








Variation in the utilization of medical devices across Germany, Italy, and the Netherlands: A multilevel approach

Stefan Rabbe¹  | Meilin Möllenkamp¹  | Benedetta Pongiglione²  |
Hedwig Blommestein³  | Pim Wetzelaer³  | Renaud Heine³  | Jonas Schreyögg¹ 

¹Hamburg Center for Health Economic, Universität Hamburg, Hamburg, Germany

²Centre for Research on Health and Social Care Management (CERGAS) Bocconi University, Milano, Italy

³Erasmus School of Health Policy & Management (ESHPM), Erasmus University Rotterdam, Rotterdam, the Netherlands

Correspondence

Rabbe Stefan, Hamburg Center for Health Economic, Universität Hamburg, Esplanade 36, Hamburg, Germany.

Email: stefan.rabbe@uni-hamburg.de

Funding information

European Union's Horizon 2020 research and innovation program, Grant/Award Number: 779306

Abstract

Variation in healthcare utilization has been discussed extensively, with many studies showing that variation exists, but fewer studies investigating the underlying factors. In our study, we used a logistic multilevel-model at the patient, hospital, and regional levels to investigate (i) the levels to which variation could be attributed and (ii) the hospital and regional factors associated with treatment decisions. To do so, we used hospital discharge records for the years 2012–2016 in Germany and Italy and for 2014–2016 in the Netherlands combined with hospital and regional characteristics in nine case studies. We used a theoretical framework to categorize these case studies into effective, preference-sensitive, and supply-sensitive care. Our results suggest that most variation in the treatment decision can be attributed to the hospital level (e.g., case volume), whereas only a minor part is explained by regional characteristics. Italy had the highest share attributable to the regional level, whereas the Netherlands had the lowest. We observed less variation for procedures in the effective-care category compared to the preference- and supply-sensitive categories. Although our results were heterogeneous, we identified patterns in line with the theoretical framework for treatment categories, underlining the need to address variation differently depending on the category in question.

KEYWORDS

Germany, Italy, medical access, medical devices, Netherlands, regional variation, treatment decisions

1 | INTRODUCTION

Variation in the utilization, costs, and outcomes of health care services across regions or providers is a well-known phenomenon in health economics and health policy. The existence of such variation raises questions about the over- or underutilization of services, as well as the efficiency and equity of a health system (Corallo et al., 2014; OECD, 2014). Variation is of special concern when it cannot be attributed to patient needs or preferences and may therefore be unwarranted. Unwarranted variation may result from demand-side factors not only influenced by the patient, such as patient access to health care or differing levels of health literacy or knowledge of treatment options, or supply-side factors, such as practice styles of the physician or hospital, supplier-induced demand, or resource availability and capacity constraints in a given region (Appleby et al., 2011; Skinner, 2012).

A large body of literature has explored and attempted to explain the effects of geographic variation in health services and medical practice (Corallo et al., 2014). The first studies on the topic were undertaken by Glover (1938), who examined the case

of tonsillectomies in England, and Lembcke (1952), who conducted a comparative study of appendectomy rates in the United States. Later, in their seminal work on the subject, Wennberg and Gittelsohn (1973) defined the term “small area variation” in their study of variation of health care delivery focusing on a broad set of variables compared to other authors, including procedures and hospital utilization. Since then, the number of studies published on this topic have been numerous. In a systematic review of the literature on variation in medical practice across member countries of the Organization for Economic Co-operation and Development (OECD), Corallo et al. (2014) identified 836 studies that showed large variation across regions and countries but rarely addressed the causes of this. Overall, however, because the body of literature on the topic is large and part of several disciplines, it is relatively unstructured and diverse, and it often lacks a theoretical foundation (Sutherland & Levesque, 2020).

Two recent streams of the literature have attempted to address the research gap on the causes of regional variation. The first investigates whether variation is driven mainly by demand- or supply-side factors. Cutler et al. (2019), who used vignettes from patients and physician surveys, found that the patient/demand side explains less variation than physician belief about treatments. This is also in line with results of Song et al. (2010) and Finkelstein et al. (2016), both of whom were able to control for the demand side by analyzing the Medicare claims of patients who moved or did not move to different regions during the study period (“movers” and “non-movers”, respectively). Indeed, most available literature from the US concludes that the supply side is the main driver. Other studies in this stream of the literature, however, especially in the European context, have argued that demand-side factors also have an important effect on variation. Moura et al. (2019), for example, also followed movers in an event study that used decomposition analysis. They found that demand-side factors accounted mainly for differences in expenditure across regions in the Netherlands. In their analysis using Swedish administrative data, Johansson et al. (2018) also concluded that the demand side is the most important factor explaining regional variation in outpatient services. Similarly, Göppfarth et al. (2016) found that only a minor part of variation in German health expenditure could be explained by supply-side factors, and most was driven by patient and demand-side factors (Göppfarth et al., 2016). While this stream of the literature provides detailed evidence on demand- and supply-side factors that help explain variation in medical practice, the findings are limited to individual countries.

The second stream of literature compares variation across European countries but does not focus primarily on identifying demand- and supply-side determinants of variation in medical practice. The EuroDRG project focused on the question of whether diagnosis-related groups (DRG) can explain variation in costs and length of stay (Street et al., 2012). The Health-BASKET project (Busse et al., 2008) used case vignettes to analyze variation in costs for different treatments, but relied only on selected hospitals for their analysis. The EUROhope project (Heijink et al., 2015) followed a comparable approach to that in the present study but focused only on diagnoses for four indications. It showed that differences in outcomes could not be explained by the demand- and supply-side factors considered in the analysis. Using an approach comparable to that in our study, Comendeiro-Maaløe et al. (2020) and the European collaboration for healthcare optimization project focused on inpatient mortality due to acute myocardial infarction (AMI) in several European countries but did not provide evidence on demand- and supply-side factors. Researchers in this stream of the literature have repeatedly called for a three-level approach to disentangle the contribution of different levels, such as that of the patient, hospital, region, or country (Schreyögg et al., 2011).

The aim of our study is to identify potential sources of variation in the use of medical treatments across three European countries by means of nine case studies grouped into different treatment categories following a theoretical framework of supply and demand for health services (Chandra & Skinner, 2012). Our study builds on both streams of the literature and enhances these by (i) focusing on several diagnosis-treatment combinations across different treatment categories, (ii) analyzing patient-level data on the treatment decision by using a three-level random intercept model over several years (2012–2016) and grouping variables into demand- and supply-side dimensions, and (iii) doing so for three different European countries: Germany, Italy, and the Netherlands.

First, we focus on several diagnoses in nine case studies, each of which we define as a diagnosis-treatment combination that includes the use of medical devices. In doing so, we are able to differentiate between the following three different treatment categories: effective care, preference-sensitive care, and supply-sensitive care (Wennberg, 2002; Wennberg et al., 2002). These treatment categories should, according to the theoretical framework developed by Chandra & Skinner (2012), be affected differently by supply- and demand-side factors. To our knowledge this study is one of the first to incorporate this structured framework.

Second, we use detailed patient, hospital, and regional data in a three-level random intercept model. With this approach, we are able to address several dimensions of the supply- and demand-side model described by Skinner (2012) and account for important patient differences, such as age, gender, and Elixhauser comorbidity groups. The approach also allows us to account for the clustering of data across regions from patients receiving inpatient treatment in a hospital – in contrast to previous cross-country studies, which have mostly used more highly aggregated data. Third, by running a comparable model across Germany, Italy, and the Netherlands, we can investigate variation not only within each country but also across all three.

The paper is structured as follows: The next section explains the underlying conceptual model, including the different treatment categories we investigate. The third section describes the data and our empirical model. The fourth section presents the descriptive and estimated results. The final section discusses the implications of the results and makes suggestions for future research.

2 | METHODS

2.1 | Conceptual framework

The conceptual framework for our regression analysis follows an economic model of regional variation in health care developed by Skinner (2012) and Chandra and Skinner (2012). Their model consists (a) on the demand side, of a two-period model of consumption and leisure in which the individual's perceived quality of life is influenced by his or her medical spending and (b) on the supply side, of a model of the physician who seeks to maximize the health of his or her patients, subject to financial considerations, resource capacity, and ethical constraints, as well as patient demand for certain procedures (Skinner, 2012).

We used this model to derive the dimensions for our regression variables. On the demand side, the model posits that differences in patient health status, patient access to care, patient optimism, and physician optimism are factors that influence the demand for health services. On the supply side, the model posits that the marginal utility of income, capacity constraints, malpractice risks, and the efficiency in health care delivery are factors that influence the supply of health services.

According to this model, treatments can be grouped into the following three categories, each of which is affected differently by supply- and demand-side factors. These categories follow those first developed by Wennberg et al. (2002).

Category I – Effective care: These treatments are effective in almost everyone or in individuals with the relevant diagnosis, and they are cost-effective from a provider perspective. Examples are evidence-based preventive measures such as vaccinations, antibiotics for bacterial infections (e.g., pneumonia or tuberculosis), or the combinations of antiretroviral drugs that delay or halt the progression of acquired immune deficiency syndrome in patients living with human immunodeficiency virus. In most individuals who need them, these interventions have large benefits compared to their costs and risks.

Category II – Preference-sensitive care: These treatments are potentially cost-effective in some patients but have declining marginal effects in others. Frequent examples are Caesarean sections, percutaneous coronary interventions (which is cost effective for AMI but not for stable coronary disease), or imaging technologies such as computed tomography. This category therefore also depends on physicians' beliefs about a treatment (e.g., about its effectiveness) and patients' beliefs about benefits and risks, but also factors such as a health system's financial constraints. Chandra and Skinner (2012) suggest in this context that physicians' beliefs about a treatment's effectiveness, in particular, can have a high impact on the total number of procedures performed and the associated costs.

Category III – Supply-sensitive care: Treatments in this category have unfavorable cost-effectiveness, are not cost-effective, or there is uncertainty about their benefits. The challenge here is to determine whether the effect of the technology is not yet fully understood (meaning that the technology could potentially be included in category II or even I in the future) or if it has almost no effect. Examples of category III treatments are arthroscopic surgery for knee osteoarthritis, which was performed frequently until a randomized controlled trial showed that it had no beneficial effect (Moseley et al., 2002), and stenting in patients with myocardial infarction, which was initially not approved for that indication and was used on an off-label basis but now is generally recommended for these patients after publication of high-quality evidence showing a favorable risk-benefit ratio (Neumann et al., 2019).

2.2 | Data

Based on this conceptual framework, we selected variables to reflect demand- and supply-side factors that might explain the treatment decision in the three treatment categories (i.e., effective, preference-sensitive and supply-sensitive care). Table 1 gives an overview of the variables based on the dimensions defined by Skinner (2012).

On the demand side, the first and probably most important dimension was “differences in health status” that were likely to affect the treatment decision. These comprise patient age, gender, and comorbidities categorized into 31 Elixhauser groups to adjust for different patient needs, as well as life expectancy, the unemployment rate, and the share of population older than 65 years at the NUTS3 (Nomenclature des unités territoriales statistiques) level. The second dimension was “patient access”, which comprises the number of hospital beds per 100,000 inhabitants, the population density, the share of non-citizen residents,

TABLE 1 Variables used in the analysis clustered into a demand and supply framework

Demand factors	Variable	Supply factor	Variable
Differences in health status	<ul style="list-style-type: none"> • Age¹ • Gender¹ • Main diagnosis¹ • 31 elixhauser groups¹ • Life expectancy³ • Unemployment rate³ • Population aged 65+³ 	Efficiency in delivery of healthcare	<ul style="list-style-type: none"> • Inpatient cases per bed² • Number of relevant cases² • Avg. length of stay for relevant cases² • Ownership^{2,b} • University hospital²
Patient access	<ul style="list-style-type: none"> • Hospital beds in region³ • General practitioner³ • Population density³ • Share of non-citizen residents³ 	Capacity constraints	<ul style="list-style-type: none"> • Less than 50 relevant cases² • Beds²
Patient “optimism”	<ul style="list-style-type: none"> • Students with tertiary education³ • School-leavers without lower secondary education³ • Household income^{3,a} • Sharevoters³ 		

Note: 1 = Patient level; 2 = Hospital level; 3 = NUTS3-level. Relevant cases are the cases of the diagnosis-procedure combination of each case study in the hospital and year.

^aHousehold income is the median household income of socially insured employees in Germany, mean household income in Italy and the Netherlands.

^bHospital ownership can vary according to the definition in each country.

and the number of general practitioners per 100,000 inhabitants at the NUTS3 level. The third and final dimension on the demand side was “patient optimism”, which comprises educational attainment (i.e., the number of students in tertiary education and the number of school-leavers without lower secondary education as a proportion of all students) and the median or mean household income at the NUTS3 level (median for Germany; mean for Italy and the Netherlands).

On the supply side, we considered several variables at the hospital level. The first dimension was “efficiency in the delivery of healthcare”, which comprised the number of inpatient cases per bed, the number of relevant cases, defined as the number of cases with the diagnosis-procedure combination for each case study in the hospital in the corresponding year, the average length of stay of relevant cases, hospital ownership, and whether the hospital was a university hospital. We could not operationalize the dimension “marginal utility of income” and “malpractice risks”. The second dimension was “capacity constraints”, which comprised a dummy variable for hospitals with fewer than 50 relevant cases, and the number of beds per hospital. There were no data to operationalize the dimension “physician optimism”.

