PARAMETER HETEROGENEITY IN BREAST CANCER COST REGRESSIONS – EVIDENCE FROM FIVE EUROPEAN COUNTRIES

JOEL SMITH[a,*], HELEN BANKS[b], HARRY CAMPBELL[c], ANNE DOUGLAS[c], EILIDH FLETCHER[d],
ALISON MCCALLUM[e], TRON ANDERS MOGER[f], MIKKO PELTOLA[g], SOFIA SVERÉUS[h], SARAH WILD[c],
LINDA J. WILLIAMS[c], JOHN FORBES[i] ON BEHALF OF THE EUROHOPE STUDY GROUP

[a]*Health Economics Research Centre, Nuffield Department of Population Health, University of Oxford, Oxford, UK*
[b]*Centre for Research on Health and Social Care Management, CERGAS, Bocconi University*
[c]*Usher Institute of Population Health Sciences and Informatics, The University of Edinburgh, Edinburgh, UK*
[d]*Lothian Analytical Services, NHS Lothian, Edinburgh, UK*
[e]*NHS Lothian, Edinburgh, UK*
[f]*Department of Health Management and Health Economics, Institute of Health and Society, University of Oslo, Oslo, Norway*
[g]*Centre for Health and Social Economics (CHESS), National Institute for Health and Welfare, Helsinki, Finland*
[h]*Medical Management Centre, Karolinska Institutet, Stockholm, Sweden*
[i]*Health Research Institute, University of Limerick, Limerick, Ireland*

ABSTRACT

We investigate parameter heterogeneity in breast cancer 1-year cumulative hospital costs across five European countries as part of the EuroHOPE project. The paper aims to explore whether conditional mean effects provide a suitable representation of the national variation in hospital costs. A cohort of patients with a primary diagnosis of invasive breast cancer (ICD-9 codes 174 and ICD-10 C50 codes) is derived using routinely collected individual breast cancer data from Finland, the metropolitan area of Turin (Italy), Norway, Scotland and Sweden. Conditional mean effects are estimated by ordinary least squares for each country, and quantile regressions are used to explore heterogeneity across the conditional quantile distribution. Point estimates based on conditional mean effects provide a good approximation of treatment response for some key demographic and diagnostic specific variables (e.g. age and ICD-10 diagnosis) across the conditional quantile distribution. For many policy variables of interest, however, there is considerable evidence of parameter heterogeneity that is concealed if decisions are based solely on conditional mean results. The use of quantile regression methods reinforce the need to consider beyond an average effect given the greater recognition that breast cancer is a complex disease reflecting patient heterogeneity. © 2015 The Authors. *Health Economics* Published by John Wiley & Sons Ltd.

1. INTRODUCTION

There has been a persistent increase in the rate of incidence of breast cancer across many European countries over the last 30 years (Bray *et al.*, 2004; Ferlay *et al.*, 2013). Although national breast cancer screening initiatives temporarily inflated rates by diagnosing a cohort of patients previously undetected, there remained a general trend towards an average annual increase of 1–2% across many European countries for much of this period (Berrino *et al.*, 2007; Hery *et al.*, 2008a; Hery *et al.*, 2008b; Westlake and Cooper, 2008). A primary concern for clinical decision making and resource allocation decisions relates to the greater demand on hospital services. Health service and system responses have focused upon the priority setting of effective healthcare

treatments and technologies. Productivity and efficiency in which services are delivered remain a significant policy concern whilst striving to maximise population health outcomes. One consequence of the renewed policy focus on productivity and efficiency gains is that the average length of stay for post-operative breast cancer has been reduced in some European countries (Marla *et al.*, 2013). It is unclear whether there are any changes in breast cancer resource use because of the greater demand from increasing incidence or whether this is offset by shorter length of stay.

There is greater recognition that breast cancer is a complex disease, and this has had a direct impact on the changing nature of treatments. Whole-genome sequencing and other technological advancements in breast cancer screening have changed the clinical pathway for many patients (Polyak, 2011). The treatment pathway for a representative, or average, patient has been displaced by clinical evidence, which explicitly accommodates the heterogeneity among breast cancer patients, prognoses and outcomes. Analysis of breast cancer resource use should reflect the individualised approach to treatment in order to offer more pertinent health policy insights.

The EuroHOPE project aims to investigate healthcare outcomes, performance and efficiency across seven European countries for five health conditions. Using routine individual level breast cancer data from Finland, the metropolitan area of Turin (Italy), Norway, Scotland and Sweden, this paper is concerned with parameter heterogeneity in the cumulative hospital costs distribution. The key aim will be to explore whether conditional mean effects provide a suitable representation of the national variation in hospital costs whilst confining the analysis to variables that are routinely recorded in administrative hospital databases.

2. ECONOMETRIC SPECIFICATION

One-year cumulative hospital costs are used as a surrogate measure of breast cancer patient resource use. Econometric analysis of cost data has been extensively applied in empirical research, and the methodological challenges in terms of estimation are, therefore, well documented (for example, Mullahy, 1998; Manning and Mullahy, 2001; Nixon and Thompson, 2004; Jones, Lomas and Rice, 2014). Much of this literature is concerned with the analysis of costs within trial-based economic evaluations. Within this setting, interest is confined to inferences about the population mean cost in the presence of skewed cost distributions. The econometric analysis of cost within this paper differs in two important regards. Firstly, data sharing restrictions among some of the EuroHOPE partner countries ruled out the pooling of data and confined analysis to remote access. Secondly, there is uncertainty regarding whether conditional mean effects are the most salient message given policy interest is often focused upon the tails of the distribution in terms of extreme costs or outliers. To address both of these points, a flexible approach to the econometric analysis of cost data is adopted in which no prior assumption is made regarding location scale and shape effects for each EuroHOPE country.

Ordinary least squares (OLS) regression models are used to estimate conditional mean effects for the 1-year cumulative hospital costs for each country. As cost data are commonly characterised by a long right-hand tail, a natural log transformation was undertaken for all cost data prior to analysis to address this distributional feature. Our linear regression model for the conditional mean is

$$Ln(Cost_i) = \alpha + \beta_1 d_i + \beta_2 p_i + u_i$$

where $Cost_i$ is the 1-year cumulative hospital cost of patient i. d_i is a vector of demographic and diagnosis-specific variables, including patient age, ICD-10 diagnosis and breast cancer severity as measured by stage. p_i is a vector of process variables, such as number of hospital procedures performed, length of stay in the previous year of diagnosis, whether the patient was treated in a university/teaching hospital and a dummy variable for patient mortality. No country-specific dummy variables are included as each regression is run separately for each country due to data sharing restrictions.

By extension, the quantile regression model for the conditional quantiles is

$$Q_{Cost_i}(\tau|x_i) = \alpha(\tau) + \beta_1 x_i(\tau) + u_i(\tau)$$

where x is a vector including demographic and diagnosis specific, d_i, and process variables, p_i. The quantile of interest is estimated for each $\tau \in (0, 1)$ (Koenker and Hallock, 2001; Koenker, 2005).

