Refreshing the Formulae for CCG Allocations

For allocations to Clinical Commissioning Groups from 2016-17

Report on the methods and modelling

Authors:
Michael Chaplin, Sam Beatson[1], Yiu-Shing Lau[2], Christina Davies,
Chris Smyth, Jeremy Burrows, Richard Weir and Roman Tatarek-Gintowt

NHS England, Analytical Services (Finance) – Allocations team

[1] Also Fellow at King's College London, director of Statisi Ltd
[2] Also University of Manchester, Manchester Centre for Health Economics

NHS England INFORMATION READER BOX

Directorate		
Medical	Commissioning Operations	Patients and Information
Nursing	Trans. & Corp. Ops.	Commissioning Strategy
Finance		

Publications Gateway Reference:	**05102**
Document Purpose	Resources
Document Name	Refreshing the Formulae for CCG Allocations for allocations to Clinical Commissioning Groups from 2016-17 - Report on the methods and modelling
Author	NHS England, Analytical Services (Finance)
Publication Date	April 2016
Target Audience	CCG Accountable Officers, Directors of Finance, Research Community
Additional Circulation List	
Description	This paper sets out how the formula for CCG core allocations was developed (used for CCG allocations from 2016-17)
Cross Reference	NHS England Financial Allocations 2016/17 to 2020/21 and the allocations published on 8 January 2016 https://www.england.nhs.uk/resources/resources-for-ccgs/
Superseded Docs (if applicable)	N/A
Action Required	N/A
Timing / Deadlines (if applicable)	**N/A**
Contact Details for further information	Michael Chaplin 8E25 Quarry House Leeds LS2 7UE 0113 825 3680

Document Status

This is a controlled document. Whilst this document may be printed, the electronic version posted on the intranet is the controlled copy. Any printed copies of this document are not controlled. As a controlled document, this document should not be saved onto local or network drives but should always be accessed from the intranet.

Refreshing the Formulae for CCG Allocations

For allocations to Clinical Commissioning Groups from 2016-17 - Report on the methods and modelling

Version number: 1.1

First published: April 2016

Prepared by: NHS England, Analytical Services (Finance)

Classification: OFFICIAL

Promoting equality and addressing health inequalities are at the heart of NHS England's values. Throughout the development of the policies and processes cited in this document, we have:

Given due regards to the need to eliminate discrimination, harassment and victimisation, to advance equality of opportunity, and to foster good relations between people who share a relevant protected characteristic (as cited under the Equality Act 2010) and those who do not share it;

Given regard to the need to reduce inequalities between patients in access to, and outcomes from, healthcare services and in securing that services are provided in an integrated way where this might reduce health inequalities.

Contents

Executive Summary

A weighted capitation formula is used to set target shares of the national health budget for Clinical Commissioning Groups (CCGs). The weighted capitation formula assesses the relative need per head for health care services across the country adjusted for differences in unavoidable costs.

During 2015, NHS England Analytical Services undertook a major project to refresh and update most of the weighted capitation formulae used to set target shares for CCG core allocations. The formulae refreshed were the general and acute, maternity, prescribing and the Emergency Ambulance Cost Adjustment.

The refresh encompassed collecting a large volume of data and re-estimating the previous models using regression analysis to give an updated set of weights per head for age, morbidity and other factors. We followed the same approach and methodology used to develop the previous formulae, and used more recent data.

We used anonymised data at the individual level on the use of hospital services and anonymised, individual level data for GP practice registered lists. We also collected data on the characteristics of their area of residence and NHS health services.

For general and acute and maternity services, we predicted cost weighted utilisation of hospital services in 2013-14. The main explanatory factors were age, sex, and morbidity. For prescribing, we modelled the cost of drugs prescribed in 2013-14. The main explanatory factors were age, sex and morbidity indicators. After adjusting for relative capacity of health services, these are taken to be estimates of need per head.

This research provided new weights per head for the general and acute, maternity and prescribing formulae, and also a new index for the relative, unavoidable costs of providing emergency ambulance services across the country

This report sets out in detail our work to refresh the formulae. The refresh of the formulae was inevitably a highly technical project. This report is intended as technical report and record of the modelling. It explains in detail the data used, issues with the data and how these were addressed, the methodologies adopted for the modelling, and how the results were assessed. This report is not a statement of allocations' policy.

A technically based overview of the refresh is in section 2 for the general and acute modelling, and similarly at the start of the sections for prescribing, maternity and the EACA.

A less technical summary of the work is provided in the *Technical Guide to Allocation Formulae and Pace of Change: For 2016-17 to 202-21 revenue allocations to Clinical Commissioning Group and commissioning areas* (NHS England, 2016). The Technical Guide also sets out how the results from the modelling were used to set target core allocations for CCGs for 2016-17 to 2020-21.

The Advisory Committee on Resource Allocation (ACRA) is an independent, expert committee responsible for recommending the formulae for allocations to NHS England. ACRA oversaw and steered the work to refresh the formulae, including agreeing the work plan, approach and methodology. Regular reports on progress and interim results were discussed at ACRA's meeting and with its Technical Advisory Group (TAG). ACRA scrutinised the results of the refresh and recommended to the Chief Executive of NHS England that the refreshed formulae are adopted for core allocations to CCGs from 2016-17.

1 Introduction

Since 1976, a weighted capitation formula has been used to set target shares of the national health budget for each local area. Since 2013, this has been for Clinical Commissioning Groups (CCGs). The weighted capitation formula assesses the relative need per head for health care services across the country and differences in unavoidable costs.

During 2015, NHS England Analytical Services refreshed and updated most of the weighted capitation formulae used to set target shares for CCG core allocations. The formulae were general and acute, maternity, prescribing and the Emergency Ambulance Cost Adjustment.

The refresh encompassed collecting a large volume of data and re-estimating the previous models using regression analysis to give an updated set of weights per head for age, morbidity and other factors. We followed the same approach and methodology used to develop the previous formulae, and used more recent data.

This report sets out in detail our work to refresh the formulae.

The Advisory Committee on Resource Allocation (ACRA) is an independent, expert committee responsible for recommending the formulae for allocations to NHS England. ACRA oversaw and steered the work to refresh the formulae, including agreeing the work plan, approach and methodology. Regular reports on progress and interim results were discussed at ACRA's meetings and with its Technical Advisory Group (TAG). ACRA scrutinised the results of the refresh and recommended to the Chief Executive of NHS England that the refreshed formulae are adopted for core allocations to CCGs from 2016-17. Our gratitude is extended to members of both groups for their input.

A common dataset was used for most of the formulae. We are grateful to the Health and Social Care Information Centre for providing us with patient level, hospital activity data from the Secondary Uses Services linked to GP practice registrations data from the Personal Demographic Service. The patient level data were anonymised – no names, postcodes, or date of birth were included. The data were accessed at a secure data facility from which no patent level data could be removed.

The weighted capitation formula sets target shares of the national budget. Actual allocations are set by pace of change policy, which determines how far CCGs' actual allocations are moved towards their target allocations each year. Pace of change policy is outside the scope of this report.

Sections 2 to 8 cover general and acute services, section 9 maternity, section 10 prescribing, and section 11 the emergency ambulance cost adjustment.

The formulae for mental health and specialised services are not covered in this report.

This is intended as a technical report, including much detail on how we managed the data as a reference document for those in the future who may refresh this work. This is not a policy document. Reference should be made to other papers for allocations policy, a summary of the formula and how allocations are calculated.

2 General & Acute: Overview

2.1 Introduction

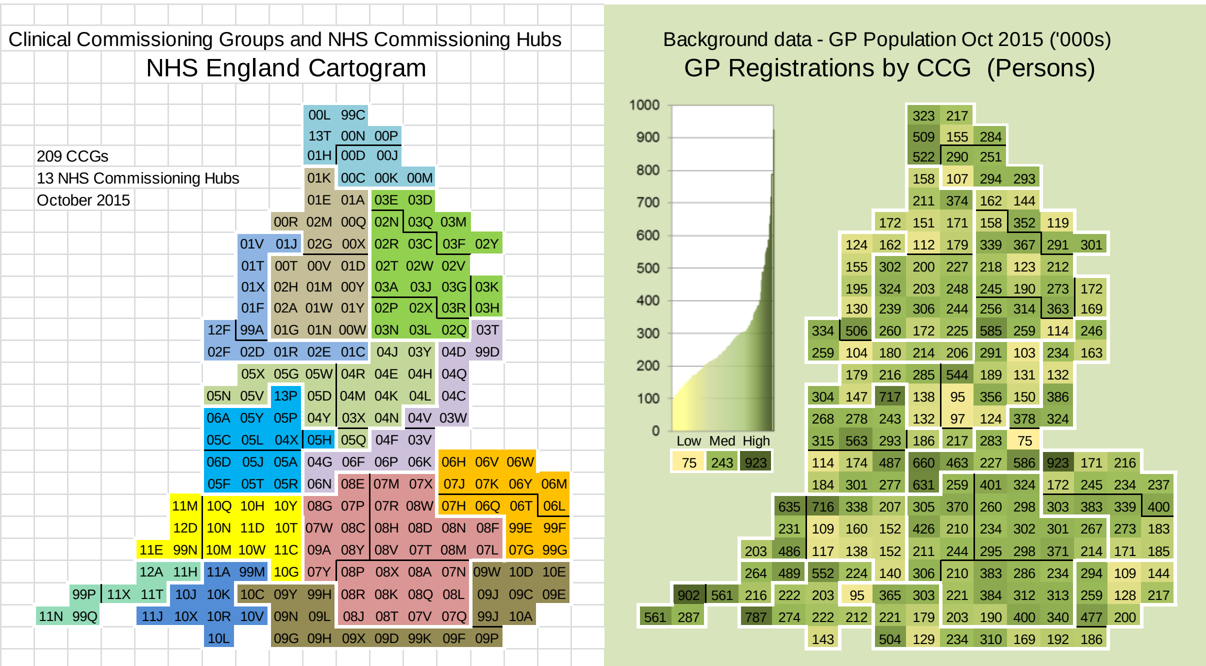
This report explains how NHS England Analytical Services updated and refreshed the Nuffield Trust's Person Based Resource Allocation (PBRA) approach to developing need weights by GP practice by age-sex group for general and acute services.

The weights from Nuffield's PBRA were used to inform funding allocations to Clinical Commissioning Groups (CCGs) in 2014-15 and 2015-16.

The work by Nuffield has been updated and refreshed to inform allocations to CCGs from 2016-17. The refresh of the general and acute formula is often referred to in this report as the PBRA 2011 formula refresh.

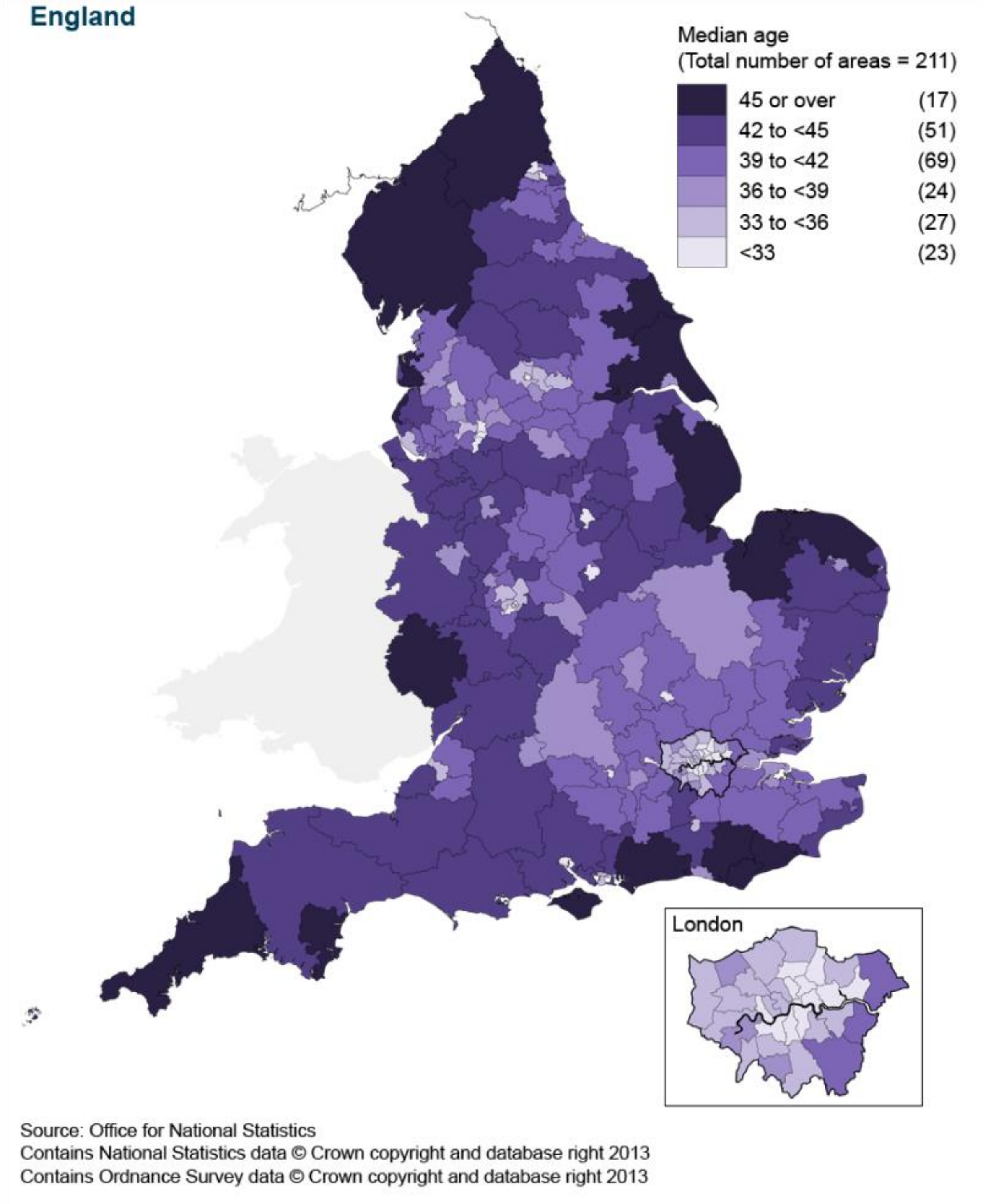
Maps of CCGs are shown in Figure 2.1 and Figure 2.2.

Figure 2.1: NHS England CCGs (cartograms)



Concise data maps, giving equal visual weighting for each CCG. Standard projection maps tend to give prominence to large rural expanses, while small urban areas disappear. CCG cartograms help to emphasise our commitment to a 'fair shares' formula, giving equal prominence to every CCG, regardless of geographical area or population.

Figure 2.2: NHS England CCGs



Source: ONS website (see below[1])

[1] As of April, 2015, three CCGs had merged, namely: 00F NHS Gateshead CCG; 00G NHS Newcastle North and East CCG, and 00H NHS Newcastle West CCG. These have formed Newcastle Gateshead CCG (code 13T) and thus there were 209 CCGs at the time of obtaining figure 1 from the ONS website.

2.2 Overview of objectives

In summary, our objectives were as follows.

2.2.1 Model health care costs

To model the effect on the cost per person of all hospital activity in 2013-14 of patient needs and supply variables for the years 2011-12 and 2012-13 using person level data. We considered a large number of variables expected to impact the cost of patient care. Some variables are likely to be associated with a higher cost of care, for example, higher levels of morbidity, whereas some are likely to be associated with a lower cost of care, for example being in a younger age group).

2.2.2 Predict health care costs

To predict the estimated cost per head of care in a validation sample using the parameters we have estimated from the model, and to measure the difference between observed and predicted costs in the validation sample. Thus, we are aiming to see how well our modelled costs are able to reflect the observed costs in a separate and exclusive sample, and in so doing, attempting to identify a preferred model for use in the final application.

2.2.3 Generate age-sex group predicted weights

To generate predicted weights by age-sex group for each GP practice registered list as at October 2015 using the model we deem to be the best predictor, taking into account plausibility, parsimony and predictive power. These weights can then be used in the allocations process to inform target allocations to CCGs.

2.3 Overview of data employed

2.3.1 Secondary Uses Service (SUS) Payments by Results data

PBRA (2011) used person level Hospital Episode Statistics (HES) for their hospital activity data to be modelled. The alternative to HES data used in this refresh is the Secondary Uses Service (SUS) data, from which HES data are derived.

We were informed through a consultative process with the Health and Social Care Information Centre (HSCIC) that SUS Payments by Results (PbR) data have a number of advantages over HES data, in particular the SUS PbR data already have costs included in the data set where there are mandatory Payment by Results tariffs.

The SUS PbR data are cleaned by the HSCIC to a lesser extent that HES, but we were able ourselves to implement many of the cleaning processes used for HES on the SUS PbR dataset.

There were a small number of fields we required that were not available in the SUS PbR data but are in HES. The HSCIC created these for us in the SUS PBR data extract.

The SUS PbR data were provided to NHS England as several segregated datasets, namely: admitted patient care (APC); outpatients (OP); critical care (CC) and accident and emergency (A&E).

The SUS PbR data were provided for the financial years 2010-11 to 2014-15. The data contain a wealth of detail for each episode; start and end dates, diagnoses operative procedures, consultant specialty, elective or non-elective admission, provider, and Health Resource Group (HRG).

2.3.2 Patient level GP Practice data

We furthermore obtained person level data from the HSCIC's personal demographic services (PDS). We were provided with PDS snapshots of the GP registration list at the person level, taken on 1 April in each financial year from 2009-10 to 2014-15. We were provided from the PDS data each registered person's age, sex, month of birth, small geographical area of residence (LSOA) and GP practice code.

2.3.3 Linking SUS PbR data and PDS data

Hospital activity from the SUS PbR data was linked to the PDS data at the person level.

The HSCIC created special identifiers (ID) for the project for each APC episode, OP attendance, A&E attendance and critical care. The HSCIC also created special identifiers for each person in the PDS data, and provided a 'bridge' file mapping the IDs in the SUS PbR data to the PDS data.

The linking of the different data sets was undertaken by NHS England to create a record for each individual in the PDS data set that also contained all their APC episodes, OP attendances, A&E attendances and critical care days in the SUS PbR dataset. This included people who had received no hospital care during the period recorded in the SUS PbR dataset.

In practice, we linked data for 2011-12 to 2013-14 at the person level, creating a record of the hospital care each person received over this three year period, for those registered with a GP practice in 2013- 2014.

The bridge file also allowed NHS England to combine hospital APC episodes into spells at the person level.

2.3.4 Attributed variables

Attributed supply and need variables refer to data that are not available at the person level, so the value of the small area where they live is given to each person in that small area. These include for example the percentage of the small area's population in each ethnic group. Other attributed variables include the distance to hospitals from each small area centroid.

We collected a wide range of attributed variables, from for example the 2011 Population Census and social security data published by the Department for Work and Pensions.

We sought to collect the same data for attributed variables tested by PBRA 2011. Some data were no longer available or had changed, and for these cases we collected the most comparable data available. We collected also additional data not tested by PBRA 2011.

We also took into account evidence to suggest the attributed variables could act as independent drivers in addition to age, sex and morbidity.

2.4 Overview of the methodology

2.4.1 Needs based formula

Conceptually, the formula aims to be need-based. Thus, the formula aims to follow the principle for allocations of equal opportunity of access to healthcare for equal need.

2.4.2 Based on the Nuffield Research

The PBRA refresh is based on, and updates, the Nuffield report *Updating and enhancing a resource allocation formula at general practice level based on individual level characteristics (person-based resource allocation)*, 2011 (PBRA 2011).

2.4.3 Purpose of the analysis

According to PBRA (2011), 'the purpose of the modelling is to estimate the contribution of supply and needs factors in explaining variations in health care costs at individual and GP practice level. The analysis is undertaken at person level but the calculation of relative expenditure needs is required at practice level…'

2.4.4 Estimate of health costs

The general and acute (G&A) formula is the modelled costs of individual persons' total G&A hospital healthcare in a given year, based on a number of explanatory variables which represent individuals' needs and the local supply of healthcare available to each person. The fundamental unit of analysis is therefore the person. The person and their prospective healthcare needs are therefore central to the formula.

2.4.5 Formula Expression

In mathematical notation, very simply we are modelling the following function:

$$C = f(N, S)$$

Equation 2.1

Where the dependent variable, C, is the cost of patient care per head in 2013-14 and N and S stand for needs and supply variables thought likely to affect patient need.

Supply variables are measures of local capacity in health care services. In actual allocations, the supply variables are set to the national average so that areas with lower costs per head due to low capacity are not penalised. This means allocations are based on estimated cost-weighted need rather than on estimated costs.

2.4.6 Dependent and independent variables

The dependent variable in the modelling is the aggregated cost of G&A hospital care in 2013-14 per person from the three datasets – admitted patient care (APC), outpatient attendances (OP) and accident and emergency (AE) attendances for each service user.

The independent variables, or right hand side variables, or explanatory variables used were age, sex, morbidity over the prior two years as recorded by diagnoses for inpatient attendances, a count of the number of morbidities recorded, interactions between morbidity Chapters (ICD10 chapters) as a co-morbidity measure, whether the care provided was on a private basis, whether the patient was newly registered at a GP practice. In addition, a large selection (initially over 200) of LSOA level, GP practice level, and hospital level ‘attributed needs and supply’ variables were tested and a sub-set included in the final model, having been selected through modelling.

Morbidity was defined by grouping ICD10 diagnoses into 152 groups, as defined by the HSCIC and following PBRA (2011).

2.4.7 Ordinary Least Squares models

Following PBRA (2011), we employed Ordinary Least Squares (OLS), using 2013-14 costs per person as our dependent variable and tested the variables described above to give the best model.

We discussed alternative models such as multi-level modelling and panel data approaches, but PBRA (2011) had found that these offered little advantage over OLS with such a large dataset.

Furthermore, the objective given to us was to update and refresh the PBRA formula, rather than develop a new model. Pragmatism also favoured OLS due to it being less demanding on computing power available; as it was the OLS regressions took many hours to run.

2.4.8 Selection of the year for the dependent variable

Providers can submit updates to the SUS PbR data set well after the end of the relevant year.

Costed activity in 2013-14 was used for our dependent variable, as this was the last year in the dataset for which we were confident that the whole year’s activity, payment by results’ tariffs, and diagnoses data had been included. For this year, updates to SUS PbR could have been submitted for just over one year prior by the date the data extract was taken for this project. We were advised by the HSCIC that for the last financial quarter of 2014-15, and particularly the last month of the financial year, many SUS PbR data fields had not yet been submitted or verified by the date the data were extracted for this project. We therefore took the decision to use the latest year with complete data for our modelling (2013-14).

2.4.9 Excluding specialised services and maternity activity

The way in which the costs of maternity, mental health and specialised services were removed during the calculation of the dependent variable can be found in section 3. Morbidity indicators for maternity, mental health and specialised services in 2011-12 and 2012-13 were included as explanatory variables for the modelling of the cost of G&A in 2013-14.

2.4.10 Estimation Sample

The estimation sample is the sample used for the selection of variables and estimation of the variable coefficients at the individual level. Separate validation samples were used to predict costs and test for the performance against observed costs.

2.4.11 Sampling Decisions

Due to the large size of the dataset and limitations in available computing power and time, we elected to use an estimation sample consisting of 15% of persons in the Personal Demographic Services (PDS) as at April 2013. PBRA (2011) used a 10% sample.

We felt that given the increased computing power available to us since PBRA (2011), taking a larger sample could be beneficial to our modelling and would certainly do no harm. Using larger samples could offer increased accuracy in terms of the parameter outputs and hence offer marginal improvements in our predictions.

2.4.12 Representativeness of estimation sample

We examined the distributions of costs, need and supply variables between the estimation sample and 100% of the cleaned dataset and found our sample to be representative of the whole data set.

2.4.13 Selecting the final model

Age, sex and morbidity markers were included in all of the models.

A plausible, but exhaustive set of attributed variables likely to affect costs was included in the initial specification of the model; this was then refined and variables excluded to give the plausible model with the best explanatory power.

The decision on which attributed variables to include was taken mainly on the results of quantitative variable selection procedures, but also involved analytical judgement.

The analytical judgement involved looking at the variables that had been selected during the quantitative variable selection procedures and their corresponding coefficients. Whether or not the sign (positive or negative) of the attributed variable aligned with prior theories about the effect of the variable on healthcare costs was an important consideration in whether to include or exclude the attributed variable.

Judgement was only exercised over the attributed need and supply variables which have a relatively small role in the models compared with age, sex and the morbidity markers.

2.4.14 Assessment of robustness of the model

The coefficients from the modelling using the estimation sample were applied to the separate validation sample of 15% of registered patients at April 2013 to calculate predicted cost at the person level. The validation sample was mutually exclusive to the 15% estimation sample.

The assessment of explanatory power of the model was undertaken on a 15% sample of GP practices, comparing predicted with observed costs per head. The robustness of the model for these 15% of GP practices was assessed using the R-squared statistic, mean absolute error (the average absolute difference between the predicted and actual costs), the proportion of GP practices where the predicted and actual costs differed by more than 10%, and a number of other measures.

2.4.15 Weights for allocations

The predicted need weights per head by age-sex group for each GP practice from the final model can be can be applied to the latest registered lists. They were applied to October 2015 registered lists as part of this project, and these weights were used to inform target allocations to CCGs. The latter is set out in *Technical Guide to Allocation Formulae and Pace of Change: For 2016-17 to 202-21 revenue allocations to Clinical Commissioning Group and commissioning areas* (NHS England, 2016).

2.4.16 Unmet Need

The question of so-called unmet need is often raised as the above methodology estimates need from observed utilisation. ACRA and NHS England recognised this issue and have taken steps to address this and other areas not captured by the formula. This is covered in *Technical Guide to Allocation Formulae and Pace of Change: For 2016-17 to 202-21 revenue allocations to Clinical Commissioning Group and commissioning areas* (NHS England, 2016).

2.5 Governance

2.5.1 Data anonymisation

Patient level data was anonymised to maintain patient confidentiality, a key concern for clinicians, patients, service users and for NHS England.

Specially constructed patient identifiers were created by the HSCIC. We were not provided with individuals' NHS numbers, exact dates of birth (only the month and year of birth were proved), or home addresses and postcodes The lowest level of identifiable geographical domicile proved was the lower level super output area (LSOA).

2.5.2 Data approval and access

Beyond this data anonymisation process, the HSCIC worked jointly with the team in NHS England Analytical Services to ensure there was a formal process for approving the data to be provided, the individuals in NHS England who had access to the data, and ensuring that individual patients could not be identified through the combination of datasets made available. There was also a formal change control process in place.

Furthermore, the data were only made available to NHS England Analytical Services at a secure data facility, from which no patient level data could be removed.

2.5.3 Oversight

These governance processes were overseen by professionals in the HSCIC at a number of levels and specialties (for example senior management, data management, information governance). Moreover, only a specific designated team of a limited number of NHS England analysts and academics contracted by NHS England Analytical Services for this project, all of whom completed the information governance training required by the HSCIC. The academics appointed had also been through a further layer of governance, examination and formal contractual confidentiality agreements before being appointed by NHS England for this project.

2.6 Quality assurance

We undertook an in-depth quality assurance of our modelling. In particular we:

- followed the Nuffield methodology and approach;
- spent considerable time on the pre-processing and cleaning of data after its collection;
- examined the data for implausible values and compared our core data at a high level with other sources, such as the number of registered patients, activity volumes and waiting times. This for example identified issues with the registrations data provided to us from the personal demographic service which were corrected. We removed clear data errors from our data set;
- tested for duplication in the data by running programmes using for example month of birth (we did not hold day of birth), admission date, diagnoses, LSOA and GP practice.
- peer reviewed the coding in the software package;
- examined the results for plausibility, and investigated outliers;
- obtained advice from experts in the data to help with interpretation and to gain knowledge of data issues;
- compared the results in detail with the Nuffield models, which prompted further testing and exploration of changes;
- used a number of approaches to the modelling to help identify the most robust models;
- calculated statistical diagnostics tests to test if the models were robust;
- examined variables for plausibility in direction of influence; and
- presented all of the modelling and outputs for review by TAG and ACRA.

3 General & Acute: Costing of the data for financial year 2013-14

3.1 Introduction

This section sets out how the data was costed for the refresh of PBRA (2011), and thus how the dependent variable was constructed from the available data for the project, and provides summary statistics of this variable.

The objective of this stage of our analysis was to estimate the 'cost of care' received by each individual. The costs here are observed, actual costs, not predicted costs. The steps to achieve this were:

- to attribute an appropriate cost to each spell, visit or attendance in each dataset; and
- to aggregate all of the accumulated costs in the year 2013-2014 for each person from all their spells and attendances in the year.

The person level costs per head could also be aggregated at, for example GP practice level, to give an observed practice level cost of secondary care activity attributable to all patients at that practice.

The aim of this section is to provide comprehensive information on the steps that were taken to apply costs to the activity data

3.2 Glossary of terms

Out of necessity we will need to make use of terms such as spell, episode, healthcare resource group (HRG), treatment function code (TFC), main specialty code (MSC), national tariff, reference cost, market forces factor (MFF) and grouper. These are briefly defined here for convenience.

3.2.1 Episode/Spell

An episode is inpatient treatment under the responsibility of one consultant. A spell refers to a complete spell of care from admission to discharge. Most spells consist of one finished consultant episode (FCE). However, a minority of spells may involve more than one consultant and therefore more than one episode within the spell.

3.2.2 HRG (Healthcare Resource Groups)

The HRG refers to a grouping of similar clinical treatments with similar costs into a 'currency' for funding purposes and is also a record of patients' treatments. There are HRGs for most spells, outpatient attendances, and A&E attendances, and also for some diagnostic procedures.

3.2.3 MSC (Main Specialty Code) / TFC (Treatment Function Code)

A MSC refers to the code identifying the specialty as designated by the Royal Colleges. They may be defined by body systems, clinical function, disease or a combination of these. MSCs are a subset of TFCs which also include approved subspecialties and treatment specialties of lead care providers, including consultants.

3.2.4 National Tariff

Payment by results national tariff prices are a set of national prices and rules set by Monitor and NHS England, which are used by commissioners to fund services on the basis of the number of pre-defined sets of treatments provided.

3.2.5 MFF (Market Forces Factor)

The MFF refers to the percentage uplift to the national tariff based on variations in unavoidable costs for staff, buildings and equipment between providers due to their geographical location alone.

As the MFF is a measure of differences in unavoidable costs across the country due to employment and premises costs, it is therefore not relevant to estimating need and is not included in the costs per patient used in the modelling.

3.2.6 Reference Costs

Reference costs are average costs per unit of healthcare services submitted by secondary healthcare providers and collected by the Department of Health. They are used in the setting of payment by results national prices.

3.2.7 Grouper

The grouper is software employing an algorithm, which according to the HSCIC website[2] is: ‘used to group data and derive HRGs, support data quality, conduct what-if-modelling, and assess local reimbursement that will be received under PbR for the 2013/14 financial year.’

3.3 Costing different services

In this project, a cost is applied to each spell, each outpatient attendance, each A&E attendance and critical care days. This was undertaken for G&A, maternity and specialised services.

This part of the report focuses on G&A care, however, since the same costing processes were applied to maternity and specialised services they are also covered here. How the costs of maternity and specialised services were removed from the 2013-14 costs for the G&A models is also covered.

This section explains how the datasets for the costing were created, how the data cleaning was carried out and how the data was costed, including summary statistics.

Inpatient episodes spells (also called known as admitted patient care) and outpatient attendances were divided into G&A for refresh of PBRA (2011), maternity (MAT) for the refresh of the maternity formula, and specialised services (SPS) for the development of the specialised services formula (which is not covered in this report).

Table 3.1 illustrates that for the G&A formula, we aggregated patients’ APC, their OP visits and their A&E attendances. For maternity and specialised services, only APC and OP were considered in the aggregation of a person’s costs. The A&E data included no information on whether the attendances were for maternity or a specialised service. All A&E costs were included in the G&A costs. In all cases a mutually exclusive dataset was formed.

---

[2] http://www.hscic.gov.uk/article/2580/HRG4-201314-Local-Payment-Grouper

Table 3.1: Datasets for pricing to be applied

Dataset	Sub-datasets
Admitted Patient Care	General and Acute
	Maternity
	Specialised Services
Outpatients	General and Acute
	Maternity
	Specialised Services
Accident and Emergency	n/a – included as part of the G&A formula

3.4 Identifying maternity and specialised services

3.4.1 Maternity

Maternity was identified as detailed in Table 3.2 in the APC and OP datasets. These spells and outpatient attendances were removed from G&A dataset.

We removed well babies episodes from our dataset for the modelling. They have a zero national tariff price as the maternity costs are attributed to the mother's episode. Should there be subsequent complications either to the baby or mother, this would be captured in a separate episode.

3.4.2 Specialised Services (SPS)

SPS were identified using a flag that identified specialised services episodes and outpatient attendances based on the 2014/15 Prescribed Services (PSS) tool[3].

This was not entirely straight forward as the toolkit gave 'error' when it was unable to determine whether the episode or outpatient attendance was a specialised service. These were mainly due to errors or missing data in the SUS PbR data, even though the field in question often had no bearing on whether the episode or attendance was specialised or not. The nature of the input error was usually obvious and corrected through an iterative process of removing some of the errors, re-running, and focusing on the causes of the remaining 'error' output rows from the PSS toolkit. Once all typos and other input errors that could be fixed had been addressed, around 1% of hospital records were unable to be classified as either specialised or not. Patients for whom the tool did not work were treated as non-specialised.

Furthermore, in principle the PSS toolkit could classify some patients as specialised based on their 14th or higher diagnoses or procedure codes. However, the SUS PbR data provided by the HSCIC contained only up to 13 ICD10 diagnostic positions and procedure codes for each episode and attendance. If an episode or attendance had not been identified as specialised within the first 13 diagnoses and procedure codes it was classified for this project as non-specialised.

Analysis for us by HSCIC, using a dataset which included all diagnostic positions and procedure codes, indicated that only around 0.004% more episodes would have been classified as specialised if our dataset had had all of these data. No more outpatient attendances would have been classified as specialised if we had had all of these data.

[3] http://www.hscic.gov.uk/casemix/prescribedspecialisedservices

Once segregated off, the 2013-14 SPS dataset was costed in the same way as other APC and OPs.

Where a spell or outpatient attendance was identified as being both maternity and specialised services, it was included in the costs of specialised services and omitted from maternity costs.

3.4.3 Remaining G&A costs

The G&A cost dataset, after segregation in this way, did not include maternity or specialised services spells of care. The maternity dataset did not include any spells flagged as being specialised services. Specialised services may include maternity spells since specialised services took precedence in terms of which dataset to assign the spell to if flagged as both SPS and maternity.

Table 3.2: How maternity was identified in each dataset

Dataset	Identifying Variable	Level	Value	Description
APC / OP	Spell Core HRG	SP / AT	NZ04C	Ante-natal or Post-natal Observation age between 16 and 40 years with length of stay 0 days
APC / OP	Spell Core HRG	SP / AT	NZ04D	Ante-natal or Post-natal Observation age under 16 or over 40 years with length of stay 0 days
APC / OP	Spell Core HRG	SP / AT	NZ05C	Ante-natal or Post-natal Investigation age between 16 and 40 years with length of stay 0 days
APC / OP	Spell Core HRG	SP / AT	NZ05D	Ante-natal or Post-natal Investigation age under 16 or over 40 years with length of stay 0 days
APC / OP	Spell Core HRG	SP / AT	NZ06Z	Ante-natal or Post-natal Full Investigation with length of stay 0 days
APC / OP	Spell Core HRG	SP / AT	NZ07C	Ante-natal or Post-natal Observation age between 16 and 40 years with length of stay 1 day or more
APC / OP	Spell Core HRG	SP / AT	NZ07D	Ante-natal or Post-natal Observation age under 16 or over 40 years with length of stay 1 day or more
APC / OP	Spell Core HRG	SP / AT	NZ08C	Ante-natal or Post-natal Investigation age between 16 and 40 years with length of stay 1 day or more
APC / OP	Spell Core HRG	SP / AT	NZ08D	Ante-natal or Post-natal Investigation age under 16 or over 40 years with length of stay 1 day or more
APC / OP	Spell Core HRG	SP / AT	NZ09Z	Ante-natal or Post-natal Full Investigation with length of stay 1 day or more
APC / OP	Spell Core HRG	SP / AT	NZ10Z	Diagnostic and Therapeutic Procedures on Foetus
APC / OP	Spell Core HRG	SP / AT	NZ11B	Normal Delivery without CC
APC / OP	Spell Core HRG	SP / AT	NZ11C	Normal Delivery with Epidural with CC
APC / OP	Spell Core HRG	SP / AT	NZ11D	Normal Delivery with Epidural without CC
APC / OP	Spell Core HRG	SP / AT	NZ11E	Normal Delivery with Induction with CC
APC / OP	Spell Core HRG	SP / AT	NZ11F	Normal Delivery with Induction without CC
APC / OP	Spell Core HRG	SP / AT	NZ11G	Normal Delivery with Post-partum Surgical Intervention
APC / OP	Spell Core HRG	SP / AT	NZ12A	Assisted Delivery with CC
APC / OP	Spell Core HRG	SP / AT	NZ12B	Assisted Delivery without CC
APC / OP	Spell Core HRG	SP / AT	NZ12C	Assisted Delivery with Epidural with CC
APC / OP	Spell Core HRG	SP / AT	NZ12D	Assisted Delivery with Epidural without CC
APC / OP	Spell Core HRG	SP / AT	NZ12E	Assisted Delivery with Induction with CC
APC / OP	Spell Core HRG	SP / AT	NZ12F	Assisted Delivery with Induction without CC
APC / OP	Spell Core HRG	SP / AT	NZ12G	Assisted Delivery with Post-partum Surgical Intervention
APC / OP	Spell Core HRG	SP / AT	NZ13A	Planned Lower Uterine Caesarean Section with CC
APC / OP	Spell Core HRG	SP / AT	NZ13B	Planned Lower Uterine Caesarean Section without CC
APC / OP	Spell Core HRG	SP / AT	NZ14A	Emergency or Upper Uterine Caesarean Section with CC
APC / OP	Spell Core HRG	SP / AT	NZ14B	Emergency or Upper Uterine Caesarean Section without CC
APC / OP	Spell Core HRG	SP / AT	NZ15Z	Caesarean Section with Eclampsia, Pre-eclampsia or Placenta Praevia
APC / OP	Spell Core HRG	SP / AT	MA17C	Dilation and Evacuation - less than 14 weeks gestation
APC / OP	Spell Core HRG	SP / AT	MA17D	Dilation and Evacuation - 14 to 20 weeks gestation
APC / OP	Spell Core HRG	SP / AT	MA20Z	Medical or Surgical Termination of Pregnancy - 20 weeks gestation or more
APC / OP	Spell Core HRG	SP / AT	MA18C	Medical Termination of Pregnancy - less than 14 weeks gestation
APC / OP	Spell Core HRG	SP / AT	MA18D	Medical Termination of Pregnancy - 14 to 20 weeks gestation
APC / OP	Spell Core HRG	SP / AT	MA19A	Vacuum Aspiration with Cannula - less than 14 weeks gestation
APC / OP	Spell Core HRG	SP / AT	MA19B	Vacuum Aspiration with Cannula - 14 to 20 weeks gestation
APC / OP	Spell Core HRG	SP / AT	MB08Z	Threatened or Spontaneous Miscarriage
APC / OP	Spell Core HRG	SP / AT	PB01Z	Major Neonatal Diagnoses

Dataset	Identifying Variable	Level	Value	Description
APC / OP	Spell Core HRG	SP / AT	PB02Z	Minor Neonatal Diagnoses
APC / OP	Spell Core HRG	SP / AT	PB03Z	Well Baby [Dropped from all datasets]
APC / OP	Well Baby Flag (TFC 424)	EP / AT	“Y”	Well Baby Episode [Dropped from all datasets]
APC / OP	Treatment Function Code	EP / AT	424	Well Baby [Dropped from all datasets]
APC / OP	Treatment Function Code	EP / AT	501	Obstetrics
APC / OP	Treatment Function Code	EP / AT	560	Midwife episode

Note: In the third column of the table, the level of identification could be at episode (EP), spell (SP), or attendance (AT) level.

3.5 Removal of other data

3.5.1 Reasons for removal

Each dataset required a stage whose purpose was the removal of unwanted and erroneous data. This took place at episode/attendance level. By unwanted it is meant out of the scope of the refresh of PBRA (2011), such as for example, mental health services (for which there is a separate formula). The rules for identification of such data are provided below. This process was the same in its essence for the now segregated G&A dataset, the maternity dataset and the specialised services dataset.

Table 3.3 outlines the broad categories for removal of data. Table 3.4 lists the activity removed in detail.

Table 3.3: Reason for removal of rows of data from the datasets (data cleaning)

Reason for Removal
Out of scope HRG or Treatment Function Code / Main Specialty Code or Well Baby flag.
Out of scope due to Military, Prisoner, Overseas or Private spell
Duplicate Episode in the Data

Note that there is a heavier reliance on treatment function code or main specialty code[4] in the OP data in Table 3.3. For outpatients, especially since 2011, a significant proportion of attendances are coded to a 'catch-all' first or follow up attendance spell HRG (beginning with 'WF'). The TFC or MSC is used as a proxy for the treatment in such cases to identify which dataset the outpatient data belonged to, or whether it was out of scope.

3.5.2 Out of scope activity

The following categories of data were removed as out of scope: dental and maxillo-facial surgery (Max-Fax), which are directly commissioned by NHS England, genito-urinary medicine (GUM), which is commissioned by local authorities and mental health, which is covered in the separate mental health services formula.

Other categories removed were military, prisoners (both directly commissioned by NHS England), and private patients, all of whom were identified as 'flags' (dummy variables) in the dataset, and patients from overseas who by definition are not registered with a GP practice.

[4] However, if one was missing, it could be replaced with the other, and in all cases both the treatment function code and main specialty code were checked to identify out of scope activity.

Table 3.4: Undesired HRGs, TFCs and MSCs

Dataset	Identifying Variable	Level	Value	Description
APC / OP	Spell Core HRG	SP / A	CZ30Y	Minor Extraction of Tooth 19 years and over
APC / OP	Spell Core HRG	SP / A	CZ31Y	Minor Dental Biopsy 19 years and over
APC / OP	Spell Core HRG	SP / A	CZ32U	Dental Fitting or Insertion Procedures 18 years and under
APC / OP	Spell Core HRG	SP / A	CZ32Y	Dental Fitting or Insertion Procedures 19 years and over
APC / OP	Spell Core HRG	SP / A	CZ33Y	Minor Dental Restoration Procedures 19 years and over
APC / OP	Spell Core HRG	SP / A	CZ34U	Minor Dental Procedures 18 years and under
APC / OP	Spell Core HRG	SP / A	CZ34Y	Minor Dental Procedures 19 years and over
APC / OP	Spell Core HRG	SP / A	CZ35U	Adjustment of Dental Device 18 years and under
APC / OP	Spell Core HRG	SP / A	CZ35Y	Adjustment of Dental Device 19 years and over
APC / OP	Spell Core HRG	SP / A	CZ36Y	Dental Excision or Biopsy Procedures 19 years and over
APC / OP	Spell Core HRG	SP / A	CZ37Y	Surgical Removal of Tooth 19 years and over
APC / OP	Spell Core HRG	SP / A	CZ38U	Creation of Dental Impression 18 years and under
APC / OP	Spell Core HRG	SP / A	CZ38Y	Creation of Dental Impression 19 years and over
APC / OP	Spell Core HRG	SP / A	CZ39U	Intermediate Dental Procedures 18 years and under
APC / OP	Spell Core HRG	SP / A	CZ39Y	Intermediate Dental Procedures 19 years and over
APC / OP	Spell Core HRG	SP / A	CZ40Y	Major Surgical Removal of Tooth 19 years and over
APC / OP	Spell Core HRG	SP / A	CZ41Y	Major Dental Procedures 19 years and over
APC / OP	Spell Core HRG	SP / A	CZ42Y	Extraction of Multiple Teeth 19 years and over
APC / OP	Spell Core HRG	SP / A	CZ16N	Minor Maxillo-Facial Procedures with CC
APC / OP	Spell Core HRG	SP / A	CZ16Q	Minor Maxillo-Facial Procedures without CC
APC / OP	Spell Core HRG	SP / A	CZ17U	Intermediate Maxillo-Facial Procedures 18 years and under
APC / OP	Spell Core HRG	SP / A	CZ17V	Intermediate Maxillo-Facial Procedures 19 years and over with CC
APC / OP	Spell Core HRG	SP / A	CZ17Y	Intermediate Maxillo-Facial Procedures 19 years and over without CC
APC / OP	Spell Core HRG	SP / A	CZ18R	Major Maxillo-Facial Procedures 19 years and over
APC / OP	Spell Core HRG	SP / A	CZ18U	Major Maxillo-Facial Procedures 18 years and under
APC / OP	Spell Core HRG	SP / A	PA51Z	Child Safeguarding (Welfare and Protection)
APC / OP	Spell Core HRG	SP / A	PA52Z	Behavioural Disorders with length of stay 1 day or less
APC / OP	Spell Core HRG	SP / A	PA52B	Behavioural Disorders with length of stay between 2 and 7 days
APC / OP	Spell Core HRG	SP / A	PA52C	Behavioural Disorders with length of stay 8 days or more
APC / OP	Spell Core HRG	SP / A	PA53A	Eating Disorders with length of stay less than 8 days
APC / OP	Spell Core HRG	SP / A	PA53B	Eating Disorders with length of stay 8 days or more
APC / OP	Spell Core HRG	SP / A	VC26Z	Rehabilitation for Drug and Alcohol Addiction
APC / OP	Spell Core HRG	SP / A	VC28Z	Rehabilitation for Other Psychiatric Disorders
APC / OP	Spell Core HRG	SP / A	WD11Z	All patients 70 years and older with a Mental Health Primary Diagnosis, treated by a Non-Specialist Mental Health Service Provider
APC / OP	Spell Core HRG	SP / A	WD22Z	All patients between 19 and 69 years with a Mental Health Primary Diagnosis, treated by a Non-Specialist Mental Health Service Provider
APC / OP	Spell Core HRG	SP / A	WD33Z	All patients 18 years and younger with a Mental Health Primary Diagnosis, treated by a Non-Specialist

Dataset	Identifying Variable	Level	Value	Description
				Mental Health Service Provider, NEC
OP	TFC / MSC	AT	659	Dramatherapy
OP	TFC / MSC	AT	656	Clinical Psychology
OP	TFC / MSC	AT	660	Art Therapy
OP	TFC / MSC	AT	661	Music Therapy
OP	TFC / MSC	AT	710	Adult Mental Illness
OP	TFC / MSC	AT	711	Child And Adolescent Psychiatry
OP	TFC / MSC	AT	712	Forensic Psychiatry
OP	TFC / MSC	AT	713	Psychotherapy
OP	TFC / MSC	AT	715	Old Age Psychiatry
OP	TFC / MSC	AT	720	Eating Disorders
OP	TFC / MSC	AT	721	Addiction Services
OP	TFC / MSC	AT	722	Liaison Psychiatry
OP	TFC / MSC	AT	723	Psychiatric Intensive Care
OP	TFC / MSC	AT	724	Perinatal Psychiatry
OP	TFC / MSC	AT	725	Mental Health Recovery and Rehabilitation Service
OP	TFC / MSC	AT	726	Mental Health Dual Diagnosis Service
OP	TFC / MSC	AT	727	Dementia Assessment Service
APC / OP	TFC / MSC	EP / A	141	Restorative Dentistry
APC / OP	TFC / MSC	EP / A	142	Paediatric Dentistry
APC / OP	TFC / MSC	EP / A	143	Orthodontics
APC / OP	TFC / MSC	EP / A	144 / 145	Max-Fax
APC / OP	TFC / MSC	EP / A	149	Surgical Dentistry
APC / OP	TFC / MSC	EP / A	360	GUM
APC / OP	TFC / MSC	EP / A	450	Dental medicine specialties
APC / OP	Admin Category	EP / A	2	Private care episode
APC / OP / AE	Overseas Patient	EP / A / V	Various	Treated as an overseas patient
APC / OP / AE	Military / Prisoner	EP / A / V	½	Military or prisoner provider episode or patient

Note: In the third column of the table, the level of identification could be at episode (EP), spell (SP), attendance (A/AT) or visit (V) level.

