, it is limited by its sole focus on revisions as the only quality indicator and the inherent endogeneity of observational research.

<sup>&</sup>lt;sup>19</sup>A project in Zurich marks the first comprehensive PROM project for hip arthroplasty in Switzerland. Since July 1, 2019, all Zurich hospitals with a cantonal mandate have been required to survey patients with hip or knee replacements and report results

## References

- Ali, Adam M and Charles ER Gibbons (2017). "Predictors of 30-day hospital readmission after hip fracture: a systematic review". In: *Injury* 48.2, pp. 243–252.
- Allison, Paul D (2009). Fixed effects regression models. SAGE publications.
- (2014). "Problems with the hybrid method". In: Statistical Horizons, September 2.
- Bell, Andrew, Malcolm Fairbrother, and Kelvyn Jones (2019). "Fixed and random effects models: making an informed choice". In: Quality & quantity 53, pp. 1051–1074.
- Best, Matthew J, Leonard T Buller, Alison K Klika, and Wael K Barsoum (2015). "Increase in perioperative complications following primary total hip and knee arthroplasty in patients with hepatitis C without cirrhosis". In: *The Journal of Arthroplasty* 30.4, pp. 663–668.
- BFS, Bundesamt für Statistik (2022). Statistik der stationären Betriebe des Gesundheitswesens - Krankenhaustypologie. Bundesamt für Statistik.
- Bilger, Jana, Mark Pletscher, and Tobias Müller (2022). "Separating the Wheat from the Chaff: How to Measure Hospital Quality in Routine Data?" In: *Research Gate*.
- Carey, Kathleen Jr and James F Burgess (1999). "On measuring the hospital cost/quality trade-off". In: *Health economics* 8.6, pp. 509–520.
- Castelli, Adriana, Silvio Daidone, Rowena Jacobs, Panagiotis Kasteridis, and Andrew David Street (2015). "The determinants of costs and length of stay for hip fracture patients". In: *PloS one* 10.7, e0133545.
- Dieleman, Joseph L and Tara Templin (2014). "Random-effects, fixed-effects and the within-between specification for clustered data in observational health studies: a simulation study". In: *PloS one* 9.10, e110257.
- Fordham, Richard, Jane Skinner, Xia Wang, John Nolan, Exeter Primary Outcome Study Group, et al. (2012). "The economic benefit of hip replacement: a 5-year follow-up of costs and outcomes in the Exeter Primary Outcomes Study". In: *BMJ open* 2.3, e000752.
- Gandhi, Rajiv, Herman Dhotar, J Roderick Davey, and Nizar N Mahomed (2010). "Predicting the longer-term outcomes of total hip replacement". In: *The Journal of rheumatology* 37.12, pp. 2573–2577.
- Gómez, Luis Fernando Useche, Hernando Gaitán-Lee, María Alejandra Duarte, Patrick Dennis Halley, Alejandro Romero Jaramillo, and Efraim Leal García (2021). "Precision and accuracy of pre-surgical planning of non-cemented total hip replacement with calibrated digital images and acetates". In: Journal of Orthopaedic Surgery and Research 16, pp. 1–6.
- Hayashi, Fumio (2011). *Econometrics*. Princeton University Press.