For each of the three countries included in our analysis, we had a separate data set. For Germany and Italy, the data sets covered the years 2012 through 2016, and for the Netherlands, the data sets covered the years 2014 through 2016.¹ The patient-level data consisted in the three countries of all hospital discharge records from all hospitals in each country. These records included information on the diagnoses and procedures undertaken during the hospital visit, as well as information on patient age, gender, residence, and a hospital identifier. For each case study, we selected all patients whose main or secondary diagnosis corresponded to one of the diagnoses in our case studies.² Using the hospital identifier and information on the NUTS3 region in which each hospital was located, we merged these data with information on hospital characteristics and regional characteristics. As part of this process, we deleted observations that could not be matched to a hospital or had missing values. To achieve a uniform analysis across countries, we first defined each variable and then identified the corresponding data in each of the three countries. Some variables differed slightly in their definition (e.g., median household income of socially insured employees in Germany vs. mean household income in Italy and the Netherlands) or were not relevant for some countries (e.g., the ownership status “not-for-profit” is relevant for Germany but not Italy, and ownership status as a variable is not relevant for the Netherlands, as all hospitals in our dataset for that country were private not-for-profit organizations). A more detailed description of the data sources can be found in Appendix A and Table A1.

We selected nine case studies for our analysis, which involved treatments that are provided mostly in inpatient care. These were selected from different medical specialities and based on whether: (i) there was evidence in the literature suggesting high

regional variation in their use, (ii) they were provided as inpatient care in all three countries, (iii) they represented common treatments, and (iv) we had potential access to data on them in each country. Furthermore, we focused on conditions that were medically important, relevant to policy and resource intensive (Corallo et al., 2014). Table 2 gives an overview of the case studies.

We selected case studies from orthopedics, cardiology, oncology, and otolaryngology and assigned each of them to the treatment categories described above. We provide a more detailed description with a short rationale for each case study in Appendix B. Because the distinction between categories II and III, in particular, is debatable and it was unclear, a priori, to which of the two categories some of the case studies belonged, we collapsed categories II and III into a joint category II/III for some cases. Therefore we differentiate in this study between three different treatment categories: effective care (category I), preference sensitive care (category II), and preference or supply sensitive care (category II/III). For each case study defined as effective care (category I), we distinguished between the relevant diagnosis and effective treatment, for example, between AMI and Percutaneous Translaminar Coronary Angioplasty (PTCA), for case study 2. For the case studies categorized as involving preference-sensitive or supply-sensitive care (category II/III), there was usually a conservative treatment option available alongside other equally valid treatment options. We therefore selected patients in two steps, first based on the diagnosis, such as malign neoplasm of the prostate for case study 7 (prostatectomy), and then by distinguishing between different treatment options (i.e., either open or laparoscopic prostatectomy for case study 7). If the alternative was a conservative or ambulatory treatment, we distinguished between the diagnosis (e.g., hearing loss for case study 9) and the procedure (e.g., cochlear implant or revision for case study 9). In all case studies, the treatment coded as one is either more resource intensive or newer compared to the alternative coded as zero. To increase the homogeneity of our sample, we restricted it to patients older than 20 years (except for case study 9) because most of the diagnoses are uncommon in young patients and can involve different treatments in children.

Case studies 1 (Femur effective (FE)) and 3 (Femur Treatment (FT)) included the same diagnosis (femur fracture). Case study 1 compared effective treatment (i.e., two options; reposition or hip implantation) with no treatment and was therefore assigned to treatment category I (effective care). In case study 3 (FT), we focus on treated patients and distinguished between

TABLE 2 Case studies for the analysis

Case	Diagnosis (ICD-9 ICD-10)	Treatment as binary variable	Treatment area	Category
1	Femoral neck fracture (820; 821 S72.0)	0 = No relevant procedure 1 = Reposition/Hip implantation	Orthopedics	I
2	Acute myocardial Infarction (410.0–2 I21.0–3; I22.0–8)	0 = No relevant procedure 1 = Percutaneous translaminar coronary angioplasty (PTCA)	Cardiology	I
3	Femoral neck fracture (820; 821 S72.0)	0 = Reposition 1 = Hip implantation	Orthopedics	II
4	Acute Myocardial Infarction (410.0–2 I21.0–3; I22.0–8)	0 = Bare metal stent 1 = Drug eluting stent	Cardiology	II
5	Knee arthrosis (715.16, 26, 36, 96 M17)	0 = No relevant procedure 1 = Knee implantation	Orthopedics	II
6	Spinal stenosis (724.00, 01, 02, 09 M48.06)	0 = Decompression (without fusion) 1 = Decompression (with fusion)	Orthopedics	II
7	Malign neoplasm prostate (185; 198.82; 233.4; 189.3; 198.1; 233.9 C61)	0 = Open prostatectomy 1 = Laparoscopic prostatectomy	Oncology	II/III
8	Benign neoplasms uterus (218; 219 D25; D26)	0 = Open hysterectomy 1 = Laparoscopic hysterectomy	Women's health/Oncology	II/III
9	Hearing impairment (389.0 H90; H91)	0 = No relevant procedure 1 = Cochlear implant or revision	Otolaryngology	II/III

Note: Table 2 gives an overview of the nine case studies, each of which represents a diagnosis-treatment combination. If the binary treatment is 1, this reflects the more resource-intensive medical device compared to a coding of 0. The ICD-9 codes and ICD-10 codes for each variable are given in parentheses.

Abbreviations: ICD-9/10 = International classification of diseases (9th/10th revision); Category I = Effective care; Category II = Preference-sensitive care; Category III = Supply-sensitive care.

the two types of effective treatment approaches, each with its own advantages or disadvantages, and therefore assigned them to category II (preference-sensitive care).

After identifying the case studies, we selected, with the help of experts in medical and procedure coding, the relevant diagnosis codes (based on International Classification of Diseases (ICD)-10 in Germany and the Netherlands and ICD-9 in Italy) and procedure codes based on the procedure coding system in each country. A more detailed description of the different case studies (Appendix B) and a complete list of procedure codes per case and country can be found in Appendix Tables B1 and B2.

2.3 | Empirical model

For our analysis, we used the following three-level logistic random intercept model for all patient cases:

$$y_{i,j,k,t} = \beta_0 + \beta_1 Pat_{i,j,k,t} + \beta_2 Hosp_{j,k,t} + \beta_3 Cnty_{k,t} + \beta_4 year_{i,j,k,t} + u_{j,k} + u_k + \varepsilon_{i,j,k}, \quad (1)$$

where $y_{i,j,k,t}$ is the binary utilization variable for an individual patient case i in hospital j and NUTS3 region k during year t . We defined the binary utilization variable to differentiate either between two different treatments (e.g., for case study 4, between bare metal stents (BMS) and drug eluting stents (DES)) or between the diagnosis and the treatment (e.g., for case study 3, between an AMI diagnosis with no PTCA treatment and PTCA treatment) (see Table 2).

The binary outcome is explained by a constant (β_0) and a vector of patient-level characteristics $\beta_1 Pat_{ijk}$ with estimators for age, gender, a binary indicator of whether the treatment under investigation was the main diagnosis upon admission, and the 32 binary Elixhauser comorbidity groups that have been coded during the hospital stay (Elixhauser et al., 1998; Quan et al., 2005). These patient-level variables, however, are not shown in the regression results but used for the risk adjustment of patients. $\beta_2 Hosp_{jk}$ is a vector of explanatory variables at the hospital level that are clustered, following Skinner (2012), into the dimensions “efficiency in health care delivery” and “capacity constraints”. At the NUTS3 level ($\beta_3 Cnty_{k,t}$), we included a vector with variables that reflected the demand side, including variables from the dimensions “differences in health status”, “patient access”, and “patient optimism”. We also introduced a year fixed effect for the years 2012 through 2016 ($\beta_4 year_{i,j,k,t}$). All continuous variables were grand-mean centered. Furthermore, we inspected the variance inflation factor (VIF) to determine whether the variables in our model were collinear, and we excluded variables that had a VIF larger or equal to 10 in any of the case studies and countries. We expected each level and variable to differ in impact depending on the case study. For example, whereas we expected hospital-level/supply-side variables to have less impact on decisions about treatments in category I (effective care) compared to other treatment categories, we expected them to have a higher impact on decisions about treatments that are supply sensitive, represented by our category II/III (preference-sensitive care/supply-sensitive care). Moreover, we expected the NUTS3-level variables to have a significant impact on the treatment decision in all three categories, but particularly in category II (preference-sensitive care).

The error terms consisted of random intercept variance $u_{j,k}$ at the hospital level (level 2), a random intercept variance u_k at the NUTS3 level (level 3), and an assumed level 1 variance component ε_{ijk} , which was constant across models.

We used the Intraclass correlation coefficient (ICC) for the hospital level ($ICC_{hospital}$) and for the NUTS3 level (ICC_{NUTS3}) to show the between-cluster variation and defined these as follows:

$$ICC_{hospital} = \frac{u_{j,k}}{u_{j,k} + u_k + \varepsilon_{i,j,k}} \quad (2)$$

$$ICC_{NUTS3} = \frac{u_k}{u_{j,k} + u_k + \varepsilon_{i,j,k}} \quad (3)$$

Our data are nested in such a way that each patient case is assigned to one hospital, which in turn is located in one region. An ICC of zero indicates perfect independence between clusters, whereas an ICC of one indicates perfect interdependence between clusters (Sommet & Morselli, 2017). We interpret these values as the proportion of variation that can be attributed to the hospital ($ICC_{hospital}$) and NUTS3 level (ICC_{NUTS3}). Comparable to the expected influence of different variables at the hospital and regional levels, described above, we also expected comparatively low values for treatment category I (effective care) and higher values for the two other treatment categories for the ICC. For case studies we group into category II (preference-sensitive care), we expected the ICC at the NUTS 3 level to be higher in magnitude compared to the other treatment categories, and for case studies grouped into category II/III (preference-sensitive/supply-sensitive care), we expected the ICC at the hospital and NUTS3 levels to be higher in magnitude compared to the other treatment categories. Furthermore, we tested the model against

a null model without covariates. In the Supporting Information S1, we additionally report the marginal R^2 , which is a measure of how much variance is explained by the fixed effects in the model, and the conditional R^2 , which is a measure of how much variance is explained by the fixed and random effects in the model, and we show the coefficients for the interested reader. We used *R* (R Core Team, 2020), especially the package lme4 (Bates et al., 2015) for our analyses and, depending on the country, *R* or Microsoft SQL Server for data preparation.

2.4 | Sensitivity analysis

Because the hospital data are self-reported, they are subject to weaknesses such as upcoding (e.g., in the number of beds or cases) or the subsuming of several hospitals under one record. We therefore performed a sensitivity analysis by running comparable analyses to those described above but excluding hospitals that were potential outliers. We defined the following exclusion criteria: (i) hospitals with more than 2,000 beds (because this might be indicative of data from a group of hospitals or a different production function); (ii) hospitals with fewer than 50 beds or fewer than 50 inpatient cases per year (because this might be indicative of a high degree of specialization and therefore a different production function); and (iii) data whose values were below or above the mean of all hospitals by more than four times the standard error (i.e., to account for mistakes in the self-reported data). Additionally, we excluded most day-patient facilities and rehabilitation facilities. Excluding these facilities was especially important for the effective care cases in which we compared the diagnosis with treatment, because the facilities in question may not have been able to perform the procedure under investigation, potentially causing bias in our results.

