**Report Supplementary Materials 5**

**ESTIMATING THE HOSPITAL COSTS OF AAA REPAIRS USING HOSPITAL EPISODE STATISTICS IN ENGLAND**

**ABSTRACT**

**Introduction:** We aim to estimate the costs of the different types of AAA repairs using routinely available hospital data in England. We also aim to investigate the variation of these costs based on the different patient and provider characteristics to accurately predict these costs on the context of economic evaluation for service reconfiguration.

**Methods:** HES data were acquired between 1st April 2008 and 15th Feb 2015. The R© programme (Version 3.4.1) (R foundation, Vienna, Austria) was used for all data manipulation and statistical analyses. The primary outcome was the total costs of the index episode including the whole hospital stay, critical care, excess bed days, devices, and other unbundled procedures. Descriptive statistics, univariate analyses, and multiple regression models were used to investigate the costs and their variation.

**Results:** 42,960 patients who were admitted and received AAA repairs between 1st Apr 2008 and 28th Feb 2015 were identified: 18.01% were ruptured, 12.18% were non-ruptured emergency, 27.85% were elective open repair, and 41.96% were elective EVAR. The mean age of the cohort was 74 years and 86% was male. Mean total costs (in 2014/15 price) for ruptured, non-ruptured emergency, elective open repair, and elective EVAR were £14779.3, £12235.8, £10607.4, and £10616.8 respectively. Costs varied with certain patient, provider characteristics, and hesyear. We found two provider characteristics that had significantly positive effects on the total costs after controlling for other case-mix factors: the annual elective volume and the market forces factor index. Higher volume centres were associated with higher costs and higher level of unavoidable costs for a location (reflected by the MFF index) were also associated with higher costs.

**Conclusion:** These findings have important implications for cost-effectiveness modelling to evaluate services reconfiguration options. On the one hand, we know that merging small units with bigger ones would improve outcomes such as in-hospital mortality. On the other hand, with the new findings from this study, merging small units with bigger ones would also increase costs significantly. Thus, it is important to properly model these two aspects of changes (outcomes and costs) to evaluate options for services reconfigurations.

# Introduction

Abdominal Aortic Aneurysm (AAA) is a common cardiovascular disease that often triggers surgical interventions. There are a number of different surgical treatments available for the AAA repair, ranging for open surgical repair to endovascular repair (EVAR). The choice of treatment depends on certain patient characteristics and provider characteristics. As such, the costs of AAA may potentially depend on a number of these patient and provider factors. Thus, using a mean (i.e. average or constant) cost for AAA repair does not take into the complexities involved in the procedure – this issue is particularly relevant in the context of economic evaluation models.

In England, the Hospital Episode Statistics (HES) datasets provide a rich source of information regarding in-patient care across England, year by year with coverage and completeness of coding well over 96% for in recent years [1-4]**.**The basic unit of activity measured in HES is the finished consultant episode (FCE), this includes a single period of care under one consultant and does not necessarily equate to all the care provided in an admission. FCE contains a primary diagnosis and up to 19 secondary diagnoses and 24 procedure fields. FCE includes information such as identification of the hospital, patient demographics, type of admission, source of admission, speciality, destination and date of discharge, dates of procedures, referral to different providers as well as length of stay in critical care (recorded in a separate linked dataset) and other important clinical and administrative information. Providers in England are required to report the costs of all their FCEs grouped by Health Resource Groups (HRGs) and other episode information (i.e. admission method, treatment specialty, length of stay, etc.). These provider-specific reference cost data are collected centrally and made available to the public. Using the reference cost data, the total costs of a FCE can be determined retrospectively by its categorisation information including its main HRG, unbundled HRGs, critical care HRG, length of critical care, level of excess bed days, and mode of admission. Nonetheless, the process is not always straightforward given changes in HRG coding and HES coding across the years. There has not been any study looking at the detailed costing of AAA episodes using HES data.

The aim of this study is to estimate the costs of the different types of AAA repair using the HES data in England between 2008 and 2015. Furthermore, the variation of these costs based on the different patient and provider characteristics is estimated to accurately predict these costs on the context of economic evaluation. We aim to conduct the most accurate costing exercise by using year-specific HRG groupers and year-specific reference cost data. The paper is organised as follows. The Methods section describes the data source, design of the dataset, the cost estimation and statistical analyses protocol. The results section presents a descriptive summary of the dataset, describes the issues identified during costing and presents the results of the statistical analyses. The key findings along with the strengths and limitations of this study are presented in the discussion section.

# Methods

## Dataset

HES data between 1st April 2008 and 15th Feb 2015 were acquired. The R© programme (Version 3.4.1) (R foundation, Vienna, Austria) was used for all data manipulation and statistical analyses. Episodes of patients who received at least a repair for infrarenal aortic aneurysm were identified and extracted from source data. Details of data extraction, cleaning, and validation were reported in another publication (Aber et al., 2018). The episodes were then sorted chronologically and grouped into NHS spells and Continuous Inpatient Stays (CIPS). The whole patient pathway was described in terms of a series of episodes, spells and CIPS that were captured in HES data. The index episode was defined as the first episode where patients received their first aneurysm repair. Each in-patient episode had a unique key to be linked with a critical care dataset that contains information about critical care episodes that took place within an in-patient episode. The primary outcome of interest was the hospital cost of these index episodes including the cost of critical care and how it was influenced by case-mix factors and volumes.

## Costing of HES data using the year-specific Groupers and Reference Cost Datasets

A two-stage process was used to cost the episodes (in-patient episodes and critical care episodes). The first stage was to assign a Health Resource Group (HRG) code to an episode using a specific Grouper that was relevant for the financial year of the episode. For example, an episode in 200809 was assigned a HRG4 code using the 200809 Grouper whereas an episode in 201415 was assigned a HRG4+ code using the 201415 Grouper. It should be noted that critical care episodes were stored in a separate dataset from the main in-patient dataset and they had their own HRG codes different from the main in-patient HRG codes. More information about HRG coding and the Groupers is referred to the NHS digital.

In the second stage, a year-specific reference cost dataset was used to cost an episode based on its HRG code, financial year, and other classification characteristics such as mode of admission (elective, emergency, day case, etc.), treatment specialty, and excessive length of stay (excess bed days). Since we were also interested in the relationship between cost and volume the provider-specific unit costs were used in our study; however, we also compare some descriptive results using the national-average unit costs. All costs were inflated to 2014/15 £ using the hospital and community health services (HCHS) index*.* For example, an episode in 200809 was costed using the provider-specific unit cost information from the 200809 reference cost dataset, and then inflated to 2014/15 £.

Missing data arose in both stages. With the HRG grouping stage, 14 (0.03%) out of 42960 episodes could not be grouped due to missing or invalid information in the episode (i.e. invalid main diagnosis). More missing data arose with the costing stage. In particular, more missing data occurred with provider-specific unit costs than with national-average unit costs. One of the main reasons was the issue with treatment specialty codes in the episode that does not correspond with those reported in the reference cost datasets. For example, many episodes recorded a treatment specialty code of 192 (critical care medicine) which did not match with the treatment specialty codes reported by the providers. To overcome the issue of missing data with provider-specific unit costs, we used the following assumptions to impute the values where possible:

(1) where a specialty code did not match with any from the reference cost dataset, an average value was used for imputation. For example, if a treatment specialty code of 192 could not be matched with any unit cost from the reference cost dataset but there are provider-specific unit costs available for other specialties codes such as 100,107, etc., their averaged value was used for imputation of the 192 specialty code (keep the same classification information on other fields i.e. HRG, department).

(2) where a provider-specific unit cost was still not available even after the imputation in (1) (only a few), it was imputed by the relevant national-average unit cost.

The total hospital cost of an aortic aneurysm repair episode includes four components: the base cost of the in-patient episode (corresponding to the in-patient HRG code), the cost of excess bed days (varied depending on the in-patient HRG, mode of admission, and the number of days that is beyond the trim point - per diem cost), the cost of critical care, and the cost of other unbundled services such as diagnostic imaging (corresponding to unbundled HRG codes).

## Market Forces Factor (MFF)

“The Market Forces Factor is an estimate of unavoidable cost differences between health care providers, based on their geographical location” - NHS England, 2015 [5].

The MFF is not included in the HES data. It is calculated separately using provider-specific information and is used to adjust cost differences between providers for the purposes of allocating commissioner funding or calculating national prices. The differences between the MFF of different providers reflect the variation in unavoidable costs (including costs of non-medical staff, medical and dental staff, land, and buildings) a provider faces. More information about the MFF and its calculation is referred to the NHS England’s guide.