3.5.3 Removing of duplicate episodes

The final step was to remove duplicate episodes. To identify these we created a 'concatenated' variable. What this means is that we took a number of variables and joined them together to create a long string identifier. We then looked for duplicates among them and deleted the duplicated episode.

We used the APC variables for: provider code, episode start date, episode end date, month and year of birth, LSOA of residence, primary diagnosis and sex.

If an episode matched exactly on all of these variables joined together, with an identical string of characters, we deemed it to be a duplicate and only one copy of the episode was included in our final dataset for the modelling.

Only a very small number of duplicate episodes were identified, around 1.2%.

3.5.4 Unregistered patients

The objective was to refresh PBRA 2011 following the same methodology. The person based approach is confined to registered patients as the eventual goal is to calculate weights for CCGs. There is a lack of reliable data on the place of residence of unregistered patients in our (and PBRA 2011) datasets, and thus we are not able to attribute their health care costs by CCG. Unregistered patients are therefore excluded from this refresh.

3.5.5 Data ready for costing

Once we had removed out-of-scope data and duplicates, we were ready to apply a costing algorithm to attach costs to the data which did not already have a national tariff price applied in the SUS PbR data. The extent to which this was needed in each dataset is shown in tables below.

3.6 Costing sequence

The order and sequence of preference followed in our approach for costing the data is shown in Figure 3.1 below.

Figure 3.1: Pricing approach order

3.6.1 Preferred method for deriving cost

The preferred way of deriving a cost in any of the datasets is simply to use the cost in the SUS PbR dataset as supplied. In these cases, which were the large majority, the only step required was to subtract the number of pounds attributable to the MFF from the price, which was provided as a separate variable - labelled 1a in Figure 3.1.

3.6.2 No final price given

In some cases there was no price given in the SUS PbR data, for example this can occur if a provider has indicated they would not like a SUS PbR cost attached to that spell. Nevertheless, a national tariff price might be available as a starting point, based on the spell or attendance's HRG code as provided in the SUS PbR data. Non-mandatory prices are not included in the SUS PbR data set but were applied in our costing.

Our approach to these cases follows a number of steps to try to reach a fair estimate of the final price given the information provided in the '2013/14 Tariff Information Spreadsheet' Excel Workbook'[5].

Reference should also be made to this workbook concerning the subsequent paragraphs. The worksheet entitled '01. APC & OPROC' provide a starting point. This worksheet contains the bulk of HRG codes along with tariff prices for day cases, combined day case/elective, electives and non-electives along with trim points after which time elective or non-elective excess bed day tariffs are applied on a day to day basis.

3.7 Admitted patient care costing

3.7.1 Spells already costed

The first and simplest step was to take those spells already costed and to subtract the pound value attributed to MFF from the PbR Final Tariff. Both variables were provided in SUS PbR (see 1a in Figure 3.1). These costs already included the appropriate excess bed days and specialist top-ups, and short stay reductions.

The proportion in each dataset which came already costed in this way varied as reflected in Table 3.11, Table 3.13: How A&E attendances were costed and per patient summary statistics and Table 3.14: How outpatient attendances were costed and per patient summary statistics below. For example, for G&A APC, more than 85% of the data was costed already, while for specialised services only around 50% of the data was already costed.

[5] https://www.gov.uk/government/publications/payment-by-results-pbr-operational-guidance-and-tariffs

The data already costed included mandatory zero prices for the HRGs in Table 3.5.

Table 3.5: Mandatory zero tariffs

HRG code	HRG description	Justification
LA08E	Chronic kidney disease with length of stay 1 day or less associated with renal dialysis	Empty core HRG for renal dialysis for chronic kidney disease
PB03Z	Healthy baby	Costs are included with the mother's care
SB97Z	Same day chemotherapy admission/attendance	Empty core HRG for chemotherapy
SC97Z	Same day external beam radiotherapy admission/attendance	Empty core HRG for external beam radiotherapy
UZ01Z	Data invalid for grouping	Organisations should not be funded for invalid data

3.7.2 Emulating tariff for missing values

The remaining spells had either a zero or a missing value for the final tariff in the SUS PbR data set. The process of emulating the tariff (1b in Figure 3.1) involved several steps as did costing using reference costs.

To emulate the tariff, the first condition was that there was a national tariff available for that HRG. There is commonly a 'Combined Day Case / Ordinary Elective Tariff' and a 'Non-Elective Tariff' provided in the spreadsheet above. In a few cases there is an 'Ordinary elective spell tariff' (but no Combined Day Case Tariff). In a few cases there is a 'Day case spell tariff'.

Where there was no combined day case/elective tariff, but there was a separate day case tariff, or elective tariff, the day case tariff was applied if the patient's stay was either 1 or 0 days.

For these cases where a national tariff exists but was not in the SUS PbR dataset, we had to consider how best to calculate the additional payments for excess bed days, short spell reductions and specialist top-ups. The PbR guidance for 2013/14[6] provides a useful reference point to these adjustments.

3.7.3 Excess bed days adjustment

The Tariff Information Spreadsheet gives a 'trim point' for each HRG and a daily cost for bed days exceeding the trim point. This varies depending on whether the spell was elective or non-elective.

Therefore, both whether the spell is elective or non-elective and the length of stay of the spell are important when applying the trim-point and costs for the excess bed days.

The next step was therefore to calculate the length of stay for these spells as follows.

1. Take the spell end date.
2. Take the admission date (spell start date).
3. Take the episode start and episode end dates.

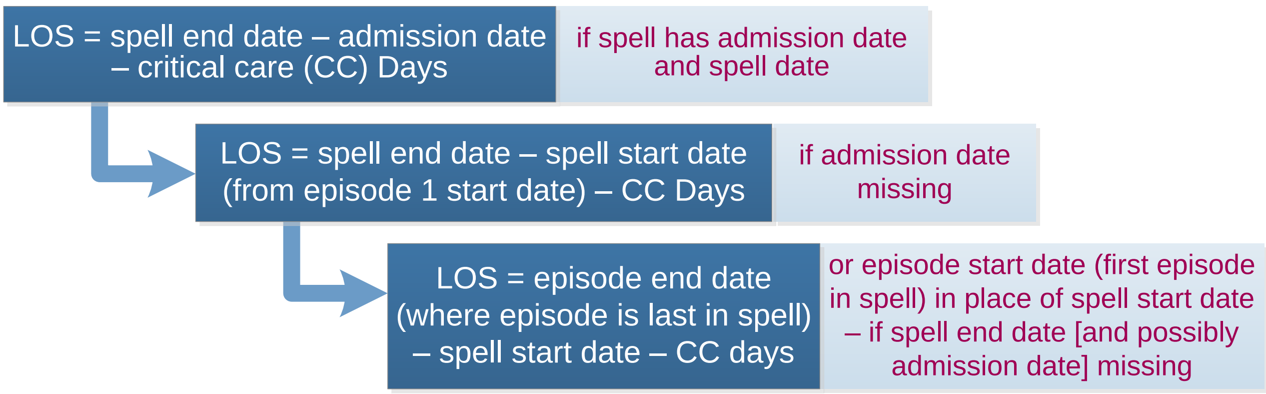
[6] https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/214902/PbR-Guidance-2013-14.pdf

4. Generate a secondary spell start date variable which is equal to the minimum (i.e. earliest) episode start date in the spell, which can be used as the spell start date, if the latter was missing.
5. Generate a first attempt at the length of stay variable equal to the spell end date minus the admission date.
6. Subtract the number of critical care days from this number of days – these are not included in length of stay calculations for the purposes of whether to apply excess bed day costs.
7. Replace the length of stay variable (where still missing) with:
    - either the spell end date minus the secondary spell start date minus critical care days; or
    - the episode end date minus the admission date minus critical care days, where there is only one episode in the spell; or
    - the episode end date minus the spell start date minus critical care where there is only one episode in the spell.
8. Across all specialty codes, generate an average length of stay (mean of length of stay for each main specialty).
9. Replace the length of stay days, if still missing, with this mean length of stay.

This approach attempts to maximise the number of spells that have a length of stay variable, given that any of the spell, admission and episode start and end dates could be missing, for the purpose of estimating a cost for the excess bed days stay.

Ideally, and in most instances, the way to calculate length of stay was the first step in Figure 3.2. The second and third steps were ways to try to maximise the use of available data, but in practice were applied to only a very small (negligible) numbers of spells.

Figure 3.2: Length of Stay (LOS) for trim point excess bed day calculations



It was then possible to calculate the excess bed days costs above the trim point for both elective and non-elective stays.

For the spells where a national tariff exists but was not applied in the SUS PbR data set, and where the excess bed days exceeded the trim point for the HRG, we were then able to add the excess bed day estimated additional cost. The value of the additional costs depended also on the elective or non-elective nature of the spell.

3.7.4 Best practice tariffs

For the HRG codes in Table 3.8, we applied the standard rather than best practice tariff for spells not already costed in the SUS PbR data.

Table 3.6: HRG codes where standard tariff applied

Spell Core HRG Code	Description
FZ54Z	Diagnostic Flexible Sigmoidoscopy 19 years and over
FZ55Z	Diagnostic Flexible Sigmoidoscopy with biopsy 19 years and over
DZ35Z	Simple Bronchodilator Studies
DZ44Z	Simple Airflow Studies
AA22A	Non-Transient Stroke or Cerebrovascular Accident, Nervous System Infections or Encephalopathy with CC
AA22B	Non-Transient Stroke or Cerebrovascular Accident, Nervous System Infections or Encephalopathy without CC
AA23A	Haemorrhagic Cerebrovascular Disorders with CC
AA23B	Haemorrhagic Cerebrovascular Disorders without CC

3.7.5 Short stay emergency reduction

Short stay emergency reductions to tariffs are applicable when spells meet specific criteria. The reductions were already applied for the spells costed in the SUS PbR data. For spells not already costed, but for which a tariff exists, a short stay reduction was therefore applied on the condition that the length of stay was 1 day or less, the patient was an adult (taken to be 19 or over on admission), the spell was non-elective and the average non-elective stay was more than 1 day.

3.7.6 Unbundled Costs

For activity not already costed in the SUS PbR data, we then looked at unbundled costs. We applied an additional cost for the spell according to the unbundled tariff element of the National Tariff Information spreadsheet 2013-14 where it existed.

In some cases the spell core HRG matched an unbundled tariff and nothing else. Therefore, we allowed for such a spell core HRG to be given the unbundled cost, and applied that cost to the spell. In other words, if the spell HRG corresponded to an unbundled HRG and the HRG does not appear elsewhere in the tariff information spreadsheet, the unbundled HRG cost becomes the spell cost.

By this point we could add to the tariff as calculated so far, the unbundled costs over the spell as well as the excess bed days and short stay emergency reductions for the HRGs in the part of the data set which did not already have a cost attached.

3.7.7 Reference costs

Spells, without a cost at this point, qualified for the reference costing procedure.

We used reference costs from 2010-11 which were the basis for 2013-14 national tariff prices. However, the reference costs include the MFF which had to be removed.

For each HRG for each provider, we divided the total reference cost for elective and non-elective HRGs by the relevant activity to give a crude average cost. We then removed the provider’s MFF from these average costs. Finally, we take the average over all providers for that HRG, after the MFF has been removed. We take this to be the national unit average reference cost for the HRG (for elective and non-elective separately). We calculate excess bed day costs in a similar way (they are identified separately in reference costs).

A compounded inflator (in reality a deflator) was applied for reference costs according to an annual tariff inflation table provided by the pricing team of NHS England – see Table 3.7. This uplifted (down-lifted) the reference costed spells by a compounded calculated by 1 x 0.985 x 0.982 = 0.96727.

Table 3.7: Tariff uplift table

	2011-12	2012-13
Tariff Uplift	-1.5%	-1.8%
(of which) Efficiency	*-4.0%*	*-4.0%*
(of which) Pay and Price Inflation	*2.5%*	*2.2%*

3.7.8 Specialist service code top-ups

For HRGs costed by emulating the tariff or by using reference costs, and topped up according to bed days beyond the trim point and respective excess bed day costs, we then applied specialist service code top-ups.

Where a specialist service code (SSC) was in the SUS PbR data (very rare) there existed a number of conditions that had to be met for the top-up to be applied. We wrote an algorithm to identify these instances and applied the necessary top-up to the spells costed through reference costs or by emulating the tariff.

An example is where the primary procedure code was “A441”, the patient’s age on admission was 19 or above, and the spell had a documented specialist service code of “SS06” in any of four populated specialist service code fields. For this spell, the provider was eligible for the price to be multiplied by 1.36 times the ‘ordinary’ price to give a 36% increase. Only designated providers are eligible for the specialist top-up, thirty four providers in our dataset.

We applied these rules for a large number of procedures and providers, based on the relevant age, procedure and specialist service code rules.

3.7.9 Specialty average

Finally, for the HRGs still without a cost, we applied the speciality average. This was based on the median costs for the main specialty code across all the spells in that specialty (3 in Figure 3.1).

Due to outliers in some specialties, such as Cochlear Implants, we assessed that the median would be a more representative cost than the mean.

Where information was not available through the main specialty code route, for example where codes were missing, we followed the same approach based on treatment function code.

We applied this median cost to all remaining HRGs not costed in the previous steps.

3.7.10 Critical care costs

Critical care costs were calculated from a critical care dataset provided as part of the SUS PbR data. We calculated the average number of critical care bed days over the different HRGs in which critical care had been recorded, and applied the average to all HRGs which had critical care recorded. This was a similar approach to that taken in PBRA (2011), due to concerns with the quality of the critical care data– see 4a in Figure 3.1. The number of bed days was multiplied by a critical cost per day, based on the average cost of critical care per day by HRG for all critical care recorded in the SUS PbR dataset.

Critical care costs were then added to the other costs by taking the sum total of estimated critical care costs over the course of the entire spell (see 4b of Figure 3.1). If there was more than one episode with days of critical care within the spell, these would be summed to give a spell level critical care cost.

3.8 Outpatient (OP) Costing

3.8.1 Outpatient costing

Outpatients were costed in a similar way as for APC, but through a simpler procedure. Firstly, specialised and maternity services were identified (see the tables above). Secondly, attendances were removed if there was no indication in the SUS PbR dataset of the patient having actually attended the appointment (see Table 3.8 below). Thirdly, already priced attendances were segmented with the MFF removed.

Table 3.8: Outpatient attendance status

Attendance status from SUS PbR dataset	In cleaned data
5 Attended on time or, if late, before the relevant CARE PROFESSIONAL was ready to see the PATIENT	INCLUDED
6 Arrived late, after the relevant CARE PROFESSIONAL was ready to see the PATIENT, but was seen	INCLUDED
7 PATIENT arrived late and could not be seen	NOT INCLUDED
2 APPOINTMENT cancelled by, or on behalf of, the PATIENT	NOT INCLUDED
3 Did not attend - no advance warning given	NOT INCLUDED
4 APPOINTMENT cancelled or postponed by the Health Care Provider	NOT INCLUDED
0 Not applicable - APPOINTMENT occurs in the future	NOT INCLUDED

Note: where the attendance status was recorded as ‘Missing’, it was treated as attended.

3.8.2 Outpatient HRG tariffs

For outpatient attendances that were not already priced in the SUS PbR dataset, but where a HRG was recorded, outpatient HRG tariffs were applied where available from the National Tariff Information Spreadsheet 2013-14.

Where an attendance had a “WF” code (single, multi-professional and non-face-to-face first and follow up attendances), the corresponding tariff was given according to treatment function code as listed in the National Tariff Information Spreadsheet 2013-14.

If there was an unbundled HRG given as the core HRG for the attendance, the HRG price for the unbundled service was applied if the outpatient HRG was recorded as an unbundled HRG with no other alternative price.

3.8.3 Reference costs

Reference costs were applied along the same lines as for APC in cases where no tariff had been attached thus far.

3.8.4 Speciality average

Finally a main specialty or treatment function code average was applied if no cost had so far been attached to the attendance, in a similar manner to the APC costing.

3.9 A&E Costing

3.9.1 Accident and Emergency attendances

A&E attendances was by far the least complicated activity to cost, in part due to small variation and range in tariff prices for A&E, and in part due to the lack of information available to work with in terms of what occurred during the attendance, for example some records contained fields recorded sporadically or not consistently.

3.9.2 A&E HRG tariffs

Where no cost had been applied in the SUS PbR dataset, but a HRG was recorded, the appropriate cost was applied from the National Tariff Information Spreadsheet 2013-14. This was the case for twelve HRG codes beginning with “VB”, and an additional ‘Dead on Arrival’ code.

3.9.3 HRG average

If the attendance remained without a cost following the above steps, an average was applied according to the HRG listed for the A&E visit, or the overall average was applied if still missing.

3.10 Linking the data for person level costs

The process set out above gave a cost for each APC spell, outpatient attendance, A&E attendance, and critical care day. These then needed to be linked at patient level to give the total cost for each patient in 2013-14, covering all the APC spells , OP attendances, A&E attendances and critical care that individual received in 2013-14. The patient level data also needed to be linked at the person level to the PDS data.

3.10.1 Linking the data at the person level

Every row of data in each of the SUS PbR datasets had an exclusive identifier, known as the PBR Generated Record ID. The PBR Generated Record ID could be linked to the bridge file on this variable.

The bridge file contained a further two variables, namely match rank and PDS generated record ID.

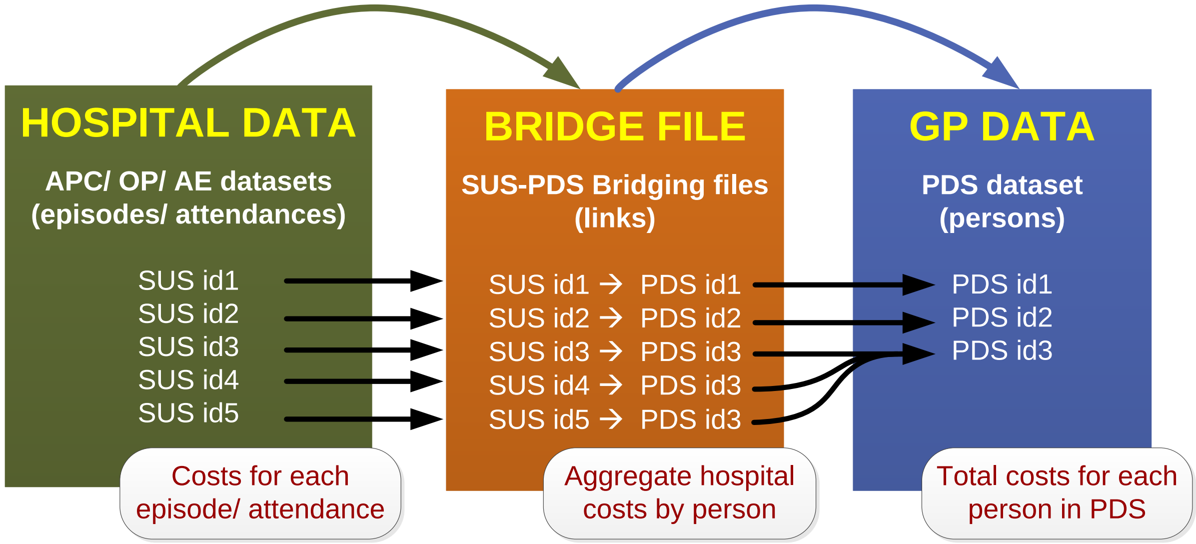
Match rank was an indicator showing whether the hospital record could be matched to an exclusive patient in the PDS data. A match rank of 1 was an exclusive match of NHS number. A match rank of 2 was an exclusive match of date of birth and postcode. A match rank of 3 indicated that such an exclusive linkage was not possible, due to two or more patients having the same date of birth and postcode, thus there being two patients (or more) in the PDS data to which the hospital data could be linked. The number of records with match rank 3 was very small.

The match rank was provided by the HSCIC as we did not have data on NHS number, date of birth and postcode.

We linked the patient level costs from the SUS PbR data to the PDS data only for those where match rank was indicated as a 1 or 2. Those where the match rank was three were excluded from the dataset used for the modelling.

Figure 3.3 illustrates how the datasets linked together to aggregate costs at the person level.

Figure 3.3: Linking core datasets for aggregating costs to the person level



3.10.2 Total cost per person

After linking the data to PDS in this way, the costs for APC, OP and A&E could be attached and summed for all three of these for each person in the case of the G&A dataset for modelling, and for APC and Outpatients for both the specialised and maternity datasets. Hence, we were able to collapse down on person ID the sum of any costs incurred through APC, OP or AE, or any combination of these.

3.11 Tables

The following tables cover:

- Table 3.9 summarises the data excluded as not relevant for the G&A modelling;
- Table 3.10 shows the removal of specialised services and the data linkage to the PDS data for each;
- Table 3.11 shows for G&A the costs obtained directly from in the SUS PbR dataset and the costs estimated using national tariffs, reference costs and specialty averages. The table also shows the distribution of the costs per patient.
- Table 3.12, Table 3.13 and Table 3.14 show the same as Table 3.11 for critical care, A&E and outpatients, included within Table 3.11.

The graphs following the tables show the average cost per patient by age-sex group, the average cost per registered patient by age-sex group, and total costs by age group. This is for various combinations of service (G&A, maternity and specialised) and treatment location (APC, OP, A&E).

Table 3.9: Data excluded

Dataset	Episodes	%	No. spells	%	Dataset	Attendances	% [1]	Dataset	Visits	%
APC (All)	**19,586,831**	**100**	**17,064,097**	**100**	**OP (All)**	**102,541,435**	**100**	**AE (All)**	**18,676,085**	**100**
Reason for exclusion:										
Military & Prisoner [2]	38,626	0.2	35,560	0.2	*Military & Prisoner*	205,587	0.1	*Military & Prisoner*	38,419	0.2
Dentistry & Max-Fax [3]	280,672	1.4	278,018	1.6	*Dentistry & Max-Fax*	2,774,801	1.4	*Overseas*	214,536	1.1
GUM [4]	570	0.0	494	0.0	*GUM*	271,853	0.1			
Well Baby	450,670	2.3	447,909	2.6	*Well Baby*	189,617	0.0			
Overseas [5]	80,945	0.4	70,097	0.4	*Overseas*	229,962	0.0			
Psychiatry [6]	245,964	1.3	208,098	1.2	*Psychiatry*	3,509,238	1.7			
Duplicates [7]	242,558	1.2	138,802	0.8	*Attendance* [8]	19,496,790	19.0			
Private Patient	88,559	0.5	88,559	0.5	*Private Pt.*	353,890	0.3			
Remaining [9]	18,228,039	93.0	15,861,883	93.0	Remaining	76,988,623	75.0	Remaining	18,423,452	98.6

[1] For each OP Dataset: Only those appointments with an attendance type specified as: missing; on time, late (but attended).

[2] Defined by an HSCIC provided Military and Prisoner flag in APC, AE and OP datasets

[3] Defined by Main Specialty Code or Treatment Function Code as follows: 141, 142 (Paediatric or Restorative Dentistry); 143 (Orthodontics); 450 (Dental); 144 (Maxillo-Facial); CZ30Y, CZ31Y, CZ32U, CZ32Y, CZ33Y, CZ34U, CZ34Y, CZ35U, CZ35Y, CZ36Y, CZ37Y, CZ38U, CZ38Y, CZ39U, CZ39Y, CZ40Y, CZ41Y, CZ42Y (Spell Core HRG - Dental)

[4] Defined by Main Specialty or Treatment Function code: 360 (GUM)

[5] Defined by three overseas patient identifier datasets provided by HSCIC (one for each of APC, AE, OP)

[6] Defined by Spell Core HRG beginning with Spell Core HRG: “WD” plus PA51Z, PA52Z, PA52B, PA52C, PA53A, VC26Z, VC28Z (Spell Core HRG)

[7] In practice these were removed at the level of the dataset in use and identified on the basis of same: provider, date of episode, LSOA, primary diagnosis, gender, month/year of birth

[8] For each OP Dataset: Only those appointments with an attendance type specified as: missing; on time, late (but attended). 17,124,324 observations dropped out of 92,325,855.

[9] Not the same as first row minus sum of rows 3-9 as rows 3-9 may overlap each other for all sets of remaining episodes, spells, attendances, visits

Table 3.10: Data linkage for the General and Acute APC, OP and A&E data 2013-14

In working dataset:	Number of episodes remaining	% Episodes of 'Remaining'	Number of spells/ attendances remaining	% Spells/ attendances of all 'Remaining'[1]	Linked Spells bridged to PDS[2]	% linked of working dataset spells	Patients traced APC to PDS	% of all PDS patients[3] (N=55,540,852)	Average Spells/Visits Per Patient
G&A APC	14,770,961	81.0	12,460,109	78.6	11,886,833	95.0	6,543,756	11.8	1.9
G&A OP			63,908,522	83.0			16,135,115	29.0	3.5
A&E			18,423,452	100	16,633,823	90.3	10,546,987	19.0	1.6

[1] The proportion of all spells including what NHS England identified as belonging to general and acute, maternity and specialised components of the formulae.

[2] This refers to spells/attendances/visits successfully mapped to the bridging file and PDS with 'match rank 1 or 2', that is no confusion as to which registered patient belonged to the corresponding care.

[3] 218 patients dropped from PDS prior to modelling due to sex being undefined.

Table 3.11: G&A – how spells were costed and per patient summary statistics

GA	Min	P1	P5	P10	P25	P50	P75	P90	P95	P99	Max	Mean	SD	Skew	Kurt	N	Totals
Costed in SUS PbR	49	251	312	380	487	712	1,858	3,631	5,280	10,519	594,966	1,590	2,514	16	1,354	10,948,844	£17,409,600,151
HRG tariff estimate	202	324	324	324	324	324	650	650	726	4,701	289,777	614	2,226	26	1,571	249,088	£152,810,464
Reference costs	38	381	587	886	1,689	2,576	3,890	4,467	4,852	10,653	113,911	2,857	2,053	7	158	421,218	£1,203,403,421
Average by Specialty	402	452	753	753	838	1,040	1,040	1,040	1,040	3,292	50,912	1,047	644	24	1,206	239,977	£251,210,077
Total	38	251	324	372	492	753	1,932	3,631	5,227	10,424	594,966	1,604	2,486	16	1,331	11,859,127	£19,017,024,112
GA – per patient	38	251	349	441	605	1,179	3,045	6,300	9,652	21,232	665,235	2,752	5,256	15	641	6,528,663	£17,969,356,988

Table Note: This table describes the way in which general and acute data was costed. The rows describe whether the data was provided as costed in SUS PbR, whether or not a tariff estimate was calculated manually, whether or not reference costs were applied or whether a specialty average was taken. The overall statistics are in the next row. The linked per-patient statistics are in the penultimate row. The columns give the statistics for each of these rows. Minimums, percentiles 1, 5, 10, 25, 50, 74, 90, 95 and 99 are given, along with maximums, standard deviations, skewness, kurtosis, the number of observations (N) and the total amount of care costed by that method, or overall in the case of the total column and overall in terms of person linked, i.e. successfully linked to PDS, in the final row. PP stands for per-person. This note applies to subsequent tables also.

Table 3.12: Critical Care cost distributions in the G&A data

	Min	P1	P5	P10	P25	P50	P75	P90	P95	P99	Max	Mean	SD	Skew	Kurt	N	Sum
GA	1,468	1,757	2,206	3,161	3,640	7,214	16,418	36,504	57,647	155,164	32,930,778	18,432	115,449	164	40,849	170,763	£3,147,562,634
GA-PP	**1,571**	**1,935**	**2,952**	**3,161**	**3,605**	**6,025**	**7,629**	**14,911**	**22,831**	**43,565**	**491,504**	**7,994**	**8,829**	**9**	**222**	**147,157**	**£1,176,379,773**

Table 3.13: How A&E attendances were costed and per patient summary statistics

A&E - How Costed	Min	P1	P5	P10	P25	P50	P75	P90	P95	P99	Max	Mean	SD	Skew	Kurt	N	Sum
Costed in SUS PbR	58	58	58	58	58	78	110	130	139	210	237	93	32	1	4	17,158,158	£1,595,555,407
HRG tariff estimate[1]	58	58	58	58	58	58	58	58	58	58	58	58	0			234,241	£13,585,978
Average	87	87	87	87	87	87	87	87	87	87	87	87	0			232,118	£20,273,653
Walk in flat tariff estimate	29	29	29	29	29	29	29	29	29	29	29	29	0			798,935	£23,169,115
Total	29	29	58	58	58	78	110	130	139	164	237	90	33	1	4	18,423,452	£1,652,584,153
AE - PP (linked)	29	29	58	58	78	110	164	258	349	635	37,362	143	142	16	1,364	10,546,987	£1,510,942,943

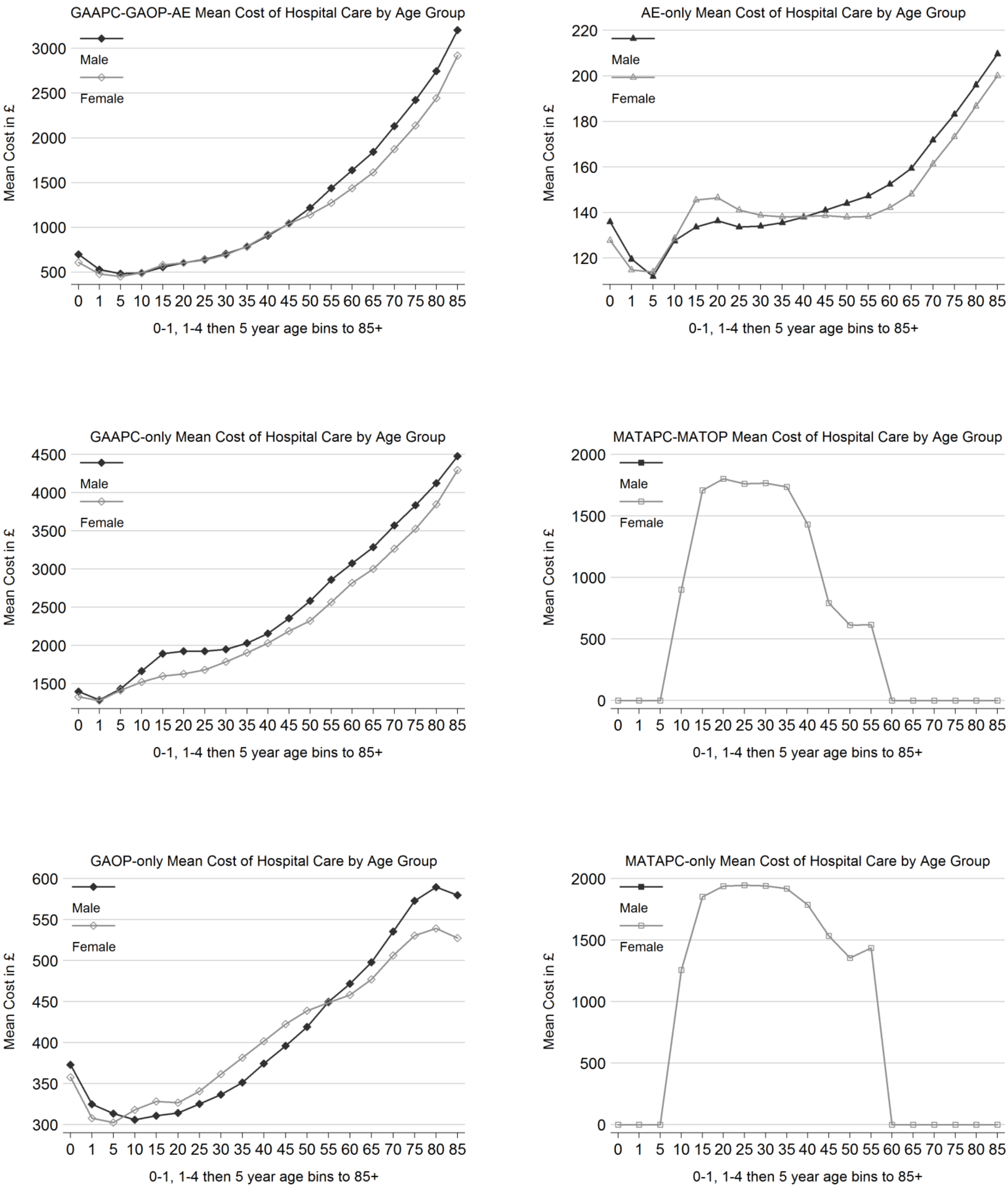
[1] Uniformity is a result of all manually estimated tariffs falling either into the 'department = 3' category - Other type of A&E/minor injury ACTIVITY with designated accommodation for the reception of accident and emergency PATIENTS. The department may be doctor led or NURSE led and treats at least minor injuries and illnesses and can be routinely accessed without APPOINTMENT. A SERVICE mainly or entirely APPOINTMENT based (for example a GP Practice or Out-Patient Clinic) is excluded even though it may treat a number of PATIENTS with minor illness or injury. Excludes NHS walk-in centres; or having a spell core HRG of "VB11Z" - No investigation with no significant treatment. Average is across all observations where a price is given. Walk-in flat tariff estimate was a team agreed flat estimate of the tariff for a walk in that would be applied.

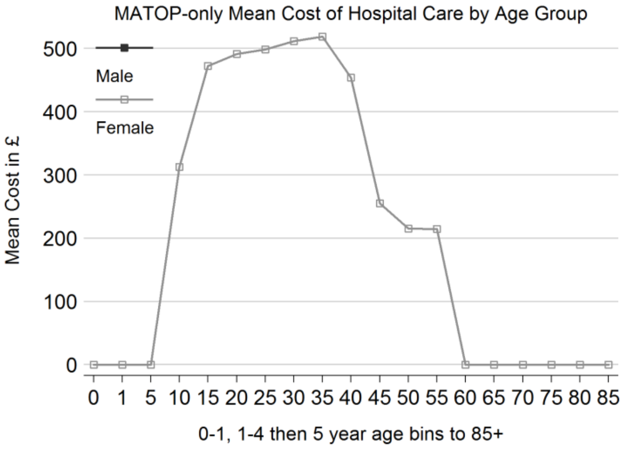
Table 3.14: How outpatient attendances were costed and per patient summary statistics

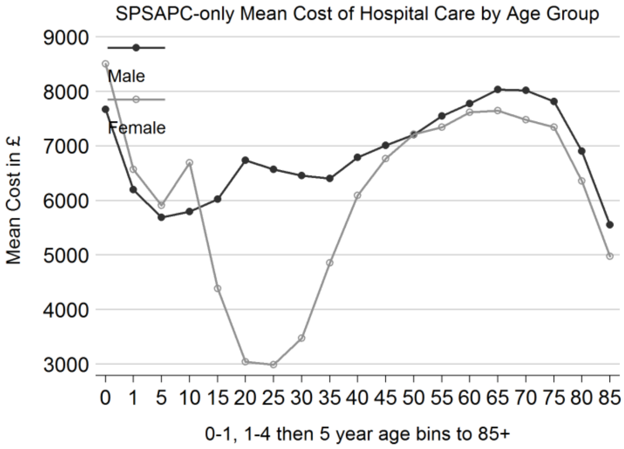
Statistic:	Min	P1	P5	P10	P25	P50	P75	P90	P95	P99	Max	Mean	SD	Skew	Kurt	N(obs)	Total Pounds
Costed in SUS PbR	45	60	60	63	71	104	132	175	215	351	1,609	114	56	3	19	43,651,394	£4,982,964,224
HRG tariff estimate	38	105	106	106	121	166	217	400	400	400	21,504[1]	194	103	27	5,296	6,679,596	£1,292,467,200
Reference costs	2	22	22	22	50	52	74	125	178	250	680	66	50	4	28	12,205,949	£806,414,656
Average by Specialty	42	69	71	74	74	150	150	171	270	270	389	133	56	1	5	1,371,583	£182,856,448
Total	2	22	45	52	71	103	138	181	225	400	21,504	114	70	11	2,579	63,908,522	£7,264,702,528
OP GA – PP (linked)	2	50	71	102	142	263	505	905	1,275	2,431	200,116	428	673	35	4,244	16,135,115	**£6,899,914,463**

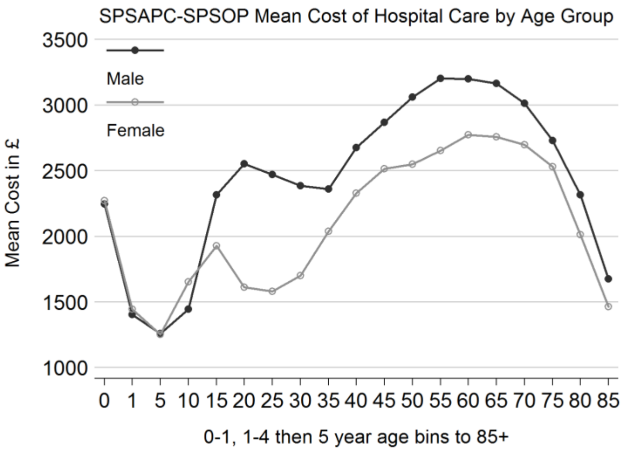
[1] Cochlear Implants.

Figure 3.4: Average cost per service user 2013-14









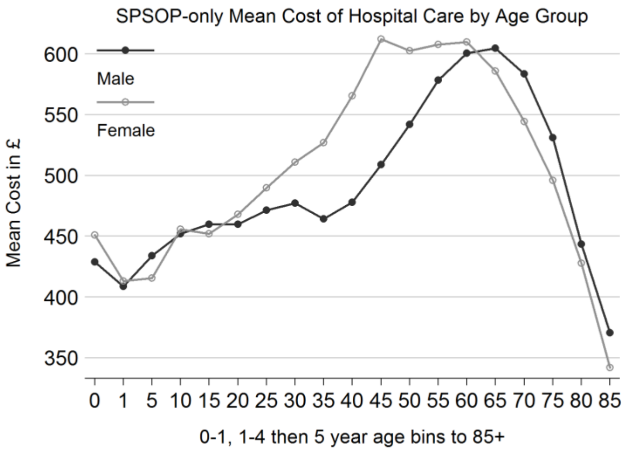
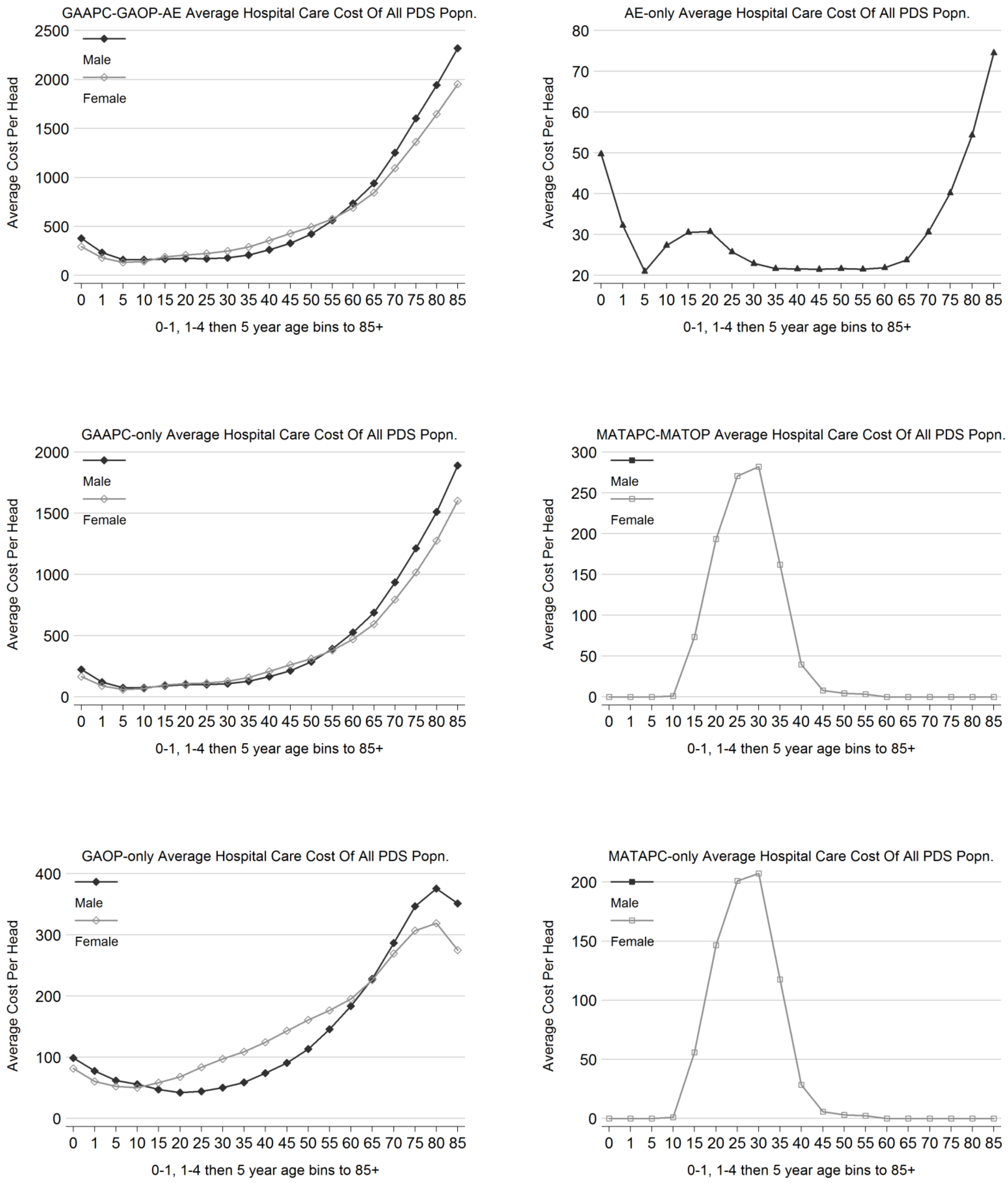
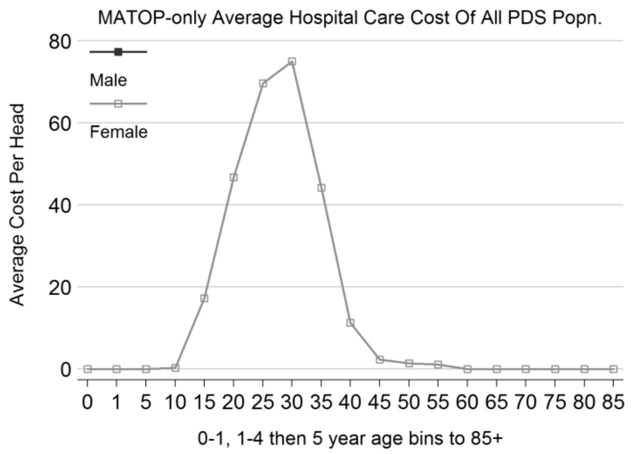
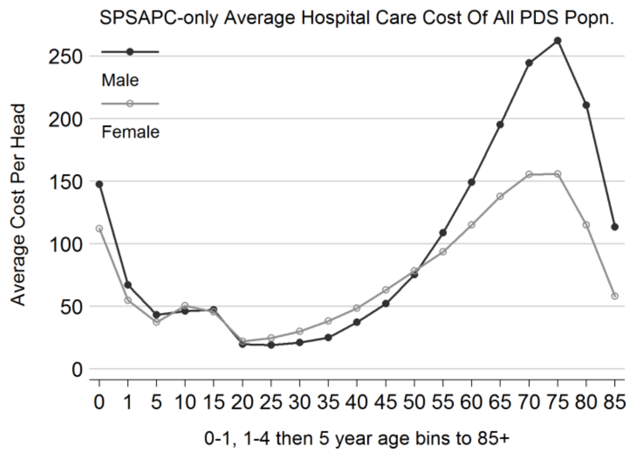


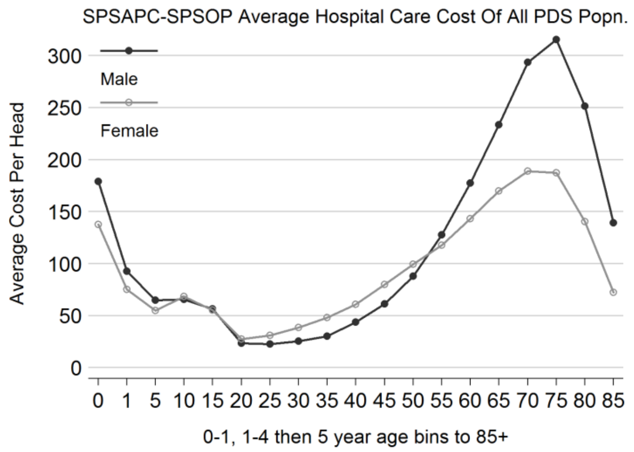
Figure Note: GAAPC-GAOP-AE indicates that general and acute APC, OP and AE have been amalgamated in this graph. Other graphs disaggregate the datasets, e.g. GAAPC only indicates only the general and acute APC data is accounted for in the graph.