3 | RESULTS

3.1 | Descriptive statistics

Figure 1 provides a descriptive overview of the nine case studies in the three countries, as well as the number of observations for all available years. For Germany, we were able to merge data from up to 1,477 hospitals and 395 regions for a total of 1,454,198 observations for case study 5 (knee prosthesis) for the years 2012 through 2016. All other case studies had fewer observations over this time period, with case study 7 (prostatectomy) having the fewest observations (110,582 in 434 hospitals in 283 regions). For Italy, we were able to merge data from up to 998 hospitals and 110 regions for a total of 584,548 observations for

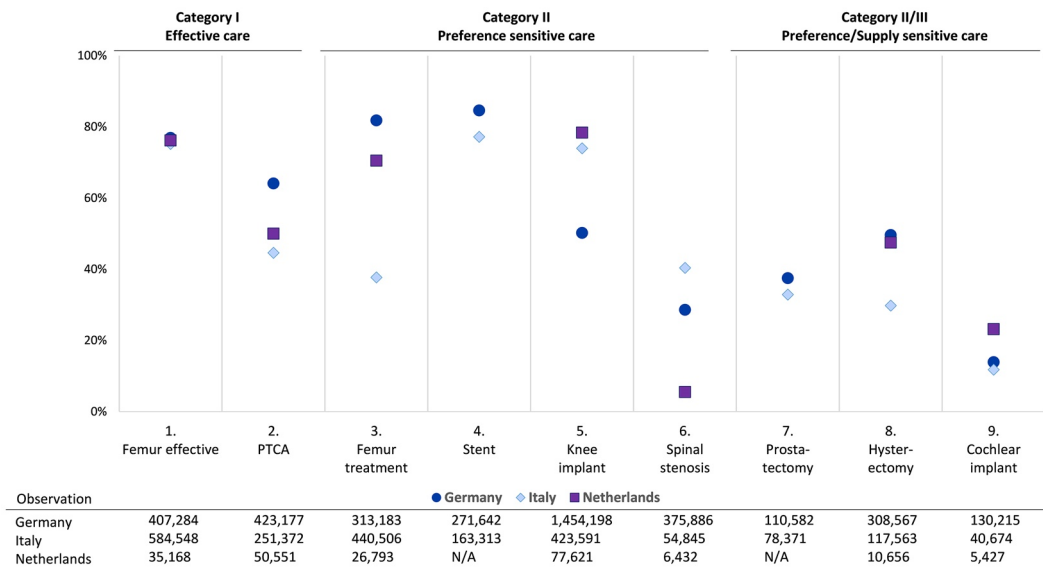


FIGURE 1 Distribution of the dependent variables in the case-studies. Distribution of the dependent 0/1 coded treatment alternatives for all years under investigation (Germany; Italy = 2012–2016; Netherlands = 2014–2016). Case studies 4 & 7 are unavailable in the Netherlands due to unavailability of procedure codes; Case studies 6 & 9 for the Netherlands will not be shown in the regression outputs due to low case-numbers or procedures. PTCA, Percutaneous transluminal coronary angioplasty.

case study 1 (FE) for the years 2012 through 2016. Other case studies had fewer observations, with case study 6 (spinal stenosis) having the fewest observations (54,845 in 392 hospitals). For the Netherlands, we were able to merge data from up to 90 hospitals and 39 NUTS3 regions for a total of 77,621 observations for case study 5 (knee prosthesis) for the years 2014 through 2016. All other case studies had fewer cases, especially case study 9 (cochlear implant or revision) with 5,433 observations in 89 hospitals in 39 regions and case study 8 (hysterectomy) with 10,656 observations in 90 hospitals in 39 regions.

The distribution of the dependent variable across the three countries was comparable for most of the case studies. Important differences were seen in case study 2 (PTCA for patients with myocardial infarction), where in Germany PTCA treatment was performed more frequently (64.1%) than in Italy (44.6%) or the Netherlands (50.0%). Differences could also be seen in case study 3 (FT), where the majority of patients in Germany (81.8%) and the Netherlands (70.5%) received a hip implantation, in contrast to Italy, where only 37.7% received a hip implantation. Furthermore, for case study 8 (hysterectomy), we saw that laparoscopic surgery was used more often in Germany (49.6%) and the Netherlands (47.5%) compared to Italy (11.8%). In the Netherlands, for case study 6 (spinal stenosis), we saw that decompression with fusion was rarely used (5.5%), whereas in Germany (28.6%) and Italy (40.4%) it was much more common. A more detailed table on the descriptive statistics of the case studies and the corresponding hospitals and NUTS3 regions can be found in the Appendix Table B2.

3.2 | Intraclass correlation coefficient

Figure 2 shows the ICC in a scatterplot with the x axis representing the ICC at the hospital level and the y axis representing the ICC at the NUTS3 level. The x axis is scaled from 0% to 90% and the y axis from 0% to 15%, which indicates that most of the clustering can be attributed to the hospital level, whereas a smaller amount can be attributed to the NUTS3 level. Because of the case study-specific distribution of the dependent variable (i.e., few hospitals that provide both options) and the comparably low case numbers for case studies 3 (spinal stenosis) and 9 (cochlear implant or revision), the regressions did not perform well and therefore we do not report results for the Netherlands for these two case studies. We also do not report results for the Netherlands for case studies 4 (stent) and case study 7 (prostatectomy) because the procedure codes in the Netherlands do not differentiate properly between the treatment options.

At the hospital level, in most case studies Germany had a higher ICC compared to Italy, whereas the Netherlands usually had the lowest ICC. The differences were especially prominent for case study 5 (knee prosthesis; Germany: 67.4%, Italy: 54.1%, Netherlands: 25.1%) and for case study 1 (FE; Germany: 62.6%, Italy: 51.8%, Netherlands: 11.2%). For case study 6 (spinal stenosis; Germany: 43.3%, Italy: 57.7%) and case study 8 (hysterectomy; Germany: 36.9%, Italy: 43.3%) Italy had a higher ICC compared to Germany. The two case studies could not be analyzed for the Netherlands. As seen in the regression results, case study 3 (FT) had a lower ICC in all countries at the hospital level (Germany: 9.9%, Italy: 4.2%, Netherlands: 10.8%), followed by case study 4 (stent; Germany: 20.6%, Italy: 20.9%). Furthermore, for case study 2 (PTCA) the Netherlands had the highest ICC (78.6%) compared to Germany (66.4%) and Italy (62.4%).

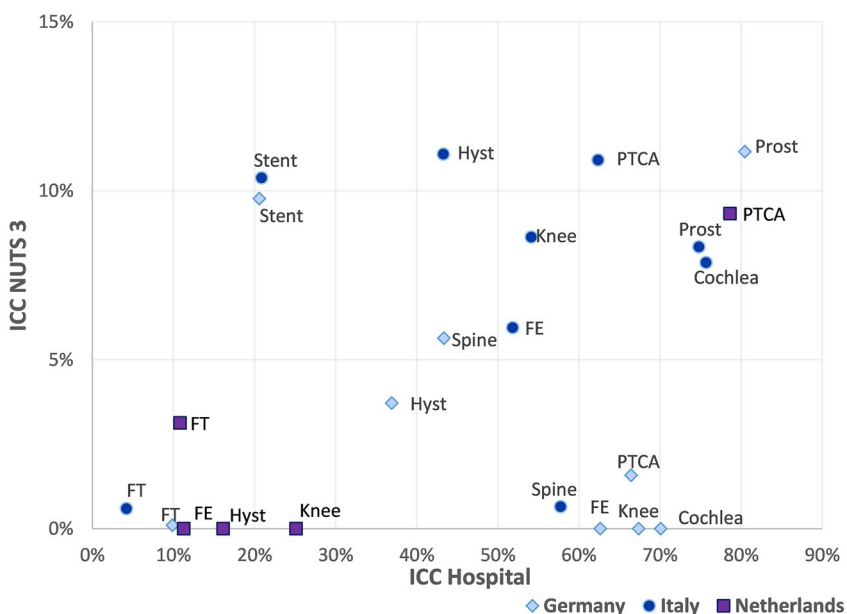


FIGURE 2 Intraclass correlation coefficient (ICC) comparison on hospital- and NUTS3-level. FE, Femur effective; FT, femur treatment; Hyst, Laparoscopic Hysterectomy; ICC, intraclass correlation coefficient; Knee, knee implantation; NUTS, Nomenclature des unités territoriales statistiques Prost, laparoscopic prostatectomy; PTCA, percutaneous transluminal coronary angioplasty; Spine, decompression with fusion

At the NUTS3 level, except for case studies 6 (spinal stenosis) and 7 (prostatectomy), Italy had a larger ICC than did Germany, indicating that more of the variance could be attributed to the regional level in Italy. In Germany, four case studies had an ICC at the NUTS 3 level that was close to zero: case study 1 (FE), case study 3 (FT), case study 5 (knee prosthesis), and case study 9 (cochlear implant or revision). For the Netherlands, the ICC at the NUTS3 level was zero for three case studies – case study 1 (FE), case study 5 (knee prosthesis) and case study 8 (hysterectomy) – whereas the other two available case studies had a large ICC at the NUTS3 level compared to the other countries. The largest difference between countries was found for case study 2 (PTCA; Germany: 1.6%, Italy: 10.9%, and the Netherlands: 9.3%), case study 5 (knee prosthesis; Germany: 0.0%, Italy: 8.6%, Netherlands: 0.0%), and case study 9 (cochlear implant or revision; Germany: 0.0%, Italy: 7.9%). Aside from these notable differences in the ICC among the case studies, the magnitude of the ICCs in the same case study were often similar at both the regional and hospital levels across the three countries.

3.3 | Regression

Table 3 gives an overview of the regression results for the different case studies in Germany, Italy, and the Netherlands organized according to different dimensions of supply (“efficiency in delivery of healthcare” and “capacity constraints”) and demand (“differences in health status”, “patient access” and “patient optimism”) and the three treatment categories (category I (effective care), category II (preference-sensitive care) and category II/III (preference-sensitive/supply-sensitive care)). All regressions included observations from all available years (Germany and Italy: 2012–2016; Netherlands: 2014–2016) and a year fixed effect. We note an Odds Ratio (OR) greater than one with a “+” and use a green color code if the predictor was significant at common thresholds. An OR less than one is indicated by a “-”, and we use a blue color code if the results were significant at common thresholds. For each case study, the first column, with the heading “GER”, shows the results for Germany, the second column, with the heading “IT”, shows the results for Italy, and the third column, with the heading “NL”, shows the results for the Netherlands.

For the case studies in category I (effective treatment), many of the hospital-level variables were significant in all three countries, except for case study 1 (FE) in the Netherlands, where these variables seem to have a significant impact on the prediction of the treatment decision in only a few case studies. In the dimension “efficiency in the delivery of healthcare”, the variables for number of relevant cases and the average length of stay for relevant cases were often highly significant. The

TABLE 3 Regression results of the available case studies for Germany, Italy and the Netherlands

HOSPITAL	Category I: Effective Care									Category II: Preference sensitive care									Category II/III: Preference sensitive care/Supply sensitive care								
	1. Femur Effective			2. PTCA			3. Femur Treatment			4. Knee			5. Knee Prosthesis			6. Spinal Stenosis			7. Prostatectomy			8. Hysterectomy			9. Cochlea Implant		
	GER	IT	NL	GER	IT	NL	GER	IT	NL	GER	IT	NL	GER	IT	NL	GER	IT	NL	GER	IT	NL	GER	IT	NL	GER	IT	NL
Efficiency in delivery of healthcare																											
No. of relevant cases †	-	-	-	+	+	+	-	-	+	-	+	-	-	-	+	-	-	+	-	-	+	+	+	-	-	-	
Avg. LOS relevant cases † [days]	+	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
Ownership private [1=private]	+	-	N/A	+	+	N/A	+	+	N/A	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
Ownership not-for profit [1=not-for profit]	-	N/A	N/A	+	N/A	N/A	+	N/A	N/A	+	N/A	N/A	+	N/A	N/A	+	N/A	N/A	+	N/A	N/A	-	N/A	N/A	-	N/A	
Inpatient cases per bed †	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
University hospital [1=university]	+	+	-	+	+	+	-	-	+	-	+	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
Capacity constraints																											
Hospital with less than 50 cases [1=50 relevant cases]	+	-	-	-	-	+	+	+	+	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+	+	+	
Beds † [Number]	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
NUTS3																											
Patient access																											
General practitioner † [per 100,000 inhab.]	-	+	+	+	+	+	-	-	-	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
Hospital beds † [per 1,000 inhab.]	-	+	-	+	+	+	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
Population density † [per km²]	+	+	+	+	+	+	-	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
Share of non-citizen residents † [% of total population]	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
Patient "optimism"																											
Students in up. sec. education † [% of total students]	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
Without lower sec. education † [% of school-leavers]	-	+	-	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
Voter turnout national election † [% of total]	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
Avg. household income † [ppp adjusted € per inhab.]	+	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
Differences in health status																											
Life expectancy of a newborn † [years]	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
Unemployment † [total employment]	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
Population aged 65+ † [% of total]	-	+	+	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
Number of observations (N)	407,284	584,548	35,168	423,177	251,372	50,551	313,652	439,773	26,793	271,642	164,423	1,454,198	423,591	77,621	375,886	94,845	110,582	78,371	308,567	117,563	10,656	130,215	41,591				
ICC Hospital	0.0%	6.0%	0.0%	1.6%	10.9%	9.3%	9.9%	4.2%	10.8%	20.6%	20.9%	67.4%	54.1%	25.1%	43.3%	57.7%	80.4%	74.8%	36.9%	43.3%	16.1%	70.1%	75.7%				
ICC Regional/NUTS3																											
		**p<0.001	**p<0.01	**p<0.01	**p<0.05	OR<1	**p<0.001	**p<0.01	**p<0.05																		

Note: Table 3 shows the results of the three-level regression for Germany, Italy, and the Netherlands. Regressions include year fixed effects for all available years (Germany & Italy: 2012–2016, Netherlands: 2014–2016). Patient-level variables (age, gender, Elixhauser groups, main diagnosis dummy) are omitted in this table and used for risk adjustment, but can be found in Appendix D. A color code in green suggest an Odds Ratio (OR) greater than one, and a blue color code suggest an OR smaller than one. † = mean-centered variables; Italy: not-for-profit ownership is not applicable, Netherlands: not-for profit and private ownership are not applicable. Case studies 4, 6, 7 and 9 are unavailable for the Netherlands due to low sample size or unavailable procedure codes. Odds ratios and standard errors, as well as information on the goodness of fit, can be found in the Supporting Information S1.

variables at the regional level had less of an impact on the prediction of the treatment decision, particularly those in the dimension “patient optimism”. The only exception was the dimension “differences in health status”, which had significant variables in all three countries. In this treatment category, the marginal R^2 was mostly relatively low, while the conditional R^2 was mostly moderate (Supporting Information S1).