There are two MFF values: underlying MFF and rebased MFF. The rebased value is just the normalised value of MFF. It is the underlying value divided through by the minimum value across all providers in a specific year. The rebased MFF score was used in our analyses. The MFF was used in our analyses as a provider factor that would explain the variation in the costs of AAA repairs.

## Statistical analyses

The outcomes of interest are costs of the index episode for AAA patients. Given the clinical differences and to reduce the effects of interactions, the whole patient cohort was divided into four groups: ruptured, non-ruptured emergency, elective open repair, and elective EVAR. This division was based on expert opinion from our clinical advisers and making sure each group still containing enough cases for the analyses.

Patient, admission, and provider characteristics were considered alongside HES data year as predictive variables. Patient characteristics included age, gender, index of multiple deprivation (IMD), distance to hospital, ruptured aneurysm, and comorbidities. Admission characteristics included mode of admission (elective versus emergency), type of intervention (open repair vs EVAR), and weekend admission. Provider characteristics included annual volume (measured as all repair cases) and their Market Forces Factor index (see information about MFF above). HES year, defined as a categorical variable from 2008/09 to 2014/15, was also considered as time trend was anticipated.

## Descriptive statistics

Descriptive summary statistics were reported for the patient population and sub-groups. The distribution of costs and the variation of these with respect to the HES year, patient, admission and provider characteristics were also examined.

## Multiple Regression Models for case-mix adjustment

To reduce the skewness of cost data, it was logarithmically transformed. The relationships between the log costs and the patient characteristics, admission characteristics, provider characteristics and HES year were examined in univariate regression models. All variables that meet the significance test of p-value less than 0.05 were considered for multiple regression analysis. Generalised linear models (GLM) with log link function and gamma distributed errors were used. Linear Model with log link function was also compared.

Backward stepwise approach was used to select the covariates for multiple regression analysis. Initially, a null model was specified without any covariates to provide a ‘baseline’ model of costs. Covariates that were significant in the univariate analysis were subsequently introduced to control for the systematic differences in these variables and also to examine their effect on the costs (or log costs). Covariates that were no longer significant, when controlled for other factors, were excluded. The models were assessed based on their goodness of fit and predictive accuracy. All analyses were conducted in R version 3.4.1.

# Results

## Data quality

Between 1st Apr 2008 and 28th Feb 2015, there were 42,960 patients who were admitted and received infrarenal AA repairs. This gave us 42,960 index episodes for costing. Through the HGR grouping stage, there were 14 (0.03%) in-patient index episodes and 379 (1.85%) critical care episodes that could not be grouped due to missing data. In the costing stage (matching with unit costs from the reference cost data), 5483 (12.76%) episodes could not be mapped directly to their provider-specific unit costs; however, after imputations based on the assumptions as discussed above, only 42 episodes (0.1%) could not be mapped to their unit costs due to missing data.

## Descriptive summary

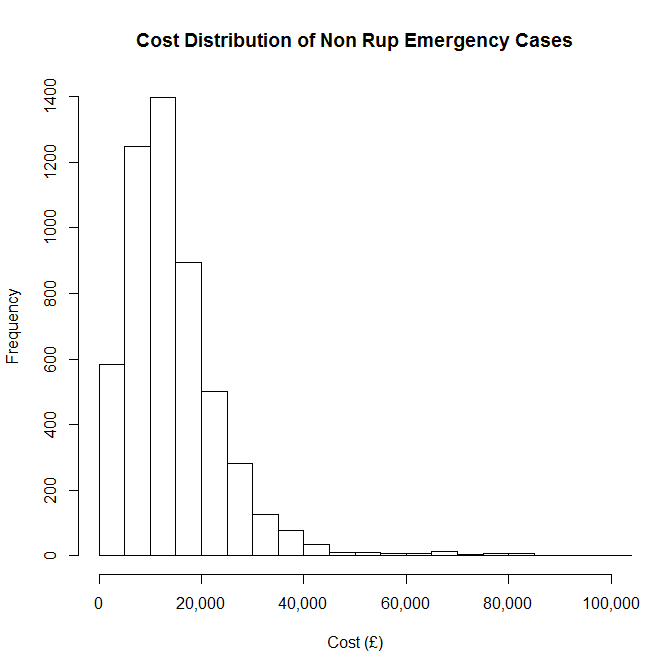
Table 1 summarises the characteristics of the whole cohort and each clinical group. The proportion of ruptured cases was 17.6%. The mean age of the cohort was 74 years old and 86% was male. Comparing the clinical groups, the elective EVAR group had older patients (mean age 76) and more men (89%). Regarding costs, the mean base cost (excluding excess bed days, unbundled procedures and critical care) of the index episode was £8,587 per patient for the whole cohort. It increased to £8,902 when the costs for excess bed days and unbundled procedures were added and increased to £ 11,562 accounting for all costs including critical care. Comparing costs between the clinical groups, if excluding the costs of critical care, elective EVAR was the most expensive group (£10,061 per patient), then non-ruptured emergency group (£9,402) followed by ruptured group (£8,031) and elective open repair group (£7,502). When the costs of critical care were included, the most expensive group was the ruptured group (£14,779), then non-ruptured emergency (£12,235) followed by elective EVAR (£10,617) and elective open repair (£10,607).