Figure 3.5 Cost per registered patient 2013-14




MATOP-only Average Hospital Care Cost Of All PDS Popn.
Male
Female
Average Cost Per Head
0
20
40
60
80
0 1 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85
0-1, 1-4 then 5 year age bins to 85+


SPSAPC-only Average Hospital Care Cost Of All PDS Popn.
Male
Female
Average Cost Per Head
0
50
100
150
200
250
0 1 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85
0-1, 1-4 then 5 year age bins to 85+


SPSAPC-SPSOP Average Hospital Care Cost Of All PDS Popn.
Male
Female
Average Cost Per Head
0
50
100
150
200
250
300
0 1 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85
0-1, 1-4 then 5 year age bins to 85+

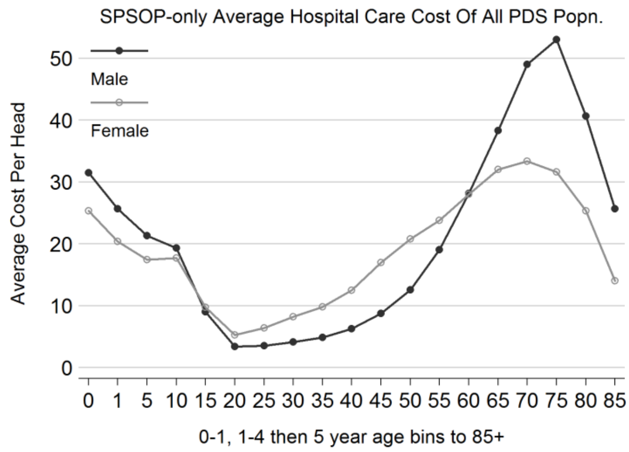
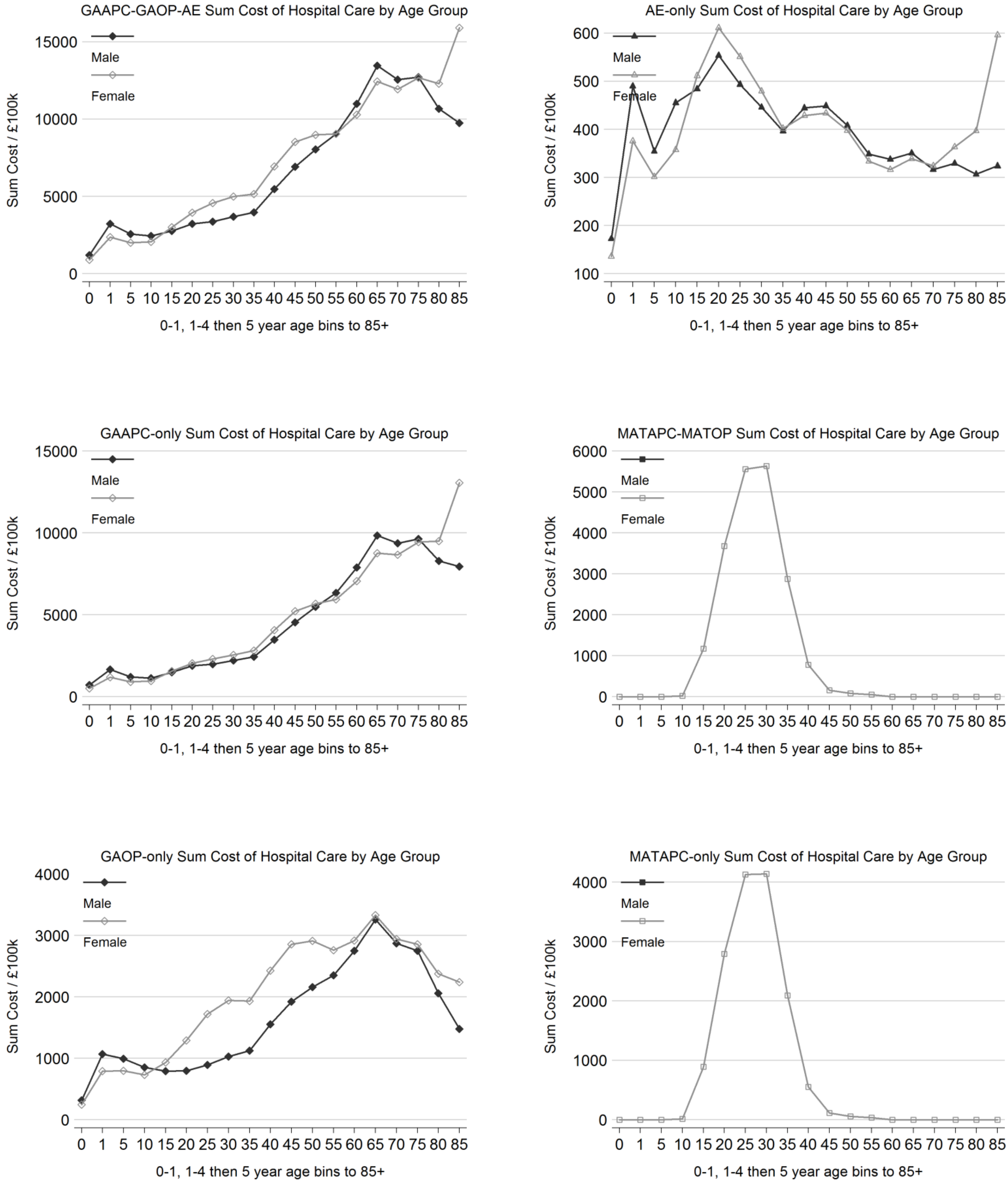
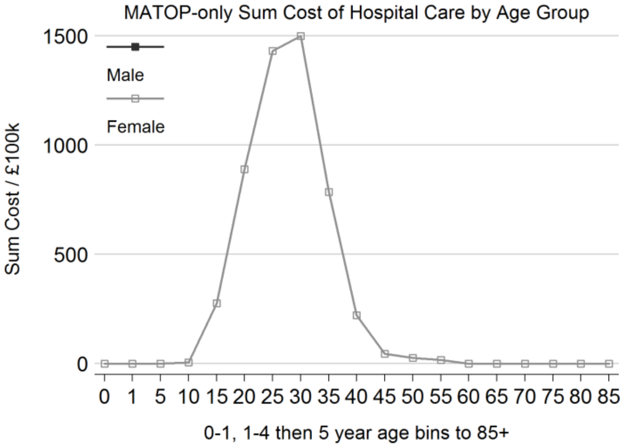
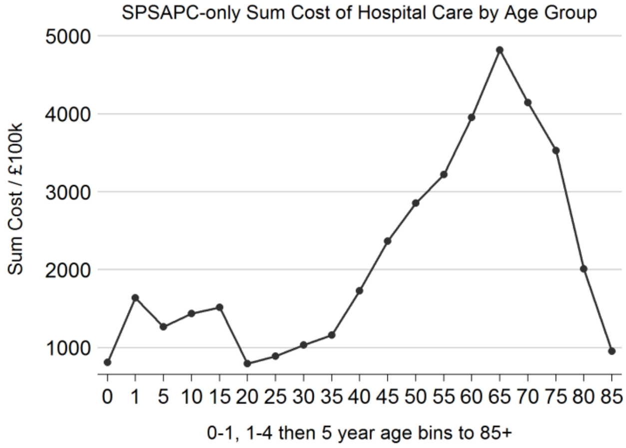
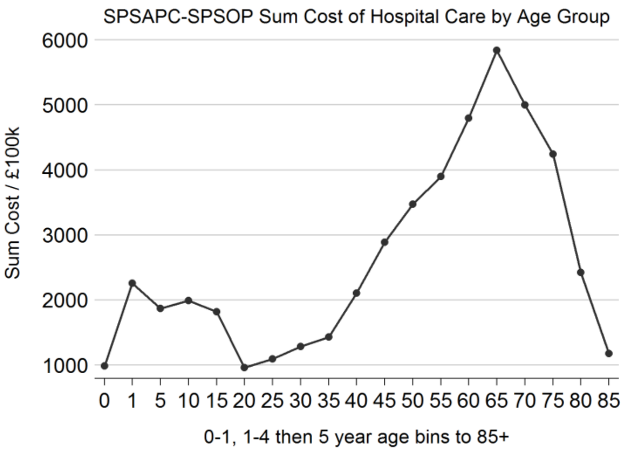

SPSOP-only Average Hospital Care Cost Of All PDS Popn.
Male
Female
Average Cost Per Head
0
10
20
30
40
50
0 1 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85
0-1, 1-4 then 5 year age bins to 85+

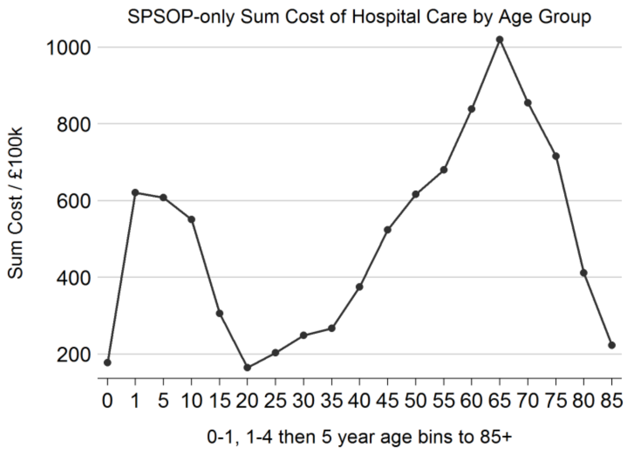
Figure 3.6: Total costs by age group




MATOP-only Sum Cost of Hospital Care by Age Group
Male
Female
Sum Cost / £100k
0
500
1000
1500
0 1 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85
0-1, 1-4 then 5 year age bins to 85+


SPSAPC-only Sum Cost of Hospital Care by Age Group
Sum Cost / £100k
1000
2000
3000
4000
5000
0 1 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85
0-1, 1-4 then 5 year age bins to 85+


SPSAPC-SPSOP Sum Cost of Hospital Care by Age Group
Sum Cost / £100k
1000
2000
3000
4000
5000
6000
0 1 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85
0-1, 1-4 then 5 year age bins to 85+


SPSOP-only Sum Cost of Hospital Care by Age Group
Sum Cost / £100k
200
400
600
800
1000
0 1 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85
0-1, 1-4 then 5 year age bins to 85+

4 General & Acute: Approach to the modelling

4.1 Person-based Resource Allocation (PBRA 2011)

The modelling for G&A services followed the same approach as PBRA (2011). The steps to estimating the weights by age-sex group by GP practice were as follows.

a. Calculating the cost per person of relevant health care received in 2013-14 (as covered in section 3).
b. Linking the costs per person data to other relevant datasets, including SUS PbR data for the two previous years and the attributed variables.
c. Calculating the values of derived variables.
d. Generating the sample files.
e. Calculating the morbidity makers, morbidity counts and co-morbidity interaction terms.
f. Model estimation.
g. Model testing.
h. Assessing the results.
i. Creating weights by age-sex group by GP practice

Each of these is covered in the following sections.

5 General and Acute: Assembling the model datasets

This section explains how the data were assembled to construct the variables used and tested in the G&A models. It also describes how the samples were drawn for estimating and validating the model. Section 3 has already described how the costs per person in 2013-14 were estimated.

5.1 Scope and data segregation

5.1.1 Services out of scope

As outlined in section 3, the cost person for G&A in 2013-14 excluded the costs of the following services in the SUS PbR dataset. They were excluded either because they are covered by a separate formula or are not funded by CCGs.

a. Maternity services. There is a separate formula for maternity services, which have a different distribution of need across the country compared with G&A services.
b. Mental health services. There is a separate formula for mental health services based on the Mental Health Minimum Dataset (MHMDS), which has broader coverage of these services than HES or SUS PbR. In addition, with relatively few patients using specialist mental health services a different approach to the modelling is required.

c. Specialised services, as they are not commissioned or funded by CCGs, but by NHS England. Separate work was also being undertaken to develop an approach for allocating funding for specialised services.

d. Secondary dental services, which are directly commissioned and funded by NHS England, not by CCGs.

e. Genito-urinary medicine (GUM), which is commissioned and funded by local authorities.

f. Health care for prisoners and the armed forces, which is funded by NHS England, not by CCGs. Prisoners and the military have their own primary medical services and were excluded from our PDS data.

g. Overseas visitors who by definition are not registered with a GP practice. Overseas patients should be charged by the provider, or if they are charge exempt funded through a separate allocation process.

h. Privately funded health care.

In addition, the cost of health care for unregister patients was excluded as the modelling is based on those registered with a GP practice. It was not possible with the data available to attribute these costs to CCGs.

The procedures by which maternity, mental health, specialised, secondary dental and GUM services were identified in the SUS PbR dataset are set out in section 3.

If those using G&A services in 2013-14 had, in the two previous years, used maternity, mental health, specialised, secondary dental or GUM inpatient services, diagnosis information from these spells are included in the morbidity markers within the set of explanatory variables used to predict the costs of individuals using G&A services in 2013-14.

5.2 Data linking

This section gives a basic description on how the main datasets were linked at the person level to create the datasets for the modelling, and the datasets for generating the weights by age-sex group by GP practice.

5.2.1 Data linkage for the estimation sample

The starting point is the Personal Demographic Dataset (PDS), which is a person level list of all patients registered with General Practices. A snapshot of GP registered lists was provided by the HSCIC for the 1st of April for each year 2009 to 2014. The PDS extract provided was pseudo-anonymised, that is it did not contain patients' names, dates of birth or postcodes.

To generate the dataset for the modelling, we used the PDS for April 2013. In this dataset we extracted patient identifiers (to link with the SUS PbR PDS datasets), age, sex, GP practice and LSOA.

The day of birth was not provided in the PDS data set, as part of the pseudo-anonymisation. We therefore calculated age from the month and year of birth data provided, and took the day of birth to be the first of the month for all individuals.

Random samples from the PDS were then taken, as described below. For the modelling, samples were taken and then data were linked from the SUS PbR and other data sets for these samples. This saved time and minimised data storage requirements and computing time, compared with the alternative option of linking the different datasets for everyone in the April 2013 PDS and then taking the samples.

Hospital activity data were from Secondary Users Service (SUS) Payment by Results (PbR) administrative data. SUS PbR was provided in four main files, Admitted Patient Care (APC), Outpatient (OP), Accident and Emergency (A&E) and Critical care.

The data to be linked to the PDS samples were as follows:

- The estimated costs of each individual patient’s use of services in 2013-14. These are described in section 3. Out of scope services were removed before linkage to PDS samples.
- Diagnostic information from the SUS PbR APC for 2011-12 and 2012-13.
- Attributed needs and supply variables.

The data linkage is summarised in Figure 5.1. The costs per person and diagnostic information were linked to the PDS sample datasets using patient identifiers specially constructed by the HSCIC for the SUS PbR data and for the PDS data. The attributed variables were linked using the GP practice code or for small area level variables by the 2011 specification of Lower Super Output Areas (LSOAs).

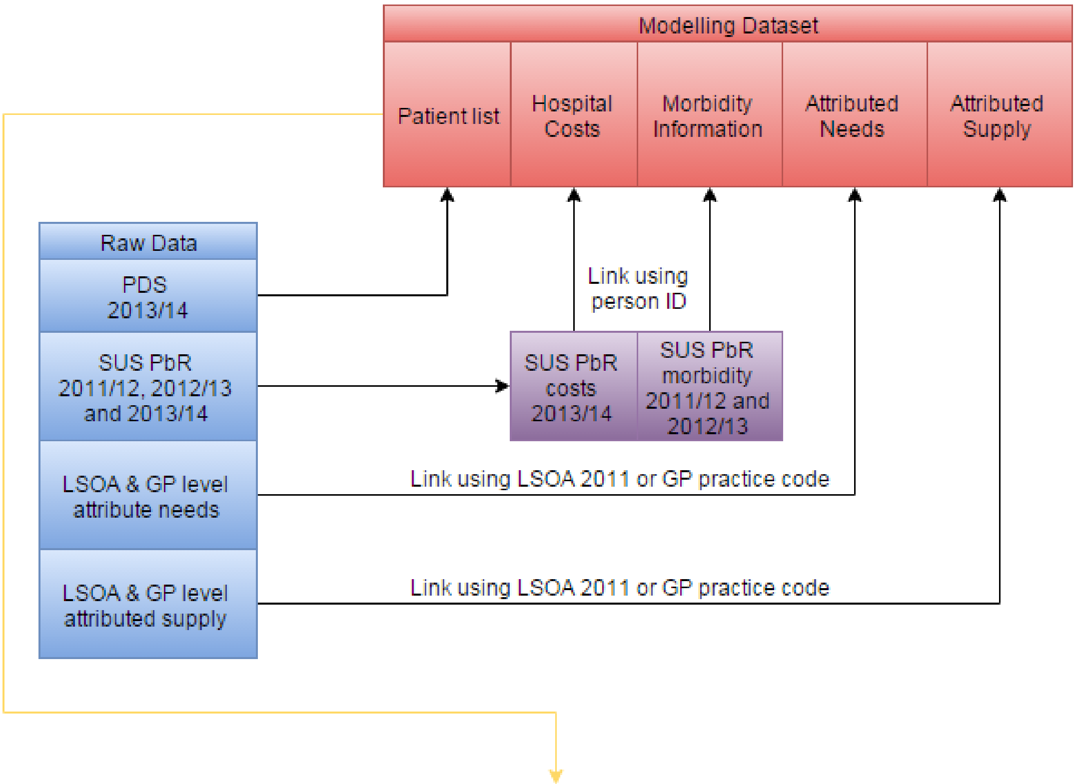
The linkage between the PDS and SUS PbR contained an added element of complexity not shown in Figure 5.1. The PDS and SUS PbR datasets were not directly linkable using a single unique patient identifying variable. The PDS dataset contained a unique patient level identifier variable and a patient year identifying variable which the SUS PbR did not; instead the SUS PbR data contained a unique episode/attendance level identifier. A bridge file was used to link both datasets together, this was the only file which linked the unique episode/attendance level identifiers with the unique patient year identifier. The bridge file also had to be used to link different episodes and attendances for the same individual.

Using a bridge file meant that there were added complexities in using the data.

Firstly, without merging the bridge, PDS and SUS PbR files there was no way to identify all data relating to the same patient. Some hospital episodes/attendances were not matched to an individual patient; these unmatched episodes were removed from the data.

Secondly, the bridge file could not be easily split into years or logically into smaller files before any merging takes place. The bridge file contained few variables but had large numbers of observations as the bridge contained combinations of linkage variables for all years of the data. Combining the bridge file to either the PDS or SUS PbR became computationally intensive in terms of the amount of RAM and time needed to complete this task.

Figure 5.1: Data linkage for estimation sample



Patient List Example					
Person ID	Age	Gender	Age and sex group	GP practice	LSOA 11
100000001	52	Male	12	A00001	E01000001
100000002	16	Male	5	A00002	E01000002
100000003	65	Female	34	A00003	E01000003
100000004	27	Female	26	A00004	E01000004

Morbidity Example							
Person ID	A	B	C	AB	AC	BC	Count
100000001	1	1	1	1	1	1	3
100000002	1	1	0	1	0	0	2
100000003	0	0	1	0	0	0	0
100000004	0	0	0	0	0	0	0

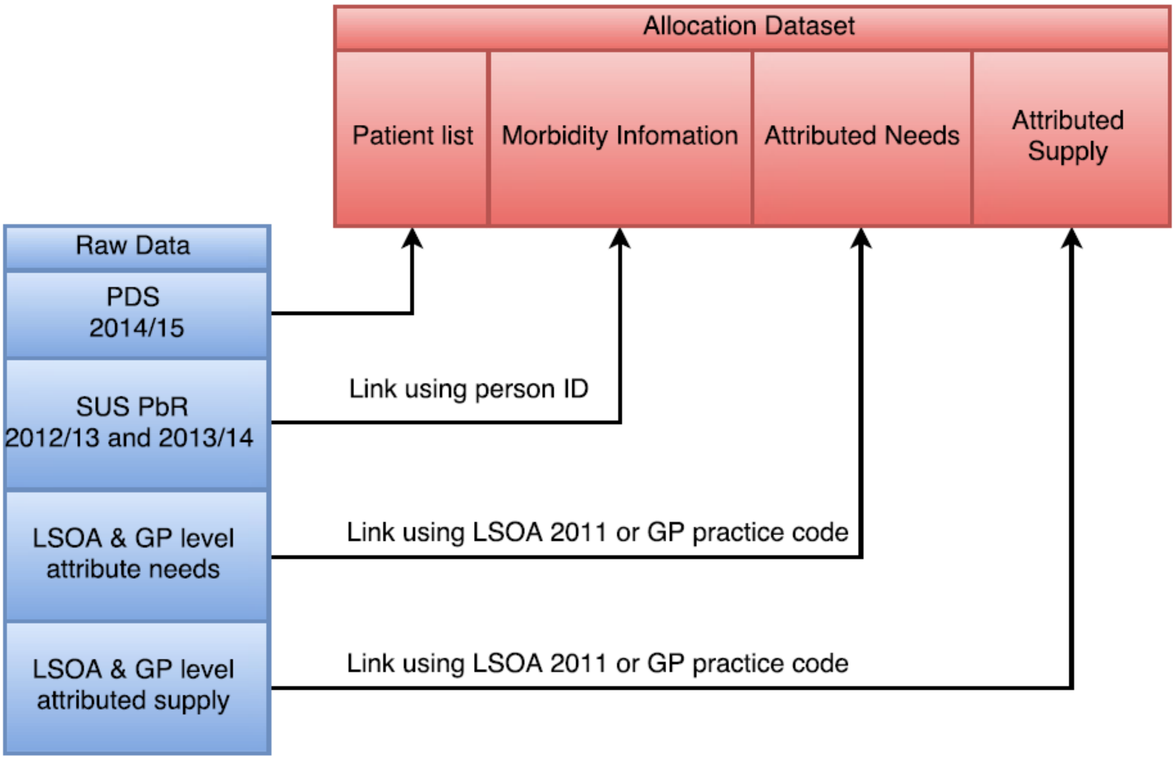
5.2.2 Data linkage to produce weights by age-sex group

Weights by age-sex group by GP practice were generated for 2014-15 using those registered with a GP practice in April 2014. This applied the coefficients from the final model to diagnostic data from SUS PbR for 2012-13 and 2013-14. These weights were used to inform allocations to CCGs from 2016-17, and are therefore known as the allocations dataset.

The starting point was the PDS for April 2014. Age was generated in the same way as above with the day of birth set to 1 for all individuals. Extracted fields were a pseudo-anonymised patient identifiers, age, sex, GP practice and LSOA. No samples were taken in the files for generating the weights by age-sex group; the weights were calculated using all those registered in April 2014.

Observed costs in 2014-15 were not relevant or calculated as we were calculating the weights for 2014-15 from the model outputs. SUS PbR APC data for financial years 2012-13 and 2013-14 were used to calculate the values of the morbidity variables. The following fields were extracted: pseudo-anonymised patient identifiers; spell and episode identifiers; and diagnostic information.

Figure 5.2: Data linkage for allocation weights dataset



The PDS dataset was divided by CCG before the morbidity information from SUS PbR was added. The dataset would have been be too large in terms of RAM requirements if the this dataset had not been split into by CCG.

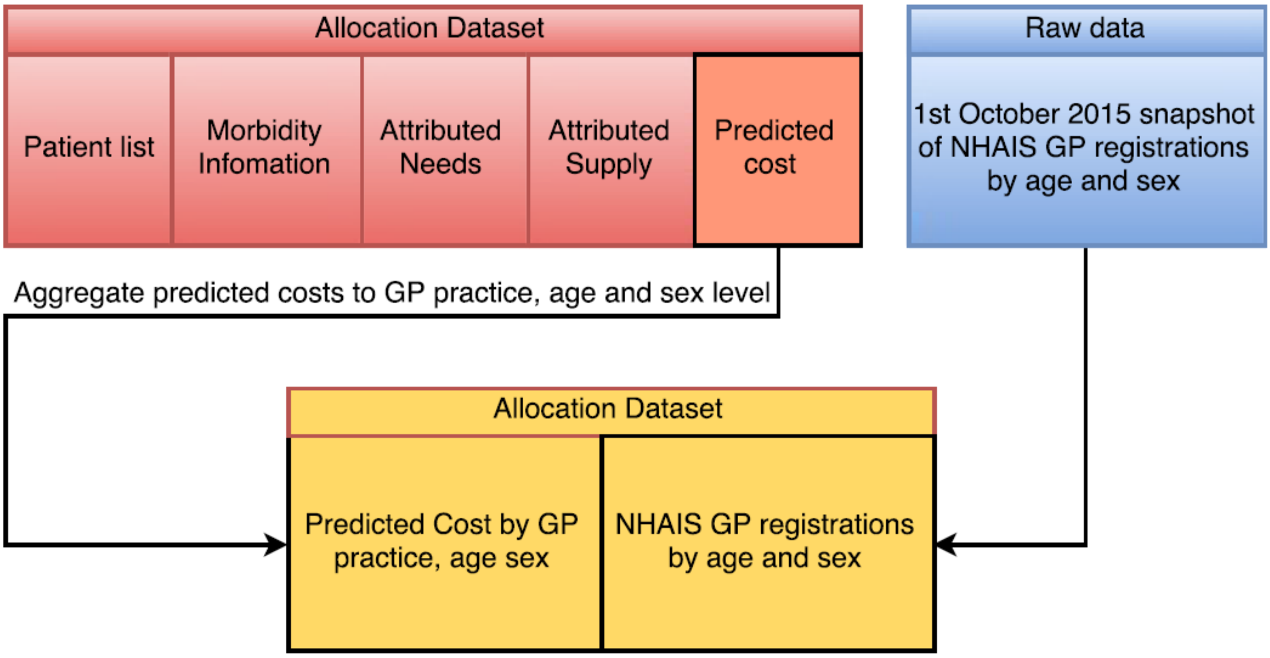
The steps for generating the allocations dataset were as follows.

- Extract all the required data fields from PDS April 2014.
- Add all the required attributed variables from the final model and CCG variables but not morbidity information.
- Freeze the supply and CCG variables by using the sample mean values as these are not indicators of need. The freezing of CCG and supply variables was conducted on the full PDS dataset to obtain correct and identical sample means. If freezing is conducted on each CCG file separately, each CCG file would have different values for each supply and CCG variable.
- Split the dataset into 211 files for each CCG and an additional file for where the CCG is not recorded.
- Using the pseudo-anonymised person identifier, add morbidity information for each separate CCG file. For each CCG file, unmatched observations from the SUS PbR dataset were discarded. These discarded observations would be present in another CCG file.

Predicted costs for each age-sex group in each GP practice were then linked to the GP registration list from NHAIS for 1st October 2015 as the most up to date data on GP registrations, CCG and age sex composition for a GP practice. This is shown in Figure 3.

NHAIS data were not linked on an individual level to SUS PbR and therefore were not used in the modelling.

Figure 5.3: Linking cost predictions to NHAIS registered lists



5.3 Variable construction

Many variables needed some manipulation to put them in the most appropriate form for the modelling. These are described in this section along with the definition of some of the other variables where this is not clear from the variable name.

5.3.1 Morbidity flags

Diagnoses for inpatient spells are used as the indicators of morbidity. 152 morbidity markers or flags were created from the diagnostic data. Each morbidity flag represented a sub-chapter of the International Classification of Diseases (ICD). Sub-chapters were used rather than individual ICD 10 codes as the number of observations for individual ICD 10 codes would be lower and which could introduce increased volatility in the model predictions.

The sub-chapters are broadly associated diagnosis codes. 147 groups are based primarily on the first three digits of the diagnosis code (the sub-chapters). There are a further 5 extra categories covering newly recognised diseases and morbidity attributable to external causes.

5.3.2 Number of years of morbidity data

Two years of data on diagnoses were used by PBRA 2011. We adopted the same approach as PBRA (2011) as the explanatory power of historical data on future hospital costs diminishes as the time lag increases. Therefore the inclusion of historical morbidity for more than the past two years may reduce the predictive performance of the models.

5.3.3 Gravity weighting for hospital supply characteristics

Hospital characteristics such as waiting times, distance from the patient's home, and numbers of beds and operating theatres could affect utilisation and therefore predicted costs per person. These supply variables were therefore tested for inclusion in the models.

A value for each of these variables needed to be attached, or attributed, to each person in the estimation sample and everyone in the calculation of the allocation weights. These needed to be attributed to those who did not use services as well as those who did uses services in 2013-14 This is achieved through gravity weighting.

Each individual in England has the potential to use hospital services anywhere in England; however, the probability of using a particular hospital reduces the further an individual resides from that hospital.

We took each individual's LSOA of residence and calculated the distance between the centroid of the LSOA and every hospital. The distances to each hospital from the centroid of the LSOA were used to give a weight to each hospital for that LSOA. This was repeated for all LSOAs where someone in the PDS data resides.

Distance between a LSOA and a hospital is calculated using Northing and Easting coordinates from the ONS. A Northing coordinate is the distance in metres north from the most southern point on the ordinance survey map of England. An Easting coordinate is the distance in metres east from the western most point on the ordinance survey map.

Northing and Easting coordinates are published by the ONS for the population weighted centroid of each LSOA. Patient postcodes are not available in the data set due to patient confidentiality and therefore the centroid of the LSOA was used.

Northing and Easting Coordinates are published by the ONS for each postcode, which were used to give the Northing and Easting coordinates for each hospital. The postcode used was the postcode of the NHS Trust as opposed to a hospital site which is more detailed and more precise. We did not use the postcode for each hospital site as the data to be attributed were for the whole Trust and not by site.

Distance between a LSOA centroid and a hospital can be obtained using Pythagoras's theorem by taking the square root of the sum of the squared differences of Northing and Easting coordinates, as shown in Equation 5.1:

$$D_{PL} = \sqrt{(N_L - N_P)^2 + (E_L - E_P)^2}$$

Equation 5.1

Where D is distance measures in kilometres, N is the northing coordinate, E is the easting coordinate, L signifies LSOA and P signifies hospital provider

Distance measured in kilometres was calculated for each LSOA and hospital combination. The formula used to create the LSOA gravity weighting of hospital level attributed variables is shown in Equation 5.2:

$$\mathrm{LSOA\ Value_L} = \sum_{P=1}^{N_P} \frac{\dfrac{i_P}{(D_{PL}+10)^G}}{\sum_{L=1}^{N_L} \dfrac{Pop_L}{(D_{PL}+10)^G}}$$

Equation 5.2

L signifies LSOA and P signifies hospital provider. $\mathrm{N_P}$ is the number of hospital providers, $\mathrm{N_L}$ is the number of LSOAs. $\mathrm{i_P}$ is the hospital attributed variable that should be a count variable, for example number of beds. $\mathrm{Pop_L}$ is the population of each LSOA. $\mathrm{D_{PL}}$ is distance measured in kilometres between each LSOA and Provider. G is the gravity weight.

A gravity weighting of 2 was applied. The higher the gravity weighting the lower the LSOA value has as the effect of distance becomes more of a barrier to healthcare than with low values of the gravity weighting. The LSOA value from gravity weighting is the proportion of hospital supply per capita of the LSOA population. A list of gravity weighted variables with descriptive statistics is shown in Table 5.1.

Table 5.1: List of gravity weighted variables and descriptive statistics (32,844 observations)

Variable	Mean	Std Deviation	Minimum	Maximum
Direct Distance	0.0000917	0.0000763	6.42E-06	0.0013118
Distance	0.0001213	0.0000976	9.10E-06	0.0016389
Driving Time	0.0002279	0.0001476	0.0000284	0.0063361
2011-12 Proportion of non-admitted patients RTT under 18 weeks	3.55E-06	1.18E-06	2.51E-07	7.87E-06
2011-12 Proportion of admitted patients RTT under 18 weeks	2.83E-06	9.62E-07	2.18E-07	6.68E-06
2012-13 Median waiting times (weeks) for non-admitted patients	0.000018	6.48E-06	1.29E-06	0.0000656
2012-13 Median waiting times (weeks) of the 95th percentile for non-admitted patients	0.0000503	0.0000165	3.84E-06	0.0001432
2012-13 Median waiting times (weeks) for admitted patients	0.000027	9.16E-06	2.20E-06	0.0000799
2012-13 Median waiting times (weeks) of the 95th percentile for admitted patients	0.0000665	0.0000233	4.92E-06	0.0001994
2012-13 Median waiting times (weeks) for Cardiology patients	0.0000169	5.20E-06	1.81E-06	0.0000875
2012-13 Median waiting times (weeks) for Cardiothoracic Surgery Patients	3.88E-06	2.82E-06	2.62E-07	0.0000282
2012-13 Median waiting times (weeks) for Dermatology Patients	0.0000172	5.53E-06	1.76E-06	0.0000705
2012-13 Median waiting times (weeks) for ENT Patients	0.0000289	8.54E-06	2.38E-06	0.0000797
2012-13 Median waiting times (weeks) for Gastroenterology Patients	0.0000203	6.40E-06	2.07E-06	0.0000776
2012-13 Median waiting times (weeks) for General Medicine Patients	9.22E-06	3.41E-06	7.43E-07	0.0000363
2012-13 Median waiting times (weeks) for General Surgery Patients	0.0000344	9.38E-06	3.28E-06	0.0001145
2012-13 Median waiting times (weeks) for Geriatric Medicine Patients	8.75E-06	3.49E-06	5.95E-07	0.000025
2012-13 Median waiting times (weeks) for Gynaecology Patients	0.0000278	8.22E-06	1.83E-06	0.000056
2012-13 Median waiting times (weeks) for Neurology Patients	0.0000163	5.80E-06	1.43E-06	0.0000531
2012-13 Median waiting times (weeks) for Neurosurgery Patients	7.74E-06	5.15E-06	4.56E-07	0.0000668
2012-13 Median waiting times (weeks) for Ophthalmology Patients	0.0000326	9.63E-06	3.05E-06	0.0000856
2012-13 Median waiting times (weeks) for Oral Surgery Patients	0.0000248	7.60E-06	2.23E-06	0.0000795
2012-13 Median waiting times (weeks) for Other Patients	0.0000263	8.32E-06	2.68E-06	0.0000894
2012-13 Median waiting times (weeks) for Plastic Surgery Patients	0.0000133	5.98E-06	8.28E-07	0.0000515
2012-13 Median waiting times (weeks) for Rheumatology Patients	0.0000165	5.10E-06	1.25E-06	0.0000587
2012-13 Median waiting times (weeks) for Thoracic Medicine Patients	0.0000113	3.79E-06	9.11E-07	0.0000387
2012-13 Median waiting times (weeks) for Trauma Orthopaedics Patients	0.0000471	0.0000122	4.34E-06	0.0001296
2012-13 Median waiting times (weeks) for Urology Patients	0.0000287	9.06E-06	2.81E-06	0.0001052
2012-13 Median waiting times (weeks) of the 95th percentile for Cardiology patients	0.000045	0.0000139	4.48E-06	0.0001979
2012-13 Median waiting times (weeks) of the 95th percentile for Cardiothoracic Surgery Patients	0.0000135	9.80E-06	8.01E-07	0.0000943
2012-13 Median waiting times (weeks) of the 95th percentile for Dermatology Patients	0.0000421	0.0000135	3.98E-06	0.0001282
2012-13 Median waiting times (weeks) of the 95th percentile for ENT Patients	0.0000654	0.0000195	5.10E-06	0.0001468
2012-13 Median waiting times (weeks) of the 95th percentile for Gastroenterology Patients	0.0000527	0.0000164	4.75E-06	0.0001646
2012-13 Median waiting times (weeks) of the 95th percentile for General Medicine Patients	0.0000329	0.0000111	2.41E-06	0.0000892

Variable	Mean	Std Deviation	Minimum	Maximum
2012-13 Median waiting times (weeks) of the 95th percentile for General Surgery Patients	0.0000883	0.000024	8.08E-06	0.0002649
2012-13 Median waiting times (weeks) of the 95th percentile for Geriatric Medicine Patients	0.0000316	0.0000108	2.20E-06	0.0000788
2012-13 Median waiting times (weeks) of the 95th percentile for Gynaecology Patients	0.0000696	0.0000197	5.00E-06	0.0001615
2012-13 Median waiting times (weeks) of the 95th percentile for Neurology Patients	0.0000359	0.0000122	3.10E-06	0.0001094
2012-13 Median waiting times (weeks) of the 95th percentile for Neurosurgery Patients	0.000019	0.0000123	1.12E-06	0.0001601
2012-13 Median waiting times (weeks) of the 95th percentile for Ophthalmology Patients	0.0000642	0.0000177	5.87E-06	0.0001929
2012-13 Median waiting times (weeks) of the 95th percentile for Oral Surgery Patients	0.0000492	0.000015	4.09E-06	0.0001476
2012-13 Median waiting times (weeks) of the 95th percentile for Other Patients	0.0000714	0.0000218	6.61E-06	0.0002247
2012-13 Median waiting times (weeks) of the 95th percentile for Plastic Surgery Patients	0.000036	0.0000155	2.09E-06	0.000117
2012-13 Median waiting times (weeks) of the 95th percentile for Rheumatology Patients	0.0000386	0.000012	2.66E-06	0.0001162
2012-13 Median waiting times (weeks) of the 95th percentile for Thoracic Medicine Patients	0.0000337	0.0000109	2.78E-06	0.0001046
2012-13 Median waiting times (weeks) of the 95th percentile for Trauma Orthopaedics Patients	0.000107	0.0000279	1.00E-05	0.0002803
2012-13 Median waiting times (weeks) of the 95th percentile for Urology Patients	0.0000724	0.0000221	6.99E-06	0.0002518
General & acute beds	0.0019402	0.000641	0.0001444	0.0050978
CT scans 2012-13	0.0861707	0.0290597	0.0072464	0.3179129
MRI scans 2012-13	0.0444471	0.0155526	0.0034196	0.1743684
General & acute day beds	0.0002161	0.0000862	0.0000147	0.0005789
Obstetric ultra-sound	0.0526194	0.020826	0.0036599	0.1499094
Non obstetric ultra-sound	0.1167138	0.0397543	0.0100033	0.4610751
Radio-isotopes	0.0109208	0.0049477	0.0009241	0.0426403
Radio-graphs no fluoroscopy	0.4124169	0.1326594	0.0335484	1.362225
Fluoroscopy	0.0239928	0.0083728	0.0021612	0.0964001
Number of operating theatres	0.0000549	0.0000193	4.25E-06	0.0001569
Number of day operating theatres	0.0000109	4.43E-06	5.05E-07	0.0000298
Number of adult critical beds January 2013	0.00007	0.000031	4.15E-06	0.0001989
Number of adult critical beds April 2013	0.00007	0.000031	4.15E-06	0.0001989

5.3.4 New GP practice marker

Using the PDS dataset, we are able to find patients that registered with a new GP practice between April 2012 and April 2013. If a patient has moved GP practice, the individual is given the value of one and zero otherwise. We do not account for the number of times a patient has moved GP practices as we only have yearly snapshots of which GP practice the patient is registered with from the PDS.

This variable seeks to test whether patients who have recently moved GP practice have higher costs due to higher need.

PBRA (2011) found new GP registrations are on average associated with higher costs in the year of moving GP practices. New GP registrations may be the result of a patient moving to reside closer to specialist healthcare centres, or patients moving into residential care homes. However, patients moving to new areas and changing GP practices may not always have higher needs, as some migration is, for example, students moving to and from universities, people moving for employment reasons, and people moving for family reasons.

5.3.5 Privately funded care marker

SUS PbR recorded episodes of care that were privately funded at an NHS healthcare facility. Using this information, a dummy variable was created taking the value of one if the individual had had privately funded care in the years 2011-12 and 2012-13, and the value of zero otherwise. This variable does not count the number of private care episodes each patient had; it only measures if a patient had received any private health care in the two years prior to the cost year.

Privately funded care was excluded from the costs in 2013-14 but was not removed from the diagnostic data for the two years prior to the cost year. Individuals may still use publicly funded healthcare and therefore morbidity information from the privately funded care is needed to predict costs for all individuals.

This variable was used by PBRA (2011) and was constructed in the same way.

5.3.6 Log population variance

The population variance variable is used to capture any GP list size reporting errors. The ONS population sizes were used to test for this, though it is recognised there legitimate reasons for GP lists and ONS populations to differ, and ONS populations also include some errors.

Log population variance is a variable which focuses on potential registered list size ‘inflation,’ due to for example unavoidable delays in removing patients from the practice’s list. To generate this variable, for each LSOA, a logarithm of the number of residents from ONS population data is subtracted from the logarithm of the number of registered patients from the PDS resident in the same area. The difference of the logged values gives the GP list size ‘inflation.’ This is shown in Equation 5.3.

$$\text{Log population variance} = \log(\text{PDS population}) - \log(\text{ONS population})$$

Equation 5.3

5.3.7 Age standardisation

Age-sex standardisation was carried out on three attributed need variables as they could otherwise create problems with the robustness of the model. These were not age-sex standardised at source and therefore had to be standardised as part of this project.

The variables we age-sex standardised were the proportion of the population with no qualifications, the proportion of the population with a limiting long term illness and the proportion not in good health. These variables were age-sex standardised using equation 5.4:

$$Std\ x_l = 100 \times \frac{\left(\frac{x_{la}}{N_{la}}\right) \times N_a}{N}$$

Equation 5.4

Where x_l= variable value as a count of people for LSOA l, a= age-sex groups and *N*=observations.

5.4 Generating the sample files

5.4.1 Reasons for using samples

The whole dataset could not be used for the modelling as it was beyond the capacity of the (high specification) computers available for this project. Furthermore, using the full population increases the amount of time needed to complete this project.

Samples of individuals were therefore used for the modelling. Regression coefficients become stable when samples of 10% or over are taken from the overall population (Morris et al, 2007[7]) and there would therefore be little advantage in undertaking the modelling using the whole population of England.

PRBA (2011) also took samples for the modelling. We used 15% samples rather than 10% as in PBRA (2011). Although estimates are stable at 10%, the extra computing power available to us enabled the use of a larger dataset. A 15% sample increases the confidence that the regression coefficients are robust.

The model design was carried out on one sample. The variables selected and coefficients were then tested on a validation sample which was used to compare predicted and observed costs.

5.4.2 Selection of the samples

The starting point is the PDS data on GP registrations for 1st April 2013. This file has 55,540,852 registered patients. This file contains a patient identifier, age, gender, GP practice and output area of individual residence (Lower Super Output Area (LSOA)).

Before taking the samples, we excluded a number of individuals from the registered patient list. The reasons were invalid data or the individual was no longer registered

[7] Morris S, Carr-Hill R, Dixon P, Law M, Rice N, Sutton M, and Vallejo-Torres L. (2007) *Combining Age Related and Additional Needs (CARAN) Report. The 2007 Review of the Needs Formulae for Hospital Services and Prescribing Activity in England.* London: Department of Health.

with a GP practice. The reason for exclusion and the number of individuals excluded are shown in Table 5.2.

Table 5.2: Exclusion of individuals from PDS registrations for 1st April 2013

Exclusion reasons	Number of people	% of all people
Patients that died before 1st April 2013	88,514	0.16%
Patients with invalid gender	217	<0.01%
GP Practice located outside of England	4,066	0.01%
Invalid GP Practice code	17,251	0.03%
GP practices with zero list size in 2011, 2012	62,586	0.11%
No longer registered	33,961	0.06%
Total removed individuals	206,595	0.37%

A total of 206,595 individuals (0.37%) were removed. This gave total a list of 55,334,257 individuals, from which the samples were taken.

Three main samples were taken as follows.

- Estimation sample (S1) was used to develop the models and also used to conduct statistical analysis. This was a 15% randomly selected sample of 8,291,907 individuals from 8,760 GP practices.
- Validation sample (S2) was used to validate the performance of the models at the individual level. Like sample S1, it is a random sample of 15% of 8,293,838 individuals[8] from 8,756 GP practices. Samples S1 and S2 were mutually exclusive, so no individual was in both samples S1 and S2.
- Validation sample (S3) was also used to validate the performance of the models. S3 was all those registered with a randomly selected sample of 15% of GP practices with 1,000 or more patients. This sample had 8,318,853 individuals from 1,218 GP practices.

The random samples were taken by using the population list, PDS, and generating a random variable from a uniform distribution. This will assign all people in the PDS a random number between 0 and 1. Estimation sample S1 were all people assigned a number between 0 and less than 0.15. Validation sample S2 were all people assigned a number between 0.15 and less than 0.3. As the random number was generated before any samples were taken, the number of observations in samples S1 and S2 should be similar.

In generating the validation sample S3, all GP practices with fewer than 1000 registered patients were removed. For the remaining GP practices, each GP practice was assigned a random number from a uniform distribution taking values between 0 and 1. Practices with an assigned number between 0 and under 0.15 were included in the S3 sample. Some patients in samples S1 and S2 will be present in the GP practice based patient sample S3.

[8] The small difference in the number of individuals in samples S1 and S2 was only due to the random number generating process.

6 General and Acute: Developing the model

In this section, we set out how we developed the model. This included determining how best to use the morbidity information from SUS PbR and how attributed needs and supply variables are selected.

The steps in developing the model were undertaken as in the sequence of this section. A different sequence could have given different results.

6.1 Number of diagnostic positions

For the estimation sample S1, 152 morbidity markers or flags were created using the diagnoses data for inpatient episodes in 2011-12 and 2012-13 in the SUS PbR data set.

The PbR SUS data provided had up to 13 diagnostic codes, where diagnostic position one in the data, the primary diagnosis, is usually the reason for admission. Each diagnosis has an International Classification of Disease version 10 (ICD10) code.

We needed to estimate the optimal number of diagnostic codes or positions for each episode to use in calculating the morbidity markers. Information on diagnostic position one is likely to be more useful at predicting costs than diagnosis position 13. While at face value using more diagnostic positions would seem preferential, the identification of the appropriate payment by results tariffs is unlikely to use all the diagnostic information. The use of all 13 diagnostic positions may, therefore, give a poorer model than the use of a smaller number of diagnostic positions.

PBRA (2011) found that using up to six diagnostic positions to create morbidity flags provided the best goodness of fit. PBRA (2011) also found that using more than six diagnostic positions resulted in a lower goodness of fit than using six diagnostic positions.

We followed the same approach as PBRA (2011) to find the optimal number of diagnostic positions to include.

Morbidity flags were calculated from diagnoses data for episodes. Patients could have multiple episodes within each spell and multiple spells in the financial years 2011-12 and 2012-13. To create the morbidity flags, we documented the lowest diagnostic position across all episodes. Therefore a patient could have multiple diagnoses in diagnostic position 1.

The method for testing the number of diagnostic positions to use is as follows.

- 152 morbidity flags were initially generated using only the first diagnostic position. The variables in the PBRA 3 Nuffield model were used to estimate the goodness of fit. A description of the PBRA 3 Nuffield model is given below.
- 152 morbidity flags were then generated using the first and second diagnostic positions and the PBRA 3 Nuffield model re-estimated and the goodness of fit statistics obtained.
- This process of accounting for an additional morbidity position was repeated until we estimated the PBRA 3 model using a set of morbidity flags which used information from all 13 available diagnostic positions.

To assess the goodness of fit we use the following statistics.

- The adjusted R squared statistic. This is based on the R squared statistic which takes values between zero and one, where zero signifies no level of fit and one signifies a perfect fit. We adopt the adjusted R squared statistic as this accounts for the number of variables included in the model.
- Mean Absolute Error (MAE) which is the absolute difference between the predicted and observed cost.
- Akaike Information Criterion (AIC). This assesses both the goodness of fit and the number of variables included in the model. The AIC statistic does not contain an upper bound and is indicative of the information loss when using the model. Therefore a lower AIC is indicative of a better goodness of fit.
- Bayesian Information Criterion (BIC). This is an alternative approach to AIC.

We estimated the models to obtain coefficients and goodness of fit metrics on the estimation sample S1 and validated the results using a sample S2. The model specification used to estimate the optimal number of diagnostic positions contains 38 age-sex group variables, CCG dummy variables, new practice marker, private care utilisation marker, PBRA Nuffield selection of attributed needs and supply variables and 152 morbidity flags.

The goodness of fit statistics for predicted 2013-2014 costs are in Figure 6.1,

Figure 6.2, Figure 6.3 and Table 6.1.