For the case studies in category II (preference-sensitive care), again the hospital-level variables had many significant results, as was the case with category I. The variables within the dimension “efficiency in delivery of health care” seemed to be important, especially the number of relevant cases and the average length of stay for relevant cases, as well as private ownership. An exception to this pattern was case study 3 (FT), where in all three countries fewer hospital variables were significant compared to the other case studies. The variables in the dimension “capacity constraints” were significant less often for case studies in treatment category II compared to the other categories, especially the variable for the number of beds. For category II at the regional level, we see different results compared to category I. The dimension “patient optimism” appeared to be important in this treatment category based on the number of significant results, especially for the variables for the share of students with tertiary education and for household income. In the dimension “patient access”, the variables for general practitioners and for the share of non-citizen residents were also significant predictors of the treatment decision at the regional level in all three countries, except for case study 3 (FT) and for the Netherlands in both of the case studies available in this category. Variables in the dimension “differences in health status”, such as life expectancy, the unemployment rate, and the population age 65+ years were also often significant predictors of the treatment decision in all three countries. In this treatment category, the marginal R^2 was for most case studies, relatively large, except for case study 6 (spinal stenosis). In contrast to case studies in the effective-care category, the difference between conditional R^2 and marginal R^2 was usually lower (Supporting Information S1).

For the category II/III (preference-sensitive care/supply-sensitive care), we see mostly similar results compared to category II except for the Netherlands. Notable differences compared to the other treatment categories were at the hospital level, where variables for the dimension “efficiency in delivery of health care” and especially for the dimension “capacity constraints” have an often significant negative impact on the predicted outcome for Germany and Italy. At the regional level we could not identify any clear differences compared to the other treatment categories, except that the variable for general practitioners in a region in all case studies had a significant impact on the predictions and the variable household income appears to have been less significant results for these case studies than they were for case studies in treatment category I and II. Case study 8 (hysterectomy) in the Netherlands had no significant results according to common thresholds, except for the average length of stay for relevant cases, which is probably due to the low sample size. In this treatment category, the marginal R^2 varied across case studies. Whereas it was relatively low for case study 7 (prostatectomy) and case study 8 (hysterectomy), for case study 9 (cochlea) it was relatively large. The difference between the conditional R^2 and marginal R^2 was quite high for most case studies (Supporting Information S1).

3.4 | Sensitivity analysis

For a sensitivity analysis, we performed the same analysis as that described above but excluded hospitals that could be regarded as outliers. With this restriction, we lost an average of 5.9% (min: 3.1%; max: 8.5%) of the observations in Germany, 3.9% (min: 0.7%; max 7.8%) of the observations in Italy, and 0.9% (min: 0.1% max: 2.1%) of the observations in the Netherlands. The results can be found in Appendix C (Table C1, Figure C1) and are mostly comparable to the model that included outliers. As expected, the results were often less extreme and we did see, in some cases, lower levels of significance. Notable differences were shown in the variable for inpatient cases per bed in the dimension “efficiency in delivery of healthcare”, which in all three countries was significant more often when excluding the outlier. The ICC at the hospital level was also reduced in all case studies, except for case study 8 (hysterectomy) in the Netherlands, where the ICC increased slightly. At the regional level, there was a slight increase in the ICC in all countries when excluding the outlier. This is in line with our hypothesis that the outliers would not substantially alter our results on the importance of most variables.

4 | DISCUSSION

Using a multilevel model approach and nine case studies grouped into the categories effective care, preference-sensitive care, and supply-sensitive care, we aimed to identify potential sources of variation in the utilization of medical treatments across three European countries. Furthermore, we sought to identify the level to which most of the unexplained variation could be attributed. Based on our ICC analysis, the results of our within-country analysis suggest that a substantial part of the variation in

the treatment decision in Germany, Italy, and the Netherlands could be predicted by differences at the hospital level and only a comparably minor part by the differences in regional (NUTS3) characteristics. These results on the magnitude of the ICC were consistent across all of our case studies in Germany, Italy, and the Netherlands. With regard to potential drivers of variation, as described in the model by Skinner (2012), the results were less unambiguous because signs and significance levels changed across countries or case studies. Nevertheless, we could identify some patterns across the case studies and treatment categories that were in line with the theoretical model. In case studies grouped into the effective-care category (category I), the regional level had less explanatory power based on the number of significant estimators, whereas in case studies grouped into category II (preference-sensitive care) or category II/III (preference-sensitive/supply-sensitive care) these variables were more often significant and the regional ICC was in several case studies larger compared to the effective-care case studies. Our regression results suggest that the hospital level plays a crucial role in explaining the variation among different treatment options. Therefore, we argue that accounting for hospital and regional factors is important when studying regional variation. However, many studies in this field are descriptive or ignore one or both of these important levels (Corallo et al., 2014).

Although our regression results are heterogeneous, we suggest possible explanations and interpretations of our findings. At the hospital level, there are some important differences in the role of variables between dimensions and treatment categories. For almost all of the case studies, variables we attributed to the category “efficiency in delivery of health care” seemed to play an important role in the treatment decision in Germany, Italy, and the Netherlands because many estimators had significant results. In some case studies, for example, case study 2 (PTCA) in Germany or case study 6 (spinal stenosis) in Germany and Italy we see that all, or almost all, estimators in the category “efficiency in delivery of health care” had a positive impact on the outcomes. This is in line with the literature on hospital efficiency, which shows that more efficient hospitals are also those that tend to have high-technology medical equipment or better information technology systems (Leleu et al., 2018; Zhivan & Diana, 2012). For category II (preference-sensitive care), the variables attributed to “capacity constraints” appeared to be less important, based on the number of significant variables and their significance level, for explaining the use of the different treatment options compared to the other categories, which is in line with the hypothesis that these treatments are influenced less by the scarcity of resources. For case studies in treatment category II/III, especially for the case studies 7 (prostatectomy) and 8 (hysterectomy), variables attributed to the category “capacity constraints” appeared to be important for Germany and Italy, suggesting that hospitals with a certain volume in the treatment area were more likely to use innovative procedures.

At the NUTS3 level, the results for the category “patient access” were inconclusive for most case studies. Surprisingly, for case study 3 (FT), the variables we classified as “patient access” had almost no significant results. Unlike the results of Torbica et al. (2017), who used a comparable methodology but did not account for hospital-level variables, education and the average age of a region's inhabitants did not seem to have a consistent significant effect across our case studies on the treatment decision. This may be due to our more detailed risk-adjustment methods and because we also addressed the hospital level.

Furthermore, we found some patterns in the results that followed our classification of treatments into three categories, especially when analyzing the ICC. Treatments classified as involving effective care had a comparably low ICC at the regional level, especially in Germany, but a high ICC at the hospital level compared to case studies involving preference-sensitive or supply-sensitive care. Case study 3 (FT), which we categorized as preference-sensitive care, had a low ICC at both the regional and hospital levels in Germany, Italy, and the Netherlands. In fact, in Germany and Italy the ICC was even lower than that for the case study we derived as an effective-care case for treating femur fracture. This indicates that although there is variation across countries with regard to a preference one of the options, each country treats patients uniformly within its own jurisdiction. This can be interpreted in different ways. It could suggest that once the patient is in a hospital that offers treatment for the diagnosis in question, he or she receives a quite similar care regardless of where he or she is within each country. While this could explain the results for the treatment of femur fracture, we could not find similar results for case study 4 (stent), for which variation was large at the regional level. This variation may result from the long debate over the benefits and risks of BMS versus DES, and changes in the body of evidence and the recommendations of professional societies over time, leading to a long period of uncertainty. In contrast, evidence on the treatment of femoral neck fractures has not changed markedly in recent years (Avdic et al., 2019). This may also imply that the variation we observed at the hospital level in case studies 1 (FE) and 2 (PTCA) in the effective-care category may be due to a potential mismatch of patients with a certain diagnosis and the capability of hospitals to treat these patients effectively. While for case study 1, this would be less severe because the treatment is less time-sensitive, for PTCA treatment of patients with an AMI this can be problematic because timely treatment is crucial.

The results for category II treatments indicate that variables we chose to operationalize the concept of patient optimism play an important role more frequently than they do in the other treatment categories. Furthermore, these cases often have a high ICC at the regional level, especially for case study 5 (knee prosthesis), a finding that is in line with the hypothesis that these treatments are preference sensitive. These results indicate that such treatments may benefit from efforts to better inform the patient about the disease or improve shared decision making (Veroff et al., 2013). Lastly, for the case studies that we expected

to be influenced either by preferences of the patient or the physician or to be influenced by the supply-side of health services (category II/III), we see that the variables we selected to operationalize to the dimension “capacity constraints” and the variable for general practitioners seem to be especially important, except for case study 8 for the Netherlands. This also fits into the classification of treatment categories that we used as it underlines that these case studies are influenced by supply-side and patient-access variables.

Our study has a number of strengths and adds to the previous research on variation in the use of medical devices. To begin with, it is one of the first studies to investigate variation in the use of medical devices across jurisdictions and was able to include a large number of observations for most countries in most case studies. Second, we used an analytic framework to group a range of case studies into treatment categories that may influence the treatment decision differently based on the evidence, preferences and supply. We assigned independent variables to different supply- and demand-side dimensions that may influence the treatment decision. In doing so, we overcame some of the limitations of primarily descriptive and exploratory approaches to analyzing patterns in variation, and linked our results to theoretical considerations related to supply and demand. Third, to our knowledge this is the first study to use a three-level logistic model to analyze patient-level data combined with hospital and regional variables in this research context. This approach enabled us to attribute differences in the treatment decision in different categories to hospital and regional characteristics while correcting for important patient-level factors.

Our study also has a number of limitations that must be considered when interpreting our findings. We used rich nationwide data sets combining data at three different levels (patient, hospital, NUTS3 level) in three countries. While this gave us a unique opportunity to study variation in the use of medical devices, we were limited in several ways. First, standardizing diagnoses and procedures across different countries is challenging. In particular, the level of detail of procedure systems varies across countries. Thus, in spite of the efforts we undertook with medical or coding experts to map procedures uniformly to the case studies, the definitions of procedures and their interpretations may vary in medical practice. Second, we found that coding patterns in ICD systems varied across countries, also due to financial incentives. For instance, Germany coded more co-morbidities than Italy, which may have impacted our risk adjustment. Furthermore, in the Netherlands some of the procedure codes we needed to distinguish among our cases were unavailable (e.g., case study 4 for the differentiation of DES and BMS) or rarely coded (e.g., case study 7 (prostatectomy)), which reduced the number of cases studies in that country. Additionally, the number of observations for the Netherlands was relatively low in some cases (e.g., for case study 8 (Hysterectomy)), which might have had an impact on the specification of the complex three-level model. Third, by using a three-level regression model, we were able to account for factors at the hospital and regional levels. However, we neglected other potentially important levels at which clustering might have taken place and affected the treatment decision. For instance, the surgeon or physician level has been found to have an important impact on the treatment decision (Fisher et al., 2016). This could be the reason for the relatively low ICC for most case studies at the regional level. Fourth, our study was not designed to identify causal relationships, and our results should be interpreted accordingly. Fifth, we determined relevant dimensions on the supply side, such as resource availability, and on the demand side, such as differences in health status following the framework proposed by Skinner (2012) in all three countries and selected variables that would reflect each of the dimensions. Although we were able to identify variables for many dimensions, others such as physician optimism, the marginal utility of income of a physician, or malpractice risks went beyond the scope of the administrative data sets at our disposal. It is important to note that distinguishing between the three categories of treatments is debatable, especially for categories II and III. A good example in this study is case study 7 (prostatectomy), where we distinguished between open and laparoscopic surgery. The treatment decision can be either preference-sensitive (because either the patient or physician has certain beliefs about effectiveness in a given case), but it can also depend on supply-sensitive factors, that is, the capability of a hospital to provide laparoscopic surgery, which then would also be a question about the adoption of medical technology (Hatz et al., 2017).

Further research could expand our analysis in several ways. As already mentioned, we were not able to identify variables for all of the dimensions proposed in the framework. To obtain a more comprehensive picture of patterns of supply and demand in the use of medical devices, researchers may wish to identify further datasets containing other relevant variables in the missing dimensions or reasonable surrogate variables that reflect those variables. These could then be merged with the datasets we used in the present study. Also, a more detailed look into fewer but more comparable case studies across treatment categories could be of benefit (e.g., in one disease area). Focusing on more detailed analyses with fewer case studies or examining a change in treatment evidence over time could facilitate an understanding of the underlying mechanism of variation in the treatment categories and may help explain results that were different (e.g., opposite sign) across countries in our study. Moreover, our analysis focused on the use of different procedures. While this indicates practice variation, the impact of utilization on outcomes such as mortality or quality of life was beyond the scope of this study and should be investigated by future research. While investigating variation in the use of medical devices reveals under- or overprovision of health services, the overall aim of the physician is to provide care in the most cost-effective way, and therefore health outcomes are of particular interest.