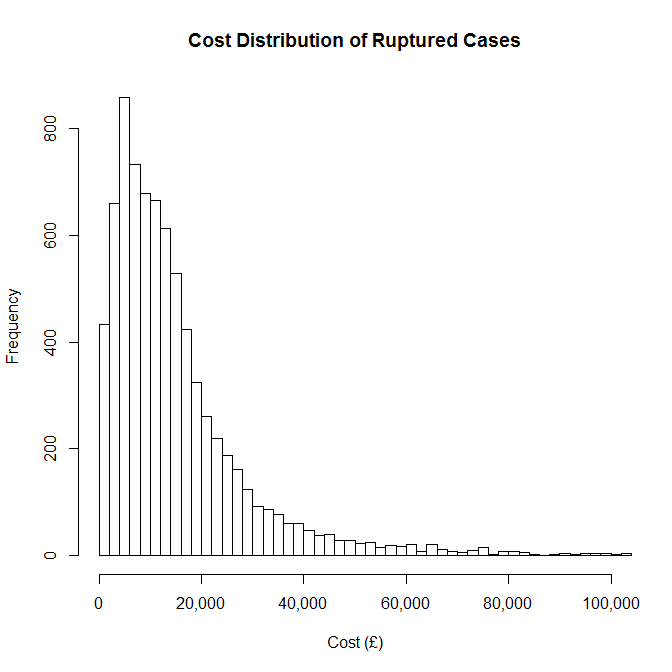
Table 1: Cohort summary

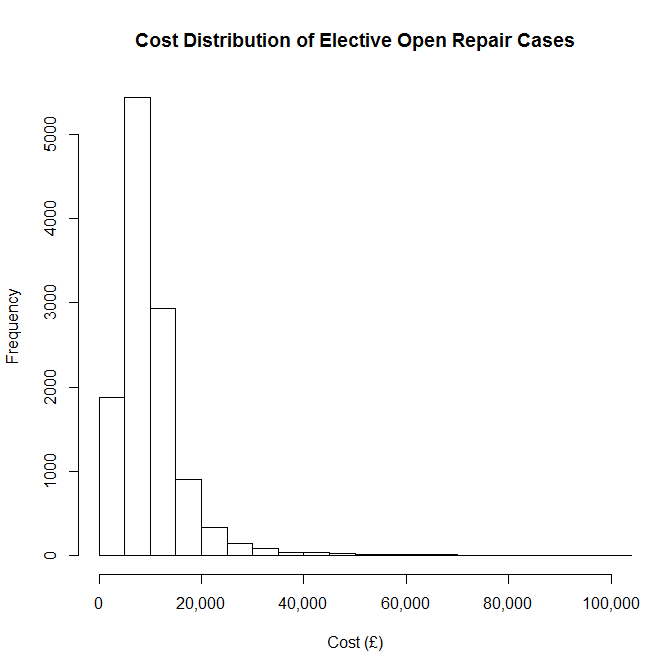
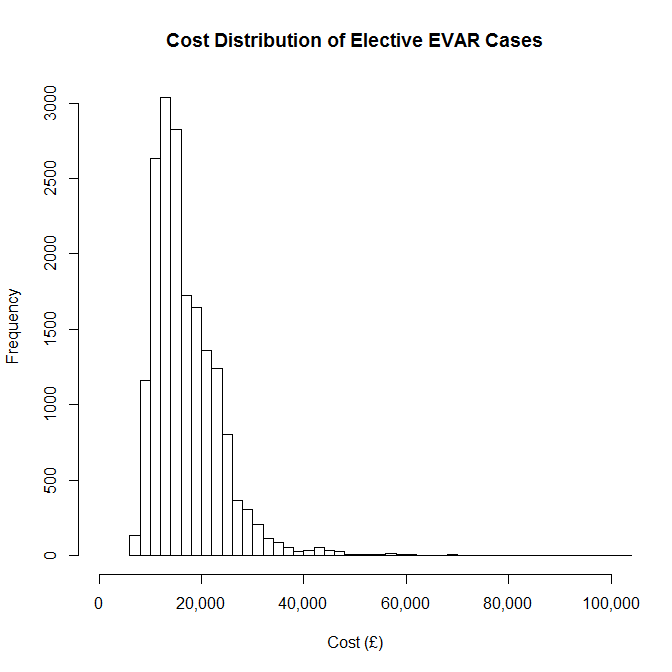
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  | All cohort | Ruptured | Non-rup emergency | Elective Open Repair | Elective EVAR |
| Cases | | 42960 | 7737 (18.01%) | 5233 (12.18%) | 11964 (27.85%) | 18026 (41.96%) |
| Age (mean, median) | | 74; 75 | 75; 76 | 73; 74 | 71; 72 | 76; 76 |
| Male | | 37024 | 6405 (82.8%) | 4320 (82.6%) | 10282 (85.9%) | 16017 (88.9%) |
| EVAR cases | | 21525 | 1362 (17.6%) | 2137 (40.8%) | NA | NA |
| LOS (days) index episode (mean, median) | | 7.74; 5 | 10.78; 6 | 11.06; 8 | 8.85; 7 | 4.73; 3 |
| LOS (days) index CIS (mean, median) | | 10.98; 6 | 19.74; 11 | 15.13; 8 | 12.08; 8 | 5.29; 3 |
| Cost of the index episode\*  (Per patient) | Base Cost (mean, median) | £8,587;  £7,250 | £7,483.6;  £6,520.8 | £8,778.9;  £7,458.7 | £7,212.4 ;  £6,688.6 | £9,919.4;  £8,320.1 |
| Total Costs with Excess Bed Days and Unbundled Procedures Added (mean, median) | £8,902;  £7,413 | £8,031.1;  £6,709.1 | £9,401.7;  £7,859.9 | £7,502.4;  £6,839.1 | £10,061.3;  £8,379.1 |
| Total Cost with Excess Bed Days, Unbundled Procedures, and Critical Care Added (mean, median) | £11,561.7;  £9,035.3 | £14,779.3;  £10,022.6 | £12,235.8;  £9,703.4 | £10,607.4;  £8,708.4 | £10,616.8;  £8,753.6 |

\*: after imputations based on assumptions (see discussion above)

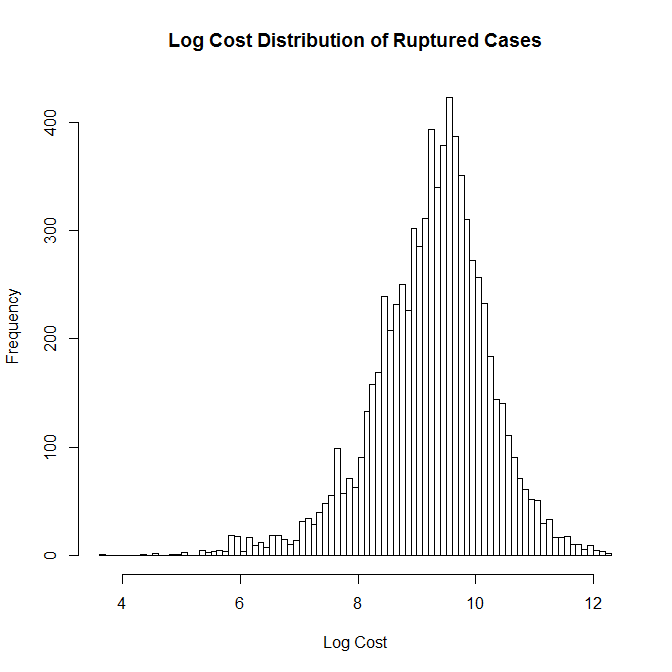
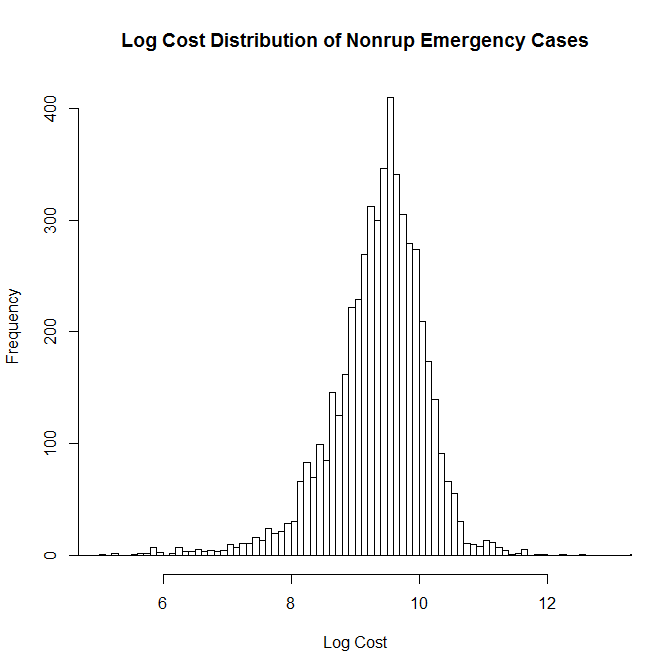
Distribution of costs data for each group:

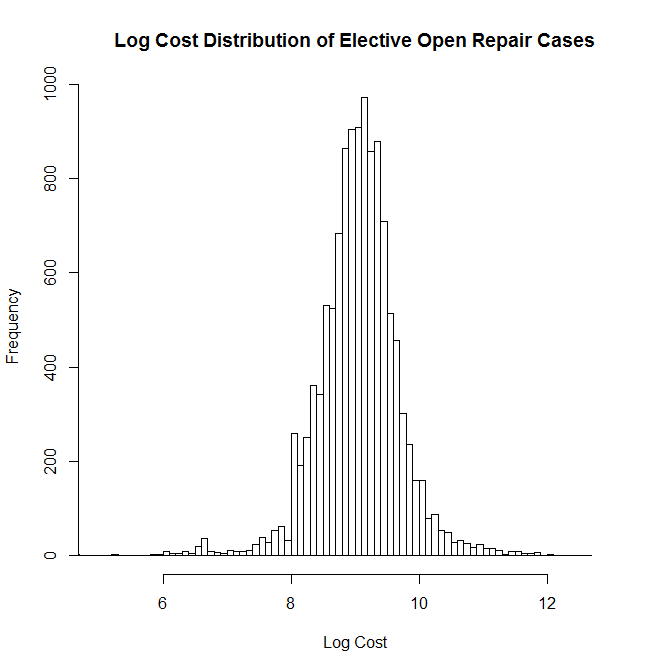
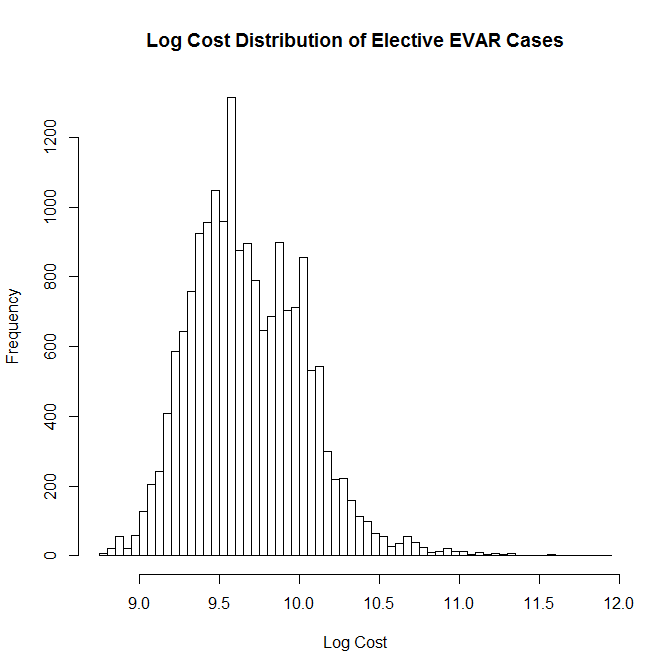