3. DATA

We use routinely collected individual breast cancer data from five European countries to derive a cohort of patients using a predefined methodology across EuroHOPE partners (Douglas, 2012). The data source registries include patient records on every breast cancer diagnosis. We focus on those patients with a primary diagnosis of invasive breast cancer (ICD-9 codes 174 and ICD-10 C50 codes) in the financial year 2005, with the exception of Norway which uses 2009 data.[1] Each individual record contains data on both demographic (age and gender) and clinical variables (stage and diagnosis and number of procedures performed) as well as linked data on administrative variables of interest (e.g. hospital type). Cumulative hospital costs cover the period from admission to 1 year follow-up using linked hospital discharge data as well as death and cancer registers. Only Finland, Norway and Sweden were able to include prescription data in the calculation of cumulative hospital costs. Using the predefined methodology, each country excluded patients that were either male or under 25 years of age at presentation and had recurrence of breast cancer of same histology/laterality that was first diagnosed before the start of data collection. In addition, only patients with complete hospital identification numbers were included in the study.

Aas (2012) outlines the episode-based costing methodology adopted by each EuroHOPE country. The first hospital episode for breast cancer is defined as the first hospital inpatient record following a diagnosis of breast cancer using the relevant ICD-9/ICD-10 codes. Cumulative hospital costs are defined as any patient breast cancer resource use as an inpatient, day patient or outpatient up to 1 year after the recorded date of diagnosis of breast cancer. This includes surgical and non-surgical treatments, such as chemotherapy, radiotherapy and hormone therapy, with the exception of Norway, which were unable to include hormone therapy. Hospital inpatient, including day surgery, is classified by diagnosis-related groups for all countries with the exception of Scotland, which is based on healthcare resource groups for all inpatient activity. Fee-based mechanisms are assigned to all outpatient visits and prescribing data, where available. Costs for all countries have been converted to 2011 UK pound sterling (£s)[2] using Organisation for Economic Co-operation and Development (OECD) purchasing power parities (OECD, 2012). Although Hungary and the Netherlands are also participating in the wider EuroHOPE project, neither country supplied data for the breast cancer cost analysis reported in this paper. The Netherlands did not contribute data for the breast cancer analysis as it was not possible to link patient records from the National cancer register to hospital discharge records. Hungary was only able to provide data on length of stay due to difficulties in record linkage to discharge registers needed to calculate hospital costs. The sample for the cost analysis is, therefore, confined to Finland, the metropolitan area of Turin (Italy), Norway, Scotland and Sweden.

One-year cumulative log hospital costs represent the dependent variable in all regression models. Independent variables include patient characteristics such as breast cancer stage, ICD-10 diagnosis, mortality, hospital length of stay in the year prior to the breast cancer diagnosis, age in years and a quadratic function for age to accommodate non-linear relationships. Dummy variables are used to control for differing ICD-10 diagnoses as well as a separate variable to capture patient mortality using survival as the reference category (e.g. survival = 0; died = 1 based on the date of death with follow-up up to December 2010 for all countries except Finland who

---

[1]Data from 2009 for Norway was used as it was not possible to link data on resource use to patient records before 2008 (the year prior to the index admission of 2009).

[2]The interpretation and magnitude of the coefficients reported in the regression analyses remain the same for other currencies, such as the euro (€), with the exception of the intercept which will be sensitive to the currency selected.

had follow-up up to December 2011). Breast cancer stage is used as a measure of severity using stage 1 as the reference category. The number of days spent in hospital in the previous year prior to diagnosis is used as a proxy for comorbidities in addition to the mortality dummy variable. We also control for the number of procedures performed to account for the possibility of variation in treatment pathways depending on hospital type and locality. Hospital type is included as a dummy variable to estimate whether there are differences between university teaching hospitals/specialist breast cancer centres relative to other hospitals in terms of length of stay and costs.

4. RESULTS

Table I presents the summary statistics for the key variables. A comparison of the mean and standard deviation for the length of stay variable illustrates the variation in the patterns of breast cancer care across countries. Despite the use of a predefined methodology for data collection and analysis for the EuroHOPE project, Table I

Table I. Descriptive statistics for the key variables of interest for Finland, Italy, Norway, Scotland and Sweden

	Mean	Std. Dev.	Min.	Max.	N
Finland					
Length of stay (LoS)	6.1	19.8	1	365	3200
Year 1 costs (UK £)	9788	6827	341	136996	3200
Age (years)	61.9	13.1	27	98	3943
LoS_{t-1}	0.38	3.3	0	118	3943
Mortality dummy	0.23	0.42	0	1	3943
No. of procedures	1.2	0.9	0	4	3943
Hospital dummy	0.30	0.46	0	1	3943
Italy					
Length of stay (LoS)	4.7	7.3	1	153	680
Year 1 costs (UK £)	10348	17934	0	311036	778
Age (years)	63.5	14.1	29	94	778
LoS_{t-1}	0.10	0.8	0	12	778
Mortality dummy	0.21	0.41	0	1	779
No. of procedures	0.9	0.5	0	3	779
Norway					
Length of stay (LoS)	2.9	4.6	1	130	2804
Year 1 costs (UK £)	34293	22747	0	255546	2790
Age (years)	62.2	14.0	26	98	2816
LoS_{t-1}	2.92	9.2	0	185	2816
Mortality dummy	0.07	0.26	0	1	2816
No. of procedures	1.3	0.8	0	4	2816
Scotland					
Length of stay (LoS)	6.3	14.0	1	365	3427
Year 1 costs (UK £)	15822	13183	886	137660	3542
Age (years)	63.1	14.5	25	102	3963
LoS_{t-1}	1.83	10.6	0	243	3963
Mortality dummy	0.33	0.47	0	1	3963
No. of procedures	0.9	0.5	0	3	3963
Hospital dummy	0.52	0.50	0	1	3427
Sweden					
Length of stay (LoS)	4.0	6.1	1	272	5896
Year 1 costs (UK £)	6917	4640	204	76507	6782
Age (years)	63.4	13.9	25	101	7164
LoS_{t-1}	0.19	1.4	0	48	7164
Mortality dummy	0.22	0.41	0	1	7164
No. of procedures	0.9	0.5	0	3	7164

LoS = first hospital episode; LoS_{t-1} = length of stay in the previous year; Mortality = patient mortality dummy variable; No. of procedures = number of procedures performed; Hospital dummy = dummy variable for whether the patient was treated in a university teaching hospital or specialist breast cancer centre.

illustrates that some variables do not appear to be directly comparable across countries. The magnitude of some of the differences across countries may simply be a manifestation of registry data access or quality and, therefore, observed differences may be spurious. Notable examples outlined in Table I include the mortality dummy variable for Norway and the costs estimates across all countries, which include a mixture of diagnosis-related groups (Finland, Italy, Norway and Sweden) and healthcare resource group (Scotland) episode-based costing methodologies. To minimise spurious conclusions, interpretation of all results presented should focus on national parameter heterogeneity rather than cross-country comparisons.

The distribution of 1-year hospital costs by country is presented in Figure 1. As illustrated, the cost distributions across countries are skewed with a long right-hand tail.

Table II presents the estimated coefficients from the OLS regression model for the 1-year cumulative hospital costs for each country. The national conditional mean effects for Finland are reported in column (1). Cost estimates for patients diagnosed using ICD-10 code C508 (overlapping lesion of breast) are 11% lower, on average, relative to the reference category of ICD-10 code C500 (nipple and areola). For patients classified with stage 2 breast cancer, estimated costs are 25% higher relative to stage 1 patients. This rises to a 38% and 31% increase in costs for stage 3 and 4 patients, respectively, compared with stage 1. The proxy for comorbidities in Finland results in a 3% increase in costs for each additional day spent in hospital in the previous year. Patient mortality is associated with an 11% increase in 1-year hospital costs relative to patients that survived.