In conclusion, the findings of our study suggest several measures that policymakers may wish to implement to reduce the amount of potential unwarranted variation. First, our findings indicate that treatment patterns may vary across countries for several reasons depending on the treatment categories and the procedures themselves. Indeed, the treatment categories may serve as a useful tool to better understand these reasons and help tailor the use of common approaches to reduce variation (McCulloch et al., 2013). For effective care, the measures to reduce variation need to be addressed at the hospital or physician level, for example, by increasing and improving efforts at disseminating clinical practice guidelines, encouraging physicians to benchmark themselves against their peers, providing financial or other incentives to use evidence-based procedures, or improving the availability of standard and, where appropriate, also newly approved technologies. Additionally, for preference-sensitive care, the patient needs to be better informed about the treatment decision. Due to information asymmetry, patients usually have less information than physicians about their disease and treatment options. The physician therefore plays a crucial role in providing the patient with the information needed to make the best decision across the patient pathway, for example, by means of standardized and detailed information on the options or shared decision making. For the cases we classified as potential supply-sensitive care, we saw that variables such as the number of general practitioners also seemed to explain part of the variation. While we cannot determine in this study whether this was because of supplier-induced demand and therefore represents unwarranted variation, it would imply that for these treatments policymakers may wish to focus on ensuring the availability of clear guidance to help patients and physicians decide whether a given treatment is applicable to their situation so that they are less affected by supply factors.

ACKNOWLEDGMENTS

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 779306.

CONFLICT OF INTEREST

Mr. Rabbe reports grants from European Union's Horizon 2020 research and innovation program under grant agreement No 779306, during the conduct of the study. Ms. Möllenkamp reports grants from European Union's Horizon 2020 research and innovation program under grant agreement No. 779306, during the conduct of the study. Dr. Blommestein reports grants from European Union's Horizon 2020 research and innovation program under grant agreement No 779306, during the conduct of the study. Dr. Wetzelaer reports grants from European Union's Horizon 2020 research and innovation program under grant agreement No 779306, during the conduct of the study. Dr. Heine has nothing to disclose. Dr. Schreyögg reports grants from European Union's Horizon 2020 research and innovation program under grant agreement No 779306, during the conduct of the study.

ETHICS STATEMENT

The project was approved by the WISO Laboratory of Universität Hamburg for compliance with terms of use and ethical standards at April 26, 2018.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the statistical offices of the corresponding country. Restrictions apply to the availability of these data, which were used under licence for this study.

ORCID

Stefan Rabbe  <https://orcid.org/0000-0002-8657-5694>

Meilin Möllenkamp  <https://orcid.org/0000-0003-1831-6954>

Benedetta Pongiglione  <https://orcid.org/0000-0001-8539-1554>

Hedwig Blommestein  <https://orcid.org/0000-0002-5846-0531>

Pim Wetzelaer  <https://orcid.org/0000-0002-7450-0046>

Renaud Heine  <https://orcid.org/0000-0002-5324-6659>

Jonas Schreyögg  <https://orcid.org/0000-0001-8030-2161>

ENDNOTES

¹ For the year 2012–2013 our data were incomplete, including in 2012 only 82% and in 2013 only 84% of cases. Moreover, the coding system (diagnoses and procedures) changed in 2012 and hospitals were permitted to adapt differently to the new system until 2014, which compromised the analysis.

- ² One exception is case study 9 (cochlear implant or revision) for which we selected only on the main diagnosis because conductive hearing loss is a common comorbidity.
- ³ Robotic assisted surgery is a third and newer approach, which will be due to comparably low case numbers in most hospitals not analysed separately in this study, but classified as part of the laparoscopic surgery (Aarts et al., 2015).

REFERENCES

- Appleby, J., Raleigh, V., Frosini, F., Bevan, G., Gao, H., & Lyscom, T. (2011). *Variations in health care: The good, the bad and the inexplicable*. The Kings Fund.
- Avdic, D., van Hinke, S., Lagerqvist, B., Propper, C., & Vikström, J. (2019). *Information shocks and provider adaptation: Evidence from interventional cardiology* (Vol. 13627). CEPR Discussion Papers.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1). <https://doi.org/10.18637/jss.v067.i01>
- Busse, R., Schreyögg, J., & Smith, P. C. (2008). Variability in healthcare treatment costs amongst nine EU countries - results from the HealthBASKET project. *Health Economics*, 17(1 Suppl), S1–S8. <https://doi.org/10.1002/hec.1330>
- Chandra, A., & Skinner, J. (2012). Technology growth and expenditure growth in health care. *Journal of Economic Literature*, 50(3), 645–680. <https://doi.org/10.1257/jel.50.3.645>
- Comendéiro-Maaløe, M., Estupiñán-Romero, F., Thygesen, L. C., Mateus, C., Merlo, J., & Bernal-Delgado, E. (2020). Acknowledging the role of patient heterogeneity in hospital outcome reporting: Mortality after acute myocardial infarction in five European countries. *PLoS One*, 15(2), e0228425. <https://doi.org/10.1371/journal.pone.0228425>
- Corallo, A. N., Croxford, R., Goodman, D. C., Bryan, E. L., Srivastava, D., & Stukel, T. A. (2014). A systematic review of medical practice variation in OECD countries. *Health Policy*, 114(1), 5–14. <https://doi.org/10.1016/j.healthpol.2013.08.002>
- Cutler, D., Skinner, J. S., Stern, A. D., & Wennberg, D. (2019). Physician beliefs and patient preferences: A new look at regional variation in health care spending. *American Economic Journal: Economic Policy*, 11(1), 192–221. <https://doi.org/10.1257/pol.20150421>
- Elixhauser, A., Steiner, C., Harris, R. D., & Coffey, R. (1998). Comorbidity measures for use with administrative data. *Medical Care*, 36(1), 8–27. <http://www.jstor.org/stable/3766985>
- Finkelstein, A., Gentzkow, M., & Williams, H. (2016). Sources of geographic variation in healthcare: Evidence from patient migration. *Quarterly Journal of Economics*, 131(4), 1681–1726. <https://doi.org/10.1093/qje/qjw023>
- Fisher, S., Yasui, Y., Dabbs, K., & Winget, M. (2016). Using multilevel models to explain variation in clinical practice: Surgeon volume and the surgical treatment of breast cancer. *Annals of Surgical Oncology*, 23(6), 1845–1851. <https://doi.org/10.1245/s10434-016-5118-2>
- Glover, A. J. (1938). The incidence of tonsillectomy in school children. *Proceedings of the Royal Society of Medicine*, 31, 95–113. <https://doi.org/10.1093/ije/dym258>
- Göpfarth, D., Kopetsch, T., & Schmitz, H. (2016). Determinants of regional variation in health expenditures in Germany. *Health Economics*, 25(7), 801–815. <https://doi.org/10.1002/hec.3183>
- Hatz, M. H. M., Schreyögg, J., Torbica, A., Boriani, G., & Blankart, C. R. B. (2017). Adoption decisions for medical devices in the field of cardiology: Results from a European survey. *Health Economics*, 26(Suppl 1), 124–144. <https://doi.org/10.1002/hec.3472>
- Heijink, R., Engelfriet, P., Rehnberg, C., Kittelsen, S. A. C., & Häkkinen, U. (2015). A window on geographic variation in health care: Insights from EuroHOPE. *Health Economics*, 24(Suppl 2), 164–177. <https://doi.org/10.1002/hec.3287>
- Johansson, N., Jakobsson, N., & Svensson, M. (2018). Regional variation in health care utilization in Sweden - the importance of demand-side factors. *BMC Health Services Research*, 18(1), 403. <https://doi.org/10.1186/s12913-018-3210-y>
- Leleu, H., Al-Amin, M., Rosko, M., & Valdmanis, V. G. (2018). A robust analysis of hospital efficiency and factors affecting variability. *Health Services Management Research*, 31(1), 33–42. <https://doi.org/10.1177/0951484817730537>
- Lembcke, P. A. (1952). Measuring the quality of medical care through vital statistics based on hospital service areas: 1. Comparative study of appendectomy rates. *American Journal of Public Health*, 42, 276–286. <https://doi.org/10.2105/ajph.42.3.276>
- McCulloch, P., Nagendran, M., Campbell, W. B., Price, A., Jani, A., Birkmeyer, J. D., & Gray, M. (2013). Strategies to reduce variation in the use of surgery. *The Lancet*, 382(9898), 1130–1139. [https://doi.org/10.1016/S0140-6736\(13\)61216-7](https://doi.org/10.1016/S0140-6736(13)61216-7)
- Moseley, B. J., O'Malley, K., Petersen, N. J., Menke, T. J., Brody, B. A., Kuykendall, D. H., Wray, N. P., Ashton, C. M., & Wray, N. P. (2002). A controlled trial of arthroscopic surgery for osteoarthritis of the knee. *New England Journal of Medicine*, 347(2), 81–88. <https://doi.org/10.1056/NEJMoa013259>
- Moura, A., Salm, M., Douven, R., & Remmerswaal, M. (2019). Causes of regional variation in Dutch healthcare expenditures: Evidence from movers. *Health Economics*, 28(9), 1088–1098. <https://doi.org/10.1002/hec.3917>
- Neumann, F.-J., Sousa-Uva, M., Ahlsson, A., Alfonso, F., Banning, A. P., Benedetto, U., Collet, J.-P., Falk, V., Head, S. J., Jüni, P., Kastrati, A., Koller, A., Kristensen, S. D., Niebauer, J., Richter, D. J., Seferović, P. M., Sibbing, D., Stefanini, G. G., Windecker, S., ... Zembala, M. O. (2019). 2018 ESC/EACTS Guidelines on myocardial revascularization. *European Heart Journal*, 40(2), 87–165. <https://doi.org/10.1093/eurheartj/ehy394>
- OECD. (2014). *Geographic Variations in health care: What do we know and what can be done to improve health system performance (OECD health policy studies)*. OECD Publishing. <https://doi.org/10.1787/9789264216594-en>
- OECD. (2019). *Health at a Glance 2019. OECD indicators*. OECD Publishing. <https://doi.org/10.1787/4dd50c09-en>
- Quan, H., Sundararajan, V., Halfon, P., Fong, A., Burnand, B., Luthi, J.-C., Ghali, W. A., Beck, C. A., Feasby, T. E., & Ghali, W. A. (2005). Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. *Medical Care*, 43(11), 1130–1139. <https://doi.org/10.1097/01.mlr.0000182534.19832.83>

- R Core Team. (2020). R: A language and environment for statistical computing [computer software]. <https://www.R-project.org/>
- Schreyögg, J., Stargardt, T., & Tiemann, O. (2011). Costs and quality of hospitals in different health care systems: A multi-level approach with propensity score matching. *Health Economics*, 20(1), 85–100. <https://doi.org/10.1002/hec.1568>
- Skinner, J. (2012). Causes and consequences of regional variations. In M. V. Pauly, P. P. Barros, & T. G. McGuire (Eds.), *Handbooks in economics. Handbook of health economics: Volume 2* (1st ed., Vol. 2, pp. 45–93). Elsevier North Holland. <https://doi.org/10.1016/B978-0-444-53592-4.00002-5>
- Sommet, N., & Morselli, D. (2017). Keep calm and learn multilevel logistic modeling: A simplified three-step procedure using stata, R, mplus, and SPSS. *International Review of Social Psychology*, 30(1), 203–218. <https://doi.org/10.5334/irsp.90>
- Song, Y., Skinner, J., Bynum, J., Sutherland, J., Wennberg, J. E., & Fisher, E. S. (2010). Regional variations in diagnostic practices. *New England Journal of Medicine*, 363(2), 198–198. <https://doi.org/10.1056/NEJMc100034>
- Street, A., Kobel, C., Renaud, T., & Thuilliez, J. (2012). How well do diagnosis-related groups explain variations in costs or length of stay among patients and across hospitals? Methods for analysing routine patient data. *Health Economics*, 21(Suppl 2), 6–18. <https://doi.org/10.1002/hec.2837>
- Sutherland, K., & Levesque, J.-F. (2020). Unwarranted clinical variation in health care: Definitions and proposal of an analytic framework. *Journal of Evaluation in Clinical Practice*, 26(3), 687–696. <https://doi.org/10.1111/jep.13181>
- Torbica, A., Banks, H., Valzania, C., Boriani, G., & Fattore, G. (2017). Investigating regional variation of cardiac implantable electrical device implant rates in European healthcare systems: What drives differences? *Health Economics*, 26(Suppl 1), 30–45. <https://doi.org/10.1002/hec.3470>
- Veroff, D., Marr, A., Wennberg, D. E. (2013). Enhanced support for shared decision making reduced costs of care for patients with preference-sensitive conditions. *Health Affairs (Project Hope)*. 32(2), S. 285–293. <https://doi.org/10.1377/hlthaff.2011.0941>
- Wang, J. T., Wang, A. Y., Psarros, C., & Da Cruz, M. (2014). Rates of revision and device failure in cochlear implant surgery: A 30-year experience. *The Laryngoscope*, 124(10), 2393–2399. <https://doi.org/10.1002/lary.24649>
- Wennberg, J. E. (2002). Unwarranted variations in healthcare delivery: Implications for academic medical centers. *BMJ British Medical Journal*, 325(7370), 961–964. <https://doi.org/10.1136/bmj.325.7370.961>
- Wennberg, J. E., Fisher, E. S., & Skinner, J. S. (2002). Geography and the debate over Medicare reform, *Health Affairs*, 21, W96–W112. <https://doi.org/10.1377/hlthaff.w2.96>
- Wennberg, J. E., & Gittelsohn, A. (1973). Small area variation in health care delivery: A population-based health information system can guide planning and regulatory decision-making. *Science*, 182(4117), 1102–1108. <https://doi.org/10.1126/science.182.4117.1102>
- Zhivan, N. A., & Diana, M. L. (2012). U.S. Hospital efficiency and adoption of health information technology. *Health Care Management Science*, 15(1), 37–47. <https://doi.org/10.1007/s10729-011-9179-2>

SUPPORTING INFORMATION

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

How to cite this article: Rabbe, S., Möllenkamp, M., Pongiglione, B., Blommestein, H., Wetzelaer, P., Heine, R., & Schreyögg, J. (2022). Variation in the utilization of medical devices across Germany, Italy, and the Netherlands: A multilevel approach. *Health Economics*, 1–23. <https://doi.org/10.1002/hec.4492>

APPENDIX A: DESCRIPTION OF DATA SOURCES

Germany

The patient level data in Germany is based on the German diagnose related group (DRG)-Data. It includes all hospital cases which are reported within the German DRG system. It includes main and all secondary diagnoses based on International Classification of Diseases (ICD-10-GM) and procedures based on the Operationen-und Prozedurenschlüssel.