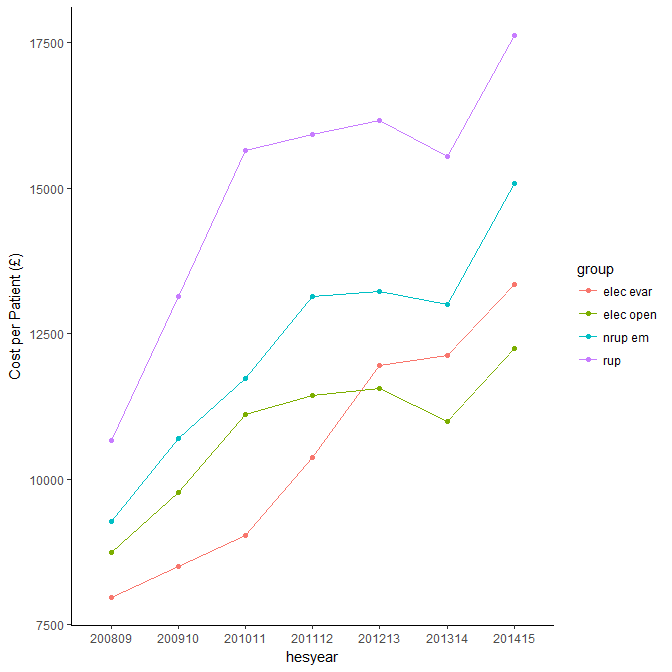
Distribution of log costs data for each group:



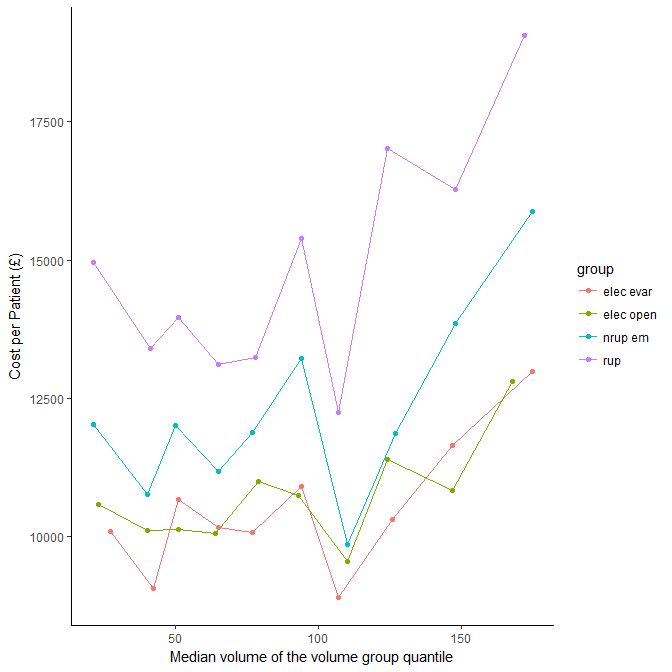


## Descriptive univariate analyses

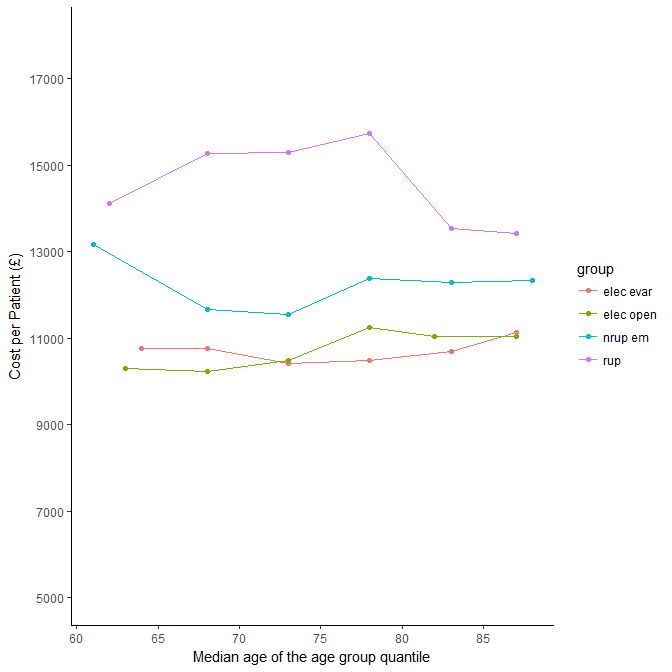
***Trends by year***

****

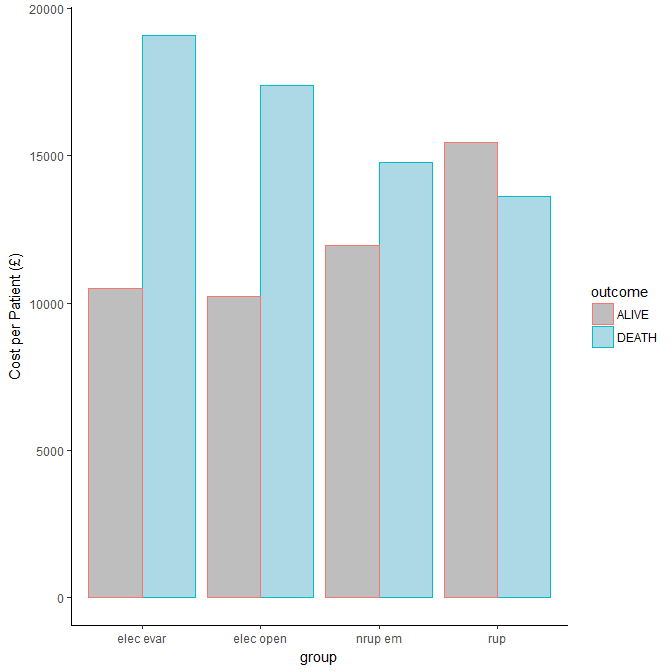
***Differences by volumes***

****

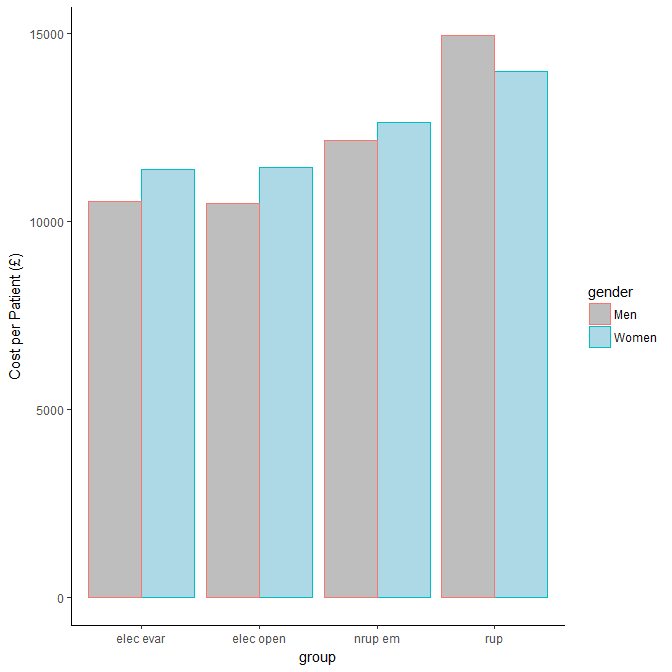
***Differences by age groups***

****

***Differences by in-hospital death outcome***

****

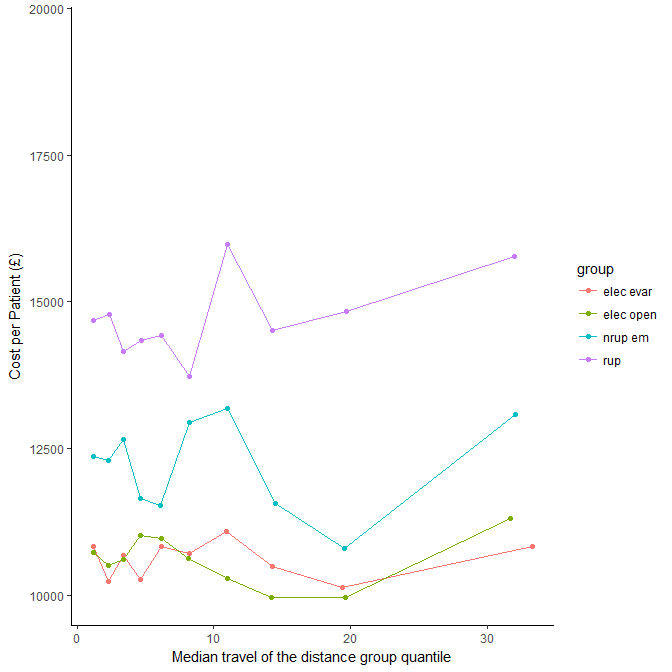
***Differences by gender***

******

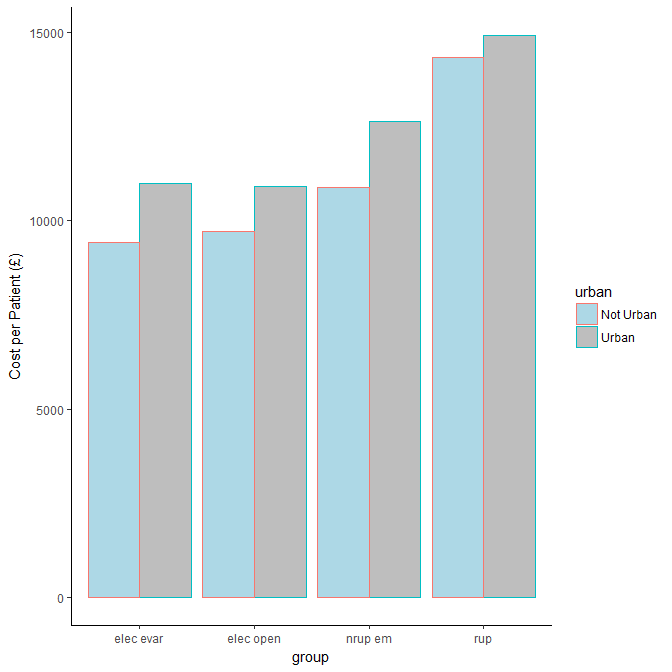
***Differences by IMD groups***

******

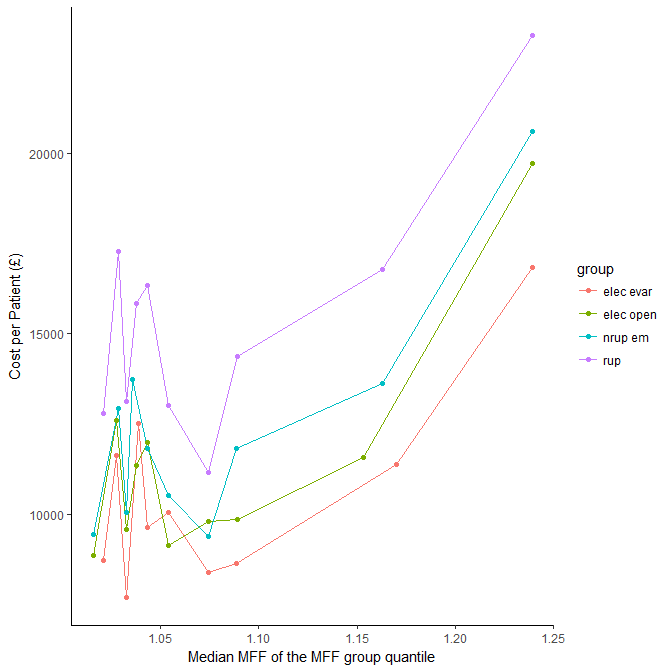
***Differences by travel distance***



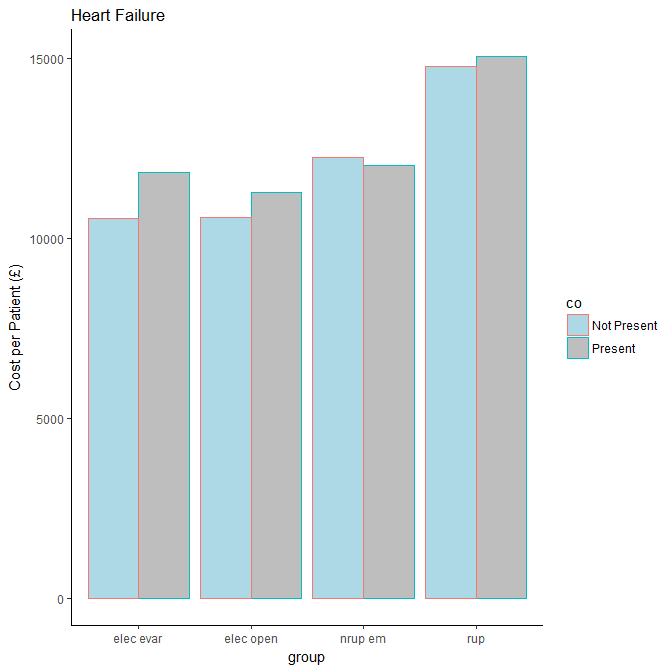
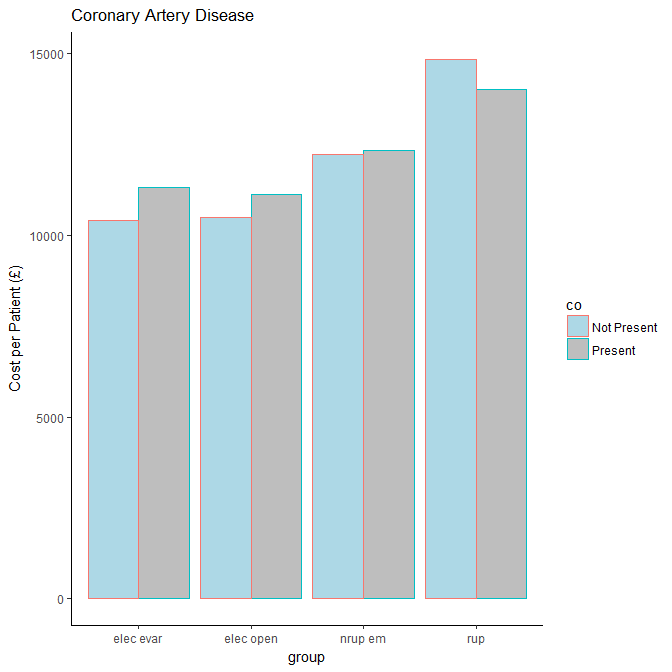
***Differences by urban and rural***

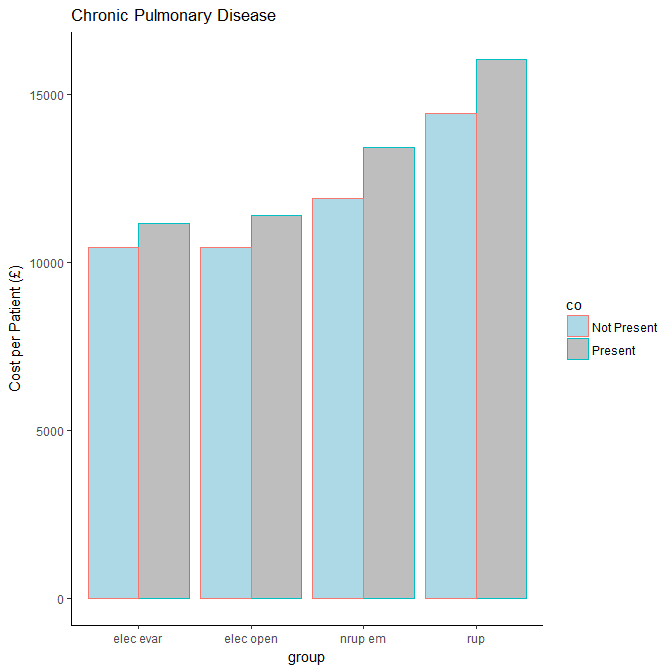
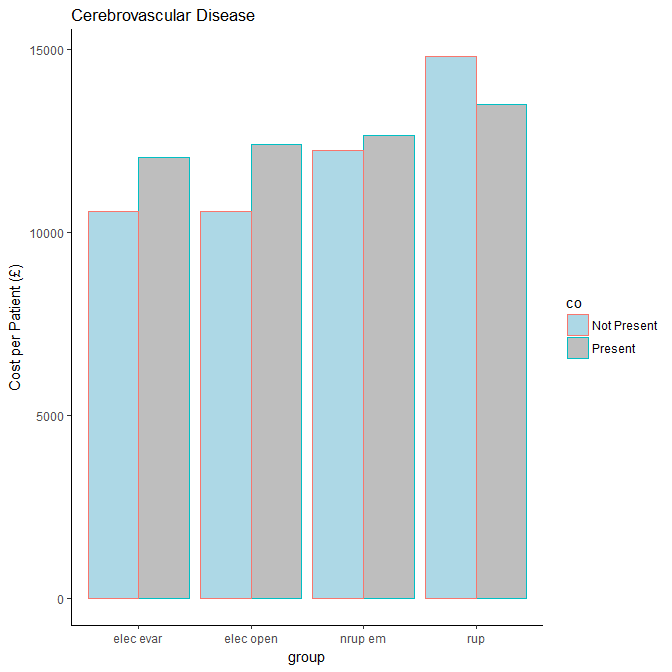