Figures 2–6 present the conditional mean and conditional quantile results for each country. The quantile regression results are used to explore evidence of parameter heterogeneity around the conditional mean effect reported in Table II. Each figure is constructed with the conditional mean effect represented by the thick black
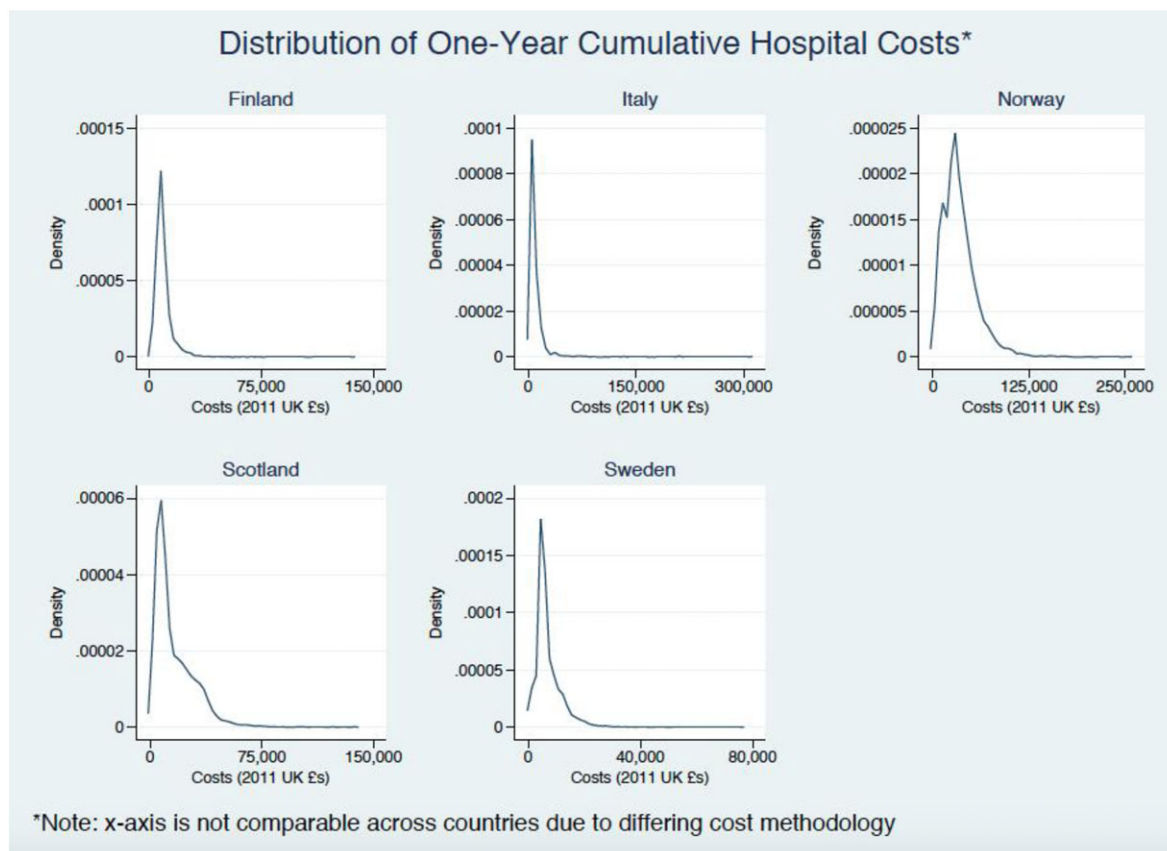

Figure 1. Kernel density of the 1-year cumulative hospital costs for Finland, Italy, Norway, Scotland and Sweden

Table II. OLS regression coefficients (conditional mean) for 1-year cumulative hospital costs for Finland, Italy, Norway, Scotland and Sweden

	(1) Finland N = 2600		(2) Italy N = 219		(3) Norway N = 2427		(4) Scotland N = 3080		(5) Sweden N = 4708	
	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE	Coef.	SE
Age	0.006	0.006	0.076*	0.029	0.033*	0.007	−0.016**	0.008	0.024*	0.005
Age^2	−0.0001**	0.00005	−0.0007*	0.0002	−0.0004*	0.00006	0.00003	0.00006	−0.0003*	0.00004
C501	−0.065	0.054	1.671*	0.479	—	—	0.031	0.130	0.001	0.132
C502	−0.049	0.051	0.085	0.233	—	—	−0.191	0.122	−0.056	0.131
C503	−0.022	0.057	0.227	0.212	—	—	−0.065	0.130	0.008	0.139
C504	−0.069	0.045	0.246	0.292	—	—	−0.162	0.110	−0.030	0.127
C505	−0.024	0.056	0.336***	0.174	—	—	−0.190	0.127	−0.067	0.132
C506	−0.082	0.170	0.192	0.248	—	—	0.086	0.194	−0.017	0.368
C508	−0.110***	0.060	−0.091	0.659	—	—	−0.148	0.111	−0.023	0.130
C509	−0.073	0.050	0.252	0.167	—	—	−0.007	0.108	−0.198	0.126
Stage 2	0.250*	0.020	0.241**	0.099	0.256*	0.027	0.520*	0.027	0.169*	0.019
Stage 3	0.382*	0.035	0.783*	0.155	0.553*	0.042	0.681*	0.046	0.296*	0.040
Stage 4	0.307*	0.069	1.509*	0.496	0.263*	0.075	0.548*	0.070	0.395*	0.056
LoS_{t-1}	0.027*	0.005	−0.037	0.064	0.007*	0.001	−0.0003	0.002	0.047*	0.007
Mortality	0.114*	0.028	0.051	0.131	0.142**	0.063	0.161*	0.032	0.040	0.025
No. of procedures	0.011	0.013	0.322*	0.119	−0.074*	0.018	0.284*	0.031	0.634*	0.020
Hospital type	0.023	0.020	−0.118	0.092	−0.070*	0.025	−0.035	0.025		—
Constant	8.924*	0.195	6.544*	0.900	9.839*	0.230	9.944*	0.254	7.887*	0.204