The hospital level data follows the publicly available structured quality reports of German hospitals following § 132 Sozialgesetzbuch (SGB) V. Data is self-reported yearly by all hospitals that treat patients of the statutory health insurance defined by § 108 SGB V but collected by the institute for quality and transparency in health care/Federal Joint Committee (G-BA). It is merged via a hospital individual id but can include several (satellite) locations of a hospital.

The NUTS3 level data is based on publicly available data collected by Federal Institute for Research on Building, Urban Affairs and Spatial Development, in German Indikatoren und Karten zur Raum- und Stadtentwicklung and the GENESIS online database of the German federal statistical office and merged via the hospitals' postal code.

Italy

The patient level data in Italy is based on the Schede di Dimissione Ospedaliera (SDO) dataset which includes all hospital cases financed by the regional National Health Service for public and accredited private hospitals. Diagnoses are coded in ICD-9 and procedures in ICD-9-CM. The data contains only the first five coded diagnoses but all coded procedures during a hospital visit.

TABLE A1 Overview of data sources

Level	Germany	Italy	The Netherlands
Patient	G-DRG-data (ICD-10-GM; OPS)	SDO of Ministro della salute; (ICD-9; ICD-9-CM)	LBZ data managed by DHD and accessed via CBS (ICD-10; ZA; CBV; CMSV)
Hospital	IQTIG; G-BA	SDO of Ministro della salute;	DHD; CBS
NUTS3	BBSR; StBA	ISTAT; ministry of finance, EUROSTAT; Ministero dell'Interno	CBS statline; NIVEL

Note: Coding system in parenthesis.

Abbreviations: BBSR, Bundesinstitut für Bau-, Stadt- und Raumforschung; CBV, Centraal Beheer Verrichtingenbestand; CM, clinical modification; CMSV, Classificatie van Medisch Specialistische Verrichtingen; DHD, Dutch Hospital Data; EUROSTAT, European Statistical Office; G-BA, Gemeinsamer-Bundesausschuss; G-DRG, German Diagnose Related Groups; GM, German modification; ICD, International Classification of Diseases; ISTAT, Istituto Nazionale di Statistica, NIVEL, Nederlands Instituut voor Onderzoek van de Gezondheidszorg; IQTIG, Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen; LBZ, Landelijke Basisregistratie Ziekenhuiszorg; NUTS, Nomenclature des unités territoriales statistiques; OPS, Operationen und Prozedurenschlüssel; SDO, Scheda di Dimissione Ospedaliera; StBA, Statistisches Bundesamt; ZA, Zorg Activiteit.

The hospital level data is based again on SDO data but complemented with data from the ministry of health. The public available data is reported to give an overview about the local health services, employed staff and care units in each area.

The NUTS3 level data is based on National Institute of statistics, in italian Istituto Nazionale di Statistica, Ministry of finance and interior, and Eurostat.

Netherlands

The patient level data in the Netherlands accessed through the Central Bureau of Statistics (CBS) and collected by Dutch Hospital Data (DHD) which includes data on all diagnoses based on ICD-10 and procedures codes based on Zorg Activiteit, Centraal Beheer Verrichtingenbestand and the Classificatie van Verrichtingen Classificatie codes. For two reasons in the Netherlands only the years 2014 to 2016 are available for analysis. First, the data provided by CBS is incomplete only including 82% in 2012 or 84% in 2013 of all cases and second the coding system (diagnoses and procedures) and the database changed until 2014 which disturbed the analysis.

The hospital level data is based on the Enquête Jaarcijfers Ziekenhuizen questionnaire provided by DHD and relied on CBS data for merging the data with the usually anonymized patient level data.

The NUTS3 level data is based on data by CBS Statline and Netherlands Institute for Health Research.

APPENDIX B: CASE-STUDY DESCRIPTION AND PROCEDURE LIST

Case 1 Femoral neck fracture | Effective treatment of femur & Case 3 Femoral neck fracture | Femoral neck fracture treatment decision

Femur fracture are a high incidence disease with a substantial burden of disease, especially for the elderly and one major public health problem in the Western nations with estimated worldwide case numbers up to 3.94 mil. in the year 2025 (Johnell & Kanis, 2004). The fracture of the femoral neck is the commonest reason of a proximal femur fracture, with about 60% of all hip fractures. For the treatment of femoral neck procedures surgery is mostly the treatment of choice especially for patients between age 40–65 (Sendtner et al., 2010). Bed or chair bound patients may also be treated conservatively (Scottish Intercollegiate Guidelines Network [SIGN], 2009). Surgical treatment in general is described as an effective treatment and aims on immediate pain relief and rapid mobilization and an accelerated rehabilitation (Hopley et al., 2010). Regarding the surgical procedure of choice, there exist two main treatment choices: Osteosynthesis/reduction (usually using several medical devices such as screws, cages or other implants) or the use of a hip prosthesis/arthroplasty (either total or partial). Several guidelines and reviews indicate that the treatment options are vigorously debated over the past decades, but agree on several indications where arthroplasty is preferred, namely age, activity level and arthrosis as comorbidity level (SIGN, 2009).

In this paper, we are focusing on this indications in two different ways: First in case 1, we are focusing on if the patient is treated with one of each surgical option. This case is often used to describe effective care with low variation. Second in case 3, we are focusing on the treatment decision (osteosynthesis vs. arthroplasty) of the effective care, which is debated in the medical field. In this case we are expecting variation. We are differentiating between osteosynthesis and hip replacement.

Case 2 ST-elevation myocardial infarction | Percutaneous Translaminar Coronary Angioplasty (PTCA) & Case 4 ST-elevation myocardial infarction | Drug Eluting (DES) and Bare Metal Stent (BMS)

Myocardial infarction is an emergency indication following a myocardial ischemia. In this case, we are focusing on ST-elevation myocardial infarction (STEMI) as a diagnosis indicating reperfusion therapy. Guidelines show that the preferred method for reperfusion is Percutaneous Translaminar Coronary Angioplasty (PTCA) performed in a cardiac catheterization lab compared to fibrinolytic therapy (Ibanez et al., 2018). PTCA should be available as fast as possible for as many patients as possible with a STEMI diagnosis (Ibanez et al., 2018; Kalla et al., 2006). Additionally to the PTCA treatment, stents can be used to increase long-term results and lower restenosis risks. For stenting, two main options are available: The bare metal stent (BMS), which treats the narrowed vessels and the drug eluting stent (DES), which additionally limits the regrowth of cells and therefore restenosis. In the past a discussion on superiority on each of the treatment options (DES vs. BMS) took place but since the 2012 European Society of Cardiology guidelines (Steg et al., 2012) DES are preferred. In the more recent guidelines and under the light of second generation DES this recommendation was strengthened to the highest level of evidence due to higher efficacy and safety compared to BMS (Ibanez et al., 2018; Neumann et al., 2019).

For the STEMI diagnose we defined two different case-studies. Case-study 2 defines the PTCA as an effective care case, meaning that every patient with the diagnosis should receive the treatment and although alternatives are available, non-treatment can be seen as an under-provision of care. We therefore looked at all patients with the diagnoses of STEMI and indicated if they received the PTCA treatment.

Case-study 4 focuses on stenting as a decision between the use of BMS and DES as a case-study for preference sensitive. While DES are now preferred by guidelines we expect that the long uncertainty still lead to some variation in the treatment choice.

Case 5 Knee arthrosis | Implantation of a total hip implant

Knee replacement surgeries are a very frequent and cost-effective (Price et al., 2018) intervention and the number of knee replacements continues to increase in most Organisation for Economic Co-operation and Development (OECD) countries with high variations in knee replacements rates across as well as within countries (OECD, 2014). As well as hip implants, total knee replacement surgeries are an effective intervention to treat patients suffering from chronic osteoarthritis with persistent severe symptoms such as pain, restrictions in mobility and/or degeneration of the joint cartilage primarily reducing the health related quality of life (OECD, 2019), however the decision on if and when the surgery is appropriate is based on preferences of the patient and surgeon (Price et al., 2018). The average age of patients undergoing a knee replacement is approximately 65 years (Price et al., 2018), with risk factors being among others genetic factors, aging, female sex, obesity and high bone density as well as local biomechanical risk factors such as joint laxity or injury (National Institute for Health and Care Excellence [NICE], 2014). Within this procedure the patients' knee will be completely or partially replaced by an artificial device. Generally, surgery is recommended if symptoms such as pain and disability affecting the quality of life persist after exhausting conservative treatments (NICE, 2014). Diagnoses are often based on patients' symptoms in combination with an X-Ray or a magnetic resonance image. However, there is often only a poor link between them due to heterogeneity in patients' noticeably symptoms and the resulting imaging test. Brownlee et al. (2012) indicate that the severity of the damage detected by an image is not a sufficient reason for recommending treatment. Price et al. (2018) also refer to a not well-defined threshold for surgery since it is influenced by many factors originating from patients' and surgeons' preferences. Hawker et al. (2001) showed that willingness to undergo surgery for eligible patients was comparably low, but higher for patients in high prevalence areas, which underlines the hypothesis that utilization is demand driven. Overall, knee replacement surgeries show a high rate and significant evidence of successful treatment outcome and long-term implant survival (Price et al., 2018).

Case 6 Lumbar Spinal Stenosis | Decompression versus decompression plus fusion

Lumbar spinal stenosis is a condition where diminished space is available for neural and vascular elements in the spinal canal (Kreiner et al., 2013). A strong debate about the treatment of spinal stenosis with no clear evidence on one solution exists. Options are conservative treatment (e.g., pain medication, mobilization or exercise) or surgery to increase the space in the spinal canal to remove spinal elements (decompression) and/or the instrumentation to tackle instabilities (fusion or posterior spacers). Evidence is insufficient on all three options leading to no clear treatment standard, but some form of medical treatment is recommended for patients with moderate or severe symptoms (Kreiner et al., 2013).

In this case we focus on the treatment differences in surgery differentiating between decompression without fusion and decompression with fusion, since we cannot properly observe the conservative treatment in our data. While the additional fusion avoids iatrogenic instability of the decompression, reoperation rates are higher as well as costs, while potential gains effectiveness are still under debate (Deyo et al., 2013; Shen et al., 2018).

Case 7 Malign neoplasm of the prostate | Radical prostatectomy (open surgery vs. laparoscopic)

Prostate cancer is the fourth most common diagnoses in both sexes and second most diagnosed form of cancer among men worldwide in 2012 with over 1.1 Mio. new cases diagnosed (Ferlay et al., 2015). The mortality rate is comparably low in contrast to many other cancer diagnoses, and decreasing in some high income countries (Bray et al., 2010). If prostate cancer is diagnosed, several treatment options, such as watchful waiting/active surveillance, radical prostatectomy, radiation- and chemo therapy are most common, but regional variation and time trends seem to exist (Cooperberg et al., 2010).

In this case, we are focusing on a subgroup of the prostate cancer treatment, the most prominent and only surgical option of the treatment: Radical prostatectomy, as the other options are not properly coded in the data. We are differentiating between laparoscopic surgery (which can include robotic surgery) and open surgery. Compared to open surgery, which has been for a long time the standard procedure and is usually a shorter procedure, laparoscopic surgery appears to have advantages regarding blood loss, catheterization time, faster recovery process and shorter hospital stays. Conflicting results are shown for the overall complication rate and the positive surgical margin rate (Ficarra et al., 2009).