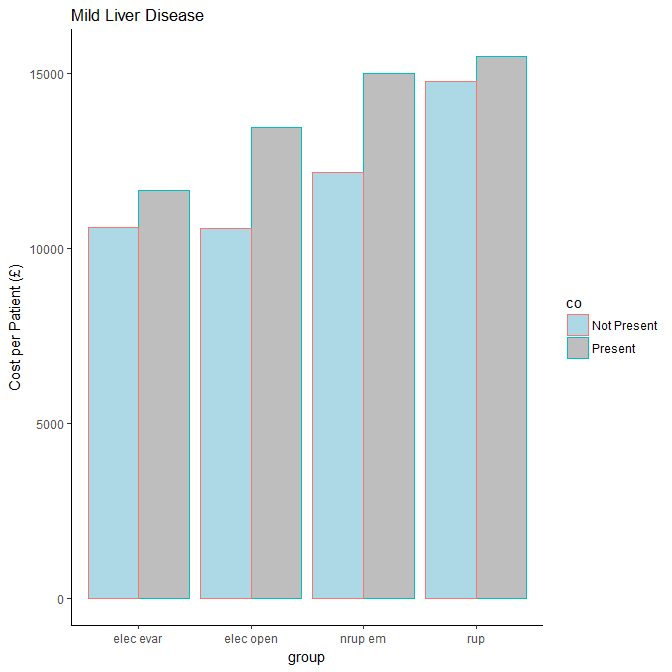
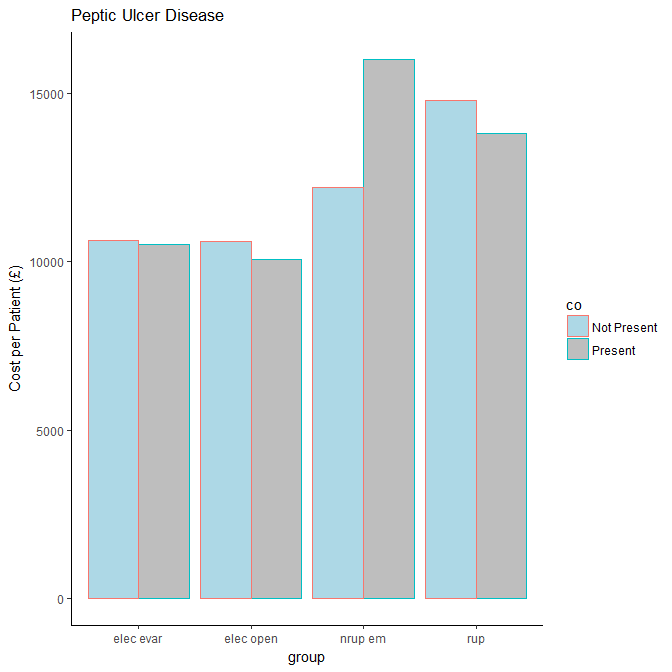
***Differences by Market Forces Factor Index***

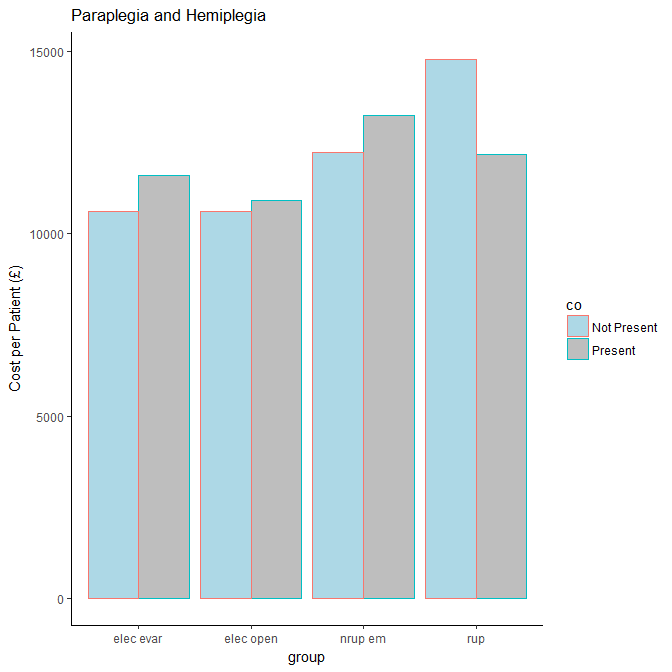
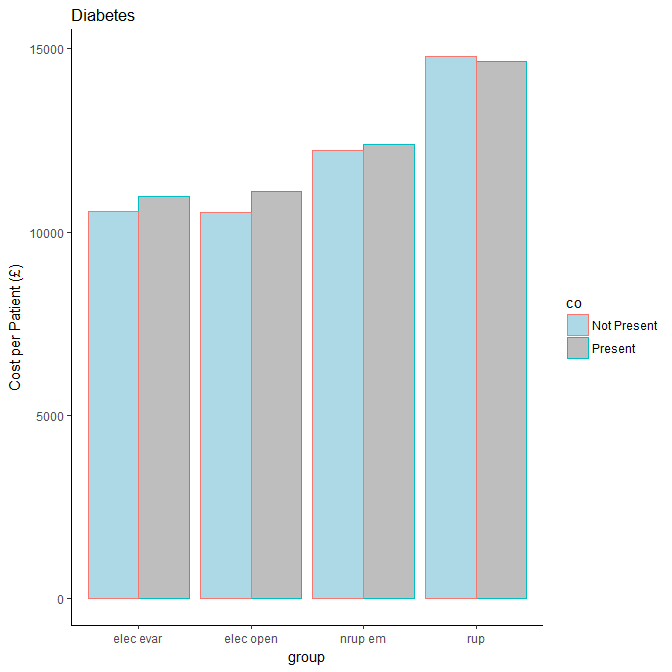
******

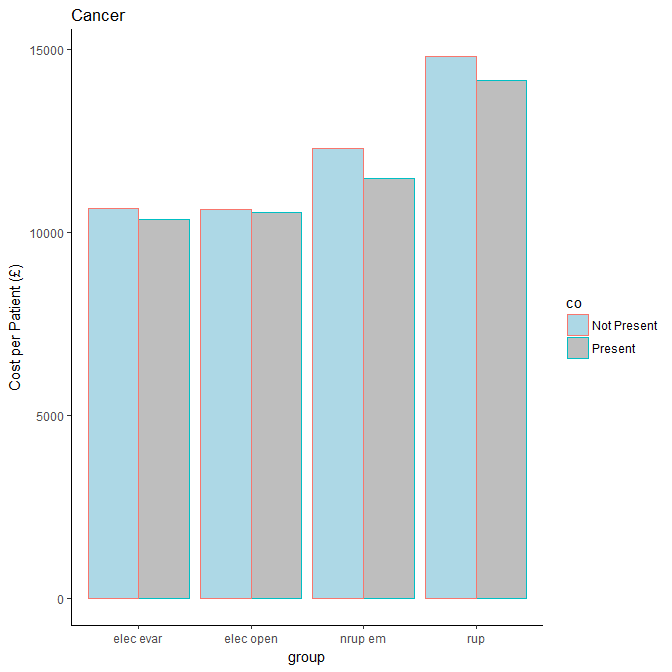
***Differences by the presence of comorbidities* (15 groups)**