SE = standard errors; C501 = central portion of breast; C502 = upper-inner quadrant of breast; C503 = lower-inner quadrant of breast; C504 = upper-outer quadrant of breast; C505 = lower-outer quadrant of breast; C506 = axillary tail of breast; C508 = overlapping lesion of breast; C509 = breast unspecified; Stage 2–4 = breast cancer severity; LoS_{t-1} = length of stay in the previous year; Mortality = patient mortality dummy; No. of procedures = no. of procedures performed; Hospital type = University/specialist centre dummy.
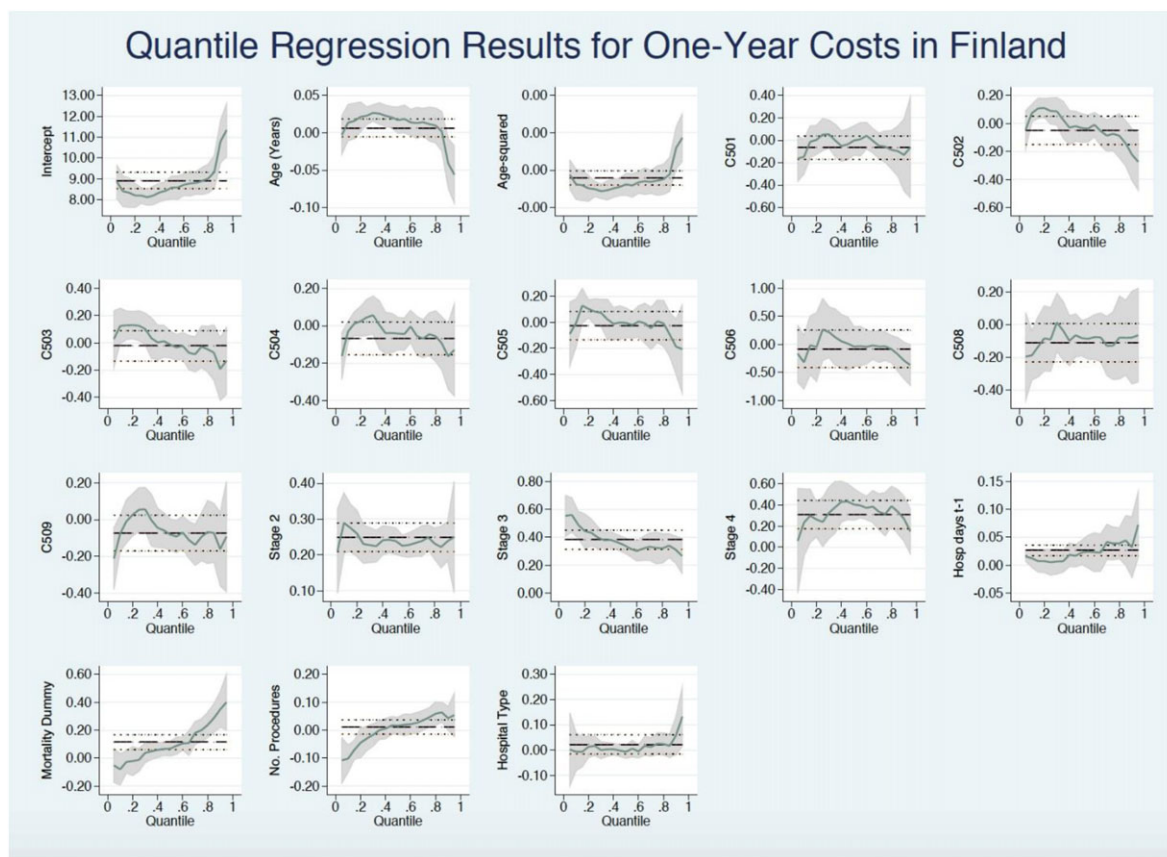*Significant at $p<0.01$. **Significant at $p<0.05$. ***Significant at $p<0.1$.

Figure 2. A comparison of the conditional mean (represented by the thick black dashed line with corresponding 95% confidence interval denoted by the upper and lower dotted lines) and the conditional quantile effects (denoted by the thick solid line and the grey shaded area corresponding to the quantile confidence interval) for 1-year cumulative hospital costs for Finland

dashed line with corresponding 95% confidence interval denoted by the upper and lower dotted lines. Heterogeneity in the response of covariates for breast cancer costs across conditional quantiles is represented by the thick solid line, and the grey shaded area denotes the corresponding uncertainty in terms of the quantile confidence interval. If the conditional quantile effect exceeds the bounds of the OLS confidence interval, then this provides support for parameter heterogeneity. To supplement Figures 2–6, the coefficients and standard errors for the quantile regression model results at the median; lower and upper quartile are presented in Tables S3–S7 in the Supporting Information.

Figure 2 presents the conditional mean and conditional quantile results for Finland. There is evidence of treatment response heterogeneity across the conditional quantiles. There is a 13% increase in the costs for patients diagnosed using ICD-10 code C504 (upper-outer quadrant of breast) at the lower conditional quartile but no significant differences at the conditional mean, median or upper quartile. The conditional mean effect for stage 2 patients provides a good approximation of the relative increase in costs compared with stage 1 patients across conditional quantiles. In contrast, the conditional mean effect for stage 3 patients masks response heterogeneity across the conditional quantiles. For example, the costs for stage 3 patients are estimated to be 43% higher at the lower conditional quartile, 32% higher at the upper quartile and 38% higher, on average, relative to stage 1 patients. Hospital costs for stage 4 patients are estimated to be 24% higher at the lower conditional quartile, 40% higher at the median and 31% higher for both the conditional mean and upper conditional quartile relative to stage 1 patients. Although the conditional mean effect for the comorbidity variable estimates each
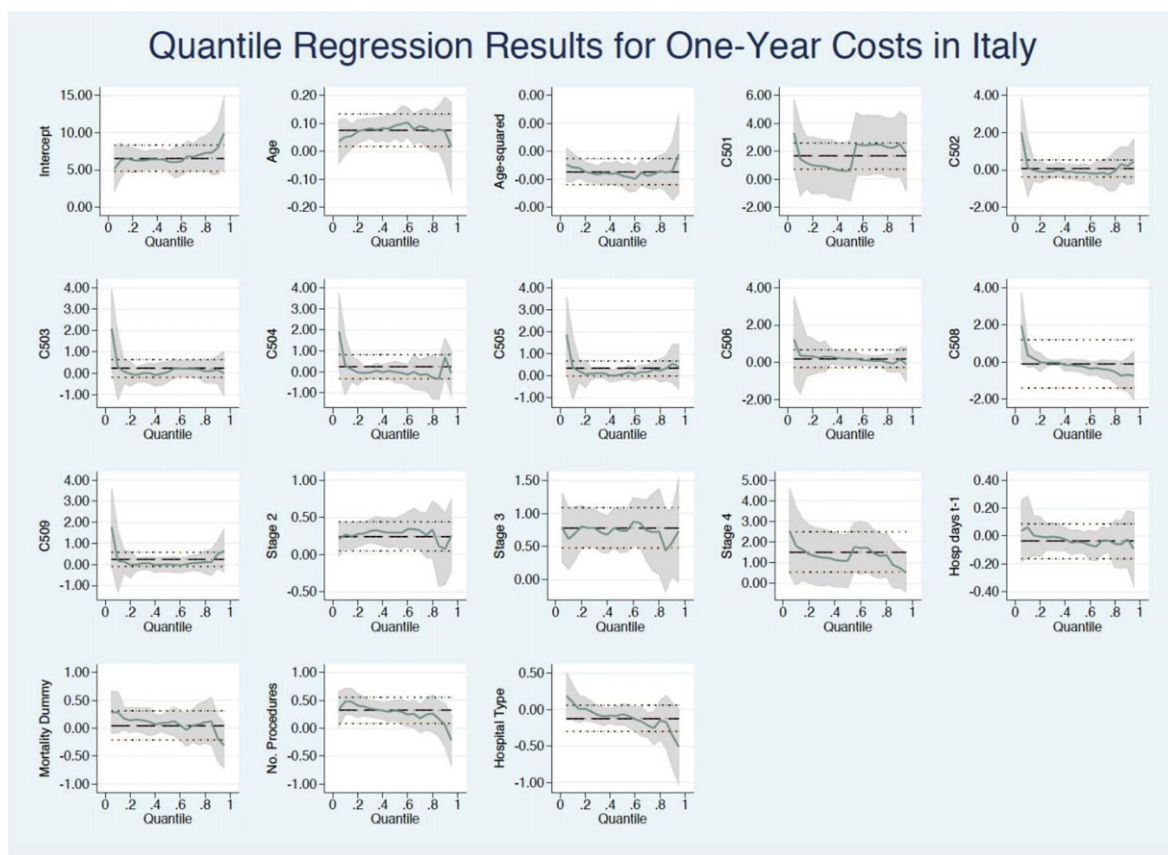
Figure 3. A comparison of the conditional mean (represented by the thick black dashed line with corresponding 95% confidence interval denoted by the upper and lower dotted lines) and the conditional quantile effects (denoted by the thick solid line and the grey shaded area corresponding to the quantile confidence interval) for 1-year cumulative hospital costs for Italy

additional day spent in hospital in the previous year to increase costs by 3%, conditional median results suggest each day is associated with a 2% increase and upper quartile estimates consider an increases in costs of 4% per day in the previous year. Patient mortality is associated with a 7% increase in 1-year hospital costs at the conditional median, 20% increase in costs at the upper quartile but the estimated coefficient is not significant at the lower conditional quartile relative to patients that survived. The number of breast cancer procedures performed is not significant at any conventional level for both the conditional median and lower quartile. However, the sign and magnitude of the coefficients change between the lower and upper quartile with the upper conditional quartile estimating each additional procedure performed to increase hospital costs by 5%.