Figure 6.1: Adjusted R-squared using different number of diagnostic positions to create morbidity flags

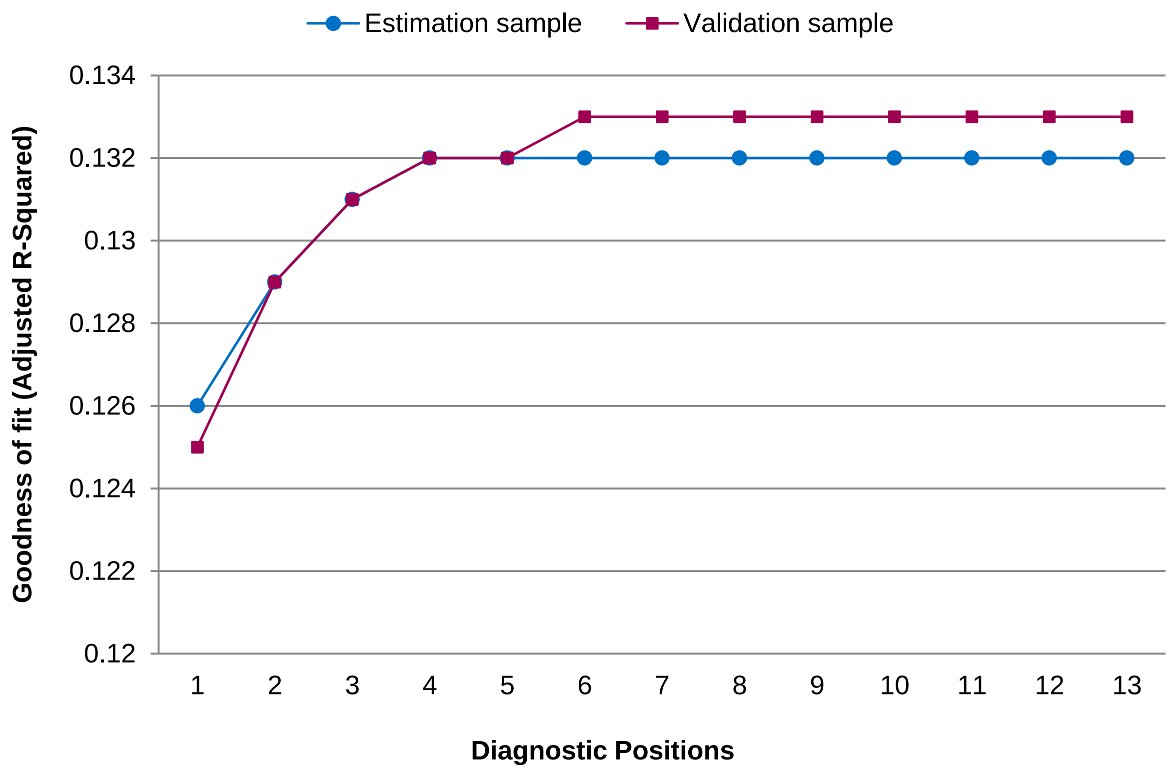


Figure 6.2: AIC using different number of diagnostic positions to create morbidity flags

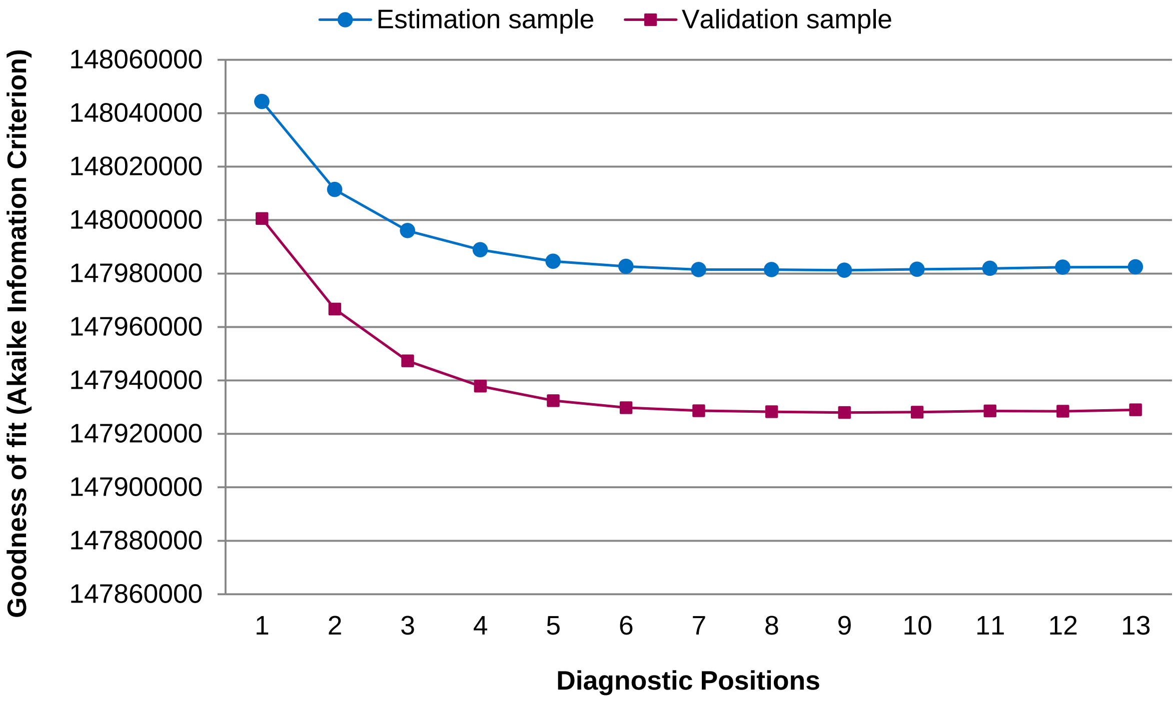


Figure 6.3: Mean absolute error using different diagnostic positions to create morbidity flags

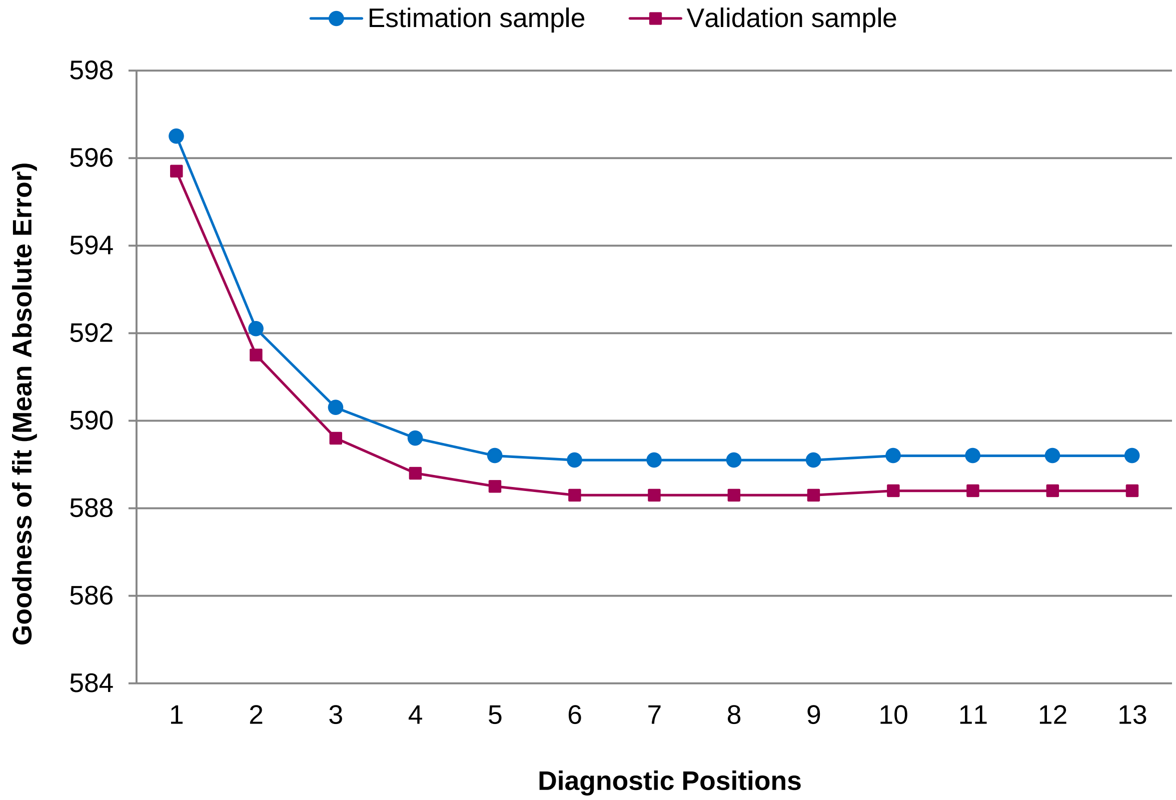


Table 6.1: Effect of number of diagnostic positions on goodness of fit

	Estimation sample				Validation sample			
	Adj Rsq	MAE	AIC	BIC	Adj Rsq	MAE	AIC	BIC
Diagnostic position 1	0.126	596.5	148044363.8	148050103.1	0.125	595.7	148000573.2	148000587.1
Diagnostic positions 1&2	0.129	592.1	148011400.5	148017153.6	0.129	591.5	147966724.9	147966738.9
Diagnostic positions 1-3	0.131	590.3	147996022.8	148001789.9	0.131	589.6	147947315.3	147947329.2
Diagnostic positions 1-4	0.132	589.6	147988886.8	147994653.9	0.132	588.8	147937848.5	147937862.5
Diagnostic positions 1-5	0.132	589.2	147984574.9	147990342	0.132	588.5	147932457	147932470.9
Diagnostic positions 1-6	0.132	589.1	147982648.9	147988416	0.133	588.3	147929806	147929819.9
Diagnostic positions 1-7	0.132	589.1	147981455.5	147987222.6	0.133	588.3	147928665.1	147928679.1
Diagnostic positions 1-8	0.132	589.1	147981441.6	147987208.7	0.133	588.3	147928283.4	147928297.3
Diagnostic positions 1-9	0.132	589.1	147981218.1	147986985.2	0.133	588.3	147927995.2	147928009.2
Diagnostic positions 1-10	0.132	589.2	147981583.5	147987350.6	0.133	588.4	147928145.2	147928159.2
Diagnostic positions 1-11	0.132	589.2	147981907.2	147987674.3	0.133	588.4	147928590.3	147928604.3
Diagnostic positions 1-12	0.132	589.2	147982344.9	147988112	0.133	588.4	147928453.2	147928467.1
Diagnostic positions 1-13	0.132	589.2	147982436.2	147988203.3	0.133	588.4	147929011.8	147929025.7

We found the six diagnostic positions adopted by PBRA Nuffield is no longer the optimal number. The optimal number of diagnostic positions to use to create the morbidity flags is between seven and nine. The Akaike and Bayesian information criteria select a range of seven to twelve diagnosis positions, however, the mean absolute error increases for diagnostic positions higher than nine.

Seven diagnostic positions were selected. Seven diagnostic positions was found to have the joint highest adjusted R-squared, joint lowest mean absolute error, and performs well when looking at the two information criteria measures.

6.2 Comorbidity interactions

The cost of health services used is likely to be affected by the type of comorbidities. The number of comorbidities is covered further below. To capture the differences in costs for each comorbidity interaction, we generated variables for the interactions of each of the main chapter ICD10 codes. The main chapters are shown in Table 6.2.

Table 6.2: Main Chapters of ICD10 Classifications

Main Chapters	Descriptions of main chapters
A00-B99	Certain infectious and parasitic diseases
C00-D48	Neoplasms
D50-D89	Diseases of the blood and blood-forming organs and certain disorders involving the immune mechanism
E00-E90	Endocrine, nutritional and metabolic diseases
F00-F99	Mental and behavioural disorders
G00-G99	Diseases of the nervous system
H00-H59	Diseases of the eye and adnexa
H60-H95	Diseases of the ear and mastoid process
I00-I99	Diseases of the circulatory system
J00-J99	Diseases of the respiratory system
K00-K93	Diseases of the digestive system
L00-L99	Diseases of the skin and subcutaneous tissue

Main Chapters	Descriptions of main chapters
M00-M99	Diseases of the musculoskeletal system and connective tissue
N00-N99	Diseases of the genitourinary system
O00-O99	Pregnancy, childbirth and the puerperium
P00-P96	Certain conditions originating in the perinatal period
Q00-Q99	Congenital malformations, deformations and chromosomal abnormalities
R00-R99	Symptoms, signs and abnormal clinical and laboratory findings, not elsewhere classified
S00-T98	Injury, poisoning and certain other consequences of external causes
V01-Y98	External causes of morbidity and mortality
Z00-Z99	Factors influencing health status and contact with health services

Comorbidity interaction terms were generated from taking one main disease chapter and interacting with one other main disease chapter. The total of 21 main disease chapters generated 210 possible comorbidity interaction terms. To select the comorbidity interaction terms to be included in the models, we ran a regression model using the PBRA Nuffield specification and initially included all 210 comorbidity interaction terms. Comorbidity interactions with a P value <0.001 were selected for the models and the others were not used. This follows the same approach as PBRA (2011). A total of 40 comorbidity interaction terms are used. These are shown in Table 6.3. In generating the comorbidity interactions we used the morbidity flags for each individual. This meant that the number of diagnostic positions used to calculate the morbidity flags must be determined before the co-morbidity interactions are generated.

Table 6.3 Co-Morbidity Interactions of main ICD10 classification chapters

Co-Morbidity Interactions of main ICD10 chapters			
A00B99-G00G99	A00B99-H00H59	A00B99-O00O99	A00B99-Q00Q99
A00B99-Z00Z99	C00D48-H00H59	C00D48-L00L99	C00D48-N00N99
C00D48-P00P96	C00D48-Z00Z99	D50D89-K00K93	D50D89-O00O99
D50D89-Z00Z99	E00E90-G00G99	E00E90-H00H59	E00E90-I00I99
E00E90-L00L99	E00E90-O00O99	E00E90-R00R99	E00E90-Z00Z99
F00F99-I00I99	F00F99-J00J99	F00F99-O00O99	F00F99-R00R99
H00H59-L00L99	I00I99-K00K93	I00I99-L00L99	J00J99-O00O99
K00K93-N00N99	K00K93-Q00Q99	L00L99-M00M99	L00L99-R00R99
M00M99-N00N99	M00M99-O00O99	N00N99-Q00Q99	N00N99-S00T98
N00N99-Z00Z99	O00O99-R00R99	O00O99-Z00Z99	V01Y98-Z00Z99

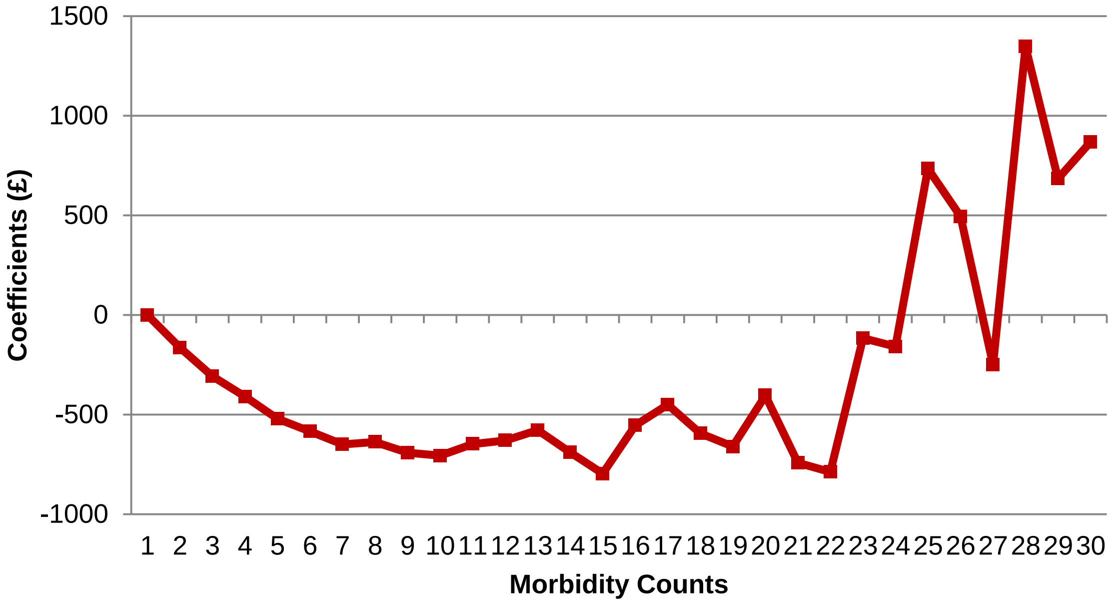
6.3 Morbidity counts

Morbidity counts were introduced to the PBRA Nuffield model as a single patient with two morbidity flags may cost less than two separate patients with a single morbidity flag each. This is because two conditions can be treated in the same hospital spell.

PBRA (2011) tested dummy variables for morbidity counts of 1 to 29 and a single variable for 30 and over. Dummy variables were used as there is a non-linearity in the relationship between the number of morbidity flags and costs.

To test the number of morbidity counts to include, we used the PBRA3 Nuffield model with our new set of comorbidity interactions terms. The coefficients (in £s) on modelling up to 30 morbidity counts are shown in Figure 6.4. The higher the number of morbidity counts that are introduced, the more volatile become the coefficients of the higher counts, as there are fewer people in each. This is the reason for the large volatility in the coefficients when counts of over 13 are included.

Figure 6.4: The coefficients of the 30 morbidity counts



The assumption is that morbidity flags from previous hospital service should be indicative of higher future costs. Therefore the morbidity flags should have a positive and significant coefficient. There will be exceptions to this assumption as some procedures may prevent further service use and therefore lead to a lower cost.

The method adopted by Nuffield PBRA team when selecting the number of morbidity count dummies to include was based on the significance (P<0.05) and sign of the coefficients of the actual morbidity flags. We used the same approach as Nuffield and tested different sets of morbidity count variables. The results comparing models with no morbidity count variables, 6, 9 and 30 morbidity count variables are shown in Table 6.4; we do not show the results from all of the models for brevity.

Table 6.4: Percentage of morbidity flags with positive or negative coefficients, and significance (P<0.05)

	Excluding morbidity counts		6 morbidity counts		9 morbidity counts		30 morbidity counts	
	Positive	Negative	Positive	Negative	Positive	Negative	Positive	Negative
Significant	67.76	5.92	80.26	1.97	80.92	1.97	79.61	1.97
Not Significant	13.82	11.84	9.87	7.24	9.87	6.58	9.21	8.55

Using up to 9 morbidity counts generated the largest percentage of positive coefficients for the morbidity flags. We found that using different morbidity counts consistently identifies more positive and significant morbidity flags than not using any morbidity count. Using up to nine morbidity counts yields the lowest percentage of negative coefficients on the morbidity flags. Table 6.5 shows the goodness of fit statistics for the overall PBRA3 model from using different sets of morbidity count variables.

Table 6.5: Goodness of fit statistics from different sets of morbidity count variables

Morbidity counts	Adjusted R-Squared	Mean absolute error	AIC	BIC	diff AIC	diff BIC
3	0.134	588.9	147967126	147973478		
5	0.134	588.7	147964880	147971260	2245.4	2217.5
6	0.134	588.7	147964490	147970884	390.6	376.7
7	0.134	588.6	147964274	147970682	215.3	201.5
8	0.134	588.6	147964227	147970649	47.5	33.4
9	0.134	588.6	147964137	147970573	89.4	75.5
10	0.134	588.6	147964127	147970576	10.8	-3.1
15	0.134	588.6	147964039	147970559	87.2	17.6
20	0.134	588.6	147963935	147970524	104.4	34.8
30	0.134	588.7	147963383	147970111	552.3	412.9

We found that the different number of morbidity counts included in the model have little impact on the goodness of fit statistics. Using AIC and BIC statistics, the higher number of morbidity counts the better the goodness of fit. However, increasing the number of morbidity counts introduces more volatility. Using morbidity counts of up to nine appears to be the optimal model when a trade-off is made between model goodness of fit and parsimony.

Morbidity counts are generated for each person by counting the number of morbidity flags each person has. Once a morbidity count variable has been generated, individuals with counts of 10 or more are truncated to 9 counts. Individuals with one morbidity count will be replaced with zero counts, this is due to co-linearity between the morbidity count variable and morbidity flags.

6.4 Variables in all models

The refresh models all contain the same set of explanatory variable except for the attributed needs and supply variables. The common explanatory dummy variables in all the models were:

- 38 age-sex groups;
- 152 morbidity flags;
- 40 comorbidity interaction variables;
- 9 morbidity count variables;
- 211 CCG dummy variables;
- the new GP practice variable; and
- a private care variable.

6.5 Selection of attributed variables

The next step was to determine which attributed variables to include in the model. Attributed variables are recorded at area level, MSOA and LSOA, GP practice level or hospital level.

While morbidity flags are only available for individuals who had inpatient hospital spells in 2011-12 and 2012-13, all individuals in the samples have a set of attributed variables. If attributed variable values were missing for certain GP practices or LSOAs, CCG averages for those missing variables were applied.

In total we have 313 attributed needs and supply variables. Not all attributed variables will be correlated with costs and therefore, not all attributed variables are suitable to be used to predict future costs. Variable selection processes must be conducted to obtain a list of attributed variables that are associated with hospital costs.

To select attributed needs and supply variables for the models, we followed the approach used by PBRA Nuffield. All variable selection models are estimated on the estimation sample, S1.

Four different variable selection procedures followed are set out below.

1. T-statistic selection method.
    a. This method estimates at the person level a ‘full model’ containing age, sex, morbidity flags, morbidity interactions, morbidity counts, CCGs, new GP practice, private utilisation, population variance and all 313 attributed needs and supply variables.
    b. After running the model though ordinary least squares regression, we remove all attributed variables with t-statistic less than 0.2.
    c. After step 1b, we re-run the model but without variables omitted in stage 1 b. We then remove all attributed variables with t-statistics lower than 0.4.
    d. We repeat this process, increasing the t-statistic by 0.2 for each iteration, until we have removed all attributed variables with a t-statistic of under 2.
    e. The next step is to remove attributed variables with implausible coefficient signs, run the model again and discard variables with t-statistic of less than 2. Implausible signs were determined by three analysts developing the model, who separately stated their expectation for the sign being positive or negative (before seeing the results). These were then compared and discussed and a final view formed.
    f. Repeat step 1e until no attributed variables are removed, then we increase the t-statistic to 2.2 and repeat the process in 1e.
    g. Repeat step 1f until all attributed variable have a t-statistic of 2.58 and have the expected signs.

2. Forward stepwise procedure.
    a. First estimate a model at the individual level containing age, sex, morbidity flags, morbidity interactions, morbidity count, new GP practice, private utilisation, CCG dummy variables and population variance.
    b. Aggregate the individual level prediction errors (difference between observed and predicted cost) to give the prediction errors at GP practice level.
    c. Perform forwards stepwise procedure at the practice level using the estimated error in step 2b, all attributed needs and supply variables.
    d. Remove attributed variables with implausible signs.

3. Backwards stepwise procedure.

a. First estimate a model on individual level containing age sex, morbidity flags, morbidity interactions, morbidity count, new GP practice, private utilisation, CCG dummies and registration list variance.

b. Aggregate the individual level prediction errors (difference between observed and predicted cost) to give the prediction errors at GP practice level.

c. Perform backwards stepwise procedure on practice level using the estimated error in step 3b all attributed needs and supply variables and fixing all CCG variables

d. Remove attributed variables with implausible signs.

4. Once a list of attributed needs and supply variables are obtained from methods 1, 2 and 3, we use this list of attributed variables and perform a forwards and backwards variable selection process at an individual level (method 4).
    a. The model contains age, sex, morbidity flags, morbidity interactions, morbidity counts, CCGs, new GP practice, private utilisation, population variance and the list of attributed variables from methods 1, 2 and 3. The resulting model is named the PBRA Refresh model.

We estimated in total three model specifications.

1. PBRA 3 Nuffield model. This model has the same attributed needs and supply variables as PBRA3, but with updated morbidity flags, morbidity counts, and comorbidity interaction variables. The distance variable has also been updated to road distance.
2. PBRA Refresh model. This includes all the attributed needs and supply variables after method 4 of the variable selection procedures as above. This is referred to as CCG refresh.
3. T-Selection Model. This model includes all attributed needs and supply variables from variable selection process 1. This model contains the lowest number of attributed needs and supply variables of all the models.

The attributed needs and supply variables from the T-Selection process and the CCG refresh model are shown in Table 6.6.

Table 6.6: Attributed needs and supply variables

T-Selection variables	CCG refresh model
Attributed needs variables	
All Usual Residents Aged 16+	All Usual Residents Aged 16+
All Usual Residents Aged 16 to 74	All Usual Residents Aged 16 to 74
Resident Population	Resident Population
Proportion Single Pensioner Households	Proportion Single Pensioner Households
Proportion of population age 16-74 in routine occupations	Proportion of population age 16-74 in routine occupations
Proportion Single (never married)	Proportion Single (never married)
Proportion Divorced	Proportion Divorced
Rented from private landlord or letting agency	Rented from private landlord or letting agency
Proportion (un standardised) with not good health (NGH)	Persons in social rented housing
All people living in the area	Owner occupiers (Owned with a Mortgage or

T-Selection variables	CCG refresh model
	Loan)
IMD 2015 Health Deprivation Domain Score	Proportion (un standardised) with not good health (NGH)
Average with (long term) medical condition for those with at least one	All people living in the area
2012-13 QOF KD Total Exceptions	2012-13 QOF KD Total Exceptions
2012-13 QOF Epilepsy Prevalence	2012-13 QOF Epilepsy Prevalence
2012-13 QOF Mental Health Prevalence	2012-13 QOF Mental Health Prevalence
Log population variance between PDS and ONS	Average with (long term) medical condition for those with at least one
	IMD 2015 Health Deprivation Domain Score
	Children and Young People Sub-domain Score
	Log population variance between PDS and ONS
Attributed supply variables	
Adult critical beds Jan 13	2012-13 QOF Obesity Weighted Achievement Score
2012-13 Median waiting times (weeks) for Dermatology Patients	Distance between patient LSOA and hospital provider
2012-13 Median waiting times (weeks) of the 95th percentile for Neurosurgery Patients	Adult critical beds Jan 13
	2012-13 Median waiting times (weeks) for Dermatology Patients
	2012-13 Median waiting times (weeks) of the 95th percentile for Neurosurgery Patients

The T-selection process selected 16 need and 3 attributed supply variables. The CCG refresh selection process which used the same methods as PBRA Nuffield selected 19 attributed need variables and 5 attributed supply variables. The T-selection method produced a more parsimonious set of attributed needs and supply variables.

Descriptive statistics for all the attributed variables tested and other variables are given in Appendix A.

A list of the stages when variables were removed from the T-selection process is given in Appendix B.

The selection processes for attributed needs or supply variables explain variations in costs over and above person level variables (age, sex, morbidity flags etc,). Correlation in a set of attributed variables may impact on coefficients for these. However, as the models used for resource allocations are predictive models, including correlated variables will not harm the overall predictive power of the models. Furthermore, this also means each variable cannot be individually interpreted or used to draw conclusions with regard to the association of the variable with the total cost variable. It is the total set of variables combined that give the predicted costs.

Regressions

6.5.1 Ordinary Least Squares (OLS)

The regressions were all carried out using estimation sample S1. We used Ordinary Least Squares (OLS) regression to model costs. OLS was chosen as the datasets contained sufficient variation in costs per person and was large enough to give robust goodness of fit statistics.

Different modelling techniques could be explored in future research, however complex models introduce more underlying assumptions which may not be realistic and the transparency of the work would reduce when disseminating to non-technical audiences. PBRA (2011) found that more complex models did not out-perform OLS.

6.5.2 Age stratified models

For each of the four model specifications, we adopted two approaches: an all age model and an age stratified model. The age groups for the three stratified models were 0-14, 15-64 and 65 and over. The age stratified models allow the coefficients for the morbidity flags, attributed variables and other variables to have different values for each of the three age groups.

The relationship between cost and age is non-linear so age was included in the models through dummy variables for each age group rather than a single age variable.

6.5.3 High cost patients

The dependent variable is the observed costs hospital utilisation in 2013-14. Costs were truncated at £100,000 to remove outliers from the sample which could unduly affect the model coefficients.

The choice of the truncation value was a trade-off between the model's robustness and the total value of the costs which were to be predicted. The value of £100,000 was also used for the Nuffield PBRA formula. The decision to truncate at £100,000 took into account the following.

- Truncating costs is advantageous as it reduces the problem of modelling outliers. If outliers were present, that is individuals with high costs, the cost prediction for the average person will be too high and therefore not result in the best overall predictions.
- Several patients in the data have costs over £100,000 who had not used hospital services in 2011-12 and 2012-13. These individuals were inpatients for long periods of time and discharged during 2013-14 (the SUS PbR data only included completed spells). With no truncation, the predicted costs of such individuals would not be accurate and the average cost of people with no morbidity information would increase. Lowering the value of truncation lowers the notional allocation for high costs patients.
- The lower the level of truncation, the lower the total costs captured in the models.
- It was recognised that a value of £100,000 is an arbitrary number, but more conventional methods to generate a truncation value such as the 99th percentile would have removed too high a proportion of total costs from the modelling.
- Cost truncation at £100,000 led to the removal of a small proportion of total costs, namely 0.33% or £86.8 million. There was some regional bias; London accounted for 45.7% of the £86.8 million truncated costs, the South of England accounted for 26.6%, the North accounted for 14.3%,the Midlands and East accounted for 13.4%, and 0.04% was not attributable to a region.

7 General and acute: Results and model performance

7.1 Model performance

The goodness of fit statistics were obtained at the person level using validation sample S2. Overall model performance was tested using validation sample S3 which contains 100% of people from 15% of GP practices for those GP practices with over 1000 patients.

We used the coefficients from each of the model specifications estimated using sample S1 to predict costs for the individuals registered with the GP practices in sample S3 data. We then obtain the average predicted cost and average observed cost per head for each GP practice by averaging over each person registered at the GP practice.

Predicted GP practice level costs are given from:

$$\widehat{C_p} = \hat{\alpha} + L_p^{-1} \sum_{i \in p} \sum_j \widehat{\beta_j} N_{ipj} + L_p^{-1} \sum_{i \in p} \sum_k \widehat{\gamma_k} S_{ipk}$$

Equation 7.1

Where: $\widehat{C_p}$ denotes predicted cost per head $\hat{C}$ for practice p in one year, financial year 2013-2014
$\hat{\alpha}$ is the predicted constant term
N_{ipj} are j number of needs variables for individual i in the prior two years to the cost year, financial years 2011-2012 and 2012-2013.
S_{ipk} are k number of supply variables
L_p is the number of individuals registered to a GP practice in one point in time, PDS snapshot on 1st April 2013.
$\widehat{\beta_j}$ and $\widehat{\gamma_k}$ are the predicted coefficients.

We obtained goodness of fit statistics for the models using the average predicted cost and average observed costs per head for each GP practice in sample S3. Model specifications were assessed through multiple criteria as follows and the goodness of fit statistics are shown in Table 7.1.

- Adjusted R-squared. The proportion of the variation in observed costs explained by predicted costs.
- Mean absolute error in £s. The absolute mean difference in observed costs and predicted costs.
- Proportion of practices with predicted costs not within 10% of observed costs.
- Redistribution index. The proportion of total costs that would be reallocated across practices comparing predicted with observed costs.
- Mean absolute percentage change in practice shares between observed costs and predicted costs.
- Proportion of Practice shares substantially affected is the proportion of practices with the calculated mean absolute percentage change in share is greater than 5%.

The redistribution index, mean absolute percentage change and proportion of practice shares substantially affected were also used to compare the predictions from each model.

Equations for the goodness of fit statistics are shown in Appendix D.

To obtain the overall R-squared statistic for the age-stratified models, we combined the predicted costs from each age-stratified model and summed to give the predicted costs for all age-groups. We then ran a regression of actual costs with observed costs to obtain the goodness of fit statistics.

7.2 Final model

All the models performed well. The age-stratified models performed better than the all age models and we therefore recommend an age-stratified model.

The PBRA Nuffield specification performs the worst in terms of all the metrics we used. This is unsurprising as it would expected with more recent data that the best fitting set of attributed variables would change. The PBRA Nuffield specification is included mainly as a baseline comparator for our new specifications.

The age-stratified PBRA CCG model is the best performing model in terms of the goodness of fit statistics, however this model is the least parsimonious in terms of the number of attributed variables included (see Table 6.6).

The T-selection model is the recommended model due to goodness of fit statistics and parsimony. The T-selection has 5 fewer attributed variables, and performs well with an R-squared of 0.8502 and mean absolute error of £30.49 at GP practice level.

At CCG level, the difference in weighted populations between T-selection and PBRA CCG is minimal. The difference in the weighted populations is in the range -1.86% to +1.24%, and the lower and upper deciles are -0.52% and +0.55%.

Table 7.1 has the goodness of fit statistics for the three models. A selection of coefficients from the final age stratified T-selection model is show in Table 7.2 -Table 7.7. A full set of coefficients are shown in Appendix C: Coefficients from age stratified T-Statistic model for G&A model.

The main drivers in the models are diagnostic information and age-sex group variables. These alone give a R-squared of 0.8157, so the attributed variables explaining an adding 3.5% of the variance. However, the attributed variables affect the distribution between GP practices as measured by mean absolute error and the percentage of predictions outside of 10% of observed practice costs.

Table 7.1: Goodness of fit statistics from all models

	PBRA Nuffield		PBRA CCG		T-Stat Selection	
	All age	Age Stratified	All age	Age Stratified	All age	Age Stratified
R-Squared	0.8404	0.8453	0.8455	0.8509	0.8442	0.8502
Mean absolute error (£)	31.6315	31.3306	30.8534	30.3755	30.9893	30.4886
Proportion not within 10%	0.2301	0.2309	0.2243	0.2161	0.2251	0.2202
Redistribution index	0.029	0.0289	0.028	0.0277	0.0281	0.0278
Mean absolute percentage change in share	7.2167	7.17	7.0605	6.9642	7.0818	6.9791
Proportion of Practice shares substantially affected	0.5111	0.516	0.4938	0.4938	0.4971	0.4922
Redistribution index	0	0.0046	0.0082	0.0094	0.0078	0.0089
Mean absolute percentage change in share	0	1.0521	1.8583	2.162	1.77	2.0551
Proportion of Practice shares substantially affected	0	0.0172	0.0451	0.0722	0.0435	0.0689

Table 7.2: Coefficients (T selection method) for age and sex groups

	Age Group 0-14				Age Group 15-64				Age Group 65+			
	Male		Female		Male		Female		Male		Female	
Age	Coefficient	T-Statistic	Coefficient	T-Statistic	Coefficient	T-Statistic	Coefficient	T-Statistic	Coefficient	T-Statistic	Coefficient	T-Statistic
<1	.	.	-84.39	(-9.700)								
1-4	-170.5	(-23.344)	-209.2	(-28.837)								
5-9	-212.2	(-29.038)	-233.3	(-32.265)								
10-14	-193.2	(-26.394)	-206.8	(-28.405)								
15-19					.	.	16.89	(6.770)				
20-24					-4.153	(-1.722)	10.16	(4.283)				
25-29					-13.69	(-5.517)	21.14	(8.672)				
30-34					-8.775	(-3.353)	41.39	(15.471)				
35-39					4.831	(1.753)	66.32	(23.021)				
40-44					35.38	(11.484)	98.88	(31.900)				
45-49					76.45	(23.298)	139.3	(41.205)				
50-54					121.3	(32.809)	175.0	(44.745)				
55-59					197.7	(41.826)	205.9	(47.554)				
60-64					287.7	(51.499)	270.7	(53.106)				
65-69									.	.	-39.60	(-4.821)
70-74									173.1	(16.336)	106.2	(11.221)
75-79									359.8	(28.979)	243.7	(23.058)
80-84									565.2	(36.060)	427.8	(34.815)
85+									813.9	(44.474)	593.6	(46.087)

Table 7.3: Examples of coefficients on morbidity flags (highest and lowest)

	Age Group 0-14		Age Group 15-64		Age Group 65+	
	Coefficient	T-Statistic	Coefficient	T-Statistic	Coefficient	T-Statistic
C81-C96 Malignant neoplasms of lymphoid, haematopoietic & rel. tiss.	2127.9	(3.682)	4188.9	(17.272)	3284.7	(21.662)
M95-M99 Other disorders of the musculoskeletal system & conn. tiss.	2174.5	(1.220)	684.8	(3.669)	1377.1	(3.730)
U Unclassified	1002	(2.851)	1461.2	(6.433)	888.5	(5.117)
C45-C49 Malignant neoplasms of mesothelial and soft tissue	2152.9	(1.301)	491.4	(1.456)	363.6	(1.934)
K70-K77 Diseases of liver	1087.5	(1.894)	1124.4	(16.987)	747.8	(8.912)
O00-O08 Pregnancy with abortive outcome	-15.01	(-0.155)	95.18	(14.108)	2632.6	(1.453)
N17-N19 Renal failure	-190.2	(-0.486)	2063.1	(19.501)	801.1	(19.584)
C15-C26 Malignant neoplasm of digestive organs	-254.8	(-0.239)	1991.9	(14.608)	880.2	(11.483)
I70-I79 Diseases of arteries, arterioles & capillaries	819.1	(1.393)	1009.5	(12.466)	722.8	(16.072)
F70-F79 Mental retardation	1265.2	(3.266)	994.3	(6.841)	229	(0.749)
E10-E14 Diabetes Mellitus	1204.1	(12.590)	583	(22.414)	567.7	(21.422)
I00-I09 Rheumatic heart disease	-762.9	(-2.523)	249.7	(1.703)	137.3	(2.150)
N99 Other disorders of the genitourinary system	-317.9	(-1.438)	52.51	(0.345)	-194.5	(-1.224)
P05-P96 Other conditions originating in the perinatal period	468	(2.745)	239.2	(0.639)	-1287.8	(-1.424)
R69 Unknown & unspecified causes of morbidity	-608.6	(-3.649)	73.34	(0.331)	-144.8	(-0.274)
G00-G09 Inflammatory diseases of the central nervous system	-99.92	(-0.549)	-110.3	(-0.678)	-470	(-1.884)
B85-B99 Other infectious and parasitic diseases	-246.2	(-1.429)	-55.16	(-0.158)	-433.7	(-1.005)
C69-C72 Malignant neoplasms of eye, brain & other parts of CNS	8.191	(0.008)	-227.2	(-1.175)	-626.2	(-3.083)
C40-C41 Malignant neoplasm of bone and articular cartilage	-524.1	(-0.473)	-788.3	(-2.356)	190.9	(0.398)
A90-A99 Arthropod-borne viral fevers & viral haemorrhagic fevers	-196.2	(-2.130)	-40.24	(-0.298)	-1212.2	(-1.319)
B50-B64 Protozoal diseases	-162.1	(-0.759)	-435.4	(-2.679)	-910.5	(-1.362)
A50-A64 Infections with predominantly sexual mode of transmission	-4.412	(-0.019)	427.2	(1.543)	-3982.9	(-15.909)

Table 7.4: Coefficients on new GP practice, private care and morbidity counts

	Age Group 0-14		Age Group 15-64		Age Group 65+	
	Coefficient	T-Statistic	Coefficient	T-Statistic	Coefficient	T-Statistic
New GP practice	28.62	(10.3)	45.98	(21.526)	-43.83	(-2.58)
Private care	-77.98	(-1.44)	-35.63	(-0.861)	-182.0	(-2.89)
Morbidity Count 9						
2 morbidities	-98.34	(-5.07)	-51.81	(-4.621)	181.2	(5.81)
3 morbidities	-273.5	(-10.8)	-209.7	(-20.649)	34.92	(1.95)
4 morbidities	-390.0	(-10.7)	-248.0	(-15.721)	120.7	(3.15)
5 morbidities	-288.7	(-6.72)	-239.8	(-16.789)	37.52	(1.08)
6 morbidities	-508.1	(-12.2)	-400.7	(-26.824)	-101.1	(-4.48)
7 morbidities	-541.6	(-9.64)	-362.1	(-18.442)	25.70	(.672)
8 morbidities	-512.7	(-6.7)	-398.9	(-17.653)	-52.20	(-1.33)
9 morbidities	-691.8	(-10.3)	-604.2	(-26.795)	-208.0	(-7.49)

Table 7.5: Coefficients on interaction terms between ICD10 chapters (Part 1)

			Age Group 0-14		Age Group 15-64		Age Group 65+	
Combination	**Short description of disease groups**		Coefficient	T-Statistic	Coefficient	T-Statistic	Coefficient	T-Statistic
a00b99-g00g99	Infectious and parasitic	Nervous system	518.4	(2.93)	234.6	(1.720)	89.64	(.751)
a00b99-h00h59	Infectious and parasitic	Eye and adnexa	194.1	(.995)	701.9	(2.968)	220.6	(1.6)
a00b99-o00o99	Infectious and parasitic	Pregnancy, child birth	2439.3	(4.1)	-533.2	(-7.121)	.	.
a00b99-q00q99	Infectious and parasitic	Congenital health problems	368.6	(2.85)	440.9	(1.150)	676.4	(1.22)
a00b99-z00z99	Infectious and parasitic	Other factors	-16.49	(-.308)	298.0	(4.548)	111.0	(1.45)
c00d48-h00h59	Neoplasms	Eye and adnexa	278.3	(.445)	-186.0	(-1.203)	-184.1	(-2.82)
c00d48-l00l99	Neoplasms	Skin and subcut.tissue	34.44	(.069)	-177.5	(-1.661)	-209.0	(-2.73)
c00d48-n00n99	Neoplasms	Genitourinary system	110.2	(.134)	-90.64	(-1.722)	-124.5	(-2.28)
c00d48-p00p96	Neoplasms	Perinatal period	116.5	(.261)	-1616.8	(-2.485)	-93.33	(-.099)
c00d48-z00z99	Neoplasms	Other factors	-65.64	(-.339)	102.4	(2.961)	129.9	(3.36)
d50d89-k00k93	Blood and blood-forming etc	Digestive system	-276.2	(-.853)	-134.0	(-1.920)	-371.2	(-6.2)
d50d89-o00o99	Blood and blood-forming etc	Pregnancy, child birth	437.0	(.703)	-634.2	(-8.222)	-5455.2	(-11.3)
d50d89-z00z99	Blood and blood-forming etc	Other factors	948.9	(3.1)	482.6	(7.185)	274.8	(4.89)
e00e90-g00g99	Endocrine and metabolic	Nervous system	164.2	(.592)	82.71	(1.792)	52.16	(1.09)
e00e90-h00h59	Endocrine and metabolic	Eye and adnexa	1100.1	(2.09)	308.0	(5.204)	46.03	(1.37)
e00e90-i00i99	Endocrine and metabolic	Circulatory system	-91.62	(-.184)	-156.5	(-5.389)	-133.1	(-4.66)
e00e90-l00l99	Endocrine and metabolic	Skin and subcut.tissue	103.7	(.396)	322.3	(4.763)	166.8	(2.61)
e00e90-o00o99	Endocrine and metabolic	Pregnancy, child birth	523.5	(1.31)	-149.1	(-5.955)	11434.9	(27.2)
e00e90-r00r99	Endocrine and metabolic	Symptoms, signs	331.8	(3.51)	11.73	(0.449)	28.33	(.992)
e00e90-z00z99	Endocrine and metabolic	Other factors	346.3	(2.9)	59.74	(2.449)	-42.39	(-1.62)

Table 7.6: Coefficients on interaction terms between ICD10 chapters (Part 2)

Combination	Short description of disease groups		Age Group 0-14 Coefficient	Age Group 0-14 T-Statistic	Age Group 15-64 Coefficient	Age Group 15-64 T-Statistic	Age Group 65+ Coefficient	Age Group 65+ T-Statistic
f00f99-i00i99	Mental disorders	Circulatory system	-53.37	(-.119)	-12.55	(-0.485)	-61.85	(-1.83)
f00f99-j00j99	Mental disorders	Circulatory system	212.6	(1.92)	-31.04	(-1.358)	-28.44	(-.682)
f00f99-o00o99	Mental disorders	Pregnancy, child birth	235.7	(1.4)	-65.39	(-5.939)	3827.9	(4.94)
f00f99-r00r99	Mental disorders	Symptoms, signs	-20.88	(-.244)	-13.28	(-0.730)	14.63	(.391)
h00h59-l00l99	Ear and mastoid process	Skin and subcut.tissue	404.6	(1.48)	564.0	(2.985)	117.1	(1.23)
i00i99-k00k93	Circulatory system	Digestive system	427.8	(1.43)	-60.55	(-2.617)	-48.69	(-2.16)
i00i99-l00l99	Circulatory system	Skin and subcut.tissue	102.9	(.195)	282.1	(4.401)	131.7	(2.32)
j00j99-o00o99	Circulatory system	Pregnancy, child birth	293.2	(1.06)	-159.6	(-10.191)	-5688.3	(-7.09)
k00k93-n00n99	Digestive system	Genitourinary system	181.3	(1.47)	38.45	(1.097)	55.48	(1.45)
k00k93-q00q99	Digestive system	Congenital health problems	310.9	(2.71)	324.8	(2.601)	111.6	(.519)
l00l99-m00m99	Skin and subcut.tissue	Musculoskeletal	454.1	(1.62)	142.2	(2.153)	86.62	(1.37)
l00l99-r00r99	Skin and subcut.tissue	Symptoms, signs	43.04	(.89)	201.1	(3.441)	159.4	(2.52)
m00m99-n00n99	Musculoskeletal	Genitourinary system	180.5	(.492)	16.25	(0.351)	-248.0	(-6.47)
m00m99-o00o99	Musculoskeletal	Pregnancy, child birth	-401.8	(-.877)	-243.0	(-10.727)	-434.1	(-.567)
n00n99-q00q99	Genitourinary system	Congenital health problems	90.88	(.818)	136.4	(1.060)	803.7	(3.5)
n00n99-s00t98	Genitourinary system	Injury, poisoning and external	115.4	(.587)	392.3	(5.676)	50.56	(.942)
n00n99-z00z99	Genitourinary system	Other factors	1.115	(.016)	-15.26	(-0.742)	113.6	(3.37)
o00o99-r00r99	Pregnancy, child birth	Symptoms, signs	-2928.3	(-6.18)	-132.3	(-10.900)	3555.2	(8.26)
o00o99-z00z99	Pregnancy, child birth	Other factors	123.2	(.628)	-0.142	(-0.010)	-6772.7	(-15.6)
v01y98-z00z99	External causes	Other factors	-102.4	(-1.75)	-2.651	(-0.107)	-33.75	(-.884)

Table 7.7: Final set of attributed variables

	Age Group 0-14		Age Group 15-64		Age Group 65+	
	Coefficient	T-Statistic	Coefficient	T-Statistic	Coefficient	T-Statistic
Attributed Needs						
Log population variance			-60.48	(-11.059)		
All Usual Residents Aged 16+					-0.659	(-7.681)
All Usual Residents Aged 16 to 74					0.721	(8.096)
Resident Population					-0.856	(-4.384)
Proportion Single Pensioner Households					10.20	(8.809)
Proportion aged 16-74 people never worked			1.840	(8.003)		
Proportion Single (never married)			0.459	(5.435)	3.105	(5.345)
Proportion Divorced			1.627	(4.598)		
Rented from private landlord or letting agency	-0.430	(-5.853)	-1.088	(-13.758)	-2.352	(-4.973)
Proportion (un standardised) with not good health (NGH)			1.651	(2.811)	15.80	(8.424)
All people living in the area					0.851	(4.412)
Average with (long term) medical condition for those with at least one			9.755	(2.524)	54.45	(2.936)
2012-13 QOF KD Total Exceptions			0.0400	(2.610)	0.110	(2.076)
2012-13 QOF Epilepsy Prevalence			32.00	(7.204)		
2012-13 QOF Mental Health Prevalence			7.240	(3.161)		
Health Deprivation and Disability Score	15.40	(15.681)	13.43	(7.393)		
Attributed Supply						
Adult critical beds Jan 13					13.41	(3.367)
2012-13 Median waiting times (weeks) for Dermatology Patients					-28.25	(-2.502)
2012-13 Median waiting times (weeks) of the 95th percentile for Neurosurgery Patients			-3.911	(-4.001)	-22.65	(-4.429)

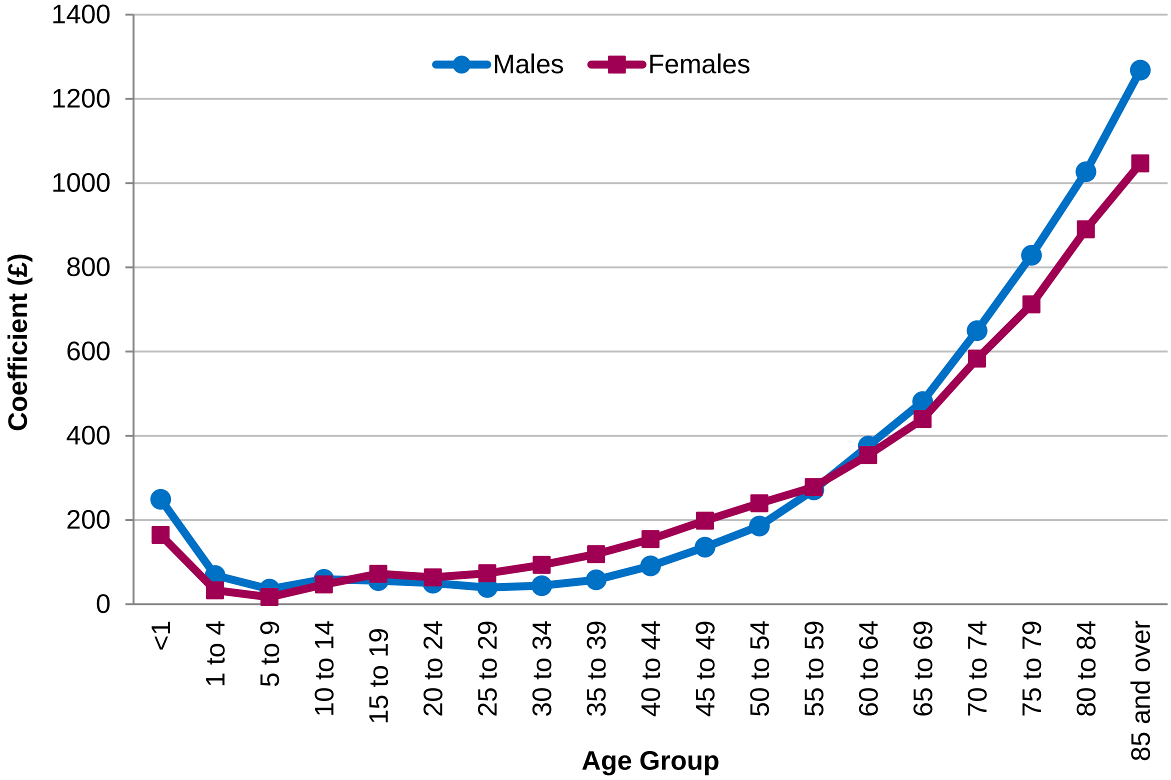
7.3 Reflection of results

The final T- selection model is the recommended model and is also ACRA's preferred model. This section examines the coefficients for this model. All the coefficients from the T- stat selection models are shown in Appendix C.