- Hayes, Andrew F (2006). "A primer on multilevel modeling". In: *Human communication* research 32.4, pp. 385–410.
- Hooper, GJ, AG Rothwell, M Stringer, and C Frampton (2009). "Revision following cemented and uncemented primary total hip replacement: a seven-year analysis from the New Zealand Joint Registry". In: *The Journal of Bone and Joint Surgery. British* volume 91.4, pp. 451–458.
- Issa, Kimona, Steven F Harwin, Arthur L Malkani, Peter M Bonutti, Anthony Scillia, and Michael A Mont (2016). "Bariatric orthopaedics: total hip arthroplasty in super-obese patients (those with a BMI of 50 kg/m2)". In: JBJS 98.3, pp. 180–185.
- Joseph, Noah M, Jared Roberts, and Michael T Mulligan (2017). "Financial impact of total hip arthroplasty: a comparison of anterior versus posterior surgical approaches". In: Arthroplasty today 3.1, pp. 39–43.
- Keeler, Emmett B (1990). "What proportion of hospital cost differences is justifiable?" In: Journal of Health Economics 9.3, pp. 359–365.
- Kiridly, Daniel N, Alexa J Karkenny, Lorraine H Hutzler, James D Slover, Richard Iorio, and Joseph A Bosco III (2014). "The effect of severity of disease on cost burden of 30-day readmissions following total joint arthroplasty (TJA)". In: *The Journal of arthroplasty* 29.8, pp. 1545–1547.
- L'Hommedieu, Coles E, James J Gera, Gerald Rupp, Jeffery W Salin, John S Cox, and Paul J Duwelius (2016). "Impact of anterior vs posterior approach for total hip arthroplasty on post-acute care service utilization". In: *The Journal of Arthroplasty* 31.9, pp. 73–77.
- Laudicella, Mauro, Kim Rose Olsen, and Andrew Street (2010). "Examining cost variation across hospital departments-a two-stage multi-level approach using patient-level data". In: *Social science & medicine* 71.10, pp. 1872–1881.
- Lenguerrand, Erik, Michael R Whitehouse, Andrew D Beswick, Setor K Kunutsor, Ben Burston, Martyn Porter, and Ashley W Blom (2018). "Risk factors associated with revision for prosthetic joint infection after hip replacement: a prospective observational cohort study". In: *The Lancet Infectious Diseases* 18.9, pp. 1004–1014.
- Lix, Lisa Marie, Mark Smith, Marshall Pitz, Rashid Ahmed, Harvey Quon, Jane Griffith, Donna Ranneris Turner, Say Hong, Heather Jane Prior, Ankona Banerjee, et al. (2016). Cancer data linkage in Manitoba: expanding the infrastructure for research. Manitoba Centre for Health Policy, College of Medicine, Faculty of Health ...
- Luo, Wen, Haoran Li, Eunkyeng Baek, Siqi Chen, Kwok Hap Lam, and Brandie Semma (2021). "Reporting practice in multilevel modeling: A revisit after 10 years". In: *Review of Educational Research* 91.3, pp. 311–355.

- Maggs, Joanna and Matthew Wilson (2017). "The relative merits of cemented and uncemented prostheses in total hip arthroplasty". In: *Indian journal of orthopaedics* 51.4, pp. 377–385.
- McKay, Niccie L and Mary E Deily (2008). "Cost inefficiency and hospital health outcomes". In: *Health economics* 17.7, pp. 833–848.
- Mundlak, Yair (1978). "On the pooling of time series and cross section data". In: *Econo*metrica: journal of the Econometric Society, pp. 69–85.
- OECD, Organisation for Economic Co-operation & Development (2019). *Hip and Knee Replacement*. OECD Paris, France.
- Pablo, Paola de, Elena Losina, Charlotte B Phillips, Anne H Fossel, Nizar Mahomed, Elizabeth A Lingard, and Jeffrey N. Katz (2004). "Determinants of discharge destination following elective total hip replacement". In: Arthritis Care & Research 51.6, pp. 1009–1017.
- Park, Jung Min, Angela R Fertig, and Paul D Allison (2011). "Physical and mental health, cognitive development, and health care use by housing status of low-income young children in 20 American cities: A prospective cohort study". In: American journal of public health 101.S1, S255–S261.
- Paxton, Elizabeth W, Maria Inacio, Jasvinder A Singh, Rebecca Love, Stefano A Bini, and Robert S Namba (2015). "Are there modifiable risk factors for hospital readmission after total hip arthroplasty in a US healthcare system?" In: *Clinical Orthopaedics and Related Research* (R) 473.11, pp. 3446–3455.
- Pennington, Mark, Richard Grieve, Jasjeet S Sekhon, Paul Gregg, Nick Black, and Jan H van der Meulen (2013). "Cemented, cementless, and hybrid prostheses for total hip replacement: cost effectiveness analysis". In: *Bmj* 346.
- Prosser, Gareth H, Piers J Yates, David J Wood, Stephen E Graves, Richard N de Steiger, and Lisa N Miller (2010). "Outcome of primary resurfacing hip replacement: evaluation of risk factors for early revision: 12,093 replacements from the Australian Joint Registry". In: Acta orthopaedica 81.1, pp. 66–71.
- Ramkumar, Prem N, Sergio M Navarro, William C Frankel, Heather S Haeberle, Ronald E Delanois, and Michael A Mont (2018). "Evidence-based thresholds for the volume and length of stay relationship in total hip arthroplasty: outcomes and economies of scale". In: *The Journal of arthroplasty* 33.7, pp. 2031–2037.
- Ramos, Nicholas L, Emily L Wang, Raj J Karia, Lorraine H Hutzler, Claudette M Lajam, and Joseph A Bosco (2014). "Correlation between physician specific discharge costs, LOS, and 30-day readmission rates: an analysis of 1,831 cases". In: *The Journal of* arthroplasty 29.9, pp. 1717–1722.
- Rice, Nigel and Andrew Jones (1997). "Multilevel models and health economics". In: *Health economics* 6.6, pp. 561–575.