Column (2) in Table II reports the estimated coefficients for the OLS cost regression model for the metropolitan area of Turin (Italy). Cost estimates for breast cancer stage monotonically increase with the level of severity. For patients classified with stage 2 breast cancer, estimated costs are 24% higher relative to stage 1 patients. This rises to a 78% and 151% increase in costs for stage 3 and 4 patients, respectively, compared with stage 1. Each additional breast cancer procedure performed in hospitals in Italy is estimated, on average, to increase costs by 32%. The conditional mean and conditional quantile results for Italy are presented in Figure 3. As illustrated in Figure 3, the conditional mean effect for all breast cancer stage variables provides a good approximation of the relative increase in costs compared with stage 1 patients across conditional quantiles. In contrast, there is evidence of parameter heterogeneity for both the number of procedures performed and hospital type across conditional quantiles. The number of breast cancer procedures performed is not significant
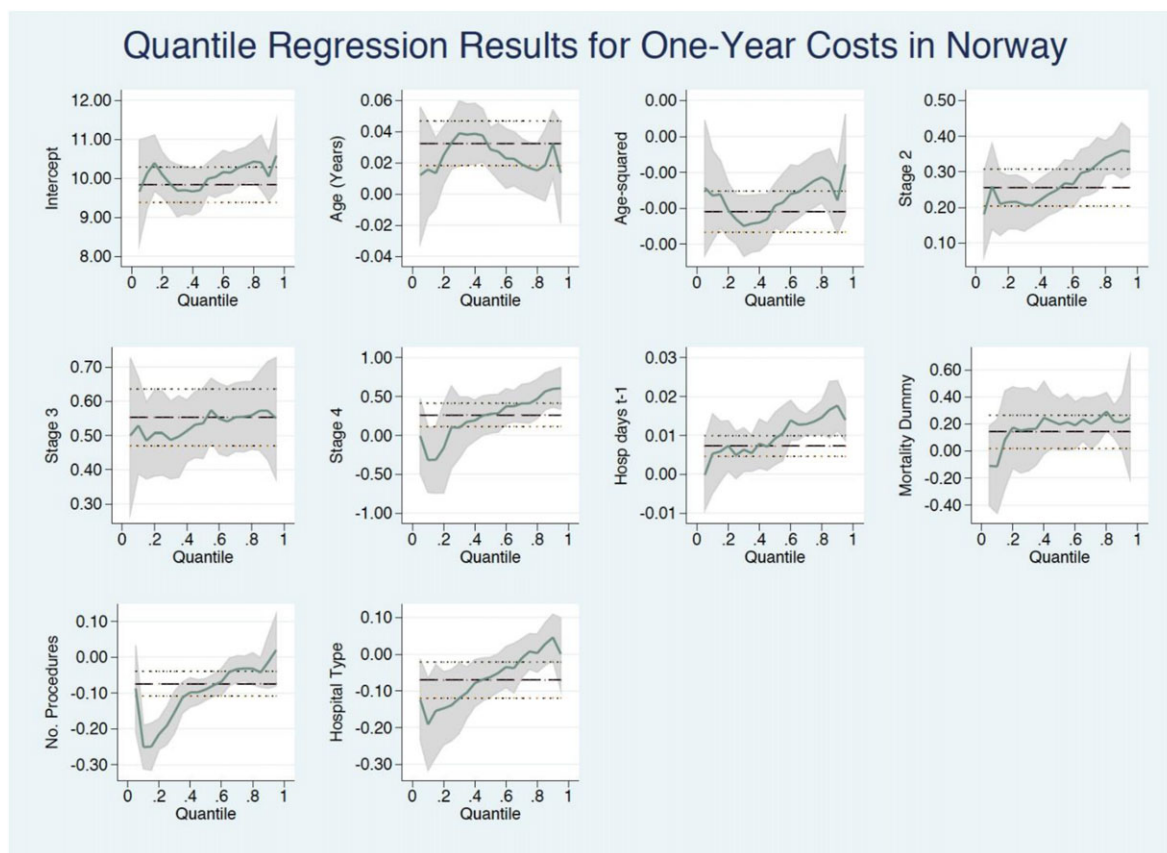
Figure 4. A comparison of the conditional mean (represented by the thick black dashed line with corresponding 95% confidence interval denoted by the upper and lower dotted lines) and the conditional quantile effects (denoted by the thick solid line and the grey shaded area corresponding to the quantile confidence interval) for 1-year cumulative hospital costs for Norway

at the upper quartile. However, costs are increased by 32% for each additional breast cancer procedure performed at the conditional mean and median as well as a 39% increase in costs at the lower quartile. The costs for patients treated in university teaching hospitals or specialist breast cancer centres in Italy is not significant at any conventional level at the conditional mean, median or lower quartile. In contrast, patients treated in university teaching hospitals or specialist breast cancer centres at the upper conditional quartile is estimated to reduce costs by 25% relative to other hospital types.

The estimated coefficients from the Norway OLS regression model for 1-year cumulative hospital costs are presented in column (3) of Table II. For patients classified with stage 2 and stage 4 breast cancer, estimated costs are 26% higher relative to stage 1 patients. This rises to a 55% increase for stage 3 patients compared with stage 1. The proxy for comorbidities in Norway results in a 0.7% increase in costs for each additional day spent in hospital in the previous year. Patient mortality is associated with a 14% increase in 1-year hospital costs relative to patients that survived. Each additional breast cancer procedure performed in hospitals in Norway is estimated, on average, to reduce costs by 7%. Furthermore, the costs for patients treated in university teaching hospitals or specialist breast cancer centres in Norway is estimated to be 7% lower relative to other hospital types.