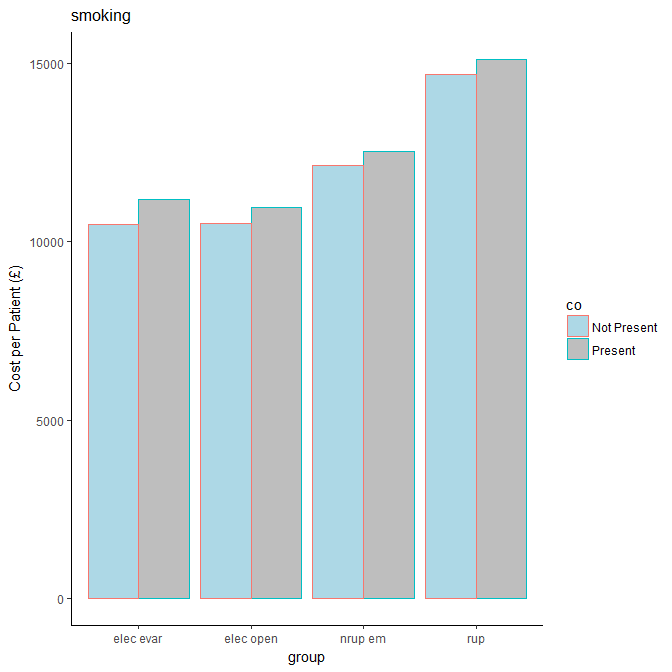
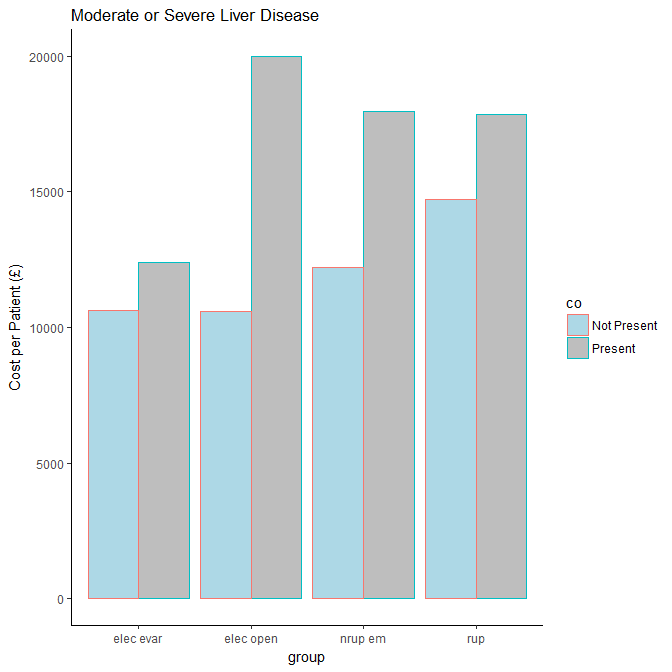


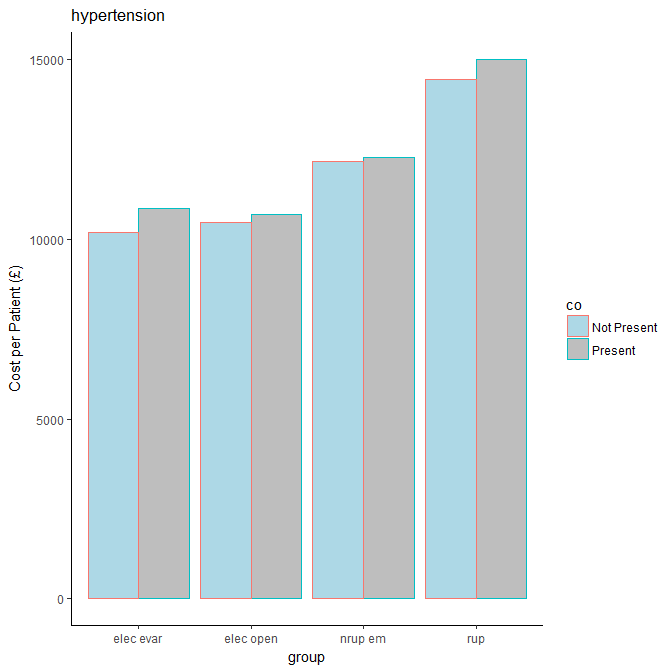
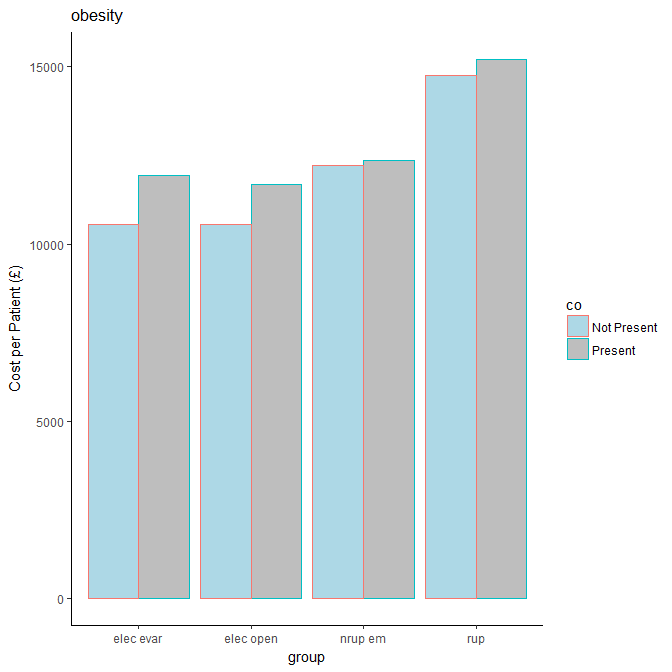


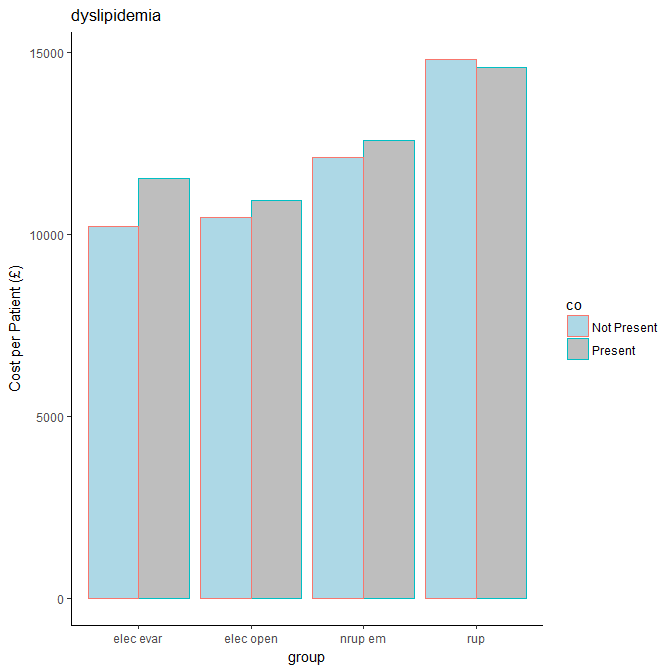












## Multiple Regression Models

The results of univariate analyses suggested that it is necessary to fit separate models for different clinical groups. Tables below presents the regression models (note, only significant covariates were retained, p < 0.05).

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Model 1: Ruptured AAA**  (GLM with Gamma Distribution and Log Link) | | Predicting total cost of index episode | | | | | |
| Coefficient | | SE | | EXP of coefficient | |
| Intercept | | 6.8837 | | 0.2412 | | 976.253 | |
| Having Open Repair | | 0.1424 | | 0.0364 | | 1.153 | |
| HES year | |  | |  | |  | |
| * 200809 | | 0 | | 0 | | 1 | |
| * 200910 | | 0.2180 | | 0.0471 | | 1.244 | |
| * 201011 | | 0.3940 | | 0.0473 | | 1.483 | |
| * 201112 | | 0.3847 | | 0.0473 | | 1.469 | |
| * 201213 | | 0.4031 | | 0.0478 | | 1.496 | |
| * 201314 | | 0.3721 | | 0.0495 | | 1.451 | |
| * 201415 | | 0.5015 | | 0.0518 | | 1.651 | |
| Died in hospital within the admission | | -0.1541 | | 0.0279 | | 0.857 | |
| IMD (continuous between 0 and 100) | | 0.0020 | | 0.0009 | | 1.002 | |
| MFF index of the hospital | | 2.0187 | | 0.2141 | | 7.529 | |
| Having COPD | | 0.0957 | | 0.0321 | | 1.100 | |
| Having Moderate/Severe Liver Disease | | 0.2412 | | 0.1082 | | 1.273 | |
| Annual Volume of the hospital | | 0.0010 | | 0.0003 | | 1.001 | |
| **Model 2: Nonrup Emergency AAA**  (GLM with Gamma Distribution and Log Link) | Predicting total cost of index episode | | | | | |
| Coefficient | | SE | | EXP of coefficient | |
| Intercept | 6.0290 | | 0.2089 | | 415.2841 | |
| Having Open Repair | 0.0574 | | 0.0281 | | 1.0590 | |
| HES year |  | |  | |  | |
| * 200809 | 0 | | 0 | | 1 | |
| * 200910 | 0.1079 | | 0.0482 | | 1.1140 | |
| * 201011 | 0.2448 | | 0.0482 | | 1.2774 | |
| * 201112 | 0.3660 | | 0.0481 | | 1.4419 | |
| * 201213 | 0.3820 | | 0.0478 | | 1.4652 | |
| * 201314 | 0.3631 | | 0.0487 | | 1.4378 | |
| * 201415 | 0.4632 | | 0.0500 | | 1.5891 | |
| In-hospital Death | 0.2015 | | 0.0439 | | 1.2232 | |
| Living in Urban Area | 0.0765 | | 0.0318 | | 1.0795 | |
| MFF index of the hospital | 2.6676 | | 0.1901 | | 14.4053 | |
| Having COPD | 0.1044 | | 0.0314 | | 1.1101 | |
| Having Moderate/Severe Liver Disease | 0.3170 | | 0.1644 | | 1.3731 | |
| Annual Volume of the hospital | 0.0006 | | 0.0003 | | 1.0006 | |