7.3.1 Age sex coefficients

The age coefficients follow the expected pattern as shown in Figure 7.1:

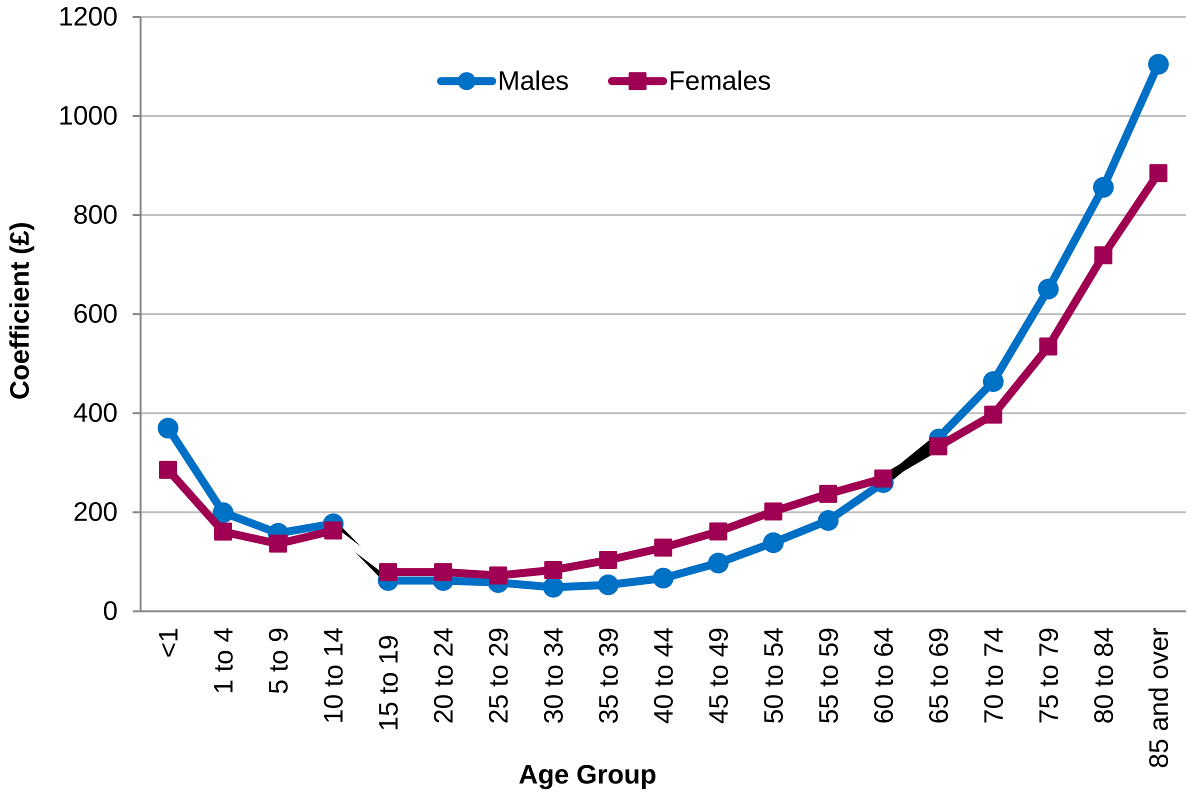
Figure 7.1: Coefficients of age sex groups from all age T-stat selection model



The model exhibits the usual non-linear age curve where expected hospital costs initially decrease, before increasing with age. We find that the predicted costs of females are lower in the early years and lower after the age of 60. Maternity costs are excluded from the G&A models.

The age stratified coefficients are shown in Figure 7.2.

Figure 7.2: Coefficients of age sex groups from age stratified T-stat selection model



7.3.2 Morbidity information

To predict costs in 2013-14, morbidity data for the two years 2011-12 and 2012-13 were used. Morbidity was captured in three ways, 152 morbidity flags, 40 comorbidity interactions and 9 morbidity count variables.

The number of positive and negative coefficients of the morbidity flags for each age strata is shown in Table 7.8:

Table 7.8: Number of positive and negative coefficients on morbidity flags

	Ages 0-14		Ages 15-64		Ages 65 and over	
	Positive	Negative	Positive	Negative	Positive	Negative
Significant	66	3	115	2	80	9
Not significant	51	26	25	8	35	25

For the statistically significant coefficients we find the majority of morbidity flags have a positive coefficient as expected. The highest number of negative coefficients are in the 65 and over model. Morbidity flags can take negative values if morbidity flags in the first two years are associated with death early in the costed year or with diagnoses that prevent the development of illness. The number of statistically significant and negative coefficients is therefore a sensible reflection of the morbidity data.

Comorbidity interactions are included as different combinations of ICD 10 Chapters may increase, decrease or not affect costs. The number of positive, negative and statistically significant comorbidity interaction coefficients are shown in Table 7.9:

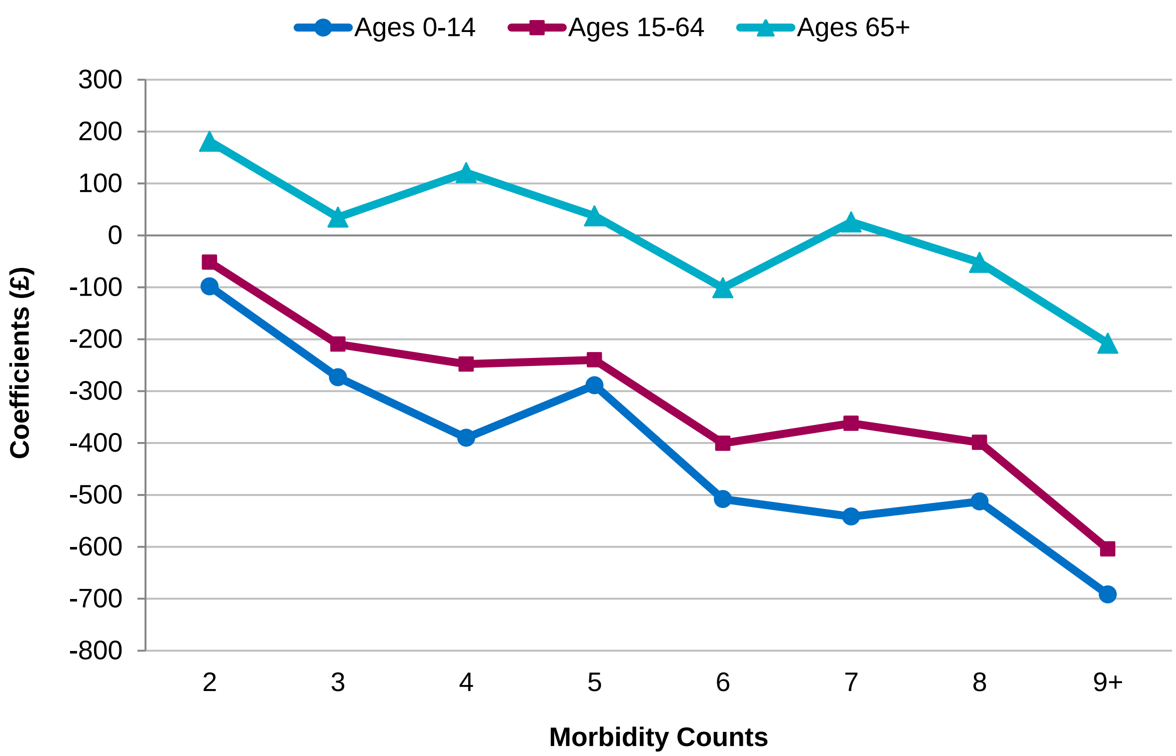
Table 7.9: Number of positive and negative coefficients from co-morbidity interaction terms

	Ages 0-14		Ages 15-64		Ages 65 and over	
	Positive	Negative	Positive	Negative	Positive	Negative
Significant	8	1	13	10	10	14
Not significant	23	8	7	10	10	6

We find that the comorbidity interaction variables affect costs differently between the age groups. As with the morbidity flags, more significant negative coefficients appear for the older age group.

The impact of the morbidity count dummy variables is negative for the age groups 0-14 and 15-64 indicating that costs are lower for treatment for two diagnoses at the same time compared with treatment in two separate spells for each diagnosis. The coefficients are not all negative for the 65 and over age group as shown in Figure 7.3. This may be because comorbidities increase the cost of treating a single diagnosis, through for example longer lengths of stay.

Figure 7.3: Coefficients of morbidity count variables



7.3.3 New practice and private care utilisation

The coefficients of the new practice indicator are positive for age groups 0-14 and 15-64 and negative for age group 65 and over. A possible rationale may be that mobile younger age groups are more likely to register with a new GP practice only when they require treatment. Therefore for people registering with a new GP practice, the use of healthcare is therefore expected to be higher than non-movers.

The coefficient for elderly patients is lower indicating a lower use of healthcare. A reason for elderly people to change GP practice is moving into a residential home which may be associated with lower hospital use because more care and support are provided in the home.

Previously privately funded healthcare has a negative coefficient for all three age models. This is expected as this is likely to suggest future use of privately funded healthcare.

7.3.4 Attributed need variables

A total of 16 attributed need variables were selected through the T-selection process.

- For the age group 0-14, two attributed need variables remained in the model. They are rented from a private landlord with a negative coefficient, which is seen as measure of (lack of) wealth of the parents, and the area level health deprivation score with an expected positive coefficient.
- For the age group 15-64, 11 attributed variables were included in the final model. The attributed variables are: the variance between the registered and ONS population, with a negative coefficient; the (poor) health of the population at both area and GP levels, all having positive coefficients; housing tenure; and marital status which is possibly related to levels of informal care.
- For the age group 65 and over, 10 attributed needs variables were selected. Three population size variables were included which effectively act as the proportion of the population in this age-group. Living arrangements, marital status and health of the population are also included.

The attributed needs variables overall indicate need not captured by the age, sex and morbidity variables. This may be because morbidity information is only available for people who have used hospital inpatient services in the two years prior to the costed year.

7.3.5 Attributed supply variables

Attributed supply variables are used to account for the supply of healthcare that may influence utilisation and the predicted costs. If the supply of healthcare is not directly controlled for, the estimated coefficients on all the other variables in the model will include aspects of supply that are correlated to health need. Supply variables are subsequently frozen out of the allocation weights as target allocations should not be based on the local supply of healthcare facilities. Therefore supply variables are present when running the models to give correct coefficients.

- No attributed supply variables were found to be significant for the 0-14 age group. As CCG dummy variables also capture the supply of healthcare, additional attributed supply variables are not necessarily needed.
- For the 15-64 age group, one attributed supply variable is included in the final model: the waiting times variable with a negative coefficient. There are a number of possible reasons for lower costs when the waiting time is higher. It may be due to a more efficient service and higher hospital bed utilisation due to the pressure to reduce waiting times. Therefore the longer the waiting list may be indicative of a shorter hospital stay leading to lower costs. Alternatively, longer waiting times may be an indication that patients decide to use non-hospital services instead or admission thresholds are higher.

- Three attributed supply variables are included for the model for ages 65 and over. Two are waiting times and one is the supply of adult critical beds. A higher number of critical beds available signifies a higher usage of critical care which is more expensive than usual inpatient or outpatient care.

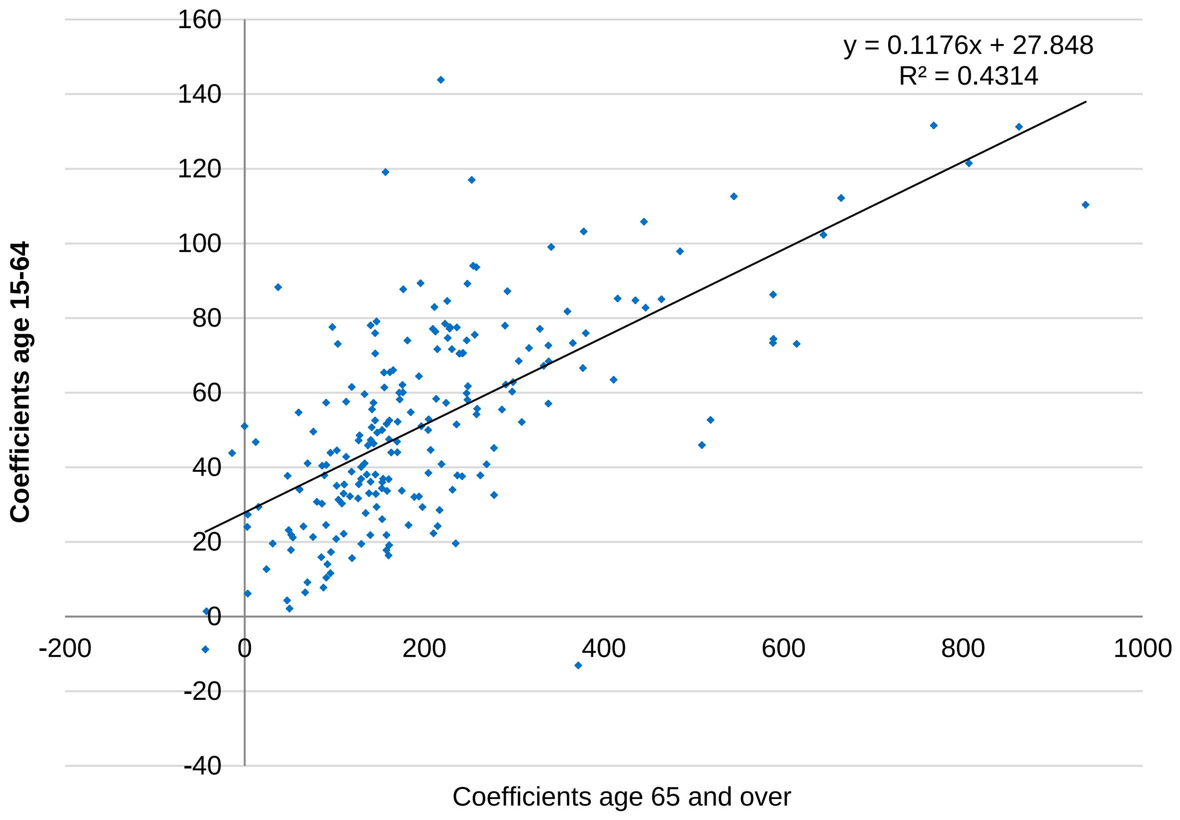
7.3.6 CCG dummy variables

CCG dummy variables are used to capture additional supply side variables that are not elsewhere covered. PBRA (2011) found that the inclusion of PCT dummy variables improved the performance of the models. We included CCG dummy variables. It is unclear, however, how much of the variation that CCG dummy variables capture are need factors or supply.

The effect of allowing supply side variables to determine the cost prediction, however, increases the potential circularity of the prediction models by allocating more money to areas which have invested in more supply of healthcare. Addressing this issue was deemed more important than potentially missing aspects of need through CCG dummy variables. The CCG dummy variables were therefore treated as supply variables and subsequently frozen out in the calculation of the weights used for allocation purposes. This is the same approach as Nuffield PBRA.

The CCG dummy variable coefficients were consistent between each age strata and were generally correlated between the age strata. The relationship between the coefficients for the CCG dummy variables for 15-64 and 65 and over strata is shown in Figure 7.4:

Figure 7.4: Scatter diagram of CCG coefficients comparing ages 15-64 and 65 and over



8 General and acute: Calculation of practice weights

This section explains how the estimated coefficients were used to generate weights for each GP practice by age-sex group which were subsequently used to calculate target allocations for CCGs.

8.1.1 Generating the weights for registered patients in April 2014

To calculate the weights we used a dataset containing all individuals in the April 2014 PDS, their linked morbidity information from SUS PbR for the years 2012-2013 and 2013-2014, and all attributed variables in our preferred model specification.

We do not use samples when generating the weights as it is important to avoid possible volatility from small numbers.

The full dataset was then split by CCG to create files which are more manageable. We also included a file for all people who were not attributed a CCG. Each CCG file is therefore a list of patients registered with the GP practices within the CCG.

The reasons for splitting the full dataset into separate CCG files were:

- The full dataset was too large for the computers to handle, at over 100GBs.
- Opening files split by CCG and predicting costs was more manageable and computationally quicker.
- Predicted costs were aggregated for each GP and age-sex combination. As each GP practice belongs to only one CCG, there was no possibility that a GP practice would be present in two CCG files which would affect the data aggregation.

We froze the supply and CCG dummy variables by obtaining all the attributed supply and CCG variables for all of the people in the April 2014 PDS list and creating a mean for each of these variables. The relevant means were used in the calculation of the predicted weights in place of the values of individual supply and CCG variables. The mean values were multiplied by the coefficients for each variable, which has the effect of changing the value of the constant term.

Population means are generated from Equation 8.1:

$$\overline{S_k} = L^{-1} \sum_k S_{ipk}$$

Equation 8.1

S_k are k number of population mean supply variables
L is the total population of the 1st April 2014 snapshot of the PDS
S_{ipk} are k number of supply variables for individual i and GP practice p

We used estimated the coefficients from our preferred model specification and predicted costs from each of the CCG files. Within each file we obtained average predicted costs for each age-sex group for each GP practice. We then combined all the CCG files together to give GP practice age sex cost predictions for all those in the PDS at the 1 April 2014.

We obtained predicted costs for each age-sex group in each GP practice from Equations 8.2 to 8.5:

$$\widehat{C_{ap}} = \hat{\alpha} + L_{ap}^{-1} \sum_{i \in ap} \sum_j \widehat{\beta_J} N_{ipj} + \sum_k \widehat{\gamma_k} \overline{S_k}$$

Equation 8.2

$$\widehat{\alpha_s} = \sum_k \widehat{\gamma_k} \overline{S_k}$$

Equation 8.3

$$\widehat{\alpha_2} = \hat{\alpha} + \widehat{\alpha_s}$$

Equation 8.4

$$\widehat{C_{ap}} = \widehat{\alpha_2} + L_{ap}^{-1} \sum_{i \in ap} \sum_{j} \widehat{\beta_J} N_{ipj}$$

Equation 8.5

Where:

$\widehat{C_{ap}}$ is the predicted cost for each age sex group a in practice p

$\widehat{\alpha_2}$ is the predicted constant term, which is the sum of the predicted frozen supply side constant and predicted constant

L_{ap} is the number of patients within each age sex group in each GP practice from the 1st April 2014 snapshot of the PDS

N_{ipj} are j number of needs variables for individual i at practice p

The person level needs variables such as 38 age and sex categories, 152 morbidity flags, 40 comorbidity interaction variables, 9 morbidity count variables, new GP practice variable and a private care utilisation variable are from financial years 2012/13 to 2013/14.

$\overline{S_k}$ are k number of population mean supply variables.

8.1.2 Weights for October 2015 registrations

We then applied the weights per head by age-sex group by GP practice to registered patient lists in October 2015 from NHAIS to create the weights that were subsequently used to calculate target allocations for CCGs. The NHAIS data contain the number of patients registered at each GP practice by the same age-sex groups as in the estimation models. We multiplied the number of patients in each age-sex group at each GP practice by the calculated weights.

There were a small number GP practices which are in NHAIS October 2015 but not in PDS April 2014 and therefore did not have predicted weights per head. Instead we gave these GP practices the CCG average weight for each age-sex group.

8.1.3 Insignificant morbidity information

In generating the GP level weights we used all the morbidity variables regardless of statistical significance. The rationale for this is that modelling using OLS aims to minimise the root mean squared error in the estimation. Estimated coefficients regardless of statistical significance are generated for all morbidity variables which are all used together to predict costs. The estimated coefficients when all are used will therefore result in the lowest root mean squared prediction error.

- The morbidity flags are generally positive as shown on Table 7.8. However a higher proportion statistically insignificant coefficients are negative. Therefore removing insignificant morbidity information will result in an over predication of actual costs.
- In conventional hypothesis testing, each coefficient is tested to be statistically different from zero. A statistically significant relationship is found when there is sufficient evidence to suggest that coefficients are not equal to zero. If a variable is not statistically significant, then there is not enough evidence to say that the coefficient should be different to zero. However, as the hypothesis test does not test if coefficients are equal to zero, we cannot assume that statistically insignificant variables would have a zero coefficient.
- To remove statistically insignificant morbidity information from the allocations weights would require the removal of morbidity flags from the estimations stage, as coefficients on statistically significant morbidity flags will change with the removal of statistically insignificant morbidity flags. This would require a variable selection process to be conducted for each age stratified estimation sample rather than only focussing on attributed needs and supply variables.

A reason for morbidity flags to be insignificant is the small number of observations for certain morbidity flags. If the number of observations is low, then there may not be enough estimation power and therefore the coefficient is less likely to become statistically significant regardless of the actual association between morbidity flags and costs. The percentage and number of morbidity flags with less than 100 observations is shown in Table 8.1:

Table 8.1: Percentage (number) of morbidity flags with under 100 observations

	Ages 0-14	Ages 15-64	Ages 65 and over
Statistically insignificant	40.26 (31)	15.15 (5)	20 (12)
Statistically significant	5.80 (4)	0.85 (1)	3.37 (3)

Table 8.1 shows that the percentage of morbidity flags with a low number of observations is higher for morbidity flags that are not statistically significant. Therefore some morbidity markers may be statistically insignificant due to estimation power rather than not having an association with costs. However, some may genuinely have no association with costs.

The removal of insignificant morbidity flags from the model will impact on the coefficients of statistically significant morbidity flags. This is due to a change in the base category. When no morbidity flags are removed, the base category is patients with no morbidity information. Therefore the coefficients of the morbidity information will be additional costs of having a certain morbidity compared with the cost of no morbidities. When morbidity flags are removed, the base category becomes all statistically insignificant morbidities and no morbidities. Coefficients of morbidity flags will therefore be the additional costs of having a certain morbidity over the cost of having statistically insignificant morbidities or indeed no morbidity.

8.2 Border Practices

Some patients who are registered with a GP practice in England receive treatment in hospitals in Wales and Scotland. The CCG with which a patient is registered is responsible for funding their hospital services, including services used in Wales and Scotland. Treatment in Scotland of patients registered in England is relatively rare due to low population densities and the distances involved. Treatment in Wales is more common.

SUS PbR does not record activity in hospitals in Wales and the costs of these are omitted from our predicted weights and could therefore lead to an underestimate of weighted populations for border CCGs. The spend on treatments in hospitals in Wales, however, is estimated to be very low. Table 8.2 shows the estimated spend on hospital services from the four CCGs along the Welsh border.

Table 8.2: Spend (2014-15) on hospital services in Wales by CCG

	CCG		Spend (£)	Persons	Per Person Spend (£)
Inpatient	02F	NHS West Cheshire CCG	1,136,950	736	1,545
	05F	NHS Herefordshire CCG	422,344	236	1,790
	05N	NHS Shropshire CCG	1,806,906	1,100	1,643
	11M	NHS Gloucestershire CCG	207,593	131	1,585
	Total		**3,573,793**	**2,203**	
Outpatient	02F	NHS West Cheshire CCG	352,343	2483	142
	05F	NHS Herefordshire CCG	137,009	937	146
	05N	NHS Shropshire CCG	703,132	4833	145
	11M	NHS Gloucestershire CCG	59,242	411	144
	Total		**1,251,727**	**8664**	
Accident and Emergency	02F	NHS West Cheshire CCG	226,291	1620	140
	05F	NHS Herefordshire CCG	103,583	712	145
	05N	NHS Shropshire CCG	386,613	2929	132
	11M	NHS Gloucestershire CCG	86,788	648	134
	Total		**803,275**	**5909**	

The number of episodes and attendances in Wales was provided to us by the NHS Wales Information Services at aggregate level by CCG of registration. These data could not be linked on an individual level to the PDS dataset. An off-model adjustment to the weights was made for these CCGs for activity in Wales for the costs in Table 8.2.

As a sensitivity analysis, we also ran the models with and without the four border CCGs. Coefficients and practice shares between the models with and without the four border CCGs changed very little.

9 Maternity Formula

9.1 Introduction and background

The formula for maternity services is a component of the overall formula for core CCG allocations. The maternity formula used for allocations up to 2015-16 was developed as part of the *CARAN* project (Morris et al, 2007[9]). This used a cost per birth approach, modelling the cost per birth at MSOA level.

The refresh of the formula followed the same approach of modelling the cost per birth. We were able in the refresh to model costs at an individual level for the first time, using the SUS PbR data described in sections 2 to 6. By using person level data we had access to a richer dataset and were able to test a wider range of potential need indicators. In particular, we were able to use morbidity flags derived from diagnostic information which are described in section 6.

We modelled the costs of maternity services in 2013-14 using a wide range of need and supply variables. As well as using more recent data, we achieved a higher performing model as measured by a higher R-squared statistic at both the individual and GP practice level compared with *CARAN* at MSOA level.

In 2013-14 the way in which maternity services are paid for changed with the introduction of the maternity pathway. Under the pathway model a flat rate is paid for each of the three maternity phases (antenatal, postnatal and delivery) with different payment tiers dependent on the characteristics and medical history of the patient. Our model is not based on pathway payments as we did not have access to the data used to make the decisions on the tier. Our model therefore was based on the cost of individual inpatient episodes and outpatient attendances.

This section sets out the development of the new maternity formula, including summary statistics and the process by which the final model was selected.

9.2 Data

For the refresh of the model we used the SUS PbR and PDS data provided by the Health and Social Care Information Centre (HSCIC) and described in the G&A sections of this report.

In addition to the SUS PbR data, we also used the wide range of attributed need and supply variables, again described in the G&A sections of this report. The attributed variables were attached to the mother’s record of hospital care recorded in the SUS PbR data. As described in section 3, this was undertaken by linking the SUS PbR data to the mother’s GP practice record using the bridge file. The practice record included the LSOA of residence and many of the attributed variables were at LSOA or GP practice level. The bridge file was also the only way of identifying all inpatient episodes and outpatient attendances that were for the same woman.

---

[9] Morris, S., Carr-Hill, R., Dixon, P., Law, M., Rice, N., Sutton, M. and Vallejo-Torres, L. *Combining Age Related and Additional Needs (CARAN) Report*.

9.3 Identifying maternity services

The first step was to identify maternity services in the SUS PbR data set. This was primarily based on the HRG and treatment function codes with the majority of the relevant activity having a “NZ” HRG code. Table 3.2: How maternity was identified in each dataset shows the codes which were classified as being in scope and a description of these procedures.

Specialised maternity services (for example neonatal activity) were excluded as they are not commissioned by CCGs. Section 3 sets out how specialised services were identified using the Prescribed Specialised Services (PSS) toolkit 2014-15.

A few spells were also excluded if either the woman’s age was implausible for maternity or the cost was so high that it was thought to be a data error.

9.4 Costing the data

After determining the maternity dataset the next stage was to cost the data. The approach used for costing the activity data for 2013-14 are set out in section 3.

As described in section 3, the activity was costed in three ways. The proportions costed in each way for maternity services were as follows:

1) For the majority of the data, there was already a cost included in the SUS PbR dataset. This was the case for around 60% of total maternity costs. Where this was the case this cost was used (and the MFF excluded).

If the data were not already costed

2) An existing national tariff was used where available. This was the case for around 7% of maternity costs.

If national tariffs were not available

3) Reference costs, excluding the MFF, for the procedure were used. This was the case for around 33% of maternity costs.

If reference costs were not available

4) The average costs for the speciality were used. This was the case for under 1% of maternity costs.

As each row in the dataset contained the costs for a particular spell of care or outpatient attendance, there were cases where a mother appears more than once, for example if there is a delivery episode and separate spell of postnatal care. As our model uses a cost per birth, we aggregated costs for each mother using the SUS PbR IDs, PDS IDs and the bridge file as described in sections 3 and 5.

The distribution of the costs per mother is shown in Table 9.1: Summary of costs per mother and Figure 9.1: Cost per maternity patient (£) (note Figure 9.1: Cost per maternity patient (£) excludes 1,500 cases where the cost exceeds £10,000).

Table 9.1: Summary of costs per mother

Summary statistics	Per person cost (£)
Minimum	74
1st Percentile	393

Summary statistics	Per person cost (£)
5th Percentile	428
10th Percentile	628
Lower Quartile	1,477
Median	1,477
Upper Quartile	2,299
90th Percentile	3,053
95th Percentile	3,876
99th Percentile	6,315
Maximum	254,413
Mean	1,916
Standard Deviation	1,381
Skewness	21
Kurtosis	2327
Observations	780,629
Sum	£1,495,868,489

Figure 9.1: Cost per maternity patient (£)

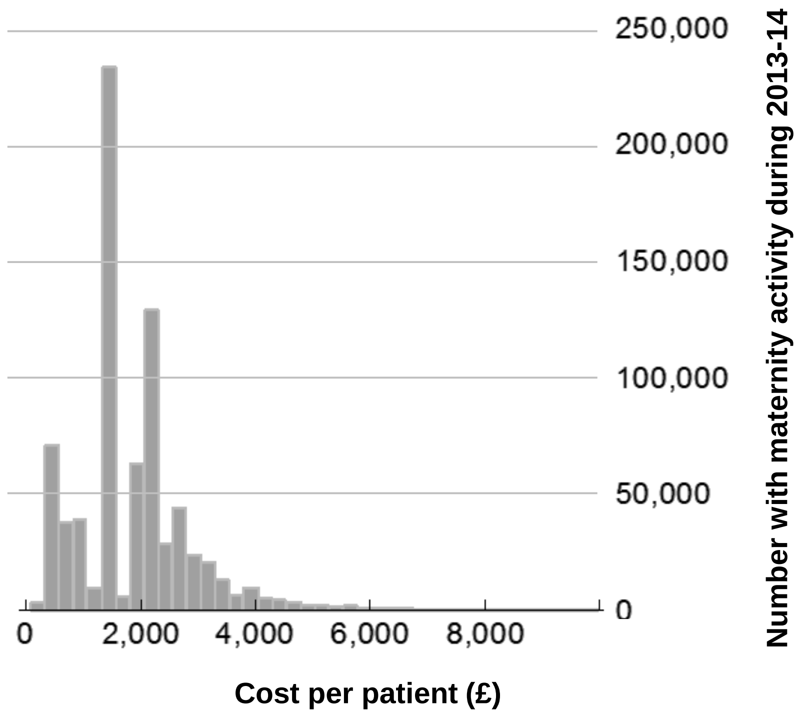


Figure 9.1 shows a distribution with two clear spikes at £1,500 and £2,000. The spike at £1,477 is the cost for a delivery episode. The higher spike, at around £2,000, includes those women where there was both a birth episode and some form of extra ante- or post-natal treatment above the normal pathway.

9.5 Modelling methods

9.5.1 Approach

We modelled costs in 2013-14 using ordinary least squares (OLS) to identify the need and supply variables influencing the cost of maternity services.

The modelling was undertaken at the individual level using explanatory variables for the prior years in our dataset. Age and morbidity flags (see section 6) were included in the models as independent variables. We tested a range of attributed need and supply variables as independent variables.

Modelling at the person level was different to *CARAN*, in which the unit of analysis was at MSOA level. By using individual level data we were able to produce a better performing model with a more complete set of explanatory variables.

The model estimated was:

$$C_{ip} = \alpha + \sum_j \beta_j N_{ipj} + \sum_k \Upsilon S_{ipk} + \varepsilon$$

Equation 9.1

Where

C_{ip} denotes the cost per birth of maternity services received by individual i in practice p in 2013-14
α is the constant term
N_{ipj} are j number of needs variables in the prior years to 2013-14
S_{ipk} are k supply variables
ε is the residual error

9.5.2 Attributed variable selection process

After cleaning the data and creating the cost variable, we used the following variable selection process to determine a preferred specification. A diagram of the process is shown in Figure 9.2: Attributed variable selection procedure and the series of regressions that produced the final model is shown in Table 9.6. The process was as follows.

1) A set of potential need and supply variables to be tested was identified. Relevant need variables were chosen on the basis of previous research and expert opinion. They included morbidity flags, attributed variables (for example from the Population Census) and variables to account for birth order (see below). A smaller set of attributed variables was tested than for the development of the G&A formula as many were not deemed relevant to maternity services. The attributed supply variables chosen included information on the provision of maternity services and dummy variables for CCGs.

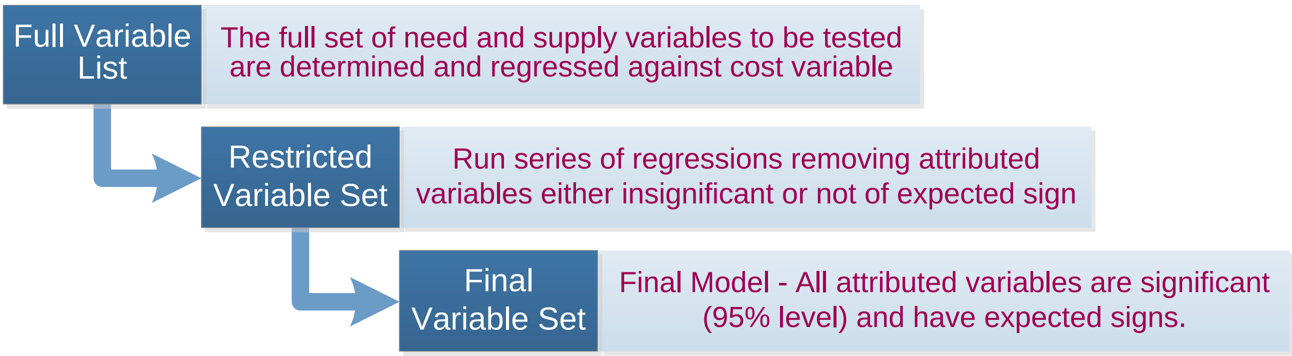
2) The set of independent variables in (1) was then regressed against the cost per mother. Attributed variables that were both significant and of the

expected sign were retained while attributed variables which were either not statistically significant or had an unexpected sign were removed.

3) The process was repeated until only those attributed variables that were both significant and of the expected sign remained in the model. This represented the most parsimonious model with high statistical power. The final specification is presented in Table 9.2: Final model. All variables included in the final model are significant at the 95% level.

Age and the morbidity flags were included throughout this procedure for selecting the attributed variables.

Figure 9.2: Attributed variable selection procedure



9.6 Final model

The final, preferred model is in Table 9.2: Final model.

Table 9.2: Final model

Variable	Description	Coefficient (T-statistic)
Cost Per Mother	Sum of maternity activity cost during 2013/14	
Aged Under 20	Dummy variable based on mothers age	26.60 (4.45)
20-24	Dummy variable based on mothers age	17.90 (4.14)
25-29	Dummy variable based on mothers age	0 (Base Category)
30-34	Dummy variable based on mothers age	1.893 (0.57)
35-39	Dummy variable based on mothers age	21.17 (5.63)
40-44	Dummy variable based on mothers age	80.53 (13.04)
45-49	Dummy variable based on mothers age	213.4 (15.26)
50+	Dummy variable based on mothers age	190.6 (10.98)
Multiple Births in 2013-14	Individual with multiple birth episodes in 2013/14	3351.1 (3.53)
0 Births by end of 2013-14	Individual without birth episode by end 2013/14	-548.1 (-93.77)
1 Birth during 2010-2014	Individual with a single delivery episode 2010-2014	0 (Base)
2+ Births during 2010-2014	Individual with more than 1 delivery episode 2010-2014	-109.9 (-34.68)
Low Birth Weight	Proportion of live births with weight <2500g	173.2 (7.78)
IMD 2015 (Overall)	IMD Overall (2015) at LSOA level	0.487 (2.97)
Pakistani	Proportion of residents with Pakistani ethnicity	1.596 (7.88)
Black African	Proportion of residents with Black-African ethnicity	3.079 (5.73)
Proportion Never Worked	Proportion of LSOA who have never worked	-13.95 (-4.31)
Proportion in Social Housing	Proportion of LSOA residents in social housing	0.261 (2.1)
QOF Diabetes Prevalence	QOF diabetes prevalence for GP practice	3.560 (3.07)
Overnight Maternity Beds	Gravity model of maternity bed supply	-5.072 (-3.98)
Obstetrics ultra-sound supply		0.0205 (6.09)
Number of Morbidity Flags	152 Morbidity flags for disease groups	

Variable	Description	Coefficient (T-statistic)
Number of CCG Dummies	211 CCG Dummies in England	
Constant		634.1 (22.7)
	Observations	774,664
	Individual R – Squared	0.301
	GP Practice R – Squared	0.6154

The R-squared for the refreshed maternity model is higher than the *CARAN* maternity model which had an R-squared of 0.29 at MSOA level. This is likely to be because we were able to model at the person level. To test the specification of the model we ran a "VIF" test which looks for multicollinearity between the explanatory variables. The test indicated multicollinearity was not an issue.

9.7 Need and supply variables

9.7.1 Need variables

Need indicators are variables which may increase (or decrease) the cost of maternity services per birth. For example people in more deprived communities may, on average, be more likely to have more comorbidities and thus be more costly.

The maternity pathway implies certain conditions increase the risk of complications and lead to more costly births. For example, if the mother has diabetes they are assigned to the "intensive" category and attract a higher payment. We included in the model 152 dummy variables based on morbidity flags used in the development of the G&A formula (see section 6). These variables are interpreted as being the additional cost associated with a birth for a mother with a particular morbidity.

Age is likely to be a determinant of cost with the relatively old and young likely to require additional care and thus be more costly. We accounted for this by using variables in five year age groups. The 25-29 age group is the base category and so the coefficients are the difference in cost compared with someone between the ages of 25 and 29.

The proportion of low birth weight births was a significant variable in the previous research and was included in this model as babies with low birth weight are more likely to require more intensive treatment and aftercare and their mothers are more likely to stay in hospital longer.

We could not derive this variable at the person level in the SUS PbR data as babies and their mothers were not linked. To produce this variable we instead obtained data from HES records for the number of births at GP practice level during 2013-14 and the number of these which were of low weight. A birth is classified as being low weight if it was a live birth with a weight of less than 2,500 grams. For each practice this produced a proportion of births which were of low weight. This proportion was then attributed to the individual patient by linking to the patient's GP practice in the linked SUS PBR and PDS data set. We explored whether it was possible to look at low weight by gestation period (to account better for premature babies), however this was not possible using HES data.

The final model included ethnicity variables for those of Pakistani or Black African origin, based on data from the 2011 Census for the proportion of these groups in the total population of each LSOA. We did not have ethnicity data at the person level and so ethnicity were attributed variables.

We included a full range of ethnicity groups within the variable selection process (Table 9.6) as it was suggested that costs could differ between ethnic groups. The variables for Pakistani and Black African origins were significant throughout the selection process and were therefore included in the final specification.

The final model also included variables related to social deprivation and living environment: the proportion of people who have never worked and those living in social housing. It should be noted that we were surprised by the sign on the variable for those who had never worked, however it was significant and was retained as it passed multicollinearity tests.

The model includes the estimated QoF prevalence rate for diabetes. Diabetes is regarded as a factor which triggers a higher payment under the maternity pathway and therefore it was recognised as something which should be included in the model. It is the only QoF variable included as the other QoF variables were highly correlated with QoF diabetes and could therefore not be included separately in the model.

9.7.2 Birth order and multiple births

In the review of modelling by ACRA, we were asked to investigate the inclusion of a birth order variable as it was suggested that second, or subsequent, births may have lower costs as the risks of complications are lower.

Our dataset did not include information on birth order therefore we developed a proxy as outlined below.

- For those patients with delivery episodes in 2013-14 we identified all hospital activity between 2010-11 and 2013-14.
- From this list we identified all delivery episodes on the basis of HRG code and the date of the episode.
- Duplicate delivery episodes (cases with matching ID and admission date) were removed.
- Improbable births were removed. Births less than 244 days (35 weeks) apart were removed as the time between births was short enough to suggest a data quality issue. 35 weeks is a conservative estimate.
- We then produced a count variable of the number of births per mother over the time period 2010-2014. Due to small numbers we limited the variable to have the values of 0 births, 1 birth, and more than 1 birth.
- A further variable was created to identify cases of multiple births within the reference year of 2013-14 and a separate dummy variable created for these cases.

The process is outlined in Figure 9.3.

Figure 9.3: Calculating births per person

As we did not hold SUS PbR data prior to 2010-11 we were unable to identify earlier births. It is probable that some mothers with a birth in 2013-14 will have given birth before April 2010, but it was not possible to access this information.

There were a small number of cases where a mother has two birth episodes within 2013-14 with more than 35 weeks between them (for example if the first episode was right at the beginning of the financial year). To account for these cases we included a separate dummy variable. This was important because these mothers have higher costs, and will also be included in the birth order variable with the value of more than one birth. If we did not include a separate dummy variable the coefficient for those with multiple births between 2010-2014 would have been positively biased.

9.7.3 Supply Variables

Supply variables are factors which influence utilisation but are related to the capacity of available services within an area rather than need. For example it is possible that distance to services will influence intensity of usage of maternity services.

Two supply variables were included in the preferred specification.

1) The number of obstetric ultra-sound scans carried out by NHS providers.
2) The number of overnight maternity beds available by NHS provider.

In both cases these were "gravity weighted". This means that the number of ultra-sound scans and the number of overnight beds for all providers were included in the variable, but those located closer to a patient's LSOA of residence received a higher weight as the closest facilities are more likely to be attended.

In addition to these we included dummy variables for each CCG in England as supply variables.

Supply variables used in the model were sterilised (set to the national average for all patients) before the model was used to produce weights for allocations. This was to ensure that areas with increased usage as a result of additional facilities do not unduly benefit.

9.8 Model evaluation

One way of assessing the performance of the model this is to look at the Mean Absolute Percentage Error (MAPE), which is the percentage difference between the estimate made by the model and the observed cost. The advantage of this method is that it equally penalises both over and under prediction. Table 9.3: MAPE at the individual level shows the MAPE statistics for: all observations; observations where the mother did not give birth in 2013-14 (or previously from 2010-11, but was receiving ante-natal care in 2013-14); observations where the mother gave birth in

2013-14 (but not also in 2010-11 to 2012-13); and observations where the mother gave birth in 2013-14 and also in 2010-11 to 2012-13.

Table 9.3: MAPE at the individual level

Statistics	All observations	First birth value = 0	First birth value = 1	First birth value = 2
Observations	774,664	223,397	398,048	153,219
25th Percentile	-12.8%	-23.8%	-10%	-10.5%
Median	8.7%	12.4%	8%	8.4%
75th Percentile	39.2%	75.8%	36.5%	32.4%
Mean	18.9%	38.9%	11.2%	10%
Observations where MAPE > 20%	467,825 (60.4%)	165,375 (74%)	216,747 (54.5%)	85,703 (55.1%)
Observations where MAPE > +/- 100%	45,240	42,386	2,093	761

The figures suggest that while the model performs well compared to the previous maternity model, there are instances where the predicted cost differs by a large amount from the observed cost. This is not surprising at the person level.

By splitting the data by the value of the First Birth variable, it is clear that the model performs less well when First Birth is equal to zero (meaning that the mother has yet to give birth). This suggests that it is less well equipped to deal with cases where there is no delivery episode, and the predictions are less reliable. This too is unsurprising, but it was felt important to include all maternity costs in 2013-14 in the modelling.

9.9 Future model development

This was the first time that an individual level dataset had been used for the maternity model and there are number of improvements which could be made to improve the performance and robustness of the model in the future.

The First Birth variable was only able to take into account births in the four years of data available to us. To create a better First Birth variable, we would need a longer time span of data.

It is noticeable from the MAPE section that the model is performing less well for those who had either not yet given birth in 2013-14, and only received ante-natal care in that year, and those who gave birth in 2012-13 and received post-natal care during 2013-14. More work is required to investigate how they can be better covered in the modelling.

This modelling was not able to utilise the definitions used to assess a patient within the maternity pathway. If we had access to these data, this would generate a costing variable that would more accurately represents payments to providers. However, the pathway payments are based on the expected individual procedures.

9.10 Translation of cost predictions to weights for allocations

After developing the model we went through the following process to apply the model outputs to produce weights which could be used for target allocations.

1) Supply variables were sterilised by setting their values to the national average, in order to ensure that differences in supply did not influence allocations.
2) A cost per birth was estimated using the coefficients of the model. While this included some cases for which there was no birth in 2013-14 (only antenatal or postnatal care in 2013-14), it was assumed that this had no geographical bias.
3) The cost per birth estimate in (2) was multiplied by the number of births in 2014-15 as measured by new GP registrations due to birth. This produced a total cost at GP practice and England level. For practices that opened since the end of 2013-14 (and do not have a modelled estimate), CCG average values were used. Practices that had closed before October 2015 were included to ensure that a CCG is not penalised if a GP practice with births during 2013-14 had closed.

Figure 9.4: Map of maternity need index

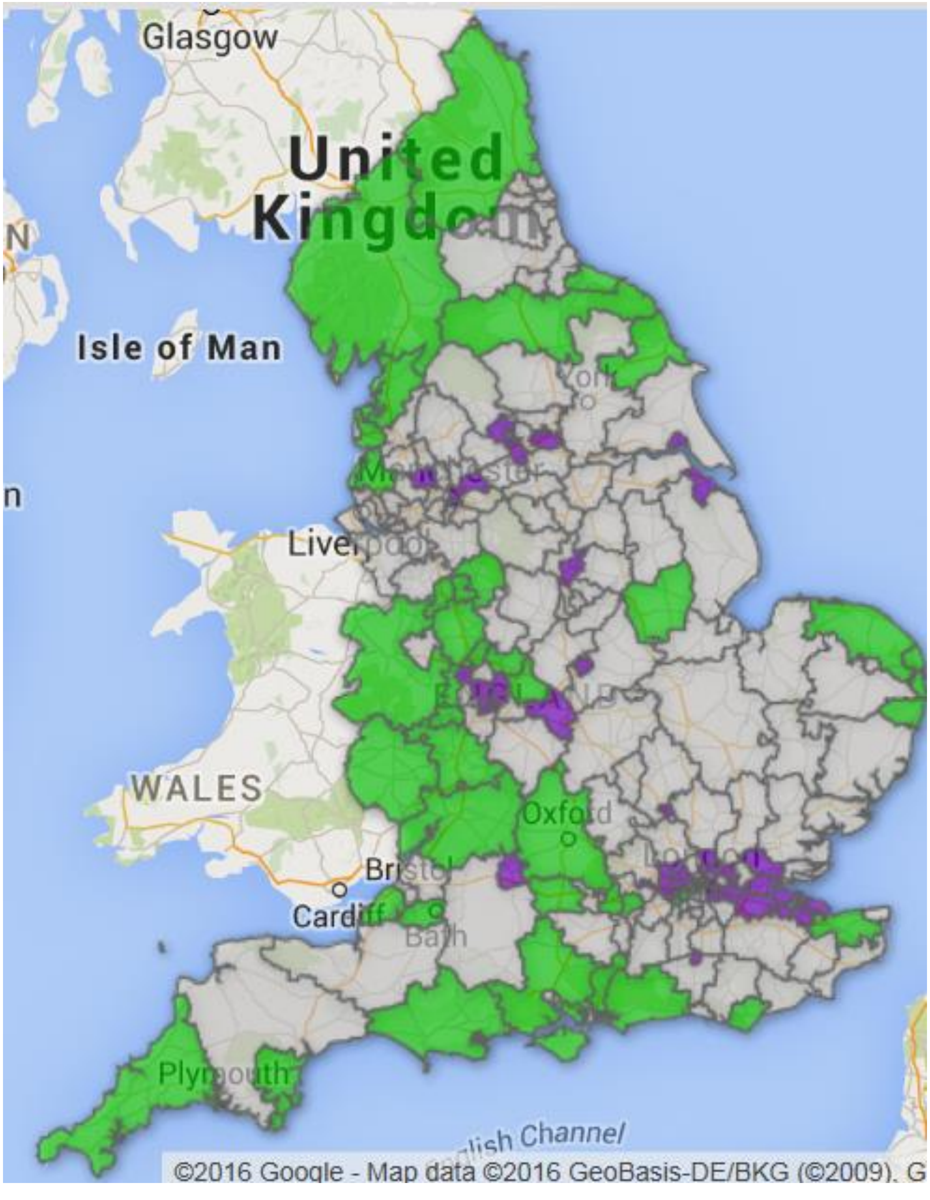


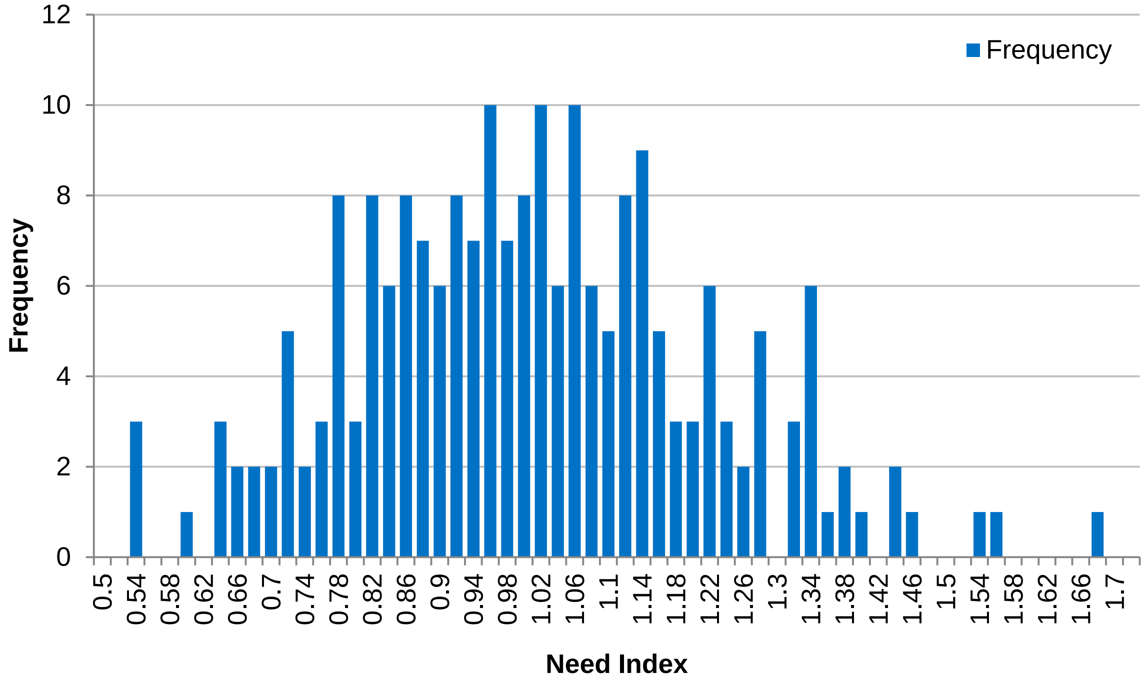
Figure 9.4 shows a map of the need index for maternity services by CCG. Areas shaded in green are in the bottom 20% of CCGs and areas shaded in purple are in the top 20% of CCGs. It suggests that the areas of highest need tend to be more urban areas, in particular around London and cities in the North of England. Areas with the lowest relative need tend to be more rural.