Figure 4 presents the conditional mean and conditional quantile results for Norway. The conditional mean effect for stage 3 patients provides a good approximation of the relative increase in costs compared with stage 1 patients across conditional quantiles. In contrast, the average treatment response based on the OLS regression
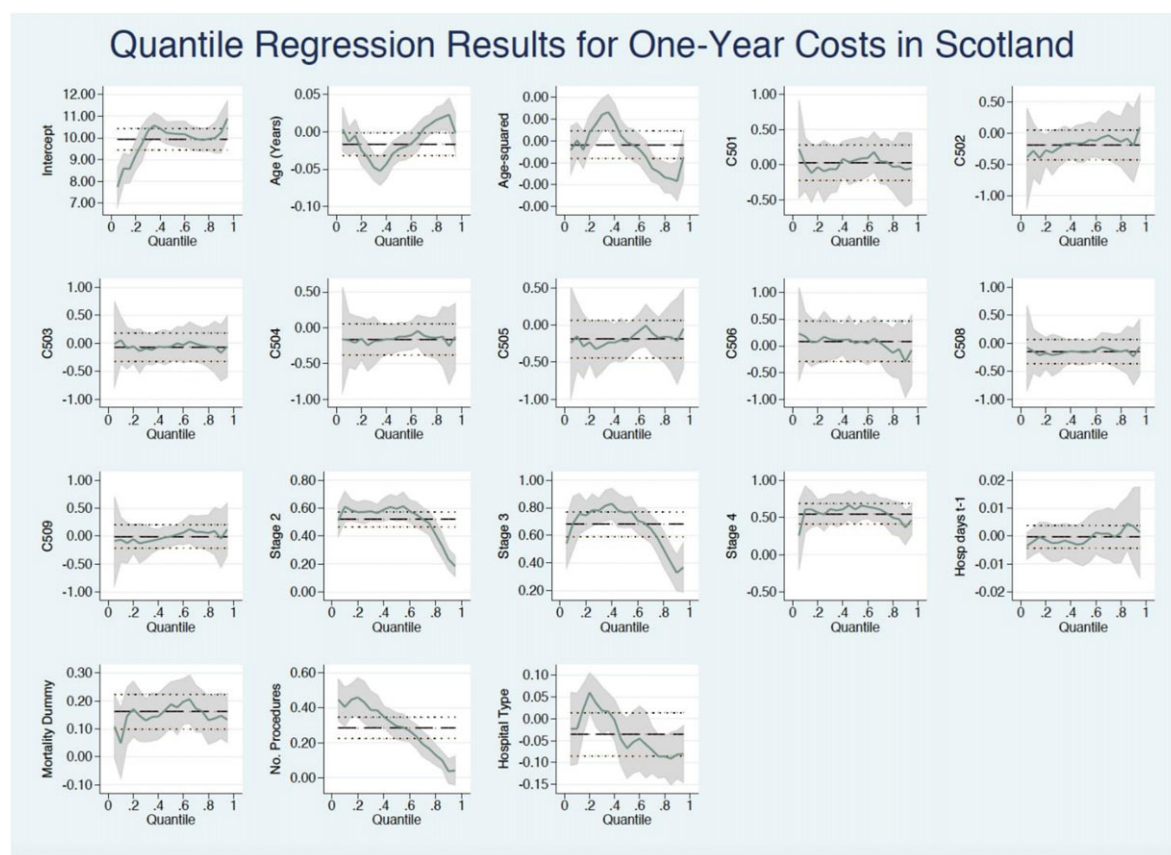
Figure 5. A comparison of the conditional mean (represented by the thick black dashed line with corresponding 95% confidence interval denoted by the upper and lower dotted lines) and the conditional quantile effects (denoted by the thick solid line and the grey shaded area corresponding to the quantile confidence interval) for 1-year cumulative hospital costs for Scotland

model for stage 2 and stage 4 patients fails to uncover important heterogeneity across the conditional quantiles. For example, the costs for stage 2 patients are estimated to be 22% higher at the lower conditional quartile, 32% higher at the upper quartile and 26% higher, on average, relative to stage 1 patients. The costs for stage 4 patients are not significant at the lower quartile but are estimated to be 28% higher at the median, 41% higher at the upper conditional quartile and 26% higher at the conditional mean relative to stage 1 patients. Although the conditional mean response is the same for both stage 2 and stage 4 patients, there is a greater variation in the most severe breast cancer category represented by stage 4. The comorbidity variable estimates each additional day in hospital in the previous year to increase costs by 0.7% for the conditional mean, 1% for conditional median and 1.4% at the upper conditional quartile. Patient mortality is associated with a 20% increase in 1-year hospital costs at the conditional median and 24% increase in costs at the upper quartile, but the estimated coefficient is not significant at the lower conditional quartile relative to patients that survived. The number of breast cancer procedures performed is not significant at the upper quartile. However, costs are reduced by 9% for each additional procedure performed at the median, reduced by 19% at the lower quartile, whilst the conditional mean effect reported a reduction of 7%. The costs for patients treated in university teaching hospitals or specialist breast cancer centres in Norway is estimated to be 6% lower at the conditional median and this increases to a reduction in costs by 14% at the lower quartile relative to other hospital types.

Column (4) in Table II reports the estimated coefficients for the cost regression model for Scotland. For patients classified with stage 2, 3 or 4 breast cancers, the estimated costs are 52%, 68% and 55% higher,
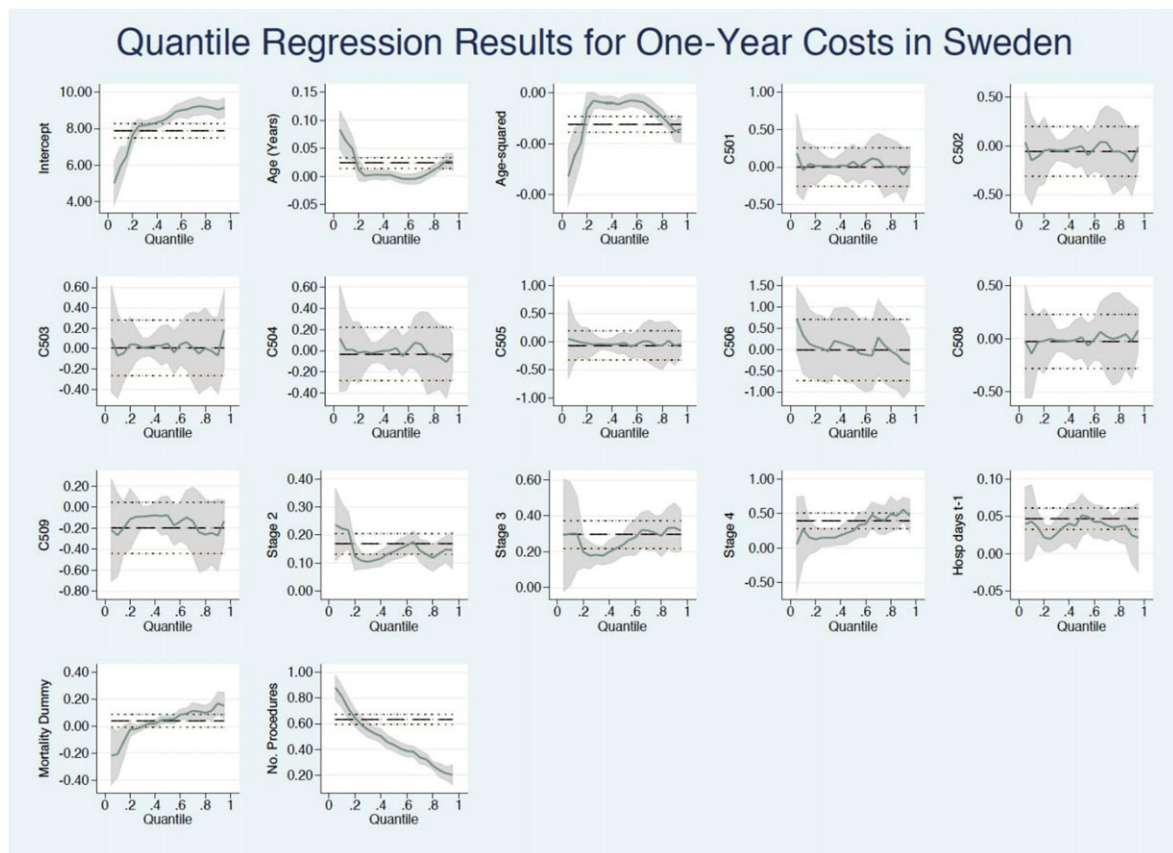
Figure 6. A comparison of the conditional mean (represented by the thick black dashed line with corresponding 95% confidence interval denoted by the upper and lower dotted lines) and the conditional quantile effects (denoted by the thick solid line and the grey shaded area corresponding to the quantile confidence interval) for 1-year cumulative hospital costs for Sweden