- Robinson, James C, Alexis Pozen, Samuel Tseng, and Kevin J Bozic (2012). "Variability in costs associated with total hip and knee replacement implants". In: *JBJS* 94.18, pp. 1693–1698.
- Rosas, Samuel, Karim G Sabeh, Leonard T Buller, Tsun Y Law, Martin W Roche, and Victor H Hernandez (2017). "Medical comorbidities impact the episode-of-care reimbursements of total hip arthroplasty". In: *The Journal of arthroplasty* 32.7, pp. 2082– 2087.
- Salzmann, Karin A (Apr. 2021). REKOLE Zertifizierungsrichtlinien. 3013 Bern.
- Schreyögg, Jonas and Tom Stargardt (2010). "The Trade-Off between costs and outcomes: The case of acute myocardial infarction". In: *Health Services Research* 45.6p1, pp. 1585–1601.
- SIRIS, schweizer implantat-register registre suisse des implants (2022). Annual Report of the SIRIS Registry Hip Knee, 2012 – 2021 Ten Years of Swiss Hip and Knee Registry.
- Sjölander, Arvid, Paul Lichtenstein, Henrik Larsson, and Yudi Pawitan (2013). "Betweenwithin models for survival analysis". In: *Statistics in medicine* 32.18, pp. 3067–3076.
- Stargardt, Tom (2008). "Health service costs in Europe: cost and reimbursement of primary hip replacement in nine countries". In: *Health economics* 17.S1, S9–S20.
- StataCorp, LP (2013). "Stata multilevel mixed-effects reference manual". In: College Station, TX: StataCorp LP 9.10.
- Street, Andrew, David Scheller-Kreinsen, Alexander Geissler, and Reinhard Busse (2010). "Determinants of hospital costs and performance variation: Methods, models and variables for the EuroDRG project". In: Working Papers in Health Policy and Management 3.
- Vries, Laura M de, Miriam C J M Sturkenboom, Jan A N Verhaar, Jan H Kingma, and Bruno H C Stricker (2011). "Complications after hip arthroplasty and the association with hospital procedure volume: a nationwide retrospective cohort study on 50,080 total hip replacements with a follow-up of 3 months after surgery". In: Acta orthopaedica 82.5, pp. 545–552.
- Werner, Brian C, Matthew D Higgins, Hakan C Pehlivan, Joshua T Carothers, and James A Browne (2017). "Super obesity is an independent risk factor for complications after primary total hip arthroplasty". In: *The Journal of arthroplasty* 32.2, pp. 402–406.
- WHO, World Health Organization (2005). Surveillance of chronic disease risk factors: country level data and comparable estimates. World Health Organization.
- Wooldridge, Jeffrey M (2019). "Correlated random effects models with unbalanced panels". In: Journal of Econometrics 211.1, pp. 137–150.
- Zegers, Marieke, Martine C De Bruijne, Peter Spreeuwenberg, Cordula Wagner, Gerrit Van Der Wal, and Peter P Groenewegen (2011). "Variation in the rates of adverse

events between hospitals and hospital departments". In: International Journal for Quality in Health Care 23.2, pp. 126–133.