Table 9.4 shows the CCGs with the highest and lowest maternity need index values and Figure 9.5 shows the distribution of the need index.

Table 9.4 Maternity index by CCG

CCG	Index	CCG	Index
NHS Bradford City CCG	1.667	NHS Fareham and Gosport CCG	0.668
NHS Greenwich CCG	1.541	NHS South Worcestershire CCG	0.656
NHS Waltham Forest CCG	1.529	NHS Isle of Wight CCG	0.650
NHS Birmingham South and Central CCG	1.444	NHS North Somerset CCG	0.638
NHS Luton CCG	1.436	NHS West Lancashire CCG	0.637
NHS Central Manchester CCG	1.433	NHS Gloucestershire CCG	0.628
NHS Thurrock CCG	1.393	NHS Wyre Forest CCG	0.581
NHS Crawley CCG	1.375	NHS Stafford and Surrounds CCG	0.539
NHS North Manchester CCG	1.371	NHS North Staffordshire CCG	0.523
NHS Leeds South and East CCG	1.356	NHS Newbury and District CCG	0.520

Figure 9.5: CCG Need Index for Maternity Services



9.11 Variables tested and model development

Table 9.5 shows the variables that were tested during the development of the model and Table 9.6 shows the attributed variables discarded in each step of the variable selection procedure.

Table 9.5: Maternity model variables tested

Variable Name	Variable Description
cen_5	Proportion not White English, Welsh, Scottish, Northern Ireland, British
Cen_6	Proportion Indian
Cen_7	Proportion Pakistani
Cen_8	Proportion Bangladeshi
Cen_9	Proportion Black Caribbean
Cen_10	Proportion Black African
Cen_11	Proportion Chinese
Cen_14	People aged 16-74 never worked
Cen_22	Persons in social rented housing
Cen_27	Proportion (un standardised) with LLTI
Cen_34	Economically Inactive; Long-Term Sick or Disabled
Cen_42	Proportion of students in population (aged 16-74)
Cen_45	Proportion of residents who are in communal establishments
Cen_51	Proportion of population aged 20-24
IMD_15_1	IMD (2015) Overall Score
GP_Sup_1	No FTE GPs per practice (excluding retainers and registrars)
GP_Sup_2	Proportion headcount GPs female (incl retainers and registrars)
Sup_20	Obstetric ultra-sound supply (Number of Scans)
Qof_98	QOF Prevalence Hypertension
Qof_103	QOF Prevalence Cancer
Qof_108	QOF Prevalence CHD
Qof_113	QOF Prevalence Kidney Disease
Qof_128	QOF Prevalence Diabetes
Qof_133	QOF Prevalence Epilepsy
Qof_153	QOF Prevalence Obesity
Qof_163	QOF Prevalence Stroke
LBW	The proportion of babies of Low Birth Weight (<2500g) by GP practice in 2013-14
Overnight_Beds	The supply of overnight maternity beds in acute providers
First Birth	The number of birth episodes between 2010-11 and 2013-14
Two Births	Dummy variable for patients with 2 birth episodes in 2013-14

Table 9.6: Maternity model development

	Spec 1	Spec 2	Spec 3	Final Spec
Aged Under 20	27.28***	27.07***	27.04***	26.60***
20-24	17.93***	18.07***	18.03***	17.90***
25-29	BASE CATEGORY (Coefficient = 0)			
30-34	2.344	2.052	2.080	1.893
35-39	22.02***	21.48***	21.54***	21.17***
40-44	81.56***	80.98***	81.05***	80.53***
45-49	214.5***	213.9***	214.0***	213.4***
Aged Over 50	191.8***	191.2***	191.4***	190.6***
gpsup_1	-0.0386			
gpsup_2	-9.790			
sup_20	0.0201***	0.0206***	0.0207***	0.0205***
Overnight_beds	-5.499***	-5.247***	-5.386***	-5.072***
0.FirstBirth	-548.5***	-548.2***	-548.3***	-548.1***

	Spec 1	Spec 2	Spec 3	Final Spec
1.FirstBirth				BASE CATEGORY (Coefficient = 0)
2.FirstBirth	-109.6***	-109.8***	-109.8***	-109.9***
TwoBirths	3351.5***	3351.3***	3351.3***	3351.1***
LBW	165.6***	168.7***	168.6***	173.2***
imd_15_1	0.543*	0.447**	0.455**	0.487**
cen_5	0.210			
cen_6	-0.383			
cen_7	1.253**	1.463***	1.436***	1.596***
cen_8	-0.307			
cen_9	0.104			
cen_10	2.625***	3.070***	2.942***	3.079***
cen_11	2.904			
cen_14	-13.61***	-14.72***	-14.49***	-13.95***
cen_22	0.394**	0.277*	0.284*	0.261*
cen_27	0.890*	0.0720		
cen_34	-3.110*			
cen_42	-2.698*	-0.592		
cen_45	-0.588	-0.205		
cen_51	114.8*			
qof_98	-2.049*			
qof_103	-2.595			
qof_108	-1.151			
qof_113	1.935			
qof_128	7.171***	5.684***	5.755***	3.560**
qof_133	5.660			
qof_153	0.575			
qof_163	-7.994	-13.79***	-13.45***	
_cons	661.1***	655.4***	655.9***	634.1***
N	774664	774664	774664	774664
r2	0.301	0.301	0.301	0.301
="* p<0.05	** p<0.01		*** p<0.001"	

Spec 1 was the initial run of the model. This included all the variables that were expected to be important in addition to morbidity variables. At this stage we removed those variables which were either not significant or had a sign which differed from our expectation.

- Cen 34 was removed as it did not seem plausible that not being economically active would reduce cost
- Cen 51 was removed as we suspected collinearity with the age group variables.
- Qof 98 was removed as increased prevalence should increase costs.
- Spec 2 shows the regression after removing variables from the initial regression. This led to three more variables losing significance and being dropped.
- Spec 3 shows this process being conducted for a second time. Any variables not significant in spec 2 were removed.
- Spec 4 shows the final model. The variable QOF Stroke was removed as a negative sign did not seem plausible.

The morbidity flags and CCG dummy variables were included in all specifications.

10 Prescribing Formula

10.1 Introduction

The prescribing formula is a component of the overall formula used to distribute funding to Clinical Commissioning Groups (CCGs) in England. It covers the costs of drugs prescribed by GP practices in England.

The prescribing component is a need based utilisation formula and was first introduced in its current form in 2000, following research by Rice et al (1999)[10]. The formula was updated by Sutton et al (2002)[11], by Morris et al (2008)[12] and by Sutton et al (2010)[13].

The aim of this work was to refresh the prescribing formula following the same methodology as previously, using more recent data and re-running the regressions to determine the new, best performing model. The data used are five years more recent than those used by Sutton et al (2010) and the modelling gave a revised set of need and supply variables.

This section of the report outlines the development of the refreshed specification, the variables which are included and the outputs of the model.

10.2 Methodology

Data on the costs of drugs prescribed are not available at the person level, only at GP practice level.

A two-stage approach was therefore followed, as was the case in the research referenced above.

10.2.1 Stage 1 – Age-sex weights

This stage takes into account that the costs for each GP practice will vary depending on the age-sex profile of their registered list. GP practices with a higher proportion of older patients will have higher spend per head. This stage used an index of the relative national costs of prescribed drugs by age-sex group. The index weights were multiplied by the number of registrations in each age-sex group in each GP practice. This gave the expected costs for each GP practice if they had the national average spend by age-sex group.

---

[10] Rice N, Dixon P, Lloyd D, Roberts D. *Derivation of a needs based capitation formula for allocation prescribing budgets.* Report to the Department of Health, 1999.

[11] Sutton M., Gravelle H., Morris S., Leyland A., Windmeijer F., Dibben C., and Muirhead M. (2002) *Allocation of resources to English areas: Individual and small area determinants of morbidity and use of healthcare resources.* Report to the Department of Health. Edinburgh: Information and Statistics and Division.

[12] Morris S, Carr-Hill R, Dixon P, Law M, Rice N, Sutton M, Vallejo-Torres L. *Combining Age Related and Additional Needs: 2007 Review of the Needs Formulae for Hospital Services and Prescribing Activity in England (CARAN Report).*

[13] Sutton et al (2010), *Report of the Resource Allocation for Mental Health and Prescribing Project* (RAMP)

10.2.2 Stage 2 – Modelling actual to expected spend

Actual spend was divided by the expected spend from stage 1. Modelling was undertaken of the ratio of actual to expected spend at GP practice level using a wide range of explanatory variables on characteristics of patients, GP practices, and their place of residence.

10.3 Data

10.3.1 Spend by GP practice

Spend by GP practice for 2013-14 was provided by the NHS Business Services Authority. They are the total Net Ingredient Costs (NICs) for each drug prescribed and dispensed. NICs are the prices listed in the Drug Tariff, British National Formulary or price lists. Using NICs ensured the same prices for drugs were used for each GP practice. Some pharmacies pay different prices to NICs, but this is not a reflection of need and so these were not used in the cost data.

For GP practices in October 2013, data on spend were provided for 7,971 practices, totalling £8.3 billion.

10.3.2 Deprivation Score

For the stage 2 modelling, we updated the variables tested by Sutton et al (2010). Where possible data were collected on the same basis and updated for the latest available data. For example the Indices of Multiple Deprivation (IMD) were updated to the 2015 indices published in September 2015. We also extended the list of variables tested by including relevant attributed variables collected for the refresh of the G&A formula.

10.3.3 Suitable replacement data

There were a small number of cases where the variables included, or tested, in the previous model had either been discontinued or their definitions had changed, such as some welfare benefits. Where this was the case we attempted to source variables which were as close to direct replacements as possible.

The Low Income Scheme Index (LISI) was included in the formula from Sutton et al. (2010). LISI is an indicator of deprivation. It is the percentage of dispensed items exempt from the prescription charge on the grounds of low income (it does not include exemptions based on age, pregnancy or health). LISI has not been updated since 2007 and was therefore out of date and not available for new GP practices formed since 2007. As LISI was, broadly, a measure of deprivation we tested alternative measures from IMD 2015 and benefits data from the Department for Work and Pensions.

As the dependent variable in stage 2 of actual to expected costs is age-sex standardised, it was important that any independent variables which substantially vary by age were also age-sex standardised. Sutton et al included age-sex standardised QoF prevalence data as indicators of health status. These were calculated as a one-off exercise by Hippisley-Cox J et al. (2007)[14]. These were for

---

[14] Hippisley-Cox J, Vinogradova Y, Coupland C, Heaps M. Quality and Outcomes Time Series Analysis in QRESEARCH 2001 to 2006. Leeds: The Information Centre, 2007.

QoF data for 2001 to 2006, and these age-sex standardised QoF data were felt to be too old to use for this refresh of the formula. No later age-sex standardised QoF prevalence data were available (this would require individual level data from GP practices). We were therefore unable to use QoF prevalence data).

The University of Plymouth has developed models predicting prevalence by age-sex group for GP practices and LSOAs for common mental health disorders and CVD/CVD-related conditions in English households (Gibson et al. 2015)[15]. The predictions are based on micro-simulating from mainly national surveys. We considered using these variables but were advised by ACRA instead to use the underlying variables from the University of Plymouth models as it was felt that using the underlying data would be preferable and reduce multi-collinearity.

We therefore calculated age-sex standardised ethnicity and the number with limiting long term illness variables and a measure, not age-sex standardised, of housing tenure. Other variables used by Plymouth University, such as age and deprivation, were already in our data set.

A full list of variables collected in stage 2 is included in Table 10.6. Not all of these variables were used in our modelling. For example, the non-standardised ethnicity variables were collected as they had been tested by Sutton et al. (2010) but were not tested as we had ethnicity with age-sex standardisation.

10.4 Ratio of actual to expected spend

10.4.1 Prescribing units (ASTRO-PUs)

The average spend per head by age-sex group used to calculate the expected costs by GP practice were the age, sex and temporary resident originating prescribing units (ASTRO-PUs) from 2013 developed by the HSCIC. ASTRO-PU(2013)[16] values are shown in Table 10.8.

10.4.2 Dependent variable

The dependent variable for the modelling in stage 2 was the ratio of actual to expected prescribing costs for every GP practice. A practice with a ratio greater than 1 has higher than expected costs and a practice with a ratio below 1 has lower than expected costs. Figure 10.1 and Table 10.1 show the distribution of this ratio across GP practices.

[15] Gibson et al (2015) - Small Area Estimates of the Prevalence of Common Mental Health Disorders and CVD/CVD-related Conditions in English Household Populations: Technical Report

[16] http://www.hscic.gov.uk/prescribing/measures

Figure 10.1: Ratio of actual to expected costs

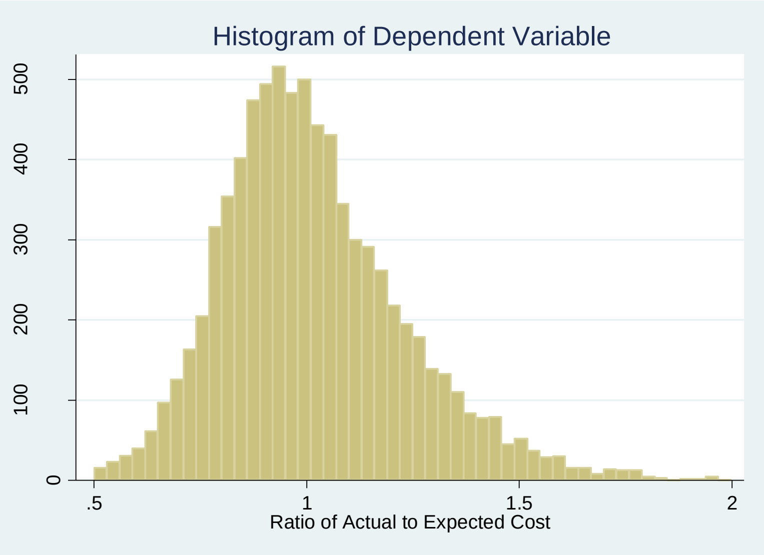


Table 10.1: - Dependent Variable Summary Statistics

Ratio of actual to age-sex standardised spend summary statistics	
1st Percentile	0.612
5th Percentile	0.723
10th Percentile	0.783
25th Percentile (Lower Quartile)	0.874
50th Percentile (Median)	0.981
75th Percentile (Upper Quartile)	1.112
90th Percentile	1.264
95th Percentile	1.365
99th Percentile	1.561
Mean	1.005

This shows that the ratio of average to expected cost follows an approximate bell curve, normal distribution with a peak around the point where the ratio of actual to expected costs equals 1. There are practices on either side of the distribution where costs were either higher or lower than would be expected based on age-sex alone.

10.5 Modelling

For stage 2, modelling the ratio of actual to expected spend at GP practice level, Ordinary Least Squares (OLS) regression was used with robust standard errors, as was the case in the development of the previous prescribing formula. The dependent variable (the ratio of actual to expected costs) was regressed against a number of supply and demand side variables. The supply variables also included dummy variables for each of the 209 Clinical Commissioning Groups (CCGs) in 2013-14 to take account of CCG wide supply factors.

The basic model was:

$$R_i = \alpha + \sum_j \beta_j N_{ji} + \sum_k \Upsilon_k S_{ki} + \varepsilon$$

Equation 10.1

Where R_i is the ratio of actual to expected costs for GP practice i
α is the constant term
N_{ji} are the need j variables for practice i
S_{ki} are the k supply variables for practice i
β_j are the coefficients on the j need variables
Υ_k are the coefficients on the k supply variables
ε is the residual error

10.5.1 Supply variables

Supply variables were included to take account of factors affecting costs that were not related to need. They are included in the models but removed, by setting their values to the national average values for all GP practices, in the calculation of the weights for CCG target allocations.

10.5.2 Practice list size as frequency weight

We used the practice list size as a frequency weight in the modelling. This means that practices with more patients had a larger influence on the coefficients compared with smaller practices. We followed this approach as it was not appropriate for a practice with 10,000 patients to have the same influence on the model as a practice with 1,001 patients. This approach also had a positive impact on the predictive power of the model.

10.5.3 Excluded practices

A small number of practices were excluded from the modelling. Practices with fewer than 1,000 patients were excluded. In addition practices with a ratio of actual to expected spend either below 0.5 or above 2.5 are removed. This is because they were considered to be practices not reflective of general practices as a whole and may serve populations with special needs. For example, one of the practices with a particularly high ratio was found to be a practice specialising in treating those with substance misuse issues, and will have received additional funding from other budgets. The outliers were removed to avoid distorting the outputs from the modelling.

Table 10.2: Excluded GP Practices

Total GP Practices	7,971
Practices with list size of under 1,000	19
Practices with ratio greater than 2.5	31
Practices with ratio less than 0.5	60
Remaining Practices (12 practices in 2 exclusion categories)	7,869

10.5.4 Model selection

We investigated the use of alternative model selection processes including modelling based on the variables in the Sutton et al. (2010) model, stepwise regression and Least Angle Regression (LARS). Stepwise regression, an automated approach whereby variables are either added or removed until only significant variables remain, was not selected as it produced models with both unexpected signs and omitted important variables in the Sutton et el. formula. Least Angle Regression (LARS) is a variant of forward stepwise regression but uses slightly different algorithms to add and remove variables. LARS was not selected as it failed to produce a parsimonious model; only one out of over fifty variables was not significant in the LARS version and we therefore rejected it.

The preferred option therefore was to base the model on the variables used in the Sutton et al. Specification, and update the variables to use the latest available data. Where updated variables were not available (for example LISI), they were replaced with the closest alternative, and finally new, relevant variables were tested.

In place of the LISI variable we considered the Index of Multiple Deprivation (IMD) and benefits data from the Department for Work and Pensions. The IMD was preferred to benefits data, as some benefits were already in the Sutton et al. specification, and data were available on a wide range of different benefits and it was not clear which to select.

We tested the overall IMD index and two domains (income and health). The results of these regressions are shown in Table 10.7. The overall IMD index was chosen as it had good R-squared statistic and had the expected sign (higher deprivation is expected to increase drug costs). IMD Health was not selected as we felt that it could be highly correlated with the SMR variable in the models. IMD income was not selected as while this variable had the expected sign it caused the social housing variable to become negative. This is a sign that there is high correlation between the two variables.

To replace the QoF prevalence data in Sutton et al we introduced age-sex standardised variables for ethnicity, the number with limiting long term illness and a measure, not age-standardised, of housing tenure.

The final stage of modelling was to test other variables that were either not significant in Sutton et al. (2010) or were newly available. This led to the inclusion of a dummy variable for the top 1% of practices with most patients aged 20-24. The negative sign on this variable is reasonable as those in younger age-groups, on average, have lower need for such medication. The CCG dummy variables were included in all the models.

10.6 Final model

The final regression model is shown in Table 10.3.

Table 10.3: Refreshed prescribing model

Variable	Description	Coefficients	T-statistic
Supply Variables			
DISPENSING	Dummy variable value 1 if practice dispenses medication (0 Otherwise)	0.019	392.81
ONE_GP	Dummy variable value 1 if Practice has single GP (0 Otherwise)	-0.018	-133.81
PROP_UKQUAL_GP	Proportion of practice GP's qualified in the UK	0.006	64.55
MEAN_AGE_GP	Mean age of GP in practice	-0.001	-171.31
GEN_SAVINGS	Potential savings based on top 20 drugs by GP practice	0.150	921.08
Distance to GP*	IMD (2015) distance to GP	-0.004	-161.17
Need Variables			
PROP85PLUS	Proportion of registrations over the age of 70 also over the age of 85	0.218	354.8
DLA_OVER70	DLA claimants over the age of 70	0.450	247.11

Variable	Description	Coefficients	T-statistic
SMR_ALL	Standardised Mortality Ratio (2008-2012)	0.001	548.14
FERTILITY_RATE_1214	Generalised Fertility Rate (2012-14)	0.599	575.13
_24_TOP1*	Top 1% of Practices with most patients aged between ages 20 and 24	-0.198	-948.91
IMDOverall*	IMD (2015) Overall	0.008	1403.59
ActivityLimited*	Standardised proportion with activity limiting health conditions	0.141	450.02
TenureSocial*	Proportion in Social Housing	0.102	230.2
EthnicNWh*	Standardised measure of those from Non White Ethnic categories	0.216	307.95
CCG Dummies	Number of CCG Areas	211	
Constant		0.093	
Observations		55,763,599	
Model R-Squared		65.2%	

* Indicates variable not included in Sutton et al. (2010)

The R-Squared is for the before sterilising supply variables. The R-squared of 65.2% is an improvement compared with Sutton et al. of 61.9%.

The characteristics of GPs were included as supply variables as they could influence prescribing decisions.

Potential savings based on top 20 drugs by GP practice is a measure of the savings which would be possible if generic drugs are used instead of branded counterparts. It has been provided by the NHS Business Services Authority and is calculated using the top 20 drugs with the greatest potential cost savings for each GP practice.

Distance to GP is the road distance from the patients' LSOA to the GP practice from the "Barriers to Housing and Services domain" in IMD 2015.

As a proxy for student population we included a dummy {0,1} variable for the top 1% of practices with the largest proportion of patients aged 20-24. Under the revised specification this variable was found to be significant and negative. We also tested similar dummy variables for the top 5% and 10%, however the 1% variable performed the best statistically.

10.7 Model Evaluation

10.7.1 Predictive power

The estimates generated by the model can be compared with the actual ratios to assess the performance of the model. Figure 10.2 shows the residuals from the model; this is the difference between the actual ratio and what was predicted by the model. Table 10.4 summarises the distribution of the residuals.

These estimates are made prior to the sterilisation of supply variables. This means that all variables, including CCG dummies, are used in the estimation.

Figure 10.2: Residuals from refreshed regression model

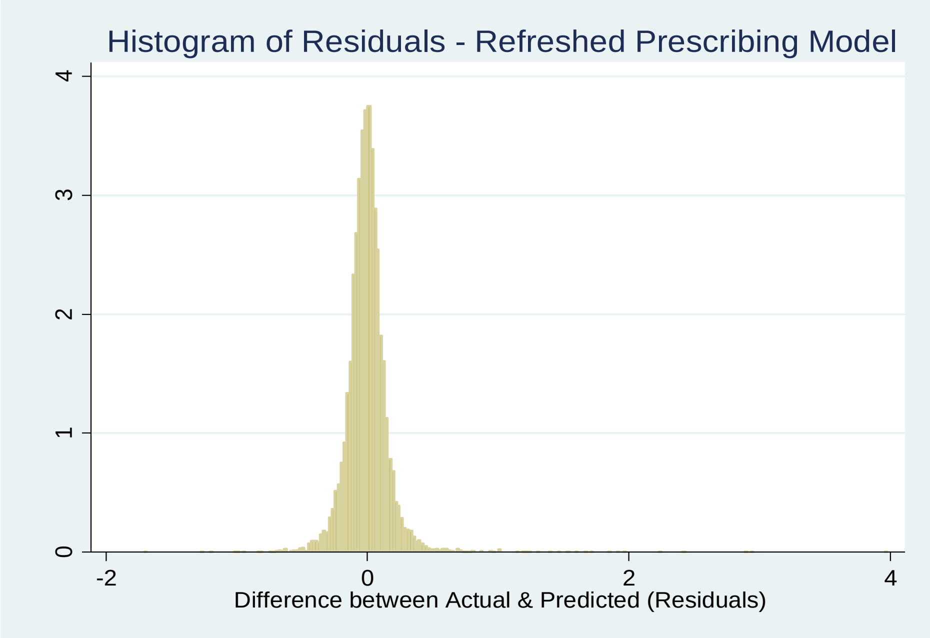


Table 10.4: Residual summary statistics

Percentile	Residual
1%	-0.407
5%	-0.229
10%	-0.162
25% (LQ)	-0.080
50% (Median)	-0.005
75% (UQ)	0.069
90%	0.152
95%	0.225
99%	0.456
Mean	0.001
Minimum	-1.706
Maximum	8.296

Figure 10.2 suggests that the model performs well at predicting the actual to expected ratio of costs at GP practice level. There are a small number of GP practices where there is a large difference between the actual and predicted totals, however these may be the outlier practices which were not included in the initial modelling or are GP practices who serve specialist communities.

10.7.2 Mean Absolute Percentage Error (MAPE)

The Mean Absolute Percentage Error (MAPE) is an alternative measure of performance. It is calculated by determining the percentage difference between the prediction made by the model and the actual value. MAPE equally penalises both under and over estimates. Table 10.5 shows the number of practices and their list size by MAPE. The table shows that fewer than 10% of patients belong to practices where the error is in excess of 20%.

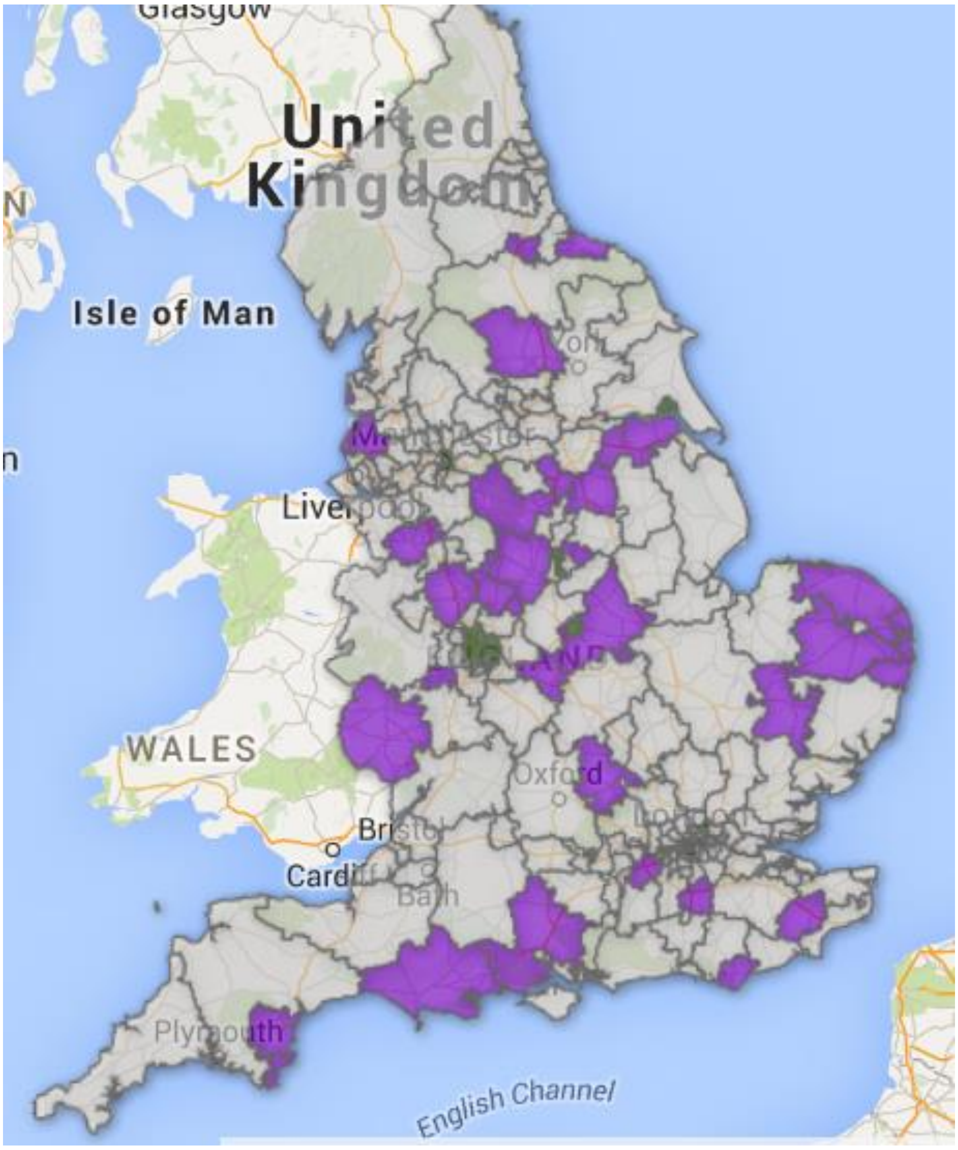
Table 10.5: Mean Absolute Percentage Error (MAPE)

MAPE	Practices	Patients	Proportion of England population
< 5% (Prediction +/- 5% of actual)	2,831	22,736,032	40.4%
< 10% (Prediction +/- 10% of actual)	4,960	38,715,214	68.8%
< 20% (Prediction +/- 20% of actual)	6,969	51,607,995	91.7%
> 20% (Prediction more than 20% from actual)	1,002	4,662,204	8.3%
Total	7,971	56,270,312	100%

To calculate a population weighted MAPE at CCG level, the MAPE for each GP practice is multiplied by the proportion of the CCG's population registered with each practice. The sum of these contributions is the MAPE at CCG level.

Figure 10.3 shows the CCGs with the 20 highest (purple) and lowest (green) MAPE. It suggests that there is no systematic bias in the model as the areas with the largest errors are dispersed across England.

Figure 10.3: Map of MAPE by CCG – highest (purple) and lowest (green)



10.8 Sterilisation of supply variables

The final specification includes both need and supply side variables. As with the other formulae, the supply variables are "sterilised" so that they do not impact the weights used for setting allocations to CCGs.

The value of each supply variable is set top the national average for all GP practices. This in practice involves modifying the constant term to incorporate the supply variables. The coefficient of each supply variable in Table 10.3 is multiplied by the national average vale for that supply variable. These values are then summed across all of the supply variables, and the result added to the original constant. This is summarised in the equation below.

$$\begin{aligned} &\text{Modified Constant} \\ &= \text{Original Constant} \\ &+ \sum_{1}^{n} \text{Supply Variable} \times \text{National Average} \\ &+ \sum_{1}^{n} \text{CCG Dummy Coefficient} \\ &\times \text{Proportion of Population in CCG} \end{aligned}$$

Equation 10.2

After performing this calculation the new constant is 0.174. This value is then used for all GP practices in calculating the weights used for allocations.

10.9 Future development

While we believe this model is an improvement on the previous specification, there are also areas for future development:

Due to time constraints we have used variables such as ethnicity and the proportion with life limiting conditions as proxies for disease prevalence. It would have been preferred to use actual prevalence figures within the model as this would have been more accurate. Similarly we would have liked to have used an updated LISI index to compare with the IMD measures which were used in the models.

The longer term aim should be to use person level data for the development of the prescribing formula, rather than as at present GP practice level data.

10.10: List of variables collected

Table 10.6: List of variables collected

Variable name	Variable label
GP_PRAC_CODE	GP Practice code
CCG_CODE	CCG code
RATIO	Ratio of the actual Net Ingredient Costs of prescribed drugs to the expected costs
GEN_SAVINGS	Potential generic savings based on the top 20 most prescribed drugs where each practice could make the most savings
MEAN_AGE_GP	Mean age of GP in practice
PROP_UKQUA~P	Proportion of practice GPs qualified in the UK
PROP_FEMAL~P	Proportion of female GPs
NUM_GP	Number of GP's in practice
ONE_GP	Dummy variable, value 1 if practice has single GP (0 otherwise)
PMS_PRAC	Dummy variable, value 1 if PMS practice (0 otherwise)
DISPENSING	Dummy variable, value 1 if practice prescribes medication (0 otherwise)
WTE_GP	Number of full-time equivalent GPs per 1,000 registered patients
PAT_JOIN	Proportion of newly registered patients per total practice registrations
PAT_LEAVE	Proportion of registered patients leaving per total practice registrations
PAT_TURNOVER	Turnover of registered patients per total practice registrations
PAT_DIST	Average patient distance from practice
PROP85PLUS	Proportion of registered patients over the age of 75 also over the age of 85
20_24_TOP1	Dummy variable, value 1 if practice in top 1% of practices with most patients aged between ages 20 and 24 (0 otherwise)
20_24_TOP5	Dummy variable, value 1 if practice in top 5% of practices with most patients aged between ages 20 and 24 (0 otherwise)
20_24TOP10	Dummy variable, value 1 if practice in top 10% of practices with most patients aged between ages 20 and 24 (0 otherwise)
YPLL_IND	Years of potential life lost indicator
COMP_ILL_IND	Comparative illness and disability indicator
WHITE	Proportion of registered patients who are white
INDIAN	Proportion of registered patients who are Indian
PAKISTANI	Proportion of registered patients who are Pakistani
BANGLADESHI	Proportion of registered patients who are Bangladeshi
BLACK_CARI~N	Proportion of registered patients who are Black Caribbean
BLACK_AFRI~N	Proportion of registered patients who are Black African
CHINESE	Proportion of registered patients who are Chinese
IBSDA_ALL	Proportion of registered patients claiming Incapacity Benefit/ Severe Disablement Allowance of all those eligible
IS_ALL	Proportion of registered patients claiming Income Support of all those eligible
PC_ALL	Proportion of registered patients claiming Pension Credit of all those eligible
DLA_ALL	Proportion of registered patients claiming Disability Living Allowance of all those eligible
DLA_OVER70	Proportion of registered patients claiming Disability Living Allowance of all those eligible aged over the age of 70
JSA_ALL	Proportion of registered patients claiming Job Seekers Allowance of all those eligible
SMR_ALL	Standardised Mortality Ratio (2008-2012) – All ages
SMR_U65	Standardised Mortality Ratio (2008-2012) - Under 65 Years
SMR_U75	Standardised Mortality Ratio (2008-2012) - Under 75 Years
BIRTH_R~1214	Practice birth rate 2012-14
FERTILI~1214	Practice generalised fertility rate 2012-14
LOWBIRT~1214	Practice proportion of births that are low birth weight 2012-2014
RURALITY_1	Dummy variable, value 1 if practice in urban conurbation (0 otherwise)
RURALITY_2	Dummy variable, value 1 if practice in urban city and town (0 otherwise)

Variable name	Variable label
RURALITY_3	Dummy variable, value 1 if practice in rural town and fringe (0 otherwise)
LIST_SIZE	Total number of practice registered patients
MALE_04	Proportion of males aged 0-4 registered with practice
MALE_514	Proportion of males aged 5-14 registered with practice
MALE_1524	Proportion of males aged 15-24 registered with practice
MALE_2534	Proportion of males aged 25-34 registered with practice
MALE_3544	Proportion of males aged 35-44 registered with practice
MALE_4554	Proportion of males aged 45-54 registered with practice
MALE_5564	Proportion of males aged 55-64 registered with practice
MALE_6574	Proportion of males aged 65-74 registered with practice
MALE_75_PLUS	Proportion of males aged 75 and over registered with practice
FEMALE_04	Proportion of females aged 0-4 registered with practice
FEMALE_514	Proportion of females aged 5-14 registered with practice
FEMALE_1524	Proportion of females aged 15-24 registered with practice
FEMALE_2534	Proportion of females aged 25-34 registered with practice
FEMALE_3544	Proportion of females aged 35-44 registered with practice
FEMALE_4554	Proportion of females aged 45-54 registered with practice
FEMALE_5564	Proportion of females aged 55-64 registered with practice
FEMALE_6574	Proportion of females aged 65-74 registered with practice
FEMALE_75_~S	Proportion of females aged 75 and over registered with practice
IMDOverall	Indices of Multiple Deprivation (IMD 2015) Overall score
Crime	IMD 2015 Crime domain score
Employment	IMD 2015 Employment domain score
Income	IMD 2015 Income domain score
Health	IMD 2015 Health domain score
Nocentralh~g	IMD 2015 No central heating indicator
Livingenvi~t	IMD 2015 Living environment domain score
Moodanxiety	IMD 2015 Mood and anxiety disorders indicator
YPLL	IMD 2015 Years of potential life lost (YPLL) indicator
Education16	IMD 2015 Staying on in education post 16 indicator
Education	IMD 2015 Entry to higher education indicator
IncomeIDACI	IMD 2015 Income Deprivation affecting Children index (IDACI)
IncomeIDAOPI	IMD 2015 Income Deprivation affecting Older People index (IDAOPI)
DistancetoGP	IMD 2015 Road distance to a GP surgery indicator
AirQuality	IMD 2015 Air Quality indicator
ActivityLi~d	Age-sex standardised proportion with activity limiting health conditions
BadHealth	Age-sex standardised proportion with bad or very bad health
TenureSocial	Proportion in social housing
EthnicNWh	Age-sex standardised proportion non-white
EthnicNWBr	Age-sex standardised proportion non-white British
EthnicABIn	Age-sex standardised proportion Asian British Indian
EthnicABPa	Age-sex standardised proportion Asian British Pakistani
EthnicABBa	Age-sex standardised proportion Asian British Bangladeshi
EthnicABCh	Age-sex standardised proportion Asian British Chinese
EthnicBBAf	Age-sex standardised proportion black British African
EthnicBBCa	Age-sex standardised proportion black British Caribbean
g218WeeksRTT	18 Weeks RTT Waiting Times

Table 10.7: IMD variable comparisons

VARIABLES	(1) IMD Overall	(2) IMD Income	(3) IMD Health
Dispensing Practice	0.0186***	0.0187***	0.0244***
One GP	-0.0176***	-0.0181***	-0.0167***
Proportion UK Qualified	0.00563***	0.00888***	0.00567***
Average Age	-0.000687***	-0.000691***	-0.000258***
Potential Generic Savings	0.150***	0.148***	0.147***
Distance to GP	-0.00378***	0.00269***	0.0115***
Proportion of those 75+ over 85+	0.218***	0.241***	0.169***
DLA Over age 70	0.450***	0.457***	0.409***
SMR	0.00132***	0.00135***	0.00117***
Fertility Rate	0.599***	0.447***	0.681***
Top 1% Aged 20-24	-0.198***	-0.169***	-0.197***
IMD Overall	0.00796***		
IMD Income		1.437***	
IMD Health			0.150***
Long Term Illness	0.141***	0.113***	0.131***
Proportion Social Housing	0.102***	-0.0274***	0.220***
Non White Ethnicity	0.216***	0.192***	0.259***
Constant	0.0926***	0.109***	0.140***
Observations	55,763,599	55,763,599	55,763,599
R-squared	0.652	0.657	0.653

***** p<0.01, ** p<0.05, * p<0.1**

Table 10.8: ASTRO-PU (2013) Values

Age Band	Male	Female
0-4	1.0	0.9
5-14	0.9	0.7
15-24	1.2	1.4
25-34	1.3	1.8
35-44	1.8	2.6
45-54	3.1	3.7
55-64	5.3	5.4
65-74	8.7	7.6
75+	11.3	9.9

11 Emergency Ambulance Cost Adjustment

11.1 Introduction

The Emergency Ambulance Cost Adjustment (EACA) is an adjustment in the core formula for core CCG allocations. The EACA compensates for unavoidable differences in the costs of providing emergency ambulance services in different parts of the country, in particular higher costs in sparsely populated areas.

The previous EACA index was introduced into the health services funding formula in 1998, and was based on modelling costs based on the volume of activity, the proportion of journeys which were emergency (case mix) and a measure of rurality[17]. This is the first refresh of the formula since it was introduced.

The refreshed EACA index is based on a regression model using data from four ambulance trusts in England. We modelled the time taken in minutes to respond to emergency incidents from when the first vehicle was dispatched. This included the times reach the scene, time at the scene, time to convey to hospital, and ambulance turnaround time at the hospital. Different models were developed depending on whether the patient was transferred to hospital or treated on scene.

The costs of call centres were not included in the refresh as it is not clear why these should unavoidably differ across the country.

We did not have data on ambulance stand-by time, which may unavoidably vary across the country. Despite this limitation, the use of data that is fifteen years more recent than used for the previous EACA index is a major improvement.

11.2 Data

11.2.1 Sources

Data were provided by four ambulance trusts across England. They were:

- The East Midlands Ambulance Service
- The North East Ambulance Service
- The London Ambulance Service
- The South West Ambulance Service

11.2.2 Coverage (time period and data items)

Each ambulance trust provided data for 2013-14 (in the case of the South West the data covered an 18 month period between 1 April 2014 and 30 September 2015). The data provided included information on the time taken to reach the scene, the time spent at the scene and the location of the incident. In addition, for see & convey cases the data included the time taken to reach the hospital, and ambulance turnaround time at the hospital. The average times covered all the vehicles which arrived on scene, including when more than one vehicle is dispatched to the same incident.

---

[17] Study of the Costs of Providing Health Services in Rural Areas, MHA and Operational Research in Health Ltd, 1997

Some of the ambulance trusts provided additional detail including emergency case-mix (whether the call was Red 1, Red 2 or Green), costs or how many vehicles were dispatched. But as these were not available for all four trusts, they could not be used in the modelling.

11.2.3 Level of geography

To produce, and implement, a model we needed to select a suitable level of geography for the analysis. Trusts were asked to use the Middle Super Output Areas (MSOA) as the location of the incident and provide average times for each MSOA. MSOAs were chosen to ensure the confidentiality of data on individual incidents, and because many potential variables for inclusion in the modelling were available at MSOA level. Derived geography

There was one ambulance trust which did not directly provide data by MSOA, but instead provided data by postcode sector (e.g. LS2 2). As postcode sectors can span multiple MSOA areas, postcode sectors were assigned to the most common MSOA for that postcode sector. Using the Postcode Directory available from the Office for National Statistics the number of MSOA areas in each postcode sector were counted and the most common MSOA area used.

11.2.4 MSOA characteristics

For the explanatory variables in the modelling, we collected data on the characteristics of MSOAs. These were selected on the basis of consultation with stakeholders and advisory groups and included ONS rurality classifications, measures of population density, and the average age of the population resident in the MSOA. We also had the road distance and travel times from the centroid of the MSOAs to the nearest major A&E department.

11.3 Number of incidents in the data

Table 11.1 shows the number of incidents in the dataset for the modelling.

Table 11.1: Number of incidents by ambulance trust

Region	Incident volume	See & convey	See & treat
London	960,960	750,794	210,166
South West	388,519	225,823	162,696
North East	444,954	349,800	95,154
East Midlands	627,160	439,915	187,245
Total	2,421,593	1,766,332 (73%)	655,261 (27%)

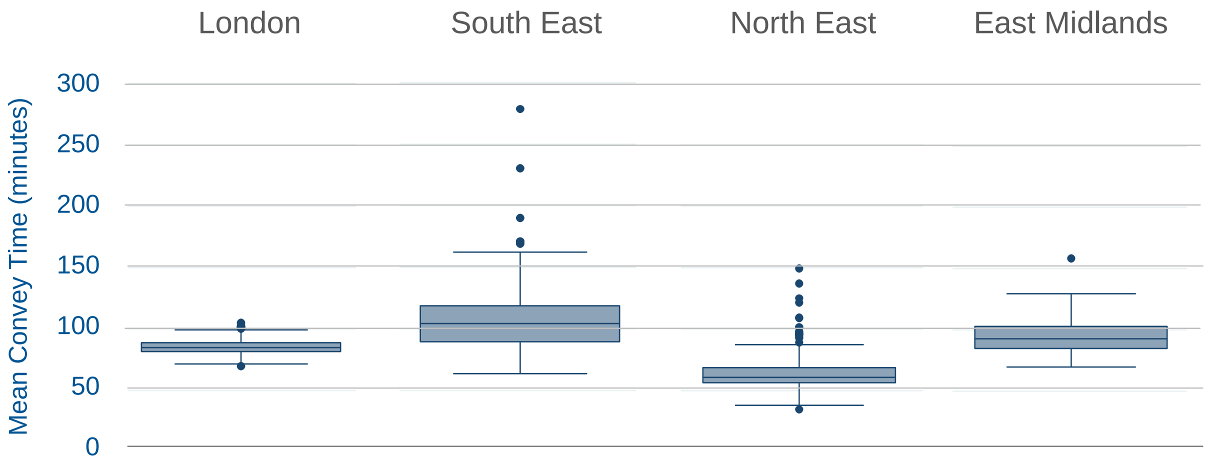
Table 11.2 shows the average times split by trust and by see & treat and see & convey.

Table 11.2: Average times by trust

Region	No of MSOAs	See & convey (mins)	See & treat (mins)
London	983	85.44	61.37
South West	631	109.63	69.51
North East	374	65.25	47.66
East Midlands	613	94.17	64.90

For see & convey cases Figure 11.1 shows the distribution of average times between MSOA areas. It shows how London has a much narrower range of average times compared with the other regions.

Figure 11.1: Average see & convey times



11.4 Modelling method

The refresh of EACA index used ordinary least squares (OLS) regression to model the times to resolve incidents (the dependent variable) as a function of independent variables such as the distance to major A&E departments. Separate models were estimated for see & treat and see & convey cases as journey times for conveying to hospital may have different determinants to the times to the scene.

The unit for the modelling was the average times by MSOA. MSOAs were weighted by the number of cases in each MSOA. This means that an area with a small number of cases had a lesser impact on the coefficients compared with a MSOA with a high volume of cases. This protected against MSOAs with low volumes of cases unduly affecting the model results.

The basic model is:

$$T_i = \alpha + \sum_j \beta_j N_{ji} + \varepsilon$$

Equation 11.1

Where

T_i is the average time to deal with incidents in MSOA i

α is the constant term

N_{ji} are the j factors affecting times to deal with incidents for MSOA i

β_j are the coefficients on the j variables

ε is the residual error

11.5 Developing the model

11.5.1 Independent (explanatory) variables

In developing the model we collected and tested a number of independent (explanatory) variables. These were selected on the basis of consultation with stakeholders and discussion with ACRA and TAG. The principal of the modelling was to develop a model which was both parsimonious and had high explanatory power.

For an independent variable to be tested, it was necessary that the data were available for all MSOA areas in England. Our data included additional information about the call (for example Red 1 / Red 2 / Green classification), however we were not able to include these in our modelling since the data were not available for all MSOAs in England, and it would, therefore, not have been possible to apply the results across England.

11.5.2 Distance to hospital

An intuitive determinant of times to resolve an incident is the distance from the location of the incident to the nearest hospital. Using travel time software we estimated the distance from the centroid of each MSOA to the closest provider with a 24/7 Accident & Emergency department.

The list of hospitals to be included was provided by NHS Choices and totalled 174 providers. The distances were road distance, not distance “as the crow flies”.

11.5.3 Age related effects

We tested different variables to account for age. The average age variable, which was suggested by TAG, was tested in case those with older average ages tended to exhibit longer incident times. We also tested the proportion of the population aged 65 and over.

11.5.4 Urban/Rural classifications

We tested different rurality classifications developed by ONS and DEFRA. These classify areas into a number of groups depending on an areas’ characteristics and whether or not the area is classed as being “built up”. For example an area is classified as urban if its population is more than 10,000 people.

Following discussions with ACRA, we ran variations of the model with different rurality classifications and dummy variables for areas classed as being either urban or sparse. The results of these regressions are shown in Table 11.6 and Table 11.7.

The preferred model excluded the rurality classifications entirely. The model with variables for all rurality/urban groups was rejected due to implausible positive signs for some rural areas. The model with a single rural dummy was rejected as this variable had an implausible, positive coefficient. The model with a single sparsity dummy was also rejected, as there were a small number of MSOAs (and cases) in this group and it did not improve the overall R-squared.