respectively, relative to stage 1 patients. Patient mortality is associated with a 16% increase in 1-year hospital costs relative to patients that survived. Each additional breast cancer procedure performed in hospitals in Scotland is estimated, on average, to increase costs by 28%. Figure 5 presents the conditional mean and conditional quantile results for Scotland. There is evidence of treatment response heterogeneity across the conditional quantiles. There is a 31% and 32% reduction in costs for patients diagnosed using ICD-10 codes C503 (lower-inner quadrant of breast) and C506 (axillary tail of breast) at the lower conditional quartile but no significant differences at the conditional mean, median or upper quartile. The conditional mean effect for all breast cancer staging variables does not provide a good approximation of the relative increase in costs compared with stage 1 patients. One-year hospital costs for stage 2 patients are estimated to be 57% higher at the lower conditional quartile, 60% higher at the median and 49% higher at the upper conditional quartile relative to stage 1 patients. Costs for stage 3 patients are estimated to be 78% higher at the lower conditional quartile, 77% higher at the median and 58% higher at the upper conditional quartile relative to stage 1 patients. Hospital costs for stage 4 patients are estimated to be 54% higher at the lower conditional quartile, 61% higher at the median and 56% higher for the upper conditional quartile relative to stage 1 patients. Patient mortality is associated with a 15%, 19% and 16% increase in 1-year hospital costs at the lower conditional quartile, median and upper conditional quartile relative to patients that survived. Hospital costs are increased by 43%, 30% and 17% for each additional breast cancer procedure performed in Scotland at the lower conditional quartile, median and upper conditional quartile, respectively. Although hospital type is not significant at the conditional

mean or lower quartile, the costs for patients treated in university teaching hospitals or specialist breast cancer centres in Scotland is estimated to be 7% lower at the conditional median and 9% lower at the upper quartile relative to other hospital types.

The national conditional mean effects for Sweden for 1-year hospital costs based on the OLS regression model are reported in column (5) of Table II. Cost estimates for breast cancer stage monotonically increase with the level of severity. For patients classified with stage 2 breast cancer, estimated costs are 17% higher relative to stage 1 patients. This rises to a 30% and 40% increase in costs for stage 3 and 4 patients, respectively, compared with stage 1. The proxy for comorbidities in Sweden results in just under a 5% increase in costs for each additional day spent in hospital in the previous year. Each additional breast cancer procedure performed in hospitals in Sweden is estimated, on average, to increase costs by 63%.

Figure 6 presents the conditional mean and conditional quantile results for Sweden. The conditional mean effect for all breast cancer staging variables does not provide a good approximation of the relative increase in costs compared with stage 1 patients. One-year hospital costs for stage 2 patients are estimated to be 11% higher at the lower conditional quartile, 14% higher at the median and 13% higher at the upper conditional quartile relative to stage 1 patients. Costs for stage 3 patients are estimated to be 18% higher at the lower conditional quartile, 24% higher at the median and 31% higher at the upper conditional quartile relative to stage 1 patients. Hospital costs for stage 4 patients are estimated to be 15% higher at the lower conditional quartile, 25% higher at the median and 40% higher for the upper conditional quartile relative to stage 1 patients. The comorbidity variable estimates each additional day in hospital in the previous year to increase costs by just under 5% for the conditional mean, just over 5% for conditional median, less than 2% at the lower quartile and over 3.5% at the upper conditional quartile. Patient mortality is associated with a 5% increase in 1-year hospital costs at the conditional median, 11% increase in costs at the upper quartile, but the estimated coefficient is not significant at the lower conditional quartile or conditional mean relative to patients that survived. Hospital costs are increased by 59%, 44% and 32% for each additional breast cancer procedure performed in Sweden at the lower conditional quartile, median and upper conditional quartile, respectively.

5. DISCUSSION AND CONCLUDING REMARKS

This paper considers whether conditional mean effects provide a suitable representation of national data on 1-year breast cancer cumulative hospital costs. Point estimates based on conditional mean effects provide a good approximation of treatment response for some key demographic and diagnostic-specific variables (e.g. age and ICD-10 diagnosis) across the conditional quantile distribution. For many policy variables of interest, however, there is a considerable evidence of parameter heterogeneity that is concealed if decisions are based solely on conditional mean results. Mortality, breast cancer stage and hospital type are all important variables in determining patient treatment, outcomes, access and service delivery. This paper has illustrated that these key variables are also important for determining the location scale and shape of breast cancer cumulative hospital costs over a 1-year period. Although there have been national policy initiatives to reduce breast cancer length of stay, heterogeneity still exists. The quantile regression results reinforce the need to consider beyond an average effect given the results of parameter heterogeneity.

A number of limitations exist in terms of the study design. Data sharing restrictions across EuroHOPE countries imposed constraints in terms of analysis as the pooling of data was not feasible. Instead, the EuroHOPE project relied on a common methodology across countries to ensure consistency in reporting and analysis of national data as outlined by Douglas (2012). There remains a large proportion of missing data across countries, particularly in terms of the variables used to derive breast cancer staging categories. The methodological challenges in terms of breast cancer staging in cross-country studies is well established, and potential solutions have been proposed (e.g. Walters *et al*., 2013). Despite this literature, attempts to follow the recommended guidelines did not address the challenges associated with imputing missing stage data. In particular, the imputation of missing stage data following the approach by Walters *et al*. (2013) still resulted in around 50%

missing stage data for both Finland and Sweden because of breast cancer stage coding discrepancies. The reported results are, therefore, confined to complete case analysis for all countries given the uncertainty in the pattern of missingness across key variables. Statistical methods for addressing missing data, such as multiple imputations by chained equations, are further inhibited by the data sharing restrictions across EuroHOPE partner countries, large percentage of missing data for some key variables (i.e. breast cancer stage) as well as uncertainty about whether the data are missing (not) at random. A number of countries were also constrained in terms of linking data to other registries which resulted in certain covariates being excluded from regression models as well as the loss of countries from the sample.[3] Beral and Peto (2010) have previously raised concerns about the quality of breast cancer registry data although Woods *et al*. (2011) posit a more optimistic outlook through a simulation study of the National Cancer Registry data from England and Wales. In light of the methodological challenges, caution is needed for the interpretation and generalisability of the regression results presented in this paper as part of the EuroHOPE project.

The costing methodology adopted by each EuroHOPE country relied on the best available data and estimates. As a result, differences exist in terms of episode-based costing estimates with a mixture of diagnosis-related groups as well as healthcare resource groups Geue *et al*. (2012) have previously demonstrated the sensitivity of results to the costing methodology adopted. It is worth reiterating that the aim of the paper was to explore parameter heterogeneity in national breast cancer cost data rather than cross-country disparities in hospital costs. However, the difference in costing methodology across countries could explain some of the differences in magnitude and sign for certain covariates included in the regression models. The number of procedures performed changes in sign and magnitude across countries and may simply be a manifestation of the underlying costing methodologies adopted. For this reason, interpretation should focus upon parameter heterogeneity across conditional quantiles within countries rather than between EuroHOPE countries.