Case 8 Benign neoplasms of uterus | Hysterectomy (open surgery vs. laparoscopic)

Hysterectomy, the surgical removal of the uterus, is one of the most common surgical procedures in women, where about 90% of treatments are undertaken for benign gynecological diseases (Aarts et al., 2015; Wright et al., 2013). While in recent years different treatment options are available (e.g., ablation techniques, intrauterine device (IUD)), hysterectomy is very common and can itself be categorized into different groups. We will differentiate in this paper between open surgery (including abdominal, the traditionally and more invasive approach, and vaginal hysterectomy, which is regarded less invasive but is discussed to have a greater risk of bladder injury) or laparoscopic surgery, where at least a part of the procedure is performed laparoscopically or robotic assisted.³ We included both total (including cervix removal) and subtotal/supracervical hysterectomy.

In this case, we want to investigate the variation between open and laparoscopic surgery for the treatment of benign neoplasm of the uterus. There exist patient factors that influence the surgeons' choice of treatment, but many studies observe large variation at hospital, regional and even surgical level. While evidence on effectiveness is still uncertain, mainly because of the vast variation in treatment options and different learning curves and lack of qualitative evidence, laparoscopic surgery is for many cases the preferred option although it is still less used (Aarts et al., 2015).

Case 9 Hearing impairment | Implantation of cochlear implant

The global prevalence of hearing loss with more than 1,500 Mill. people (430 Mi. with disabling hearing loss) in 2019 is substantially high with a growing projection (World Health Organization [WHO], 2021). Cochlear implants (CI) are the surgical alternative to traditional amplifications and the first effective cranial nerve stimulator in order to restore the hearing of deafened adult and pediatric patients with potential positive medical as well as societal impacts (Naples & Ruckenstein, 2020). But, to date patients with residual hearing in the lower frequencies also benefit from treating with CI (Skarzynski et al., 2006). In addition, Borges et al. (2021) support the feasibility of CI as a new perspective in effectively treating patients with hearing loss who also suffer from tinnitus. It has been shown that CI have a positive impact for pediatric patients on spoken language competence (Nicholas & Geers, 2006). Several studies suggest CI procedures to be safe and cost-effective (Bond et al., 2009; NICE). In our study we cannot differentiate between unilateral, bilateral, sequential or simultaneous treatment of CI. However, as it is frequently discussed and evaluated in the literature, it is considered that every treatment option of CI is cost-effective comparing to bilateral hearing aids (Foteff et al., 2016; NICE). Furthermore, in our data we could not differentiate between primary procedures and revision surgeries, nevertheless it is important to mention here, that revisions of other than natural complication risk, such as device failure (Wang et al., 2014) could bear a risk of unwarranted variation.

TABLE B1 Procedure codes for each case study in the four countries

Diagnosis	Procedures	Germany (OPS-code)	Italy (ICD-9-CM codes)	Netherlands (ZA codes)
Case 1 & 3 femoral neck fracture	Hip implantations - total & partial	5-820.0; 5-820.2; 5-820.3-5; 5-820.7-9	81.51; 81.52; 00.85; 00.86; 00.87	038,567; 038,565; 038,562
	Femur reposition	5-791g; 5-792g; 5-793e-h; 5-794e-h; OPS 5-790e-h	79.55; 79.35; 79.25; 78.15; 79.05; 79.15; 79.45 Without 81.51, 81.52, 00.85, 00.86, 00.87	038,533; 038,528; 038,530

TABLE B1 (Continued)

Diagnosis	Procedures	Germany (OPS-code)	Italy (ICD-9-CM codes)	Netherlands (ZA codes)
Case 2 & 4 acute myocardial infarction (ST-elevation)	PTCA	8–837.1; 8–837.2; 8–837.5; 8–837.6; 8–837.m; 8–837.t	00.66; 36.04	033,232; 88,370; 88,374; 033,231; 033,238; 080,827; 88,373; 39,689; 80,829
	Drug eluting stent	8–837.0	36.07	N/A
	Bare metal stents	8–837.k	36.06	N/A
Case 5 Knee arthrosis	Implantations	5–822	81.54	38,663
Case 6 spinal stenosis	Decompression without fusion	5–839.6	3.09 without 81.00–81.09; 81.30–81.39	30,326; 30,327; 30,328; 30,329; 38,437; 38,438; 38,440; 38,447; 38,448; 38,449; 038,467
	Decompression with fusion	5–83b.0–7; 5–83.9	3.09 with 81.00–81.09; 81.30–81.39	38,458; 38,459; 38,460; 38,462; 38,463; 38,464; 38,467; 38,468; 38,469; 38,471
Case 7 Malign neoplasm of the prostate	Laparoscopic	5–604.4; 5.604.5	60.5 with 54.21	36,556
	Non-laparoscopic	5–604.0–3	60.5 without 54.21	36,553
Case 8 Benign neoplasms of uterus	Laparoscopic	5–682.01-02; 5–682.11- 12; 5–682.x1-x2; 5–683.02-03; 5–683.12-13; 5–683.22-23; 5–683. x2-x3	68.31; 68.41; 68.51	37,113
	Non-laparoscopic	5–682.00; 5–682.03- 0x; 5–682.10-11; 5–682. 13-1x; 5–682. x0; 5–682.x3-x5; 5–682.y; 5–683.00- 01; 5–683.04-0x; 5–683.10-11; 5–683.14-1x; 5–683.20-21; 5–683.24-2x; 5–683.5- 7; 5–683.x0-x1; 5–683. x4-y	68.31; 68.41; 68.51; 68.61	037,100; 037,111; 037,131
Case 9 Hearing impairment	Implantation or replacement of a cochlear implant	5–209.2; 5–209.7	20.96; 20.97; 20.98	031,903; 031,905; 031,907; 031,908

Note: For the Netherlands, besides Zorg Activiteit(ZA), also Centraal Beheer Verrichtingenbestand (CBV) and the Classificatie van Verrichtingen Classificatie (CVV) codes were used.

Abbreviations: CM, clinical modification; OPS, Operationen und Prozedurenschlüssel; ICD, International Classification of Diseases; PTCA, percutaneous translaminar coronary angioplasty.

Descriptive statistics for the 10 case studies

TABLE B2 Descriptive statistics for the dependent variable of the ten case-studies

Case	Germany				Italy				The Netherlands			
	Mean	N	Hospitals	NUTS 3	Mean	N	Hospitals	NUTS 3	Mean	N	Hospitals	NUTS 3
1. Femur effective	0.769	407,284	1489	395	0.752	584,548	998	110	0.762	35,168	90	39
2. PTCA	0.641	423,177	1425	395	0.446	251,372	813	110	0.500	50,551	90	39
3. Femur treatment	0.818	313,183	1120	393	0.377	440,506	688	110	0.705	26,793	90	39
4. Stent	0.846	271,642	748	357	0.772	163,313	367	105	N/A	N/A	N/A	N/A
5. Knee implant	0.502	1,454,198	1477	395	0.740	423,591	943	110	0.784	77,621	90	39
6. Spinal stenosis	0.286	375,886	814	364	0.404	54,845	392	98	0.055	14,383	78	35
7. Prostatectomy	0.375	110,582	434	283	0.329	78,371	501	107	N/A	N/A	N/A	N/A
8. Hysterectomy	0.496	308,567	910	388	0.298	117,563	709	110	0.475	10,656	90	39
9. Cochlea implant	0.139	130,215	961	387	0.118	40,674	369	107	0.232	5433	89	39

Note: All years (Germany/Italy: 2012–2016; Netherlands: 2014–2016) are included. All dependent variables are coded as a dummy variable (0, 1); Hospitals and NUTS3 reflect the number entities in which at least one of the cases occurs. Case study 4 and 6 could not be operationalized in the Netherlands due to unavailability of procedure codes.

Abbreviations: N, number of observations; NUTS, Nomenclature des Unités Territoriales Statistiques; PTCA, percutaneous translaminar coronary angioplasty.

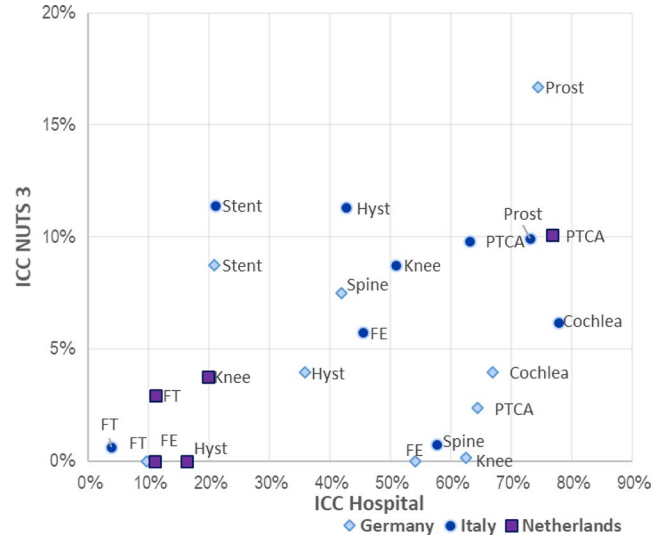
APPENDIX C: RESULTS OF THE SENSITIVITY ANALYSIS

TABLE C1 Regression results of all available case-studies for Germany, Italy and the Netherlands excluding potential outlier

	Category I: Effective Care						Category II: Preference sensitive care						Category III: Preference sensitive care/ Supply sensitive care																	
	1. Femur Effective		2. PTCA		3. Femur Treatment		4. Stent		5. Knee Prosthesis		6. Spinal Stenosis		7. Prostatectomy		8. Hysterectomy		9. Cochlea Implant													
	GER	IT	NL	GER	IT	NL	GER	IT	GER	IT	NL	GER	IT	GER	IT	NL	GER	IT												
Hospital																														
Efficiency in delivery of healthcare																														
No. of relevant cases †	-	-	-	+	-	+	-	-	+	-	-	+	-	+	-	+	-	-												
Avg. LOS relevant cases †	+	-	+	+	-	-	+	+	-	+	-	+	+	-	-	-	-	-												
Ownership private [1=private]	+	-	N/A	+	+	N/A	-	+	N/A	+	+	N/A	+	+	-	-	N/A	+												
Ownership not-for profit [1=not-for profit]	-	N/A	N/A	+	N/A	N/A	+	N/A	N/A	-	N/A	+	N/A	N/A	+	N/A	+	N/A												
Inpatient cases per bed †	+	+	+	+	+	-	-	-	+	+	+	-	+	+	+	+	+	+												
University hospital [1=university]	+	-	-	+	+	+	-	+	+	-	-	+	+	+	+	+	+	+												
Capacity constraints																														
Hospital with less than 50 cases [1=<50 relevant cases]	+	-	-	-	-	-	+	+	-	-	+	-	+	-	-	-	+	-												
Beds † [Number]	+	+	+	+	+	-	-	-	+	+	+	+	+	-	-	-	+	-												
NUTS3																														
Patient access																														
General practitioner † [per 100,000 inhab.]	-	+	+	+	+	-	-	+	+	-	-	-	+	-	+	+	+	+												
Hospital beds † [per 1,000 inhab.]	-	-	-	+	+	-	-	+	+	-	+	-	+	+	-	-	-	+												
Population density † [per km²]	+	+	+	+	+	+	-	-	+	+	+	+	+	+	-	-	-	-												
Share of non-citizen residents † [% of total population]	-	-	+	+	+	+	+	+	-	-	+	+	-	+	+	-	+	+												
Patient "optimism"																														
Students in up. sec. education † [% of total students]	-	-	+	+	+	+	+	+	-	-	-	+	+	+	-	-	+	-												
Without lower sec. education † [% of school-leavers]	-	+	-	+	+	-	-	-	-	+	+	-	-	-	-	-	+	+												
Voter turnout national election † [% of total]	-	+	+	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+												
Avg. household income per inhab. † [PPP adjusted €]	+	-	-	+	-	-	-	+	+	+	+	+	+	+	+	+	+	-												
Differences in health status																														
Avg. life expectancy of a newborn † [years]	-	-	+	-	+	+	-	+	-	-	+	+	+	-	-	+	+	+												
Unemployment † [%total employment]	-	-	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+	-												
Population aged 65+ † [% of total]	-	+	+	+	-	+	-	+	+	+	-	+	+	+	+	+	+	+												
Number of observations (N)	392,583	562,285	34,860	398,859	247,745	50,531	304,034	432,532	26,557	255,548	163,313	1,355,207	390,708	76,542	352,073	51,686	101,964	76,603	289,063	110,573	10,438	119,165	40,674							
ICC Hospital	54.2%	45.4%	11.1%	64.9%	63.0%	76.8%	9.6%	3.7%	11.1%	21.0%	21.1%	62.5%	50.9%	19.9%	42.3%	57.6%	75.9%	73.0%	35.8%	42.6%	16.3%	66.6%	77.8%							
ICC Regional/NUTS3	0.0%	5.7%	0.0%	1.5%	9.8%	10.1%	0.1%	0.6%	2.9%	8.4%	11.4%	0.1%	8.7%	3.8%	6.9%	0.7%	15.0%	9.9%	4.3%	11.3%	0.0%	4.5%	6.2%							
	OR>1						**p<0.001						**p<0.01						*p<0.05											
							OR<1						**p<0.001						**p<0.01						*p<0.05					

Note: Table 3 shows the results of the three-level model regression for Germany, Italy and the Netherlands. Regressions include year fixed effects for all available years (Germany & Italy: 2012–2016, Netherlands: 2014–2016) but exclude potential outlier. Patient level variables (age, gender, Elixhauser groups, main diagnose dummy) are omitted in this table and used for risk adjustment; † = mean centered variables; Italy: Ownership Not-for-profit not applicable, Netherlands: Ownership Not-for profit and privat not applicable. Case 4, 6, 9 is unavailable in the Netherlands due to low sample size or unavailable procedure codes.