11.5.5 Differences between trusts

Dummy variables were included for each ambulance trust that supplied data. This took account of any systematic differences, for example, in call handling techniques or operating procedures. The impact of these dummies was sterilised (by using a national average) before the results were implemented in the CCG formula.

11.5.6 Dependent variable - average time by MSOA

Initial modelling combined both see & treat and see & convey cases. This meant that the dependent variable was the average time for all cases within the MSOA. When examining the results, however, it was clear that see & treat and see & convey cases were different and had very different time distributions. Figure 11.2 and Figure 11.3 show histograms of actual times by MSOA for different types of case. They show that the see & treat distribution appears to be bimodal. As the two distributions were so different separate models were developed for see & treat and see & convey.

Figure 11.2: See & convey distribution

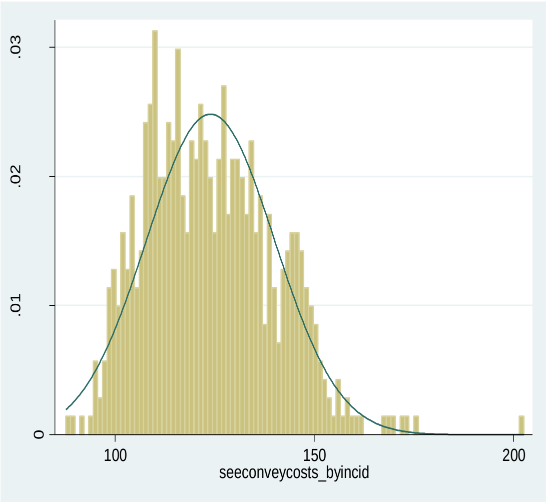
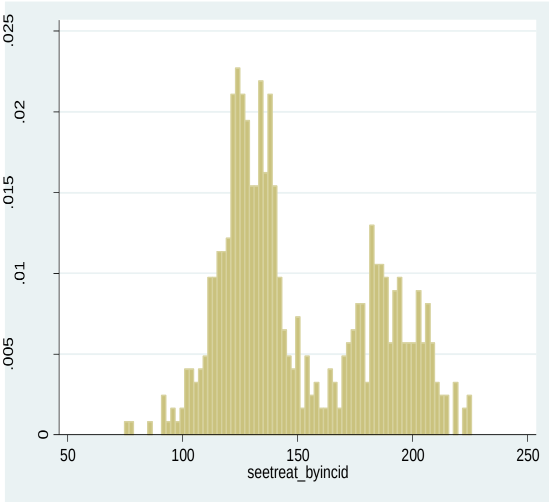


Figure 11.3: See & treat distribution



11.5.7 Stepwise approach

We adopted a stepwise style approach where by all variables were initially included and then removed if the coefficient was either insignificant or the sign was not plausible.

11.6 Model outputs

11.6.1 Final models

Table 11.3 shows the final models for both types of cases. While the models were estimated separately, the same variables were in both models.

Table 11.3: EACA regression output

Variable	Description	See & convey	See & treat
Distance to A&E	Estimated distance to hospital with 24/7 A&E	1.653***	1.147***
Distance to A&E (Sq)	Estimated distance to hospital (squared)	-0.013***	-0.025***
Log of Population Density	Natural log of population density (2011 Census)	-1.407***	-0.728***
Average Age in MSOA	Average age in MSOA area (2011 Census)	0.222***	0.426***
Trust 1 (Base region)	Trust dummies	0	0
Trust 2		7.317***	-2.299***
Trust 3		-29.847***	-19.834***
Trust 4		-0.767***	-2.684***
Constant	Constant	78.656***	45.459***
Observations		1,766,332	655,261
R-squared		0.808	0.46
Adj. R2		0.808	0.46

Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

11.6.2 Model performance

The predicted times from the models can be compared with the actual average times as a way to assess the performance of the model. These estimates were taken prior to sterilising the trust dummy variables.

Figure 11.4: Model performance - see & convey

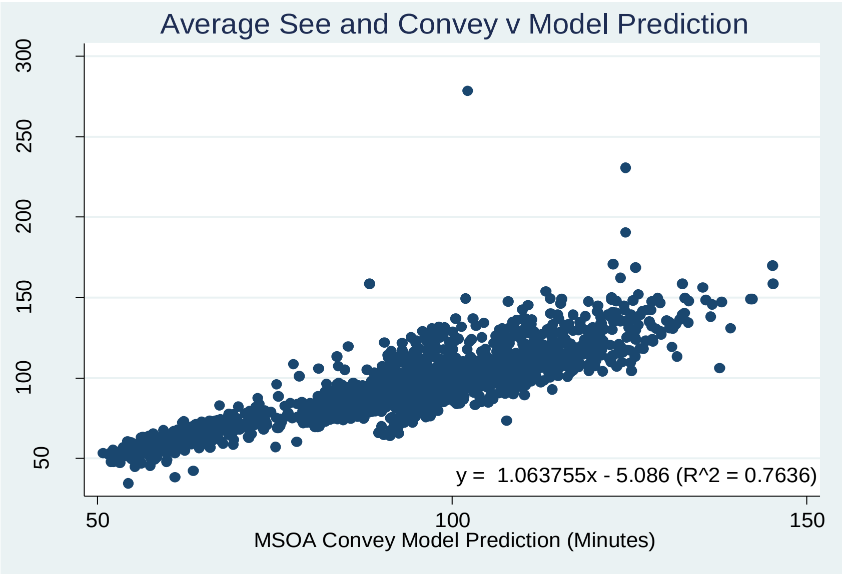
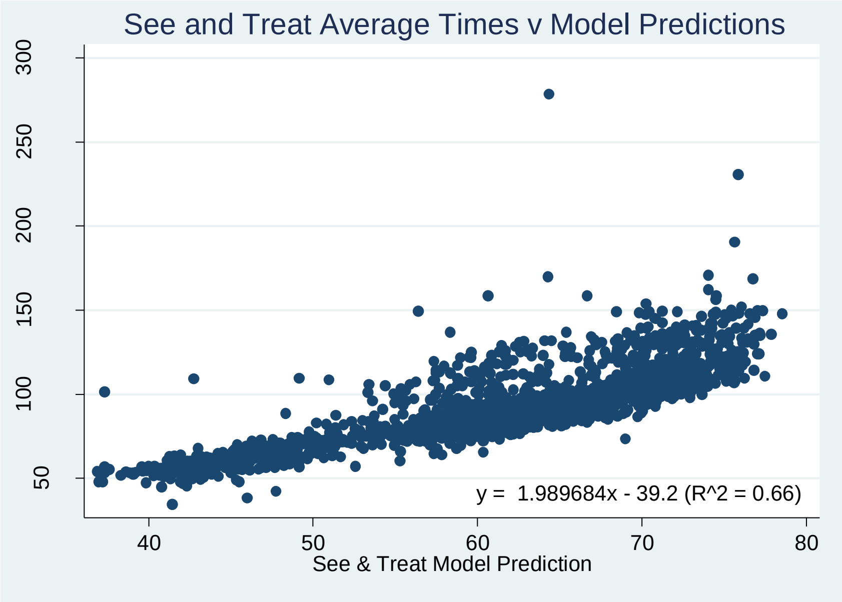


Figure 11.5: Model performance - see & treat



These figures demonstrate the performance of the model at MSOA level. In both cases they show a strong, positive correlation between the actual time and the time predicted by the model. The see & convey model performed better than the see & treat model, as also shown by the higher R-squared statistic for see & convey.

For both types of cases, there were a small number of outlier values where there was a large difference between the predicted and actual value. These, however, tended to be MSOAs with a very small number of incidents and it is possible that they had a small number of cases with extremely long times which are difficult to model.

11.6.3 Alternative measure - MAPE

An alternative, related measure of performance is the Mean Absolute Percentage Error (MAPE). This is the percentage error between the observed figure and that estimated by the model. Figure 11.6: See & convey MAPE, Figure 11.7 and Table 11.4 show the distribution of MAPE for see & convey and see & treat cases.

Figure 11.6: See & convey MAPE

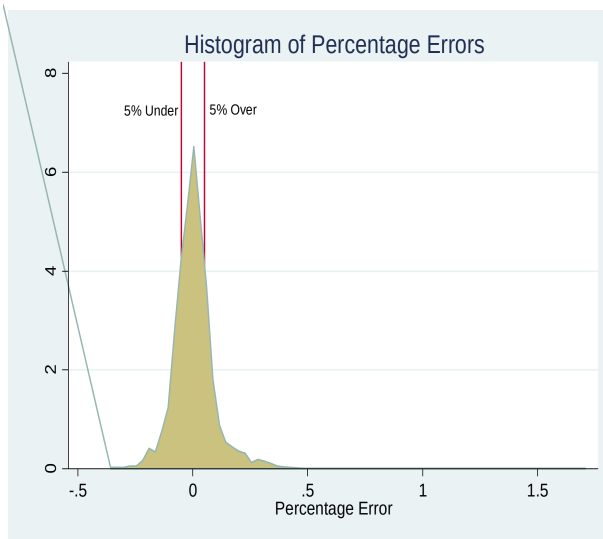


Figure 11.7: See & treat MAPE

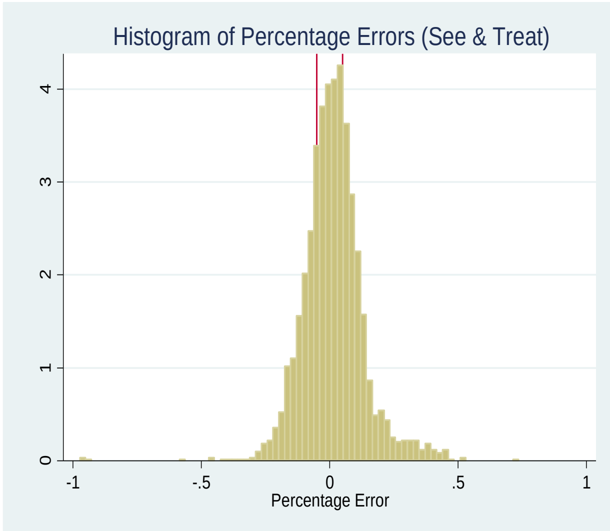


Table 11.4: MAPE summary statistics

Summary statistics	See & convey MAPE	See & treat MAPE
10^{th} Percentile	-8.87%	-11.78%
Lower Quartile	-4.43%	-5.25%
Median	0.01%	1.23%
Upper Quartile	4.50%	7.52%
90^{th} Percentile	9.56%	14.08%
Mean Error	0.57%	1.77%

The figures and table suggest that for see & convey cases there was a lower MAPE compared with see & treat cases. This would be expected given the better R-squared for the see & convey model.

11.7 Creating the EACA index

To create the EACA index for the whole of England, the models in Table 11.3 had to be applied to every MSOA in England.

After collecting data for each of the variables included in the regression for every MSOA, an average time for each MSOA was calculated for both categories of cases. The average time was calculated by multiplying the value of the explanatory variable at MSOA level by the model coefficient and summing as shown in Equation 11.2.

$$\text{MSOA Prediction Convey} = \text{Constant} + \sum_{1}^{n} \text{Convey Coefficient} \times \text{MSOA Value}$$

Equation 11.1

As the Trust variable was considered to pick up differences in procedures within the Trusts' own control, this variable was neutralised prior to implementation. In line with other allocation formulae this involved setting value of the Trust variable to the national average for all MSOA areas.

For each MSOA area an overall average modelled time was found by combining the estimates for see & convey and see & treat cases. The two modelled times were combined in proportion to the volumes of the two different cases for the four Trusts all together. This split was 73% / 27% in favour of see & convey as shown in Table 11.1.

$$\text{MSOA Average Time} = (0.73 \times \text{See \& Convey Time}) + (0.27 \times \text{See \& Treat Time})$$

Equation 11.2

The average times for CCGs are weighted averages for their MSOAs, where the weights are the MSOA population sizes. These were then turned into an index at CCG level relative to the average time for England. The England average index value was set to the value of 1.0.

To create the final EACA index, the index values were multiplied by the proportion of the CCG total spend on ambulance services. This was around 3.2%.

The CCGs with the highest and lowest EACA index values are shown in

Table 11.5:.The areas with the highest values are typically more rural areas with longer journey times to hospitals. The areas with the lowest EACA values tend to be inner city areas where journey times are lower and population density is higher.

Table 11.5: EACA Index (Top / Bottom 10)

Highest EACA Index			Lowest EACA Index		
CCG	**Name**	**Index**	**CCG**	**Name**	**Index**
07K	NHS West Suffolk	1.007	09A	NHS Central London (Westminster)	0.997
06V	NHS North Norfolk	1.007	00W	NHS Central Manchester	0.997
11N	NHS Kernow	1.006	08Y	NHS West London (Kensington & Chelsea)	0.996
10M	NHS Newbury and District	1.006	08Q	NHS Southwark	0.996
06Y	NHS South Norfolk	1.005	08H	NHS Islington	0.996
99D	NHS South Lincolnshire	1.005	08M	NHS Newham	0.996
09X	NHS Horsham & Mid Sussex	1.005	07R	NHS Camden	0.996
09E	NHS Canterbury & Coastal	1.005	07T	NHS City & Hackney	0.996
10A	NHS South Kent Coast	1.005	08C	NHS Hammersmith & Fulham	0.996
03T	NHS Lincolnshire East	1.004	08V	NHS Tower Hamlets	0.996

11.7.1 Areas for development

While we believe that the new EACA is an improvement on the previous formula as it uses recent data and utilises real times to see, treat and convey, there remain a number of areas for future development:

1. We were only able to obtain data from 4 ambulance trusts. Using data from all nine ambulance trusts would have made it possible to derive a more complete picture of the ambulance service in England.
2. Using our data it was only possible to model the characteristics of ambulance calls. We were not able to collect data on the full costs of providing ambulance services in different areas, for example is there a difference in the amount of stand-by time in certain areas?
3. Our preliminary analysis suggested a bimodal distribution for see & treat cases – we would like further investigate this finding and its implications for modelling.

11.8 Impact of rurality variables

Table 11.6: Alternative rurality flags (see & convey)

Rurality flags (see & convey)	No rurality	Rural dummy	Sparsity dummy	ONS rurality
DistancetoAE	1.653***	1.645***	1.651***	1.632***
DistSq	-0.013***	-0.013***	-0.013***	-0.012***
Log of Population Density	-1.407***	-1.228***	-1.409***	-1.094***
Average Age	0.222***	0.230***	0.222***	0.251***
Region 1	0	0	0	0
Region 2	7.317***	7.365***	7.312***	5.965***
Region 3	-29.847***	-29.746***	-29.848***	-30.324***
Region 4	-0.767***	-0.664***	-0.772***	-2.486***
Rural		1.014***		
Sparse			-0.192**	
Rural Town & Fringe				0
Rural Town & Fringe (Sparse)				-1.246***
Rural Village & Dispersed				2.365***
Rural Village & Dispersed (Sparse)				1.563***
Urban City & Town				-0.629***
Urban City & Town (Sparse)				-1.051***
Urban Major Conurbation				-1.982***
Urban Minor Conurbation				2.677***
_cons	78.656***	77.661***	78.672***	78.356***
r2	0.808	0.808	0.808	0.81

Table 11.7: Alternative rurality flags (see & treat)

Rurality flags (see & treat)	No rurality	Rural dummy	Sparsity dummy	ONS rurality
DistancetoAE	1.147***	1.152***	1.127***	1.133***
DistSq	-0.025***	-0.025***	-0.024***	-0.024***
Log of Population Density	-0.728***	-0.846***	-0.752***	-0.882***
Average Age	0.426***	0.421***	0.427***	0.432***
Region 1	0	0	0	0
Region 2	-2.299***	-2.344***	-2.334***	-1.312***
Region 3	-19.834***	-19.912***	-19.829***	-19.272***
Region 4	-2.684***	-2.764***	-2.713***	-2.398***
Rural		-0.623***		
Sparse			-1.728***	
Rural Town & Fringe				0
Rural Town & Fringe (Sparse)				-2.628***

Rurality flags (see & treat)	No rurality	Rural dummy	Sparsity dummy	ONS rurality
Rural Village & Dispersed				0.676***
Rural Village & Dispersed (Sparse)				-0.521***
Urban City & Town				0.556***
Urban City & Town (Sparse)				-2.166***
Urban Major Conurbation				1.854***
Urban Minor Conurbation				4.435***
Constant	45.459***	46.114***	45.578***	44.060***
r2	0.46	0.46	0.46	0.465

12 Appendix A: Descriptive statistics for G&A model

Descriptive statistics shown in sections 12.1 to 12.7 are descriptive statistics taken from the estimation sample, S1. The descriptive statistics from sections 12.8 to 12.15 are taken from the source data files.

Total Cost

Table 12.1: Total cost in estimation sample S1

	Observations	Mean	Standard deviation	Minimum	Maximum
Total Cost	8,291,908	463.9533	1959.427	0	100,000

12.1.1 Age sex groups

Table 12.2: Descriptive statistics of age sex categories in sample S1

Age	Male		Female	
Group	Observations	Mean	Observations	Mean
<1	43,583	0.53	41,193	0.5
1-4	205,646	2.48	196,782	2.37
5-9	240,763	2.9	227,818	2.75
10-14	228,253	2.75	217,306	2.62
15-19	249,614	3.01	238,421	2.88
20-24	279,412	3.37	284,746	3.43
25-29	298,240	3.6	308,334	3.72
30-34	305,067	3.68	299,300	3.61
35-39	284,690	3.43	266,252	3.21
40-44	313,070	3.78	293,763	3.54
45-49	316,809	3.82	298,640	3.6
50-54	284,786	3.43	271,497	3.27
55-59	242,220	2.92	233,814	2.82
60-64	223,287	2.69	223,737	2.7
65-69	212,963	2.57	220,633	2.66
70-74	149,578	1.8	163,907	1.98
75-79	117,587	1.42	138,414	1.67
80-84	81,018	0.98	110,436	1.33
85+	61,170	0.74	119,158	1.44

12.1.2 Morbidity Flags

Table 12.3: Descriptive statistics of morbidity flags in sample S1

Morbidity (IDC10)	Observations	Mean
A00-A09 Intestinal infectious diseases	40,419	0.0049
A15-A19 Tuberculosis	1,064	0.0001
A20-A49 Certain bacterial diseases	13,867	0.0017
A50-A64 Infections with predominantly sexual mode of transmission	61	0.0000
A65-A79 Other infectious and parasitic disorders	177	0.0000
A80-A89 Viral infections of the central nervous system	873	0.0001
A90-A99 Arthropod-borne viral fevers & viral haemorrhagic fevers	30	0.0000
B00-B09 Viral infections characterized by skin & mucous mem. lesns.	5,392	0.0007
B15-B19 Viral hepatitis	3,667	0.0004
B20-B24 Human imrnunodeficiency virus [HIV] disease	0	0.0000
B25-B34 Other viral diseases	21,983	0.0027
B35-B49 Mycoses	9,161	0.0011
B50-B64 Protozoal diseases	436	0.0001
B65-B83 Helminthiases	498	0.0001
B85-B99 Other infectious and parasitic diseases	306	0.0000
C00-C14 Malignant neoplasm of liporal cavity and pharynx	1,742	0.0002
C15-C26 Malignant neoplasm of digestive organs	12,300	0.0015
C30-C39 Malignant neoplasms of respiratory & intrathoracic organs	4,767	0.0006
C40-C41 Malignant neoplasm of bone and articular cartilage	291	0.0000
C43-C44 Malignant neoplasms of skin	26,192	0.0032
C45-C49 Malignant neoplasms of mesothelial and soft tissue	1,069	0.0001
C50 Malignant neoplasm of breast	13,487	0.0016
C51-C58 Malignant neoplasms of female genital organs	4,399	0.0005
C60-C63 Malignant neoplasms of male genital organs	14,280	0.0017
C64-C68 Malignant neoplasms of urinary tract	8,845	0.0011
C69-C72 Malignant neoplasms of eye, brain & other parts of CNS	1,257	0.0002
C73-C80, C97 Malignant neoplsm. of thyroid and oth. endo. Glands etc.	133	0.0000
C81-C96 Malignant neoplasms of lymphoid, haematopoietic & rel. tiss.	10,140	0.0012
D00-D48 In situ & benign neoplasms and others of uncertainty	85,713	0.0103
D50-D64 Anaemias	73,906	0.0089
D65-D89 Diseases of the blood and blood-forming organs	29,616	0.0036
E00-E07 Disorders of thyroid gland	64,439	0.0078
E10-E14 Diabetes Mellitus	137,237	0.0166
E15-E90 Endocrine nutritional and metabolic diseases	212,871	0.0257
F00-F03 Dementia	22,755	0.0027
F04-F09 Other organic including symptomatic mental disorders	7,222	0.0009
F10-F19 Mental and behavioural disorders due to psychoactive subst.	213,969	0.0258
F20-F29 Schizophrenia, schizotypal and delusional disorders	10,846	0.0013
F30-F39 Mood [affective] disorders	69,889	0.0084
F40-F69 Neurotic, behavioural & personality disorders	40,567	0.0049
F70-F79 Mental retardation	1,899	0.0002
F80-F99 Other mental and behavioural disorders	15,103	0.0018
G00-G09 Inflammatory diseases of the central nervous system	1,906	0.0002
G10-G13, G30-G32 Other degenerative diseases (incl. Alzheimer).	12,363	0.0015
G20-G26 Extrapyramidal & movement disorders (incl. Parkinsonism).	10,375	0.0013
G35-G37 Demyelinating diseases (incl Multiple Sclerosis) of the CNS.	5,845	0.0007
G40-G47 Epilepsymigraine & other episodic disorders	57,824	0.0070
G50-G73 G90-G99 Other diseases & disorders of the nervous syst.	52,062	0.0063
G80-G83 Cerebral palsy & other paralytic syndromes	13,275	0.0016
H00-H06, H15-H22, H30-H36, H43-H59 Other disorders of the eye etc.	72,403	0.0087
H10-H13 Disorders of conjunctiva (including conjunctivitis)	3,457	0.0004
H25-H28 Disorders of lens (including cataracts)	72,164	0.0087

Morbidity (IDC10)	Observations	Mean
H40-H42 Glaucoma	17,796	0.0021
H60-H95 Diseases of the ear and mastoid process	40,466	0.0049
I00-I09 Rheumatic heart disease	11,202	0.0014
I10-I15 Hypertensive diseases	367,087	0.0443
I20-I25 Ischaemic heart diseases	144,045	0.0174
I26-I28 Pulmonary heart disease & diseases of pulmonary circulation	11,651	0.0014
I30-I52 Other forms of heart disease	140,604	0.0170
I60-I69 Cerebrovascular diseases	37,756	0.0046
I70-I79 Diseases of arteries, arterioles & capillaries	29,566	0.0036
I80-I89 Diseases of veins & lymphatic system nec.	64,397	0.0078
I95-I99 Other & unspecified disorders of the circulatory system	20,997	0.0025
J00-J06 Acute upper respiratory infections	36,404	0.0044
J10-J18 Influenza & pneumonia	42,760	0.0052
J20-J22 Other acute lower respiratory infections	37,533	0.0045
J30-J39 Other diseases of upper respiratory tract	36,379	0.0044
J40-J47 Chronic lower respiratory diseases	211,693	0.0255
J60-J70 Lung diseases due to external agents	3,917	0.0005
J80-J99 Other diseases of the respiratory system	39,664	0.0048
K00-K14 Diseases of oral cavity, salivary glands & jaws	70,155	0.0085
K20-K31 Diseases of oesophagusstomach & duodenum	134,651	0.0162
K35-K38 Diseases of appendix	12,010	0.0014
K40-K46 Hernia	98,721	0.0119
K50-K52 Noninfective enteritis & colitis	52,761	0.0064
K55-K63 Other diseases of intestines	162,267	0.0196
K65-K67 Diseases of peritoneum	9,902	0.0012
K70-K77 Diseases of liver	16,962	0.0020
K80-K87 Disorders of gall bladder, biliary tract & pancreas	41,809	0.0050
K90-K93 Other diseases of the digestive system	43,966	0.0053
L00-L14 L55-L99 Other infections and disorders of the skin	83,600	0.0101
L20-L30 Dermatitis and eczema	15,354	0.0019
L40-L45 Papulosquamous disorders (including Psoriasis)	7,554	0.0009
L50-L54 Urticaria and erythems	3,688	0.0004
M00-M25 Arthropathies	210,840	0.0254
M30-M36 Systemic connective tissue disorders	13,221	0.0016
M40-M54 Dorsopathies	92,203	0.0111
M60-M79 Soft tissue disorders	75,950	0.0092
M80-M94 Osteopathies and chondropathies	43,334	0.0052
M95-M99 Other disorders of the musculoskeletal system & conn. tiss.	1,897	0.0002
N00-N08, N10-N16 Diseases of the kidney	26,525	0.0032
N17-N19 Renal failure	51,830	0.0063
N20-N23 Urolithiasis	16,249	0.0020
N25-N29 Other disorders of kidney & ureter	7,959	0.0010
N30-N39 Other diseases of the urinary system	93,107	0.0112
N40-N51 Diseases of male genital organs	49,964	0.0060
N60-N64 Disorders of breast	5,875	0.0007
N70-N77 Inflammatory diseases of female pelvic organs	14,864	0.0018
N80-N98 Noninflammatory disorders of female genital tract	76,945	0.0093
N99 Other disorders of the genitourinary system	1,823	0.0002
O00-O08 Pregnancy with abortive outcome	31,177	0.0038
O10-O75, O85-O92, O95-O99 Complications of labour and delivery	177,487	0.0214
O80-O84 Delivery	19,897	0.0024
P00-P04 Complications of foetus/neonate affected by maternal	9	0.0000
P05-P96 Other conditions originating in the perinatal period	749	0.0001
Q00-Q89 Congenital malformations	31,681	0.0038
Q90-Q99 Chromosomal abnormalities nec.	2,463	0.0003
R00-R09 Symptoms & signs inv. the circulatory/respiratory system	154,697	0.0187
R10-R19 Symptoms & signs inv. the digestive system & abdomen	178,642	0.0215

Morbidity (IDC10)	Observations	Mean
R20-R23 Symptoms & signs inv. the skin & subcutaneous tissue	23,800	0.0029
R25-R29 Symptoms & signs inv. the nervous & musculoskeletal sys.	34,256	0.0041
R30-R39 Symptoms & signs involving the urinary system	65,768	0.0079
R40-R46 Symptoms & signs inv. Cognition, perception etc.	44,829	0.0054
R47-R49 Symptoms & signs inv. speech & voice	7,732	0.0009
R50-R68 General symptoms & signs	13,3610	0.0161
R69 Unknown & unspecified causes of morbidity	245	0.0000
R70-R89 Abnormal findings of bodily fluids or samples without diag.	22,128	0.0027
R90-R94 Abnormal findings on diagnostic imaging/function studies	23,970	0.0029
R95-R99 Ill-defined & unknown causes of mortality	3	0.0000
S00-S09 Injuries to the head	47,927	0.0058
S10-S19 Injuries to the neck	2,713	0.0003
S20-S29 Injuries to the thorax	7,293	0.0009
S30-S39 Injuries to abdomen, lower back, lumbar spine & pelvis	11,315	0.0014
S40-S49 Injuries to the shoulder & upper arm	13,937	0.0017
S50-S59 Injuries to the elbow & forearm	21,158	0.0026
S60-S69 Injuries to the wrist & hand	24,946	0.0030
S70-S79 Injuries to the hip & thigh	18,846	0.0023
S80-S89 Injuries to the knee & lower leg	22,239	0.0027
S90-S99 Injuries to the ankle & foot	6,009	0.0007
T00-T07 Injuries involving multiple body regions	1,526	0.0002
T08-T14 Injuries to unspecified part of trunk limb or body	1,216	0.0001
T15-T19 Effects of foreign body entering through natural orifice	3,992	0.0005
T20-T32 Burns and corrosions	2,864	0.0003
T33-T35 Frostbite	12	0.0000
T36-T50 Poisonings by drugs medicaments & biological substances	24,122	0.0029
T51-T65 Tox. effcts. of substances. chiefly non-medicinal as to source	8,732	0.0011
T66-T78 Other and unspecified effects of external causes	4,722	0.0006
T79 Certain early complications of trauma	1,406	0.0002
T80-T88 Complications of surgical & medical care nec.	55,812	0.0067
T90-T98 Sequelae of injuries of poisoning & other consequences	7,594	0.0009
VVV	14,676	0.0018
WWW	115,994	0.0140
XXX	40,310	0.0049
YYY	88,421	0.0107
Z00-Z13 Examination and investigation	98,265	0.0119
Z20-Z29 Potential health hazards related to communicable diseases	16,614	0.0020
Z30-Z39 Health services in circumstances related to reproduction	192,537	0.0232
Z40-Z54 Persons encountering health services for specific care	181,933	0.0219
Z55-Z65 Potential health hazards reltd. to socioeconomic & psychosoc.l	16,867	0.0020
Z70-Z76 Persons encountering health services in other circs.	43,917	0.0053
Z80-Z99 Persons with potential health hazards related to family	525,977	0.0634
U Unclassified	4,245	0.0005

12.1.3 Co-morbidity interactions

Table 12.4: Descriptive statistics of co-morbidities in sample S1

Co-morbidities	Observations	Mean
A00B99-G00G99	10,542	0.0013
A00B99-H00H59	6,572	0.0008
A00B99-O00O99	4,175	0.0005
A00B99-Q00Q99	2,779	0.0003
A00B99-Z00Z99	43,702	0.0053
C00D48-H00H59	11,915	0.0014
C00D48-L00L99	13,934	0.0017
C00D48-N00N99	41,267	0.0050
C00D48-P00P96	18	0.0000
C00D48-Z00Z99	96,609	0.0117
D50D89-K00K93	48,253	0.0058
D50D89-O00O99	4,988	0.0006
D50D89-Z00Z99	69,442	0.0084
E00E90-G00G99	46,161	0.0056
E00E90-H00H59	45,995	0.0055
E00E90-I00I99	222,062	0.0268
E00E90-L00L99	29,564	0.0036
E00E90-O00O99	16,895	0.0020
E00E90-R00R99	140,660	0.0170
E00E90-Z00Z99	204,819	0.0247
F00F99-I00I99	111,166	0.0134
F00F99-J00J99	85,623	0.0103
F00F99-O00O99	32,789	0.0040
F00F99-R00R99	125,848	0.0152
H00H59-L00L99	8,211	0.0010
I00I99-K00K93	181,099	0.0218
I00I99-L00L99	41,368	0.0050
J00J99-O00O99	18,788	0.0023
K00K93-N00N99	76,582	0.0092
K00K93-Q00Q99	7,991	0.0010
L00L99-M00M99	27,698	0.0033
L00L99-R00R99	35,361	0.0043
M00M99-N00N99	62,883	0.0076
M00M99-O00O99	7,718	0.0009
N00N99-Q00Q99	6,963	0.0008
N00N99-S00T98	40,866	0.0049
N00N99-Z00Z99	150,021	0.0181
O00O99-R00R99	37,661	0.0045
O00O99-Z00Z99	186,910	0.0225
V01Y98-Z00Z99	121,030	0.0146

12.1.4 Morbidity counts

Table 12.5: Descriptive statistics of morbidity count in sample S1

Morbidity count	Observations	Mean
No morbidities	6,841,297	82.51
2 morbidities	62,259	0.75
3 morbidities	292,476	3.53
4 morbidities	63,888	0.77
5 morbidities	75,358	0.91
6 morbidities	177,681	2.14
7 morbidities	52,469	0.63
8 morbidities	39,919	0.48
9 morbidities	686,560	8.28

12.1.5 New GP practice, private care and log population variance

Table 12.6: Descriptive statistics of new GP practice, private care and log population variance in sample S1

	Observations	Mean	Std deviation	Minimum	Maximum
New GP practice	8,291,907	0.0867			
Private care	8,291,907	0.0014			
Log population variance	8,291,908	0.0371	0.1054	-4.382	1.593

12.1.6 CCG descriptive statistics

Table 12.7: Descriptive statistics of CCG variables in sample S1

Clinical Commissioning Group	Observations	Mean
NHS Darlington CCG	15,900	0.0019
NHS Durham Dales, Easington and Sedgefield CCG	42,696	0.0052
NHS Gateshead CCG	30,832	0.0037
NHS Newcastle North and East CCG	24,747	0.0030
NHS Newcastle West CCG	20,074	0.0024
NHS North Durham CCG	36,222	0.0044
NHS Hartlepool and Stockton-On-Tees CCG	43,611	0.0053
NHS Northumberland CCG	47,645	0.0057
NHS South Tees CCG	43,281	0.0052
NHS South Tyneside CCG	22,922	0.0028
NHS Sunderland CCG	37,217	0.0045
NHS Blackburn With Darwen CCG	25,457	0.0031
NHS Blackpool CCG	24,730	0.0030
NHS Bolton CCG	44,914	0.0054
NHS Bury CCG	29,483	0.0036
NHS Central Manchester CCG	31,053	0.0037
NHS Chorley and South Ribble CCG	24,004	0.0029
NHS Oldham CCG	33,874	0.0041
NHS East Lancashire CCG	55,096	0.0066
NHS Eastern Cheshire CCG	30,560	0.0037
NHS Heywood, Middleton and Rochdale CCG	33,687	0.0041
NHS Greater Preston CCG	31,490	0.0038
NHS Halton CCG	17,171	0.0021
NHS Salford CCG	38,265	0.0046
NHS Cumbria CCG	77,713	0.0094
NHS Knowsley CCG	24,062	0.0029
NHS Lancashire North CCG	24,092	0.0029

Clinical Commissioning Group	Observations	Mean
NHS North Manchester CCG	26,115	0.0032
NHS South Manchester CCG	24,339	0.0029
NHS South Cheshire CCG	26,412	0.0032
NHS South Sefton CCG	22,676	0.0027
NHS Southport And Formby CCG	17,629	0.0021
NHS Stockport CCG	42,102	0.0051
NHS St Helens CCG	28,068	0.0034
NHS Tameside And Glossop CCG	35,872	0.0043
NHS Trafford CCG	34,692	0.0042
NHS Vale Royal CCG	15,528	0.0019
NHS Warrington CCG	29,006	0.0035
NHS West Cheshire CCG	38,427	0.0046
NHS West Lancashire CCG	15,641	0.0019
NHS Wigan Borough CCG	48,135	0.0058
NHS Fylde and Wyre CCG	21,955	0.0026
NHS Airedale, Wharfedale and Craven CCG	23,172	0.0028
NHS Barnsley CCG	36,580	0.0044
NHS Bassetlaw CCG	16,762	0.0020
NHS Bradford Districts CCG	49,243	0.0059
NHS Calderdale CCG	27,379	0.0033
NHS Leeds North CCG	29,883	0.0036
NHS Bradford City CCG	17,986	0.0022
NHS Doncaster CCG	44,446	0.0054
NHS East Riding of Yorkshire CCG	44,829	0.0054
NHS Greater Huddersfield CCG	32,502	0.0039
NHS Leeds West CCG	53,151	0.0064
NHS Hambleton, Richmondshire and Whitby CCG	19,946	0.0024
NHS Harrogate and Rural District CCG	23,845	0.0029
NHS Hull CCG	45,060	0.0054
NHS Leeds South and East CCG	39,005	0.0047
NHS North East Lincolnshire CCG	25,471	0.0031
NHS North Kirklees CCG	28,100	0.0034
NHS North Lincolnshire CCG	25,569	0.0031
NHS Rotherham CCG	38,454	0.0046
NHS Scarborough and Ryedale CCG	17,474	0.0021
NHS Sheffield CCG	81,557	0.0098
NHS Vale of York CCG	50,803	0.0061
NHS Wakefield CCG	52,521	0.0063
NHS Lincolnshire East CCG	35,876	0.0043
NHS Corby CCG	10,521	0.0013
NHS East Leicestershire and Rutland CCG	47,124	0.0057
NHS Erewash CCG	13,692	0.0017
NHS Hardwick CCG	14,990	0.0018
NHS Leicester City CCG	54,436	0.0066
NHS Lincolnshire West CCG	32,553	0.0039
NHS Mansfield and Ashfield CCG	27,575	0.0033
NHS Milton Keynes CCG	38,795	0.0047
NHS Nene CCG	95,390	0.0115
NHS Newark and Sherwood CCG	19,157	0.0023
NHS North Derbyshire CCG	42,676	0.0051
NHS Nottingham City CCG	52,126	0.0063
NHS Nottingham North and East CCG	21,592	0.0026
NHS Nottingham West CCG	13,883	0.0017
NHS Rushcliffe CCG	17,939	0.0022
NHS South West Lincolnshire CCG	19,152	0.0023
NHS Southern Derbyshire CCG	78,731	0.0095
NHS West Leicestershire CCG	54,919	0.0066

Clinical Commissioning Group	Observations	Mean
NHS Birmingham South and Central CCG	36,212	0.0044
NHS Cannock Chase CCG	19,563	0.0024
NHS Coventry and Rugby CCG	70,880	0.0086
NHS Dudley CCG	46,326	0.0056
NHS East Staffordshire CCG	20,078	0.0024
NHS Herefordshire CCG	26,469	0.0032
NHS North Staffordshire CCG	31,493	0.0038
NHS Warwickshire North CCG	27,272	0.0033
NHS Redditch and Bromsgrove CCG	23,876	0.0029
NHS Sandwell and West Birmingham CCG	80,495	0.0097
NHS Shropshire CCG	42,971	0.0052
NHS Solihull CCG	35,407	0.0043
NHS South East Staffordshire and Seisdon Peninsula CCG	31,229	0.0038
NHS South Warwickshire CCG	39,562	0.0048
NHS South Worcestershire CCG	42,083	0.0051
NHS Stafford And Surrounds CCG	22,034	0.0027
NHS Stoke on Trent CCG	40,918	0.0049
NHS Telford and Wrekin CCG	23,882	0.0029
NHS Walsall CCG	38,846	0.0047
NHS Wolverhampton CCG	38,620	0.0047
NHS Wyre Forest CCG	16,608	0.0020
NHS Bedfordshire CCG	66,076	0.0080
NHS Cambridgeshire and Peterborough CCG	131,753	0.0159
NHS East and North Hertfordshire CCG	85,440	0.0103
NHS Ipswich and East Suffolk CCG	57,550	0.0069
NHS Great Yarmouth and Waveney CCG	34,150	0.0041
NHS Herts Valleys CCG	90,245	0.0109
NHS Luton CCG	32,533	0.0039
NHS Mid Essex CCG	56,776	0.0069
NHS North East Essex CCG	47,178	0.0057
NHS North Norfolk CCG	24,898	0.0030
NHS Norwich CCG	30,887	0.0037
NHS South Norfolk CCG	33,563	0.0041
NHS Thurrock CCG	24,287	0.0029
NHS West Essex CCG	43,434	0.0052
NHS West Norfolk CCG	24,694	0.0030
NHS West Suffolk CCG	35,394	0.0043
NHS Barking and Dagenham CCG	30,502	0.0037
NHS Barnet CCG	57,526	0.0069
NHS Bexley CCG	34,129	0.0041
NHS Brent CCG	56,436	0.0068
NHS Bromley CCG	49,324	0.0060
NHS Camden CCG	39,386	0.0048
NHS City and Hackney CCG	40,153	0.0048
NHS Croydon CCG	58,002	0.0070
NHS Ealing CCG	61,382	0.0074
NHS Enfield CCG	44,268	0.0053
NHS Hounslow CCG	43,598	0.0053
NHS Greenwich CCG	42,217	0.0051
NHS Hammersmith and Fulham CCG	30,680	0.0037
NHS Haringey CCG	41,695	0.0050
NHS Harrow CCG	37,712	0.0046
NHS Havering CCG	38,706	0.0047
NHS Hillingdon CCG	43,522	0.0053
NHS Islington CCG	34,015	0.0041
NHS Kingston CCG	29,959	0.0036
NHS Lambeth CCG	55,947	0.0068

Clinical Commissioning Group	Observations	Mean
NHS Lewisham CCG	45,185	0.0055
NHS Newham CCG	54,421	0.0066
NHS Redbridge CCG	43,101	0.0052
NHS Richmond CCG	29,984	0.0036
NHS Southwark CCG	46,122	0.0056
NHS Merton CCG	32,207	0.0039
NHS Sutton CCG	26,307	0.0032
NHS Tower Hamlets CCG	40,090	0.0048
NHS Waltham Forest CCG	43,027	0.0052
NHS Wandsworth CCG	57,331	0.0069
NHS West London CCG	38,713	0.0047
NHS Central London (Westminster) CCG	31,675	0.0038
NHS Ashford CCG	18,511	0.0022
NHS Brighton and Hove CCG	45,153	0.0054
NHS Canterbury and Coastal CCG	29,858	0.0036
NHS Eastbourne, Hailsham and Seaford CCG	27,835	0.0034
NHS Coastal West Sussex CCG	72,406	0.0087
NHS Crawley CCG	18,179	0.0022
NHS Dartford, Gravesham and Swanley CCG	37,673	0.0045
NHS East Surrey CCG	25,700	0.0031
NHS Guildford and Waverley CCG	32,220	0.0039
NHS Hastings and Rother CCG	27,021	0.0033
NHS Medway CCG	42,584	0.0051
NHS Horsham And Mid Sussex CCG	33,184	0.0040
NHS North West Surrey CCG	53,691	0.0065
NHS South Kent Coast CCG	29,737	0.0036
NHS Surrey Heath CCG	13,593	0.0016
NHS Swale CCG	14,885	0.0018
NHS Thanet CCG	20,716	0.0025
NHS Bracknell and Ascot CCG	20,330	0.0025
NHS Chiltern CCG	47,690	0.0058
NHS North Hampshire CCG	32,616	0.0039
NHS Fareham and Gosport CCG	29,532	0.0036
NHS Isle of Wight CCG	21,009	0.0025
NHS Newbury and District CCG	17,001	0.0021
NHS North and West Reading CCG	16,256	0.0020
NHS Oxfordshire CCG	103,532	0.0125
NHS Portsmouth CCG	31,076	0.0038
NHS Slough CCG	21,342	0.0026
NHS South Eastern Hampshire CCG	30,583	0.0037
NHS South Reading CCG	19,755	0.0024
NHS Southampton CCG	39,790	0.0048
NHS Aylesbury Vale CCG	30,009	0.0036
NHS West Hampshire CCG	80,734	0.0097
NHS Windsor, Ascot and Maidenhead CCG	22,365	0.0027
NHS Wokingham CCG	23,381	0.0028
NHS Bath and North East Somerset CCG	30,298	0.0037
NHS Bristol CCG	73,607	0.0089
NHS Dorset CCG	115,333	0.0139
NHS Gloucestershire CCG	92,895	0.0112
NHS Kernow CCG	83,027	0.0100
NHS North Somerset CCG	31,981	0.0039
NHS Somerset CCG	81,865	0.0099
NHS South Gloucestershire CCG	38,722	0.0047
NHS Swindon CCG	30,644	0.0037
NHS Wirral CCG	50,044	0.0060
NHS Birmingham Crosscity CCG	109,465	0.0132

Clinical Commissioning Group	Observations	Mean
NHS Liverpool CCG	71,408	0.0086
NHS North Tyneside CCG	32,119	0.0039
NHS South Lincolnshire CCG	23,517	0.0028
NHS Basildon and Brentwood CCG	39,609	0.0048
NHS Castle Point and Rochford CCG	25,515	0.0031
NHS Southend CCG	26,993	0.0033
NHS Surrey Downs CCG	44,193	0.0053
NHS West Kent CCG	69,161	0.0083
NHS High Weald Lewes Havens CCG	24,788	0.0030
NHS North East Hampshire and Farnham CCG	32,424	0.0039
NHS Wiltshire CCG	70,041	0.0085
NHS Northern, Eastern and Western Devon CCG	133,572	0.0161
NHS South Devon and Torbay CCG	42,512	0.0051

12.1.7 Attributed variables from the 2011 population census

Table 12.8: Descriptive statistics of variables from the population census (LSOA)

Observations 34,753	Mean	Std Dev	Min	Max
All Usual Residents Aged 16+	1312	259	686	8037
All Usual Residents Aged 16 to 74	1183	250	638	8148
Resident Population	1614	303	983	8300
Proportion Non White	13.26	18.39	0	99.28
Proportion not White English, Welsh, Scottish, Northern Ireland, British	18.57	21.88	0.28	99.37
Proportion Indian	2.41	5.83	0	85.5
Proportion Pakistani	1.84	6.16	0	84.97
Proportion Bangladeshi	0.74	3.19	0	90.35
Proportion Black Caribbean	1.01	2.30	0	27.86
Proportion Black African	1.66	3.62	0	48.38
Proportion Chinese	0.66	1.22	0	32.6
Proportion One Person Households	29.52	8.51	0.5	93.6
Proportion Single Pensioner Households	12.31	4.92	0	47.3
Proportion aged 16-74 never worked	0.71	0.70	0	9.4
Proportion aged 16-74 in routine occupation	11.29	5.62	0.25	37.16
Proportion aged 16-74 in semi-routine occupation	5.57	4.54	0.23	36.77
Proportion Single (never married)	34.07	11.57	8.3	97.1
Proportion Separated (but still legally married)	2.64	1.05	0.1	9.5
Proportion Divorced	9.09	2.62	0.3	23.4
Proportion Widowed	7.03	2.80	0	26.2
Rented from private landlord or letting agency	14.69	11.16	0.9	89.57
Persons in social rented housing	17.35	17.10	0	95.64
Owner occupiers (Total)	64.47	20.56	0.95	97.87
Owner occupiers (Owned Outright)	31.26	13.83	0.19	77.34
Owner occupiers (Owned with a Mortgage or Loan)	33.21	10.78	0.76	82.09
Over 16s who are separated, widowed or divorced.	18.77	4.59	0.49	42.42
Proportion (un standardised) with LLTI	18.12	5.59	2.18	48.76
Proportion (standardised) with LLTI	15.18	1.06	2.04	18.8
Proportion of working age with LLTI	5.74	3.10	0.18	26.21
Proportion (un standardised) with not good health (NGH)	5.39	2.51	0.23	23.22
Proportion (standardised) reporting both LLTI and NGH	2.05	1.06	0.049	9.12
Long-term unemployed.	1.73	1.16	0	10.22
Economically Inactive; Long-Term Sick or Disabled	4.23	2.90	0	25
Proportion aged 16+ in low grade work, long term unemployed	17.68	8.84	1.25	52.9

Observations 34,753	Mean	Std Dev	Min	Max
or never worked.				
Proportion in Approximated social grade DE	25.33	13.11	1.23	71.46
Persons aged 16-74 with no qualifications - age standardised	4.88	1.97	0.104	12.9
All usual residents aged 16 and over with no qualifications (un standardised)	23.01	9.26	0.53	60.08
Proportion of all households without either a car or a van	24.99	16.34	0.4	86.3
All people living in the area (from ONS mid-2011-lsoa-quinary-estimates(Census based))	1616	306.4	987	8159
Proportion of students in population (aged 16-74)	3.25	2.38	0.34	38.16
Student as a proportion of all people living in the area	2.43	2.09	0.26	36.77
All Schoolchildren and Full-Time Students Aged 4 and Over at their Non Term-Time Address	19.72	13.58	0	186
Proportion of residents who are in communal establishments	1.53	4.74	0	86.74
Persons aged 65 and over living alone (MSOA level)	12.38	3.44	0.78	29.93
Persons aged under 65 living alone (MSOA level)	17.47	6.58	5.79	56.15
Proportion of students living away from home (MSOA variable)	0.21	0.15	0.01	2.8
Proportion of population aged 20-24	0.07	0.05	0.01	0.807

12.1.8 Attributed variables from Department of Work and Pensions

Table 12.9: Descriptive statistics of variables from the Department of Work and Pensions