A common methodology was adopted for the EuroHOPE project in terms of model selection based on available data across all countries. Important covariates were included in the regression models to control for patient case-mix and characteristics that may influence hospital costs. It was not possible to include a common measure of socio-economic status across EuroHOPE countries, which may have an impact on study results. In addition, there may be other forms of unobserved patient heterogeneity that the regression models are unable to accommodate. This may include patient frailty, engagement and treatment preferences which could bias results in either direction in terms of LoS as well as hospital costs. For many forms of unobserved heterogeneity, there is an inevitable interaction with observable characteristics. The use of quantile regression methods, in this setting, becomes particularly salient as a means of exploring potential exogenous shifts in the regressors across the conditional distribution (Miranda, 2008). One final consideration relates to the presence of potential endogenous regressors in the model specification, such as the mortality dummy variable. The methodological challenges associated with the identification of suitable instruments to address the endogeneity problem with cross-sectional observational data are well established in the literature (Staiger and Stock, 1997). As a result, a pragmatic approach was adopted by focusing on the OLS estimates in light of the absence of suitable instruments collected in national breast cancer registers. Although tests for weak instruments are available for single (Stock and Yogo, 2005) or multiple endogenous regressors (Sanderson and Windmeiejer, 2013), the consequences of weak instruments can induce greater bias. In this setting, the cure to the endogeneity problem can be worse than the diagnosis itself (Hahn and Hausman, 2003).

The direction of future research should focus upon the possible determinants of heterogeneity in outcomes as well as treatment response across the conditional distribution. One important contribution to the literature would be to consider the dependency between hospitals within countries in light of the heterogeneity reported in this paper. Gravelle, Santos and Siciliani (2013) have recently demonstrated the significance of spatial dependence between hospitals in determining quality. Cross-country studies, such as EuroHOPE, need to consider the appropriate spatial unit to capture any dependency between hospitals. For example, geographic,

---

[3]For example, information relating to the fourth digit of the ICD-10 diagnosis variable for Norway was not collated at the time of data retrieval and, therefore, absent from the final analysis dataset.

administrative or health system units may all represent a clustering of outcomes in terms of healthcare performance, quality and efficiency. Jones and Spiegelhalter (2011) acknowledge the methodological challenges in the identification of extreme hospital outcomes or outliers and offer possible strategies to overcome these difficulties. The incorporation of national treatment guidelines with expert opinion from clinicians and healthcare providers across all participating countries offers a fruitful means of benchmarking European hospital outcomes prior to the analysis of data to identify extremes or outliers.

CONFLICTS OF INTEREST

The authors have declared that there is no conflict of interest.

ACKNOWLEDGEMENTS

REFERENCES

Aas E. 2012. Cost measurement and estimation of cost functions. *EuroHOPE Discussion Paper* **No.2**: 1–58.

Beral V, Peto R. 2010. UK cancer survival statistics are misleading and make survival look worse than it is. *BMJ* **341**: c4112.

Berrino F, De Angelis R, Sant M, Rosso S, Bielska-Lasota M, Coebergh JW, Santaquilani M. 2007. Survival for eight major cancers and all cancers combined for European adults diagnosed in 1995–99: results of the EUROCARE-4 study. *Lancet Oncology* **8**: 773–783.

Bray F, McCarron P, Parkin D. 2004. The changing global patterns of female breast cancer incidence and mortality. *Breast Cancer Research* **6**: 229–239.

Douglas A. 2012. EuroHOPE breast cancer: material, methods and indicators. *EuroHOPE Discussion Paper* **No.8**: 1–41.

Ferlay J, Steliarova-Foucher E, Lortet-Tieulent J, Rosso S, Coebergh JW, Comber H, Forman D, Bray F. 2013. Cancer incidence and mortality patterns in Europe: estimates for 40 countries in 2012. *European Journal of Cancer* **49**: 1374–1403.

Geue C, Lewsey J, Lorgelly P, Govan L, Hart C, Briggs A. 2012. Spoilt for choice: implications of using alternative methods of costing hospital episode statistics. *Health Economics* **21**: 1201–1216.

Gravelle H, Santos R, Siciliani L. 2013. Does a hospital's quality depend on the quality of other hospitals? A spatial econometric approach to investigating hospital quality competition. *Centre for Health Economics Research Paper* **82**: 1–33.

Hahn J, Hausman J. 2003. Weak instruments: diagnosis and cures in empirical econometrics. *American Economic Review* **93**(2): 118–125.

Hery C, Ferlay J, Boniol M, Autier P. 2008a. Changes in breast cancer incidence and mortality in middle-aged and elderly women in 28 countries with caucasian majority populations. *Annals of Oncology* **19**: 1009–1018.

Hery C, Ferlay J, Boniol M, Autier P. 2008b. Quantification of changes in breast cancer incidence and mortality since 1990 in 35 countries with caucasian-majority populations. *Annals of Oncology* **19**: 1187–1194.

Jones H, Spiegelhalter D. 2011. The identification of "unusual" health-care providers from a hierarchical model. *The American Statistician* **65**: 154–163.

Jones A, Lomas J, Rice N. 2014. Applying beta-type size distributions to healthcare cost regressions. *Journal of Applied Econometrics* **29**: 649–670.

Koenker R. 2005. *Quantile Regression*, Econometric Society Monograph Series, Cambridge University Press: Cambridge, UK.

Koenker R, Hallock K. 2001. Quantile regression. *Journal of Economic Perspectives* **15**: 143–156.

Manning W, Mullahy J. 2001. Estimating log models: to transform or not to transform? *Journal of Health Economics* **20**: 461–494.

Marla S, McMillan D, Stallard S. 2013. Factors influencing postoperative length of hospital stay after breast cancer surgery. *The Breast* **22**: 289–294.

Miranda A. 2008. Planned fertility and family background: a quantile regression for counts analysis. *Journal of Population Economics* **21**: 67–81.

Mullahy J. 1998. Much ado about two: reconsidering transformation and the two-part model in health econometrics. *Journal of Health Economics* **17**: 247–281.

Nixon R, Thompson S. 2004. Parametric modelling of cost data in medical studies. *Statistics in Medicine* **23**: 1311–1331.

OECD. 2012. Eurostat-OECD methodological manual on purchasing power parities. In *Organisation for Economic Co-operation and Development*, Publications Office of the European Union: Luxembourg.

Polyak K. 2011. Heterogeneity in breast cancer. *Journal of Clinical Investigation* **121**: 3786–3788.

Sanderson E, Windmeiejer F. 2013. A weak instrument F-Test in linear IV models with multiple endogeneous variables. *CMPO Working Paper* **13**(315): 1–34.

Staiger D, Stock JH. 1997. Instrumental variables regression with weak instruments. *Econometrica* **65**: 557–586.

Stock JH, Yogo M. 2005. Testing for weak instruments in linear IV regression. In *Identification and inferences for econometric models, essays in honor of Thomas Rothenberg*, Andrews DWK, Stock JH (eds.), Cambridge University Press: New York; 80–108.

Walters S, Maringe C, Butler J, Brierley JD, Rachet B, Coleman MP. 2013. Comparability of stage data in cancer registries in six countries: lessons from the international cancer benchmarking partnership. *International Journal of Cancer* **132**: 676–685.

Westlake S, Cooper N. 2008. Cancer incidence and mortality: trends in the United Kingdom and constituent countries, 1993 to 2004. *Health Statistics Quarterly* **38**: 33–46.

Woods LM, Coleman MP, Lawrence G, Rashbass J, Berrino F, Rachet B. 2011. Evidence against the proposition that "UK cancer survival statistics are misleading": simulation study with national cancer registry data. *BMJ* **342**: d3399.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article at publisher's website.