# APPENDIX

# A Reduced sample: "standard patients" (n 3,079)

Figure 6: Departmental Average Total Costs per THA



Note: perpendicular lines represent the averages



Figure 7: Departmental Revision Rates

Note: perpendicular lines represent the averages

# **B** Revision Odds: FE Logit Estimation

	Revision Odds:	FE Logit Estimation
Male	1.012	(0.13)
Age (Bl: 31-40 years)		
0-30 years old	0.253	(-1.30)
41-50 years old	0.719	(-0.98)
51-60 years old	0.667	(-1.28)
61-70 years old	0.576	(-1.76)
71-80 years old	0.574	(-1.76)
81-90 years old	0.552	(-1.80)
91+ years old	0.453	(-1.17)
BMI (Bl: norm. range)		
Underweight	0.620	(-0.93)
Overweight	1.024	(0.22)
Obese	$1.498^{***}$	(3.65)
Obese (severe)	$3.317^{***}$	(5.94)
ASA classification	1.006	(0.13)
Pathology (Bl: Osteo.)		
Inflammatory arthritis	2.035	(1.74)
Osteonecrosis	0.873	(-0.76)
Charlson indicators		
AMI (Acute Myocardial)	0.904	(-0.29)
CHF (Congestive Heart)	$1.734^{*}$	(2.37)
PVD (Peripheral Vascular)	1.317	(1.18)
CEVD (Cerebrovascular	0.352	(-1.83)
Dementia	0.918	(-0.16)
COPD (Chronic Obstructive Pulmonary)	$1.643^{**}$	(3.22)
Rheumatoid Disease	1.484	(1.59)
PUD (Peptic Ulcer)	$3.331^{*}$	(2.15)
Mild LD (Liver)	$2.856^{**}$	(2.65)
Diabetes	0.962	(-0.28)
Diabetes + Complications	0.866	(-0.30)
HP/PAPL (Hemiplegia or Paraplegia)	$4.449^{**}$	(3.12)
RD (Renal)	$1.948^{***}$	(4.34)
Cancer	1.000	(-0.00)
Moderate/Severe LD (Liver)	2.338	(0.74)
Metastatic Cancer	1.384	(0.28)
AIDS	0.000129	(-0.02)
Observations	20632	

Table 11: Revision Odds: FE Logit Estimation

Exponentiated coefficients; t statistics in parentheses

Patient level coefficients omitted

\* p < 0.05, \*\* p < 0.01, \*\*\* p < 0.001

## C RE Model: Null, Level-one and Full Model

	Quality Model			Cost Model			
	null	risk-adj.	full	null	risk-adj	full	
Observations	37617	37617	37617	21013	21013	21013	
ICC	3.1%			22.1%			
AIC	10512.5	10396.6	10072.9	1369.8	-1464.7	-1477.0	
$\sigma_{\alpha}^2$	0.108***	0.0958***	0.0811**	0.0183***	0.0118***	0.00831***	

Table 12: Cost and Quality Model Statistics — RE

t statistics in parentheses

\* p < 0.05, \*\* p < 0.01, \*\*\* p < 0.001

## D Observed versus predicted plots

Figure 7 and Figure 8 show observed costs divided by predicted costs for each department and revision rates, respectively. Graphically, there are still large cost differences between departments in Figure 4. The department with the highest ratio has a ratio of 1.8, which indicates that this department faces costs related to THA almost twice as high as one would expect based on its patient mix. The two extreme outliers with ratios of over 1.5 are smaller departments in terms of THA case volume. However, some larger departments also have much higher observed than expected costs,

Figure 8: Department Ratio: Observed Total Costs Divided by Predicted Costs Based on Patient Mix



indicating no obvious link between costs and volume.

Figure 9: Department Ratio: Observed Revisions Divided by Predicted Number of Revisions based on Patient Mix



Figure 5 shows salient differences regarding revision rates across departments after accounting for patient health factors. The department with the highest ratio has a ratio of 3, which indicates that this department has a revision rate three times higher than would be expected based on its patient mix. About one quarter of all departments have observed revision rates at least 1.5 times higher than expected. Four departments have not recorded any revisions, so their ratio is zero. Graphically, there is no obvious relationship between volume and ratio: some low- and high-volume departments have relatively many revisions.

## E Link between Cost and Quality

	Total Cost	Material	Other	Nursing	Anesthesia	Other	Facility
			Direct		& OP room	Overhead	Fee
Link	2.73***	1.86***	0.13***	1.33***	1.71***	2.04***	1.40***
Std. Dev.	(0.19)	(0.31)	(0.04)	(0.11)	(0.20)	(0.16)	(0.21)
n	20916	20827	20916	20816	20802	20863	20916
	with FE	with FE	with FE	with FE	with FE	with FE	with FE
Link	3.06***	2.56***	0.53***	1.61***	2.86***	2.72***	2.35***
Std. Dev.	(0.13)	(0.16)	(0.05)	(0.08)	(0.12)	(0.11)	(0.11)
n	20632	20545	20632	20534	20520	20581	20632

Table 13: Link between Cost Categories and Quality

\* p < 0.05, \*\* p < 0.01, \*\*\* p < 0.001

Table 14: Link between Direct Physician Costs (part of Other Direct Costs in Table 13) and Quality

	Direct			
	Physician Costs			
Link	-0.02			
Std. Dev.	(0.04)			
n	20916			
	with FE			
Link	0.05			
Std. Dev.	(0.04)			
n	20632			
* $p < 0.05$ , ** $p < 0.01$ , *** $p < 0.001$				