|  |  |  |  |
| --- | --- | --- | --- |
| **Model 3: Elective Open Repair**  (GLM with Gamma Distribution and Log Link) | Predicting total cost of index episode | | |
| Coefficient | SE | EXP of coefficient |
| Intercept | 6.4060 | 0.1761 | 605.4853 |
| HES year |  |  |  |
| * 200809 | 0 | 0 | 1 |
| * 200910 | 0.1054 | 0.0262 | 1.1111 |
| * 201011 | 0.2596 | 0.0276 | 1.2964 |
| * 201112 | 0.2800 | 0.0279 | 1.3231 |
| * 201213 | 0.2740 | 0.0290 | 1.3153 |
| * 201314 | 0.2479 | 0.0292 | 1.2813 |
| * 201415 | 0.3321 | 0.0311 | 1.3939 |
| In-hospital Death | 0.5027 | 0.0360 | 1.6532 |
| Age (continuous) | 0.0023 | 0.0010 | 1.0023 |
| IMD (continuous between 0 and 100) | 0.0026 | 0.0006 | 1.0026 |
| Distance from home to the hospital | 0.0018 | 0.0007 | 1.0018 |
| Living in Urban Area | 0.0733 | 0.0197 | 1.0760 |
| MFF index of the hospital | 2.1551 | 0.1469 | 8.6285 |
| Having COPD | 0.0581 | 0.0204 | 1.0598 |
| Having Mild Liver Disease | 0.1553 | 0.0717 | 1.1681 |
| Having Renal Disease | 0.1300 | 0.0442 | 1.1388 |
| Annual Volume of the hospital | 0.0002 | 0.0002 | 1.0002 |
| **Model 4: Elective EVAR**  (GLM with Gamma Distribution and Log Link) | Predicting total cost of index episode | | |
| Coefficient | SE | EXP of coefficient |
| Intercept | 6.1897 | 0.0685 | 487.6916 |
| Male Patient | -0.0402 | 0.0141 | 0.9606 |
| HES year |  |  |  |
| * 200809 | 0 | 0 | 1 |
| * 200910 | 0.0652 | 0.0179 | 1.0673 |
| * 201011 | 0.1478 | 0.0176 | 1.1593 |
| * 201112 | 0.3084 | 0.0171 | 1.3613 |
| * 201213 | 0.4694 | 0.0173 | 1.5990 |
| * 201314 | 0.4638 | 0.0171 | 1.5901 |
| * 201415 | 0.5474 | 0.0176 | 1.7288 |
| In-hospital Death | 0.5211 | 0.0387 | 1.6838 |
| IMD (continuous between 0 and 100) | 0.0021 | 0.0003 | 1.0021 |
| Distance from home to the hospital | 0.0009 | 0.0003 | 1.0009 |
| Living in Urban Area | 0.0620 | 0.0114 | 1.0640 |
| MFF index of the hospital | 2.3655 | 0.0612 | 10.6492 |
| Having Coronary Artery Disease | 0.0458 | 0.0108 | 1.0469 |
| Having COPD | 0.0284 | 0.0104 | 1.0288 |
| Having Renal Disease | 0.0784 | 0.0182 | 1.0816 |
| Annual Volume of the hospital | 0.0007 | 0.0001 | 1.0007 |

It should be noted that a log link was used in these models. A positive coefficient means the factor has a positive impact on the outcome (i.e. having the factor or positively increase the factor would increase the total cost). The predicted total costs given patient and provider characteristics can be calculated from the models as follow:

For example, for a ruptured AAA patient 74 years old with an IMD score of 20, no comorbidity, received an Open Repair in 200809 from a vascular centre which had an annual AAA volume of 60 and MFF index of 1, and the patient did not die within the admission; the predicted total costs of the index episode for this patient is £ £9,374.42 in 2014/15 price.

To give the model results a more meaningful interpretation, we can look at the exponential of the coefficients. These give us an idea of the magnitude of the effect of a factor. For example, taking a look at the model for ruptured cases, the coefficient for in-hospital death is -0.1541. The exponential of this is 0.857. This means if a patient died in hospital, the total costs would decrease by a factor of 0.857 compared to the scenario if the patient did not die.

As can be seen, the effect of volume is significantly positive in all models. This means higher volume centres are associated with higher costs in all cases after controlling for other case-mix factors.

# Discussion

This study represents the first study in England to use the HES data to estimate the hospital costs of the index episode for AAA repairs. It used individual patient data on 42,960 AAA patients treated in 188 hospitals across England between 1st Apr 2008 and 28th Feb 2015 to examine the variation of costs by patient and provider characteristics. Mean total costs (in 2014/15 price) for ruptured, non-ruptured emergency, elective open repair, and elective EVAR were £14779.3, £12235.8, £10607.4, and £10616.8 respectively. Costs varied with certain patient, provider characteristics, and hesyear as shown in descriptive univariate analyses above. We found two provider characteristics that had significantly positive effects on the total costs after controlling for other case-mix factors: the annual elective volume and the market forces factor index. Higher volume centres were associated with higher costs and higher level of unavoidable costs for a location (reflected by the MFF index) were also associated with higher costs. This certainly has important implications for cost-effectiveness modelling to evaluate services reconfiguration options. On the one hand, we know that merging small units with bigger ones would improve outcomes such as in-hospital mortality. On the other hand, with the new findings from this study, merging small units with bigger ones would also increase costs significantly. Thus, it is important to properly model these two aspects of changes (outcomes and costs) to evaluate different services reconfigurations.

The large sample size of AAA patients and the broad range of information captured from the HES data allowed us to account for a wide range of covariates on the costs in the regression models. This type of analysis is only possible with the use of routine datasets as the traditional study designs (such as randomised controlled trials) are much smaller in size. Also, given the aim is to estimate the costs for use in economic models, the worked examples estimating the costs are presented in Chapter 6. Furthermore, these equations were converted into online tools (using the R/Shiny software – see xxx) where the users can input patient and provider characteristics to estimate the costs for specific AAA procedures.

However, there are some limitations associated with our study. There were missing fields and coding changes in the HES data across the years that made it challenging to generate the HRG codes for the episodes and match them with relevant unit costs. Nonetheless, thanks to the use of year-specific HRG grouping and year-specific reference cost datasets, the proportion of records that could not be grouped or costed was small. We also attempted to impute the missing data with minimum amount of assumptions. There seems no bias in using the results from analyses based on the imputed data as they were similar to those in the complete-case analyses.

Whilst this study is quite comprehensive in estimating detailed individual costs and analysing the costs against a range of patient and provider covariates, it is possible that all the relevant factors were not included. Similarly, there could be other patient factors that were not present in the HES data that could have an impact on the costs. Also, provider factors such as surgeon-specific volume and the teaching status could have an impact on costs and could be considered as covariates in the regression models.

The potential for the use of routine data to estimate costs is recognised and there have been many similar studies in other disease areas. This study encountered similar challenges to those studies in terms of dealing with missing data and issues with HRG grouper software; and the methods used statistical analyses are broadly similar. More detailed costing estimates such as this study can help understand the effect of patient and provider characteristics on costs and help improve the accuracy of economic models.

**References**

1. Health and Social Care Information Centre, NHS Digital. Admitted Patient Care Data Quality Note. (HES 2014-2015 Annual Refresh). <http://content.digital.nhs.uk/catalogue/PUB19124/hosp-epis-stat-admi-dq-note-2014-15-rep.pdf> (Accessed 20 August 2017)
2. The Information Centre, Department of Health (December 2003). Hospital Episode Statistics, Admitted Patient Care – England 2002-2003, Headline Figures. <http://content.digital.nhs.uk/catalogue/PUB03906/hosp-epis-stat-admi-head-figs-02-03-rep.pdf> (Accessed 20 August 2017)
3. Lakhani A, Coles J, Eayres D, Spence C, Sanderson C. Creative use of existing clinical and health outcomes data to assess NHS performance in England: part 2--more challenging aspects of monitoring. BMJ. 2005 Jun 25;330(7506):1486-92.
4. Holt PJ, Poloniecki JD, Thompson MM. Multicentre study of the quality of a large administrative data set and implications for comparing death rates. Br J Surg. 2012 Jan;99(1):58-65.
5. A guide to the market force factor 2014/15, NHS England https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\_data/file/300859/A\_guide\_to\_the\_Market\_Forces\_Factor.pdf