FIGURE C1 Intraclass correlation coefficient (ICC) comparison on hospital and NUTS3 level excluding outlier. Figure shows the ICC on Hospital and NUTS3 level calculated by the three level logistic regression excluding predefined outlier. FE, Femur effective; FT, femur treatment; Hyst, Laparoscopic Hysterectomy; ICC, intraclass correlation coefficient; Knee, knee Implantation; NUTS, Nomenclature des unités territoriales statistiques; Prost, laparoscopic prostatectomy; PTCA, percutaneous transluminal coronary angioplasty; Spine, decompression with fusion



APPENDIX D: PATIENT LEVEL CHARACTERISTICS OF TABLE 3 (OMITTED IN MAIN TEXT TABLE)

Patient	Category I: Effective Care						Category II: Preference sensitive care						Category III/IV: Preference sensitive care/ Supply sensitive care												
	1. Femur Effective			2. PTCA			3. Femur Treatment			4. Stent			5. Knee Prosthesis			6. Spinal Stenosis		7. Prostatectomy			8. Hysterectomy			9. Cochlea Implant	
	GER	IT	NL	GER	IT	NL	GER	IT	NL	GER	IT	NL	GER	IT	NL	GER	IT	GER	IT	NL	GER	IT	GER	IT	
Intercept	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
Age*[Year]	+	-	-	-	+	-	+	+	+	-	+	+	+	+	+	+	+	-	-	-	-	-	-	-	
Sex(male=1)	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	N/A	N/A	N/A	N/A	N/A	N/A	N/A	
Main Diagnosis(Main diagnosis=1)	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
Elixhauser Group																									
01. Congestive Heart Failure	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
02. Cardiac Arrhythmia	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
03. Valvular Disease	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
04. Pulmonary Circulation Disorders	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
05. Peripheral Vascular Disorders	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
06. Hypertension Uncomplicated	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
07. Hypertension complicated	+	N/A ¹	-	+	N/A ¹	-	+	N/A ¹	-	+	N/A ¹	-	+	N/A ¹	-	+	N/A ¹	+	N/A ¹	+	+	+	+	N/A ¹	
08. Paralysis	+	-	-	-	-	-	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
09. Other Neurological Disorders	+	-	-	-	-	-	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
10. Chronic Pulmonary Disease	+	-	-	-	-	-	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
11. Diabetes Uncomplicated	+	-	-	+	+	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
12. Diabetes Complicated	+	-	-	+	+	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
13. Hypothyroidism	+	-	-	+	+	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
14. Renal Failure	+	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
15. Liver Disease	-	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
16. Peptic Ulcer Disease excl. bleeding	-	+	-	-	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
17. AIDS/HIV	+	+	-	+	+	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
18. Lymphoma	+	-	-	+	+	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
19. Metastatic Cancer	+	-	-	+	+	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
20. Solid Tumor without Metastasis	+	-	-	+	+	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
21. Rheumatoid Arthritis/collagen	+	+	-	+	+	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
22. Coagulopathy	+	+	-	+	+	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
23. Obesity	+	-	-	+	+	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
24. Weight Loss	-	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
25. Fluid and Electrolyte Disorders	+	-	-	+	+	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
26. Blood Loss Anemia	-	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
27. Deficiency Anemia	+	-	-	+	+	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
28. Alcohol Abuse	+	-	-	+	+	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
29. Drug Abuse	+	-	-	+	+	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
30. Psychoses	+	-	-	+	+	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	
31. Depression	-	-	-	-	-	-	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	

Note: N/A = not applicable for the case (e.g., because there are only women in sample); N/A¹: not coded for this case study.

REFERENCES

Aarts, J. W. M., Nieboer, T. E., Johnson, N., Tavender, E., Garry, R., Mol, B. W. J., & Kluijvers, K. B. (2015). Surgical approach to hysterectomy for benign gynaecological disease. *The Cochrane Database of Systematic Reviews*, (8), CD003677. <https://doi.org/10.1002/14651858.CD003677.pub5>

Bond, M., Mealing, S., Anderson, R., Elston, J., Weiner, G., Taylor, R. S., ... Stein, K. (2009). The effectiveness and cost-effectiveness of cochlear implants for severe to profound deafness in children and adults: A systematic review and economic model. *Health Technology Assessment (Winchester, England)*, 13(44), 1–330. <https://doi.org/10.3310/hta13440>

Borges, A. L. d. F., Duarte, P. L. E. S., Almeida, R. B. S. de, Ledesma, A. L. L., Azevedo, Y. J. de, Pereira, L. V., & Bahmad, F. (2021). Cochlear implant and tinnitus-a meta-analysis. *Brazilian Journal of Otorhinolaryngology*, 87(3), 353–365. <https://doi.org/10.1016/j.bjorl.2020.11.006>

Bray, F. [F.], Lortet-Tieulent, J., Ferlay, J. [J.], Forman, D. [D.], & Auvinen, A. (2010). Prostate cancer incidence and mortality trends in 37 European countries: An overview. *European Journal of Cancer (Oxford, England: 1990)*, 46(17), 3040–3052. <https://doi.org/10.1016/j.ejca.2010.09.013>

Brownlee, S., Wennberg, J. E., Barry, M. J., Fisher, E. S., Bynum, J. P., & Goodman, D. C. (2012). Improving patient decision-making in health care: A 2012 Dartmouth Atlas report highlighting the Middle Atlantic Region.

- Cooperberg, M. R., Broering, J. M., & Carroll, P. R. (2010). Time trends and local variation in primary treatment of localized prostate cancer. *Journal of Clinical Oncology: Official Journal of the American Society of Clinical Oncology*, 28(7), 1117–1123. <https://doi.org/10.1200/JCO.2009.26.0133>
- Deyo, R. A., Martin, B. I., Ching, A., Tosteson, A. N. A., Jarvik, J. G., Kreuter, W., & Mirza, S. K. (2013). Interspinous spacers compared with decompression or fusion for lumbar stenosis: Complications and repeat operations in the Medicare population. *Spine*, 38(10), 865–872. <https://doi.org/10.1097/BRS.0b013e31828631b8>
- Ferlay, J. [Jacques], Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., ... Bray, F. [Freddie] (2015). Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. *International Journal of Cancer*, 136(5), E359–86. <https://doi.org/10.1002/ijc.29210>
- Ficarra, V., Novara, G., Artibani, W., Cestari, A., Galfano, A., Graefen, M., ... van Poppel, H. (2009). Retropubic, laparoscopic, and robot-assisted radical prostatectomy: A systematic review and cumulative analysis of comparative studies. *European Urology*, 55(5), 1037–1063. <https://doi.org/10.1016/j.eururo.2009.01.036>
- Foteff, C., Kennedy, S., Milton, A. H., Deger, M., Payk, F., & Sanderson, G. (2016). Cost–utility analysis of cochlear implantation in Australian Adults. *Otology & Neurotology*, 37, 454–461.
- Hawker, G. A., Wright, J. G., Coyte, P. C., Williams, J. I., Harvey, B., Glazier, R., ... Badley, E. M. (2001). Determining the need for hip and knee arthroplasty: The role of clinical severity and patients' preferences. *Medical Care*, 39(3), 206–216. <https://doi.org/10.1097/00005650-200103000-00002>
- Hopley, C., Stengel, D., Ekkernkamp, A., & Wich, M. (2010). Primary total hip arthroplasty versus hemiarthroplasty for displaced intracapsular hip fractures in older patients: Systematic review. *BMJ (Clinical Research Ed.)*, 340, c2332. <https://doi.org/10.1136/bmj.c2332>
- Ibanez, B., James, S., Agewall, S., Antunes, M. J., Bucciarelli-Ducci, C., Bueno, H., ... Widimský, P. (2018). 2017 ESC Guidelines for the management of acute myocardial infarction in patients presenting with ST-segment elevation: The Task Force for the management of acute myocardial infarction in patients presenting with ST-segment elevation of the European Society of Cardiology (ESC). *European Heart Journal*, 39(2), 119–177. <https://doi.org/10.1093/eurheartj/ehx393>
- Johnell, O., & Kanis, J. A. (2004). An estimate of the worldwide prevalence, mortality and disability associated with hip fracture. *Osteoporosis International: A Journal Established as Result of Cooperation Between the European Foundation for Osteoporosis and the National Osteoporosis Foundation of the USA*, 15(11), 897–902. <https://doi.org/10.1007/s00198-004-1627-0>
- Kalla, K., Christ, G., Karnik, R., Malzer, R., Norman, G., Prachar, H., ... Huber, K. (2006). Implementation of guidelines improves the standard of care: The Viennese registry on reperfusion strategies in ST-elevation myocardial infarction (Vienna STEMI registry). *Circulation*, 113(20), 2398–2405. <https://doi.org/10.1161/CIRCULATIONAHA.105.586198>
- Kreiner, D. S., Shaffer, W. O., Baisden, J. L., Gilbert, T. J., Summers, J. T., Toton, J. F., ... Reitman, C. A. (2013). An evidence-based clinical guideline for the diagnosis and treatment of degenerative lumbar spinal stenosis (update). *The Spine Journal: Official Journal of the North American Spine Society*, 13(7), 734–743. <https://doi.org/10.1016/j.spinee.2012.11.059>
- Naples, J. G., & Ruckenstein, M. J. (2020). Cochlear implant. *Otolaryngologic Clinics of North America*, 53(1), 87–102. <https://doi.org/10.1016/j.otc.2019.09.004>
- National Institute for Health and Care Excellence. Cochlear implants for children and adults with severe to profound deafness.
- National Institute for Health and Care Excellence (2014). Osteoarthritis: Care and management in adults.
- Neumann, F.J., Sousa-Uva, M., Ahlsson, A., Alfonso, F., Banning, A. P., Benedetto, U., ... Zembala, M. O. (2019). 2018 ESC/EACTS Guidelines on myocardial revascularization. *European Heart Journal*, 40(2), 87–165. <https://doi.org/10.1093/eurheartj/ehy394>
- Nicholas, J. G., & Geers, A. E. (2006). Effects of early auditory experience on the spoken language of deaf children at 3 years of age. *Ear and Hearing*, 27(3), 286–298. <https://doi.org/10.1097/01.aud.0000215973.76912.c6>
- OECD. (2014). *Geographic Variations in Health Care: What do we know and what can be done to improve health system performance* (OECD Health Policy Studies). Retrieved from <http://dx.doi.org/10.1787/9789264216594-en>
- Price, A. J., Alvand, A., Troelsen, A., Katz, J. N., Hooper, G., Gray, A., ... Beard, D. (2018). Knee replacement. *The Lancet*, 392(10158), 1672–1682. [https://doi.org/10.1016/S0140-6736\(18\)32344-4](https://doi.org/10.1016/S0140-6736(18)32344-4)
- Scottish Intercollegiate Guidelines Network. (2009). *Management of hip fracture in older people: A national clinical guideline*.
- Sendtner, E., Renkawitz, T., Kramny, P., Wenzl, M., & Grifka, J. (2010). Fractured neck of femur--internal fixation versus arthroplasty. *Deutsches Arzteblatt International*, 107(23), 401–407. <https://doi.org/10.3238/arztebl.2010.0401>
- Shen, J., Xu, S. [Shen], Xu, S. [Shenxi], Ye, S., & Hao, J. (2018). Fusion or Not for Degenerative Lumbar Spinal Stenosis: A Meta-Analysis and Systematic Review. *Pain Physician*, 21(1), 1–8.
- Skarzynski, H., Lorens, A., Piotrowska, A., & Anderson, I. (2006). Partial deafness cochlear implantation provides benefit to a new population of individuals with hear loss. *Acta Oto-Laryngologica*, 126(9), 934–940. <https://doi.org/10.1080/00016480600606632>
- Steg, P. G., James, S. K., Atar, D., Badano, L. P., Blömstrom-Lundqvist, C., Borger, M. A., ... Zahger, D. (2012). Esc Guidelines for the management of acute myocardial infarction in patients presenting with ST-segment elevation. *European Heart Journal*, 33(20), 2569–2619. <https://doi.org/10.1093/eurheartj/ehs215>
- World Health Organization. (2021). *World report on hearing*. Geneva.
- Wright, J. D., Herzog, T. J., Tsui, J., Ananth, C. V., Lewin, S. N., Lu, Y.S., ... Hershman, D. L. (2013). Nationwide trends in the performance of inpatient hysterectomy in the United States. *Obstetrics and Gynecology*, 122(2 Pt 1), 233–241. <https://doi.org/10.1097/AOG.0b013e318299a6cf>