DWP variables	Observations	Mean	StdDev	Min	Max
All disability living allowance (DLA) claimants (proportion)	34,745	0.459	0.227	0.10	1.00
DLA claimants under 16 years	31,204	0.338	0.160	0.10	1.25
DLA claimants 16 to 24 years	27,008	0.421	0.181	0.10	1.00
DLA claimants 25 to 49 years	34,189	0.407	0.223	0.10	1.00
DLA claimants 50 to 59 years	33,529	0.419	0.296	0.10	25.00
DLA claimants 60 to 69 years	34,168	0.400	0.321	0.10	25.00
DLA claimants 70 years and over	32,275	0.389	0.241	0.10	5.00
Population claiming DLA over 5 years	34,738	0.393	0.208	0.10	1.00
Population claiming DLA mobility award at higher rate	34,678	0.324	0.238	0.10	1.00
All Income Support claimants	33,999	0.412	0.245	0.10	1.00
Income Support claimants lone parents	34,753	0.012	0.014	0.00	0.12
Income Support claimants carer	21,415	0.470	0.232	0.10	1.00
Income Support claimants aged 16 to 24	22,902	0.430	0.214	0.10	1.00
Income Support claimants aged 25 to 49	32,574	0.408	0.254	0.10	1.00
Income Support claimants aged 50 to 59	26,036	0.396	0.267	0.10	15.00
Income Support claimants aged over 60	6,128	0.363	0.163	0.10	5.00
Population rate income support single	33,778	0.393	0.234	0.10	1.00
Population claiming income support 2 to 5 years	26,225	0.420	0.253	0.10	1.00
Population claiming income support over 5 years	31,303	0.388	0.231	0.10	1.00
All Job Seekers Allowance claimants	34,592	0.385	0.236	0.10	1.00
Population claiming Job Seekers Allowance	34,592	0.391	0.247	0.10	1.00
Population claiming Job Seekers Allowance aged 16 to 24	29,377	0.448	0.250	0.10	1.00
Population claiming Job Seekers Allowance aged 25 to 49	33,292	0.386	0.243	0.10	1.00
Population claiming Job Seekers Allowance aged 50 to 59	27,214	0.380	0.209	0.10	5.00
All Incapacity Benefit/Severe Disablement Allowance Claimants - count persons	34,622	0.354	0.229	0.10	1.00
Incapacity Benefit/Severe Disablement Allowance Claimants - count persons aged over 60	28,196	0.341	0.237	0.10	4.29
All Pension Credit Claimants	34,730	0.420	0.213	0.10	1.00
Population claiming incapacity benefit and severe disablement allowance (age 16-64)	34,622	0.388	0.219	0.10	1.00
Population claiming incapacity benefit (age 16-64)	34,492	0.370	0.219	0.10	1.00
Population claiming SDA (age 16-64)	24,410	0.468	0.227	0.10	1.00

DWP variables	Observations	Mean	StdDev	Min	Max
Population claiming incapacity and disability 2 to 5 years	24,431	0.492	0.226	0.10	1.00
Population claiming incapacity and disability over 5 years	34,504	0.370	0.215	0.10	1.00
Population claiming incapacity and disability mental	32,301	0.371	0.243	0.10	1.00
Population claiming incapacity and disability nervous	17,859	0.505	0.150	0.11	1.00
Population claiming incapacity and disability respiratory or circulation	13,961	0.499	0.159	0.10	1.00
Population claiming incapacity and disability muscoskeletal	25,742	0.461	0.244	0.10	1.00
Population claiming incapacity and disability injury or poisoning	11,102	0.500	0.122	0.10	1.00
Population claiming incapacity and disability aged 16 to 24 year	7,491	0.309	0.111	0.10	1.00
Population claiming incapacity and disability aged 25 to 49 year	32,614	0.375	0.234	0.10	1.00
Population claiming incapacity and disability aged 50 to 59 year	32,157	0.408	0.288	0.10	25.00
Population claiming incapacity and disability aged 60 and over	28,195	0.418	0.238	0.10	6.00
Population claiming pension credit	34,749	0.042	0.025	0.00	0.26
Population claiming pension credit (prop of over 65s)	34,749	0.221	0.379	0.00	27.86
Population claiming pension credit single	34,691	0.396	0.683	0.10	82.50
Population claiming pension credit for guaranteed credit	34,272	0.412	0.538	0.10	62.50
Population claiming pension credit for savings credit	33,108	0.466	0.291	0.10	25.00
Population claiming pension credit for guaranteed and savings pension credit	34,211	0.437	0.447	0.10	65.00
Population claiming Income Support - incapacity	28,763	0.419	0.253	0.10	1.00
Proportion providing more than 19 hours unpaid care per week	34,753	0.039	0.014	0.00	0.11

12.1.9 Attributed variables from the Department for Education

Table 12.10: Descriptive statistics of variables from the Department for Education

Department for Education variables	Observations	Mean	StdDev	Min	Max
16-18 year old students entered for Level 3 qualifications	30,823	11.62	5.65	0	92
Average Level 3 QCDA point score per student	32,712	700.75	112.16	75	1581
Average Level 3 QCDA point score per entry	32,712	209.50	16.00	53	286
Students achieving 2 or more passes of A level equivalent size	14,543	98.16	7.37	40	100
All Free School Meals pupils at the end of KS4	28,083	14.32	12.29	0	107
All Free School Meals pupils at the end of KS4 (proportion)	28,067	0.16	0.12	0	1
Percentage authorised absence in all schools	32,839	4.19	0.74	2	14
Percentage unauthorised absence in all schools	32,839	1.01	0.69	0	7
Percent pupils persistently absent (imputed missing)	24,385	5.08	2.71	0	26
16-18 Year Old Students Entered for Level 3 Qualifications; Total	30,494	11.44	5.10	0	92
Average Level 3 QCDA Point Score Per Student	32,355	700.98	112.20	75	1581
Average Level 3 QCDA Point Score Per Entry; All Students	32,355	209.52	16.03	53	287
Students Achieving 2 or More Passes of A Level Equivalent Size	19,927	89.41	6.37	46	100

12.1.10 Attributed variables from Office for National Statistics

Table 12.11: Descriptive statistics of variables from the Office for National Statistics

Standardised mortality ratio (SMR)	Observations	Mean	StdDev	Min	Max
SMR for deaths from all causes, aged under 75	6,791	103.0	33.1	36.8	277.8
SMR for deaths from all cancer, aged under 75	6,791	101.9	26.0	36.1	228.2
SMR for deaths from coronary heart disease, aged under 75	6,791	105.5	52.8	0	599
SMR for deaths from circulatory diseases, aged under 75	6,791	104.9	44.4	17	362
SMR for deaths from respiratory diseases, all ages	6,791	103.5	37.2	13.8	349.4
SMR for deaths from stroke, all ages	6,791	99.5	37.2	0	380.8

12.1.11 Attributed variables from Index of Multiple Deprivation

Table 12.12: Descriptive statistics of variables from the Index of Multiple Deprivation 2015

IMD2010 Variables	Observations	Mean	StdDev	Min	Max
Index of Multiple Deprivation (IMD) Score	32,844	21.669	15.59	0.477	92.601
Income Score (rate)	32,844	0.145	0.10	0.005	0.639
Employment Score (rate)	32,844	0.119	0.08	0.003	0.58
Education, Skills and Training Score	32,844	21.691	18.53	0.006	99.502
Health Deprivation and Disability Score	32,844	0.000	0.89	-3.329	3.458
Crime Score	32,844	0.000	0.78	-3.227	3.277
Barriers to Housing and Services Score	32,844	21.691	10.57	0.439	72.586
Living Environment Score	32,844	21.691	15.94	0.151	93.352
Children and Young People Sub-domain Score	32,844	0.000	0.80	-2.861	2.906
Geographical Barriers Sub-domain Score	32,844	0.000	0.79	-2.925	3.247

IMD2010 Variables	Observations	Mean	StdDev	Min	Max
Wider Barriers Sub-domain Score	32,844	-0.002	2.41	-9.06	7.721
Indoors Sub-domain Score	32,844	0.000	0.86	-3.429	2.997
Outdoors Sub-domain Score	32,844	0.000	0.85	-3.839	3.212
Staying on in education post 16 indicator	32,844	0.185	0.07	0.038	0.575
Entry to higher education indicator	32,820	0.898	0.04	0.576	0.989
Adult skills and English language proficiency indicators - combined	32,844	0.307	0.11	0.031	0.754
Years of potential life lost indicator	32,844	61.116	14.14	25.951	258.228
Comparative illness and disability ratio indicator	32,844	112.895	41.43	37.516	373.238
Acute morbidity indicator	32,844	106.387	25.18	41.795	284.326
Mood and anxiety disorders indicator	32,844	0.000	0.79	-2.852	3.093
Road distance to a post office indicator (km)	32,844	1.157	0.73	0.13	9.715
Road distance to a primary school indicator (km)	32,844	0.893	0.55	0.051	8.421
Road distance to general store or supermarket indicator (km)	32,844	0.748	0.83	0.033	12.17
Road distance to a GP surgery indicator (km)	32,844	1.587	1.53	0.108	19.004
Household overcrowding indicator	32,844	0.086	0.09	0.007	0.705
Homelessness indicator	32,844	0.002	0.00	0	0.009
Housing affordability indicator	32,844	0.000	1.93	-6.171	6.579
Housing in poor condition indicator	32,844	0.236	0.11	0.026	0.955
Houses without central heating indicator	32,844	0.027	0.02	0.002	0.338
Air quality indicator	32,844	1.045	0.26	0.371	2.301
Road traffic accidents indicator	32,844	0.743	0.35	0.157	7.204
Income Deprivation Affecting Children Index (IDACI) Score (rate)	32,844	0.185	0.13	0.004	0.916
Income Deprivation Affecting Older People Index (IDAOPI) Score (rate)	32,844	0.186	0.13	0.007	0.98

12.1.12 Attributed variables from Quality and Outcomes Framework

Table 12.13: Descriptive statistics of variables from the Quality and Outcomes Framework

2012-13 QOF Measures	Observations	Mean	StdDev	Min	Max
Atrial Fibrillation Weighted Achievement Score	8,020	98.066	8.602	0	100
Atrial Fibrillation Unweighted Achievement Score	7,824	84.411	5.624	36	100
Atrial Fibrillation Prevalence	8,020	1.458	0.754	0	23
Atrial Fibrillation Exception Rate	7,824	8.770	5.532	0	64
Atrial Fibrillation Total Exceptions	8,009	14.990	14.407	0	153
Asthma Weighted Achievement Score	8,020	96.882	10.905	0	100
Asthma Unweighted Achievement Score	7,963	80.605	7.253	20	100
Asthma Prevalence	8,020	5.955	1.376	0	15
Asthma Exception Rate	8,005	5.888	7.084	0	100
Asthma Total Exceptions	8,009	35.007	57.379	0	695
Hypertension Weighted Achievement Score	8,020	97.451	5.886	0	100
Hypertension Unweighted Achievement Score	8,006	39.708	8.387	0	450
Hypertension Prevalence	8,020	13.813	3.829	0	61
Hypertension Exception Rate	8,006	2.808	2.919	0	100
Hypertension Total Exceptions	8,009	53.475	71.413	0	1873
Cancer Weighted Achievement Score	8,020	96.785	10.448	0	100
Cancer Unweighted Achievement Score	7,955	91.678	13.691	0	100
Cancer Prevalence	8,020	1.872	0.759	0	15
Cancer Exception Rate	7,955	1.691	5.708	0	100
Cancer Total Exceptions	8,009	0.315	0.966	0	31
CHD Weighted Achievement Score	8,020	94.140	16.558	0	100
CHD Unweighted Achievement Score	7,994	74.950	12.223	0	100
CHD Prevalence	8,020	2.322	0.949	0	10

2012-13 QOF Measures	Observations	Mean	StdDev	Min	Max
CHD Exception Rate	7,994	9.712	7.312	0	86
CHD Total Exceptions	8,009	12.083	11.303	0	126
KD Weighted Achievement Score	8,020	87.795	6.942	0	90
KD Unweighted Achievement Score	7,915	81.860	6.116	0	100
KD Prevalence	8,020	3.311	1.838	0	43
KD Exception Rate	7,915	5.905	5.039	0	100
KD Total Exceptions	8,009	26.371	38.863	0	694
COPD Weighted Achievement Score	8,020	97.520	8.716	0	100
COPD Unweighted Achievement Score	7,963	79.857	8.557	4	100
COPD Prevalence	8,020	1.795	0.931	0	20
COPD Exception Rate	7,963	12.159	7.047	0	70
COPD Total Exceptions	8,009	45.083	44.130	0	532
Dementia Weighted Achievement Score	8,020	89.684	16.503	0	100
Dementia Unweighted Achievement Score	7,728	53.871	12.150	0	100
Dementia Prevalence	8,020	0.569	0.997	0	58
Dementia Exception Rate	7,728	18.843	13.732	0	100
Dementia Total Exceptions	8,009	6.144	7.111	0	135
Diabetes Weighted Achievement Score	8,020	96.287	6.473	0	100
Diabetes Unweighted Achievement Score	7,974	81.567	4.296	41	94
Diabetes Prevalence	8,020	4.969	1.406	0	22
Diabetes Exception Rate	5,579	10.633	4.892	1	67
Diabetes Total Exceptions	8,009	267.330	230.752	0	2498
Epilepsy Weighted Achievement Score	8,020	94.596	12.207	0	100
Epilepsy Unweighted Achievement Score	7,901	71.399	12.574	0	100
Epilepsy Prevalence	8,020	0.618	0.234	0	5
Epilepsy Exception Rate	7,901	18.873	11.713	0	67
Epilepsy Total Exceptions	8,009	13.610	14.377	0	162
Heart Failure Weighted Achievement Score	8,020	98.836	7.218	0	100
Heart Failure Unweighted Achievement Score	7,950	63.149	21.758	3	267
Heart Failure Prevalence	8,020	0.715	0.374	0	11
Heart Failure Exception Rate	7,950	13.673	7.591	0	70
Heart Failure Total Exceptions	8,009	9.412	8.977	0	120
Learning Disabilities Weighted Achievement Score	8,020	82.170	24.358	0	100
Learning Disabilities Unweighted Achievement Score	5,602	85.496	27.025	0	100
Learning Disabilities Prevalence	8,020	0.377	0.257	0	5
Learning Disabilities Exception Rate	5,602	0.093	0.224	0	1
Learning Disabilities Total Exceptions	8,009	0.171	0.483	0	8
Mental Health Weighted Achievement Score	8,020	92.558	11.562	0	100
Mental Health Unweighted Achievement Score	7,433	76.517	7.583	23	100
Mental Health Prevalence	8,020	0.889	0.588	0	18
Mental Health Exception Rate	7,433	15.007	6.340	0	50
Mental Health Total Exceptions	8,009	50.279	46.268	0	514
Obesity Weighted Achievement Score	8,020	99.963	1.934	0	100
Obesity Prevalence	8,020	9.123	3.074	0	36
Palliative Care Weighted Achievement Score	8,020	93.791	19.505	0	100
Palliative Care Prevalence	8,020	0.241	0.439	0	32
Stroke Weighted Achievement Score	8,020	97.898	6.650	0	100
Stroke Unweighted Achievement Score	7,963	83.119	5.673	0	100
Stroke Prevalence	8,020	1.667	0.790	0	24
Stroke Exception Rate	7,963	5.881	3.975	0	54
Stroke Total Exceptions	8,009	29.704	31.372	0	338
Thyroid Weighted Achievement Score	8,020	99.771	2.851	0	100
Thyroid Unweighted Achievement Score	8,004	95.142	3.295	60	100
Thyroid Prevalence	8,020	3.173	1.074	0	15
Thyroid Exception Rate	8,004	0.493	1.066	0	40
Thyroid Total Exceptions	8,009	1.006	1.902	0	39

12.1.13 Remaining attributed need variables at GP level

Table 12.14: Descriptive statistics of remaining attributed needs variables at GP Level

Attributed needs variables	Observations	Mean	StdDev	Min	Max
No FTE GPs per practice (excluding retainers and registrars)	8,063	3.898	2.556	0.1	28.5
Proportion headcount GPs female (including retainers and registrars)	8,063	0.437	0.269	0	1
Proportion FTE GPs in practice who are not providers	8,063	0.272	0.282	0	1
Proportion of GPs aged 50 and over in practice (headcount, including retainers and registrars)	8,051	0.441	0.311	0	1
Number of registrations per FTE GP (excluding retainers and registrars)	8,063	1964.5	1234.9	0	46943
Proportion GPs qualified in UK	8,063	0.679	0.356	0	1
% carer	8,056	18.427	5.081	0	57
% seen GP in last 3 months	8,059	55.417	6.865	0	94
% Long term health condition	8,059	53.519	7.861	8	100
% Not British, Irish or other white	8,129	16.747	23.181	0	100
% permanently sick or disabled	8,053	4.927	3.636	0	51.4
% unemployed	8,053	6.592	5.659	0	70.6
% Full-time education	8,053	3.467	5.219	0	90.7
% long term mental health condition	8,052	4.610	3.261	0	52.5
% with (long term) medical condition	8,052	56.968	8.297	10	100
Average with (long term) medical condition for those with at least one	8,052	1.768	0.205	1	3.94

12.1.14 Remaining attributed supply variables at GP Level

Table 12.15: Descriptive statistics of remaining attributed supply variables at GP level

Attributed supply variables	Observations	Mean	StdDev	Min	Max
2012-13 by 1st birthday Diphtheria, Tetanus, Polio, Pertussis, Hib(DTaP/IPV/Hib) %	9,298	94.5	3.56	79	99
2012-13 Pneumococcal disease (PCV) %	9,298	93.6	3.79	75.9	98.8
% immunised Meningitis C 2012-13	9,298	94.2	3.55	78.7	99
Percentage of persons aged 65 and over immunised against Influenza (seasonal flu) by Primary Care Organisation, 2012-13	9,298	73.1	2.68	65.5	80.8
2012-13 by 2nd birthday Diphtheria, Tetanus, Polio, Pertussis, Hib(DTaP/IPV/Hib) %	9,298	96.1	2.69	81.9	99.4
% receiving MMR 1st dose by second birthday 2013-14	9,298	94.8	3.54	76.8	99.5

13 Appendix B: Variables removed in T-Selection method for G&A model

The T-selection method used to select the attributed variables to include in the final model is described in section 6.5.

The tables in this appendix list the variables excluded at each step of the T-selection procedure. They include variables omitted due to having an incorrect sign on the coefficients.

Table 13.1: Variables removed with t-statistics less than 0.2

Variables
All Free School Meals pupils at the end of KS4 (proportion)
16-18 Year Old Students Entered for Level 3 Qualifications; Total
Students Achieving 2 or More Passes of A Level Equivalent Size
Standardised mortality ratio for deaths from coronary heart disease aged under 75, 2008-12
General & acute day beds
Number of day operating theatres
Number of adult critical beds April 2013
Independent provider of NHS funded imaging (1 if independent, 0 if not)
Proportion of those answering the question saying seen their GP in last 3 months
Proportion of those answering the question saying they are unemployed
Proportion of those answering the question saying they in full-time education
Proportion of those answering the question saying they have a long term mental health condition
2012-13 QOF Asthma Weighted Achievement Score
2012-13 QOF Hypertension Total Exceptions
2012-13 QOF Cancer Exception Rate
2012-13 QOF Cancer Total Exceptions
2012-13 QOF Dementia Exception Rate
2012-13 QOF Dementia Total Exceptions
2012-13 QOF Diabetes Weighted Achievement Score
2012-13 QOF Diabetes Total Exceptions
2012-13 QOF Epilepsy Weighted Achievement Score
2012-13 QOF Epilepsy Un-weighted Achievement Score
2012-13 QOF Mental Health Un-weighted Achievement Score
2012-13 QOF Mental Health Total Exceptions
Persons in social rented housing
2012-13 Median waiting times (weeks)of the 95th percentile for Cardiothoracic Surgery Patients
2012-13 Median waiting times (weeks)of the 95th percentile for Neurology Patients
2012-13 Median waiting times (weeks)of the 95th percentile for Plastic Surgery Patients
Persons aged 65 and over living alone (MSOA level)
Persons aged under 65 living alone (MSOA level)
2012-13 Pneumococcal disease (PCV) %
Income Support claimants lone parents
Income Support claimants aged over 60
Population claiming income support over 5 years
Children and Young People Sub-domain Score
Adult skills and English language proficiency indicators - combined
Road distance to a GP surgery indicator (km)
Homelessness indicator
Population claiming incapacity and disability respiratory or circulation
Population claiming incapacity and disability muscloskeletal
Population claiming incapacity and disability aged 16 to 24 year
Population claiming incapacity and disability aged 25 to 49 year
Population claiming pension credit

Variables
Population claiming pension credit for savings credit
Proportion providing more than 19 hours unpaid care per week

Table 13.2: Variables removed with t-statistics less than 0.4

Variables
Percentage of pupils persistently absent (imputed missing)
Proportion One Person Households
Proportion of GPs aged 50 and over in practice (headcount, including retainers and registrars)
2012-13 QOF Atrial Fibrillation Prevalence
2012-13 QOF Asthma Un-weighted Achievement Score
2012-13 QOF CHD Un-weighted Achievement Score
2012-13 QOF CHD Exception Rate
2012-13 QOF COPD Exception Rate
2012-13 QOF Learning Disabilities Un-weighted Achievement Score
2012-13 QOF Obesity Weighted Achievement Score
2012-13 QOF Obesity Prevalence
Owner occupiers (Total)
2011-12 Proportion of admitted patients RTT under 18 weeks
2012-13 Median waiting times (weeks) for Cardiothoracic Surgery Patients
Owner occupiers (Owned Outright)
Owner occupiers (Owned with a Mortgage or Loan)
2012-13 Median waiting times (weeks)of the 95th percentile for Gynaecology Patients
2012-13 Median waiting times (weeks)of the 95th percentile for Urology Patients
Proportion of working age with LLTI
Proportion of all households without either a car or a van
All disability living allowance (DLA) claimants (proportion)
Men C 2012-13 %
Population claiming DLA mobility award at higher rate
All Income Support claimants
Population claiming Job Seekers Allowance aged 50 to 59
Employment Score (rate)
Outdoors Sub-domain Score
Road distance to a primary school indicator (km)
Road traffic accidents indicator
Population claiming incapacity and disability 2 to 5 years
Population claiming incapacity and disability injury or poisoning
Population claiming pension credit for guaranteed and savings pension credit
Average Level 3 QCDA point score per entry

Table 13.3: Variables removed with t-statistics less than 0.6

Variables
Standardised mortality ratio for deaths from coronary heart disease aged under 75, 2008-12
Proportion of those answering the question saying they have a long standing health condition
Proportion of those answering the question saying they have a (long term) medical condition
People aged 16-74 never worked
2012-13 QOF Atrial Fibrillation Un-weighted Achievement Score
2012-13 QOF Cancer Un-weighted Achievement Score
Proportion in semi-routine occupations aged 16-74
2012-13 QOF Dementia Weighted Achievement Score
Proportion of students living away from home (MSOA variable)
2012-13 QOF Heart Failure Exception Rate
2012-13 QOF Learning Disabilities Weighted Achievement Score
2012-13 Median waiting times (weeks) for non-admitted patients
2012-13 Median waiting times (weeks) for Plastic Surgery Patients
2012-13 Median waiting times (weeks) for Trauma Orthopaedics Patients

Variables
Proportion (un standardised) with LLTI
Proportion of students in population (aged 16-74)
by 2nd birthday
Proportion Non White
Index of Multiple Deprivation (IMD) Score
Income Score (rate)
Geographical Barriers Sub-domain Score
Population claiming incapacity benefit
Population claiming incapacity and disability mental
Population claiming incapacity and disability aged 60 and over
Population claiming pension credit (prop of over 65s)

Table 13.4: Variables removed with t-statistics less than 0.8

Variables
Proportion of population aged 20-24
2012-13 QOF Atrial Fibrillation Weighted Achievement Score
2012-13 QOF KD Weighted Achievement Score
2012-13 QOF KD Exception Rate
2012-13 QOF COPD Weighted Achievement Score
2012-13 QOF COPD Total Exceptions
2012-13 QOF Learning Disabilities Total Exceptions
2012-13 QOF Mental Health Weighted Achievement Score
2012-13 QOF Mental Health Exception Rate
2012-13 QOF Stroke Exception Rate
2012-13 QOF Stroke Total Exceptions
2012-13 QOF Thyroid Total Exceptions
2012-13 Median waiting times (weeks) for General Medicine Patients
2012-13 Median waiting times (weeks) for General Surgery Patients
2012-13 Median waiting times (weeks)of the 95th percentile for Dermatology Patients
2012-13 Median waiting times (weeks)of the 95th percentile for Gastroenterology Patients
2012-13 Median waiting times (weeks)of the 95th percentile for General Surgery Patients
Student as a proportion of all people living in the area
DLA claimants 16 to 24 years
DLA claimants 60 to 69 years
Percentage of persons aged 65 and over immunised against Influenza (seasonal flu) by Primary Care Organisation, 2012-13
Income Support claimants aged 16 to 24
Proportion (standardised) with LLTI
Driving time
Road distance
All Pension Credit Claimants
Entry to higher education indicator
Population claiming pension credit for guaranteed credit
Population claiming Income Support - incapacity
Average Level 3 QCDA point score per student
All Free School Meals pupils at the end of KS4

Table 13.5: Variables removed with t-statistics less than 1.0

Variables
Percentage authorised absence in all schools
Average Level 3 QCDA Point Score Per Student
Average Level 3 QCDA Point Score Per Entry; All Students
Proportion Chinese
Radio-graphs no fluoroscopy

Variables
Proportion GPs qualified in UK
2012-13 QOF Atrial Fibrillation Total Exceptions
2012-13 QOF Cancer Weighted Achievement Score
2012-13 QOF Heart Failure Un-weighted Achievement Score
2012-13 QOF Palliative Care Prevalence
2012-13 QOF Stroke Weighted Achievement Score
2012-13 QOF Stroke Prevalence
2012-13 QOF Thyroid Weighted Achievement Score
2012-13 Median waiting times (weeks) of the 95th percentile for non-admitted patients
2012-13 Median waiting times (weeks)of the 95th percentile for ENT Patients
2012-13 Median waiting times (weeks)of the 95th percentile for General Medicine Patients
2012-13 Median waiting times (weeks)of the 95th percentile for Oral Surgery Patients
2012-13 Median waiting times (weeks)of the 95th percentile for Thoracic Medicine Patients
All usual residents aged 16 and over with no qualifications (un standardised)
DLA claimants under 16 years
Persons aged 16-74 with no qualifications - age standardised
Direct Distance
Population claiming income support 2 to 5 years
Incapacity Benefit/Severe Disablement Allowance Claimants - count persons aged over 60
Acute morbidity indicator
Population claiming incapacity benefit and severe disablement allowance (age 16-64)
Population claiming SDA (age 16-64)
Population claiming incapacity and disability nervous
Population claiming pension credit single

Table 13.6: Variables removed with t-statistics less than 1.2

Variables
Percentage unauthorised absence in all schools
Proportion Black African
Number of operating theatres
MRI scans 2012-13
Non obstetric ultra-sound
Proportion of those answering the question saying they have a carer responsibility
Proportion FTE GPs in practice who are not providers
Number of registrations per FTE GP (excluding retainers and registrars)
2012-13 QOF Hypertension Weighted Achievement Score
2012-13 QOF Cancer Prevalence
2012-13 QOF CHD Weighted Achievement Score
2012-13 QOF CHD Total Exceptions
2012-13 QOF KD Un-weighted Achievement Score
2012-13 QOF Diabetes Un-weighted Achievement Score
2012-13 QOF Learning Disabilities Exception Rate
2012-13 QOF Palliative Care Weighted Achievement Score
2012-13 QOF Thyroid Un-weighted Achievement Score
2012-13 QOF Thyroid Exception Rate
DLA claimants 50 to 59 years
Proportion (standardised) reporting both LLTI and NGH
Population claiming Job Seekers Allowance aged 16 to 24
Population claiming Job Seekers Allowance aged 25 to 49
Wider Barriers Sub-domain Score
Indoors Sub-domain Score
Comparative illness and disability ratio indicator
Road distance to a post office indicator (km)
Population claiming incapacity benefit (age 16-64)
Population claiming incapacity and disability over 5 years

Table 13.7: Variables removed with t-statistics less than 1.4

Variables
Proportion Bangladeshi
Standardised mortality ratio for deaths from all cancer aged under 75, 2008-12
Standardised mortality ratio for deaths from respiratory disease, all ages, 2008-12
Fluoroscopy
Proportion of those answering the question saying they are permanently sick or disabled
2012-13 QOF Atrial Fibrillation Exception Rate
2012-13 QOF Asthma Exception Rate
2012-13 QOF Hypertension Exception Rate
2012-13 QOF Dementia Un-weighted Achievement Score
2012-13 QOF Diabetes Exception Rate
2012-13 QOF Heart Failure Prevalence
Proportion Widowed
2012-13 Median waiting times (weeks) for Cardiology patients
2012-13 Median waiting times (weeks) for Gastroenterology Patients
2012-13 Median waiting times (weeks) for Gynaecology Patients
2012-13 Median waiting times (weeks) for Urology Patients
2012-13 Median waiting times (weeks)of the 95th percentile for Cardiology patients
2012-13 Median waiting times (weeks)of the 95th percentile for Other Patients
Economically Inactive; Long-Term Sick or Disabled
Proportion in Approximated social grade DE
DLA claimants 70 years and over
Population claiming DLA over 5 years
Income Support claimants carer
Income Support claimants aged 25 to 49
All Incapacity Benefit/Severe Disablement Allowance Claimants - count persons
Household overcrowding indicator
Housing in poor condition indicator
Students achieving 2 or more passes of A level equivalent size

Table 13.8: Variables removed with t-statistics less than 1.6

Variables
Standardised mortality ratio for deaths from all causes aged under 75, 2008-12
2012-13 QOF Asthma Prevalence
2012-13 QOF Asthma Total Exceptions
2012-13 QOF Hypertension Un-weighted Achievement Score
Proportion Separated (but still legally married)
2012-13 QOF Epilepsy Total Exceptions
2012-13 QOF Heart Failure Weighted Achievement Score
2012-13 Median waiting times (weeks) for Oral Surgery Patients
2012-13 Median waiting times (weeks)of the 95th percentile for Trauma Orthopaedics Patients
Long-term unemployed.
DLA claimants 25 to 49 years
Income Support claimants aged 50 to 59
Population rate income support single
Road distance to general store or supermarket indicator (km)
Houses without central heating indicator
Air quality indicator
Proportion Indian
Population claiming incapacity and disability aged 50 to 59 years

Table 13.9: Variables removed with t-statistics less than 1.8

Variables
Proportion Black Caribbean
2012-13 QOF COPD Prevalence
2012-13 QOF Epilepsy Exception Rate
2012-13 QOF Learning Disabilities Prevalence
2012-13 QOF Stroke Un-weighted Achievement Score
2011-12 Proportion of non-admitted patients RTT under 18 weeks
2012-13 by 1st birthday Diphtheria Tetanus, Polio, Pertussis, Hib(DTaP/IPV/Hib) %
All Job Seekers Allowance claimants
Barriers to Housing and Services Score
Living Environment Score
Staying on in education post 16 indicator
Housing affordability indicator

Table 13.10: Variables removed with t-statistics less than 2.0

Variables
2012-13 QOF Hypertension Prevalence
2012-13 QOF KD Prevalence
2012-13 QOF COPD Un-weighted Achievement Score
All Schoolchildren and Full-Time Students Aged 4 and Over at their Non Term-Time Address
by 2nd birthday
Education, Skills and Training Score
Crime Score

Table 13.11: Variables removed with t-statistics less than 2.0 and coefficients exhibit incorrect signs, 1

Variables	T<2	Incorrect sign
Proportion not White English, Welsh, Scottish, Northern Ireland, British	0	1
Over 16s, who are separated, widowed or divorced.	1	1
Proportion aged 16+ in low grade work, long term unemployed or never worked.	0	1
Proportion of residents who are in communal establishments	0	1
Population claiming Job Seekers Allowance	0	1
16-18 year old students entered for Level 3 qualifications	0	1
Standardised mortality ratio for deaths from stroke all ages	0	1
General & acute beds	0	1
CT scans 2012-13	0	1
Obstetric ultra-sound	0	1
2012-13 Median waiting times (weeks) for admitted patients	0	1
2012-13 Median waiting times (weeks) for Geriatric Medicine Patients	0	1
2012-13 Median waiting times (weeks) for Neurology Patients	0	1
2012-13 Median waiting times (weeks) for Neurosurgery Patients	0	1
2012-13 Median waiting times (weeks) for Thoracic Medicine Patients	0	1
2012-13 Median waiting times (weeks)of the 95th percentile for Ophthalmology Patients	0	1
2012-13 Median waiting times (weeks)of the 95th percentile for Rheumatology Patients	0	1
IMD 2015 Income deprivation of elderly	0	1
Years of potential life lost indicator	0	1
Mood and anxiety disorders indicator	0	1
% Not British, Irish or other white	1	0
No FTE GPs per practice	0	1
Proportion headcount GPs female	0	1
2012-13 QOF CHD Prevalence	1	0
2012-13 QOF Dementia Prevalence	0	1

Variables	T<2	Incorrect sign
2012-13 QOF Diabetes Prevalence	1	0
2012-13 QOF Heart Failure Total Exceptions	0	1
2012-13 QOF Thyroid Prevalence	1	0

Table 13.12: Variables removed with t-statistics less than 2.0 and coefficients exhibit incorrect signs, 2

Variables	T<2	Incorrect sign
Proportion Pakistani	1	0
Radio-isotopes	1	0
2012-13 Median waiting times (weeks) of the 95th percentile for admitted patients	0	1
2012-13 Median waiting times (weeks) for ENT Patients	1	1
2012-13 Median waiting times (weeks) for Ophthalmology Patients	1	1
2012-13 Median waiting times (weeks) for Other Patients	0	0
2012-13 Median waiting times (weeks) for Rheumatology Patients	1	1
2012-13 Median waiting times (weeks)of the 95th percentile for Geriatric Medicine Patients	1	1
Income Deprivation Affecting Children Index (IDACI) Score (rate)	1	0
Adult Skills Sub-domain Score	0	0

After this step, of removing these variables and re-estimating the model, no variables were found to have an incorrect sign, or a t-statistic of less than 2.58. The remaining attributed needs and supply variables from t-statistic selection process are:

Table 13.13: Attributed variables from t-statistic selection criteria

Variables
All Usual Residents Aged 16+
All Usual Residents Aged 16 to 74
Resident Population
Proportion Single Pensioner Households
Proportion in routine occupations aged 16-74
Proportion Single (never married)
Proportion Divorced
Rented from private landlord or letting agency
Proportion (un standardised) with not good health (NGH)
All people living in the area (from ONS mid-2011-lsoa-quinary-estimates(Census based))
Adult critical beds Jan 13
2012-13 Median waiting times (weeks) for Dermatology Patients
2012-13 Median waiting times (weeks)of the 95th percentile for Neurosurgery Patients
Health Deprivation and Disability Score
Average with (long term) medical condition for those with at least one
2012-13 QOF KD Total Exceptions
2012-13 QOF Epilepsy Prevalence
2012-13 QOF Mental Health Prevalence

14 Appendix C: Coefficients from age stratified T-Statistic model for G&A model

Table 14.1: All coefficients from the age stratified T-statistic model

	Age Group 0-14		Age Group 15-64		Age Group 65+	
	Coefficient	T-Statistic	Coefficient	T-Statistic	Coefficient	T-Statistic
Males						
<1	.	.				
1-4	-170.5	(-23.344)				
5-9	-212.2	(-29.038)				
10-14	-193.2	(-26.394)				
15-19			.	.		
20-24			-4.153	(-1.722)		
25-29			-13.69	(-5.517)		
30-34			-8.775	(-3.353)		
35-39			4.831	(1.753)		
40-44			35.38	(11.484)		
45-49			76.45	(23.298)		
50-54			121.3	(32.809)		
55-59			197.7	(41.826)		
60-64			287.7	(51.499)		
65-69					.	.
70-74					173.1	(16.336)
75-79					359.8	(28.979)
80-84					565.2	(36.060)
85+					813.9	(44.474)
Females						
<1	-84.39	(-9.700)				
1-4	-209.2	(-28.837)				
5-9	-233.3	(-32.265)				
10-14	-206.8	(-28.405)				
15-19			16.89	(6.770)		
20-24			10.16	(4.283)		
25-29			21.14	(8.672)		
30-34			41.39	(15.471)		
35-39			66.32	(23.021)		
40-44			98.88	(31.900)		

	Age Group 0-14		Age Group 15-64		Age Group 65+	
	Coefficient	T-Statistic	Coefficient	T-Statistic	Coefficient	T-Statistic
45-49			139.3	(41.205)		
50-54			175.0	(44.745)		
55-59			205.9	(47.554)		
60-64			270.7	(53.106)		
65-69					-39.60	(-4.821)
70-74					106.2	(11.221)
75-79					243.7	(23.058)
80-84					427.8	(34.815)
85+					593.6	(46.087)
Morbidity flags						
A00-A09 Intestinal infectious diseases	144.3	(6.002)	292.8	(6.151)	217.7	(3.183)
A15-A19 Tuberculosis	702.7	(0.729)	150.9	(0.706)	1193.4	(1.723)
A20-A49 Certain bacterial diseases	239.4	(1.615)	695.9	(4.682)	388.7	(3.680)
A50-A64 Infections with predominantly sexual mode of transmission	-4.412	(-0.019)	427.2	(1.543)	-3982.9	(-15.909)
A65-A79 Other infectious and parasitic disorders	92.26	(0.542)	179.8	(0.480)	-186.2	(-0.444)
A80-A89 Viral infections of the central nervous system	56.22	(0.178)	-4.570	(-0.041)	324.3	(0.542)
A90-A99 Arthropod-borne viral fevers & viral haemorrhagic fevers	-196.2	(-2.130)	-40.24	(-0.298)	-1212.2	(-1.319)
B00-B09 Viral infections characterized by skin & mucous mem. lesns.	105.2	(1.413)	226.4	(1.457)	65.98	(0.362)
B15-B19 Viral hepatitis	-341.9	(-1.076)	711.7	(5.496)	1367.2	(2.481)
B20-B24 Human imrnunodeficiency virus [HIV] disease	.	.	.	.	.	.
B25-B34 Other viral diseases	141.4	(6.129)	207.4	(2.600)	137.5	(0.694)
B35-B49 Mycoses	394.3	(1.866)	423.7	(3.498)	114.1	(1.038)
B50-B64 Protozoal diseases	-162.1	(-0.759)	-435.4	(-2.679)	-910.5	(-1.362)
B65-B83 Helminthiases	336.9	(2.097)	-47.77	(-0.379)	73.53	(0.119)
B85-B99 Other infectious and parasitic diseases	-246.2	(-1.429)	-55.16	(-0.158)	-433.7	(-1.005)
C00-C14 Malignant neoplasm of liporal cavity and pharynx	-471.6	(-0.338)	518.1	(3.104)	471.4	(2.371)
C15-C26 Malignant neoplasm of digestive organs	-254.8	(-0.239)	1991.9	(14.608)	880.2	(11.483)
C30-C39 Malignant neoplasms of respiratory & intrathoracic organs	-379.2	(-0.268)	868.5	(5.715)	583.2	(5.306)
C40-C41 Malignant neoplasm of bone and articular cartilage	-524.1	(-0.473)	-788.3	(-2.356)	190.9	(0.398)
C43-C44 Malignant neoplasms of skin	-178.6	(-0.563)	227.0	(5.519)	125.9	(3.631)
C45-C49 Malignant neoplasms of mesothelial and soft tissue	2152.9	(1.301)	491.4	(1.456)	363.6	(1.934)
C50 Malignant neoplasm of breast	.	.	833.9	(12.480)	281.4	(4.319)
C51-C58 Malignant neoplasms of female genital organs	-59.26	(-0.056)	891.2	(7.530)	682.1	(6.308)
C60-C63 Malignant neoplasms of male genital organs	64.11	(0.027)	343.1	(4.084)	376.4	(6.793)
C64-C68 Malignant neoplasms of urinary tract	-340.5	(-0.447)	211.5	(1.543)	321.5	(3.304)

	Age Group 0-14		Age Group 15-64		Age Group 65+	
	Coefficient	T-Statistic	Coefficient	T-Statistic	Coefficient	T-Statistic
C69-C72 Malignant neoplasms of eye, brain & other parts of CNS	8.191	(0.008)	-227.2	(-1.175)	-626.2	(-3.083)
C73-C80, C97 Malignant neoplsm. of thyroid and oth. endo. Glands etc.	.	.	1469.0	(1.586)	386.0	(0.472)
C81-C96 Malignant neoplasms of lymphoid, haematopoietic & rel. tiss.	2127.9	(3.682)	4188.9	(17.272)	3284.7	(21.662)
D00-D48 In situ & benign neoplasms and others of uncertainty	145.0	(1.802)	182.7	(8.605)	389.4	(11.508)
D50-D64 Anaemias	718.8	(5.049)	430.1	(8.721)	471.6	(8.985)
D65-D89 Diseases of the blood and blood-forming organs	-625.3	(-0.698)	229.2	(1.994)	162.8	(1.875)
E00-E07 Disorders of thyroid gland	342.1	(1.359)	118.8	(4.591)	129.2	(4.277)
E10-E14 Diabetes Mellitus	1204.1	(12.590)	583.0	(22.414)	567.7	(21.422)
E15-E90 Endocrine nutritional and metabolic diseases	271.1	(5.400)	289.9	(13.482)	135.5	(5.055)
F00-F03 Dementia	.	.	127.2	(0.595)	-338.1	(-7.853)
F04-F09 Other organic including symptomatic mental disorders	-220.9	(-1.355)	312.8	(1.297)	59.84	(0.763)
F10-F19 Mental and behavioural disorders due to psychoactive subst.	65.09	(0.722)	203.0	(17.487)	302.8	(8.551)
F20-F29 Schizophrenia, schizotypal and delusional disorders	-102.1	(-0.427)	132.9	(2.965)	215.2	(2.261)
F30-F39 Mood [affective] disorders	345.5	(1.693)	232.2	(13.059)	208.0	(4.922)
F40-F69 Neurotic, behavioural & personality disorders	121.7	(1.434)	215.8	(9.283)	293.9	(5.403)
F70-F79 Mental retardation	1265.2	(3.266)	994.3	(6.841)	229.0	(0.749)
F80-F99 Other mental and behavioural disorders	430.0	(7.901)	439.2	(9.027)	87.85	(0.653)
G00-G09 Inflammatory diseases of the central nervous system	-99.92	(-0.549)	-110.3	(-0.678)	-470.0	(-1.884)
G10-G13, G30-G32 Other degenerative diseases (incl. Alzheimer).	1282.1	(2.225)	535.2	(3.605)	72.67	(1.407)
G20-G26 Extrapyramidal & movement disorders (incl. Parkinsonism).	394.2	(1.031)	776.0	(6.379)	855.2	(13.724)
G35-G37 Demyelinating diseases (incl Multiple Sclerosis) of the CNS.	85.19	(0.158)	1181.2	(15.835)	997.2	(6.668)
G40-G47 Epilepsy, migraine & other episodic disorders	449.7	(9.710)	377.6	(16.373)	344.4	(8.229)
G50-G73 G90-G99 Other diseases & disorders of the nervous syst.	299.1	(2.559)	402.5	(16.548)	380.5	(8.788)
G80-G83 Cerebral palsy & other paralytic syndromes	972.1	(9.439)	432.4	(6.292)	231.7	(2.847)
H00-H06, H15-H22, H30-H36, H43-H59 Other disorders of the eye etc.	248.8	(6.707)	202.2	(9.641)	227.5	(9.296)
H10-H13 Disorders of conjunctiva (including conjunctivitis)	-2.099	(-0.016)	150.1	(1.038)	-93.84	(-0.719)
H25-H28 Disorders of lens (including cataracts)	-324.8	(-1.214)	161.8	(3.948)	-46.84	(-2.288)
H40-H42 Glaucoma	1114.2	(1.442)	122.9	(1.846)	58.45	(1.531)
H60-H95 Diseases of the ear and mastoid process	326.8	(13.565)	236.3	(8.024)	3.292	(0.082)
I00-I09 Rheumatic heart disease	-762.9	(-2.523)	249.7	(1.703)	137.3	(2.150)
I10-I15 Hypertensive diseases	195.6	(0.511)	257.1	(13.959)	91.14	(5.683)
I20-I25 Ischaemic heart diseases	-568.8	(-0.812)	266.7	(9.496)	262.9	(14.317)
I26-I28 Pulmonary heart disease & diseases of pulmonary circulation	765.5	(1.157)	559.1	(5.877)	293.4	(4.001)
I30-I52 Other forms of heart disease	291.3	(1.823)	461.4	(15.101)	394.4	(20.540)
I60-I69 Cerebrovascular diseases	392.8	(0.858)	132.2	(2.121)	13.33	(0.392)

	Age Group 0-14		Age Group 15-64		Age Group 65+	
	Coefficient	T-Statistic	Coefficient	T-Statistic	Coefficient	T-Statistic
I70-I79 Diseases of arteries, arterioles & capillaries	819.1	(1.393)	1009.5	(12.466)	722.8	(16.072)
I80-I89 Diseases of veins & lymphatic system nec.	63.88	(0.735)	249.9	(11.133)	201.5	(5.939)
I95-I99 Other & unspecified disorders of the circulatory system	-299.7	(-0.661)	743.9	(7.060)	318.2	(6.136)
J00-J06 Acute upper respiratory infections	163.9	(10.207)	154.5	(4.588)	482.3	(2.515)
J10-J18 Influenza & pneumonia	265.0	(5.187)	477.8	(7.819)	437.8	(9.981)
J20-J22 Other acute lower respiratory infections	311.1	(9.805)	658.8	(10.814)	351.3	(7.286)
J30-J39 Other diseases of upper respiratory tract	72.29	(2.767)	154.8	(8.326)	124.1	(1.876)
J40-J47 Chronic lower respiratory diseases	283.0	(12.157)	309.5	(21.736)	465.4	(23.169)
J60-J70 Lung diseases due to external agents	1288.5	(1.845)	891.1	(3.492)	123.4	(1.014)
J80-J99 Other diseases of the respiratory system	755.7	(5.201)	521.3	(8.356)	362.8	(8.399)
K00-K14 Diseases of oral cavity, salivary glands & jaws	138.2	(7.336)	135.5	(10.085)	281.3	(4.310)
K20-K31 Diseases of oesophagusstomach & duodenum	287.7	(4.592)	293.9	(18.851)	165.9	(7.380)
K35-K38 Diseases of appendix	14.71	(0.513)	-10.08	(-0.513)	-305.4	(-2.766)
K40-K46 Hernia	73.28	(1.421)	132.2	(8.021)	66.51	(2.810)
K50-K52 Noninfective enteritis & colitis	34.71	(0.706)	595.0	(23.746)	247.1	(5.853)
K55-K63 Other diseases of intestines	247.5	(5.015)	282.0	(17.870)	166.0	(7.766)
K65-K67 Diseases of peritoneum	588.4	(1.679)	240.5	(3.556)	-30.85	(-0.272)
K70-K77 Diseases of liver	1087.5	(1.894)	1124.4	(16.987)	747.8	(8.912)
K80-K87 Disorders of gall bladder, biliary tract & pancreas	610.6	(1.822)	337.2	(12.382)	180.3	(4.310)
K90-K93 Other diseases of the digestive system	400.4	(3.831)	319.6	(11.119)	150.8	(3.480)
L00-L14 L55-L99 Other infections and disorders of the skin	97.74	(2.486)	163.6	(8.424)	154.2	(3.523)
L20-L30 Dermatitis and eczema	160.2	(4.051)	100.8	(1.529)	272.8	(2.132)
L40-L45 Papulosquamous disorders (including Psoriasis)	1221.7	(2.402)	360.3	(3.916)	73.77	(0.746)
L50-L54 Urticaria and erythems	234.6	(1.738)	262.1	(1.543)	-72.90	(-0.357)
M00-M25 Arthropathies	667.0	(8.239)	525.8	(42.227)	378.2	(22.925)
M30-M36 Systemic connective tissue disorders	663.1	(2.221)	775.6	(8.295)	375.9	(6.969)
M40-M54 Dorsopathies	515.7	(2.949)	352.8	(19.719)	377.1	(13.702)
M60-M79 Soft tissue disorders	141.7	(2.271)	323.1	(19.917)	273.3	(9.123)
M80-M94 Osteopathies and chondropathies	615.2	(4.620)	515.0	(13.584)	299.5	(8.889)
M95-M99 Other disorders of the musculoskeletal system & conn. tiss.	2174.5	(1.220)	684.8	(3.669)	1377.1	(3.730)
N00-N08, N10-N16 Diseases of the kidney	249.7	(2.519)	685.8	(10.655)	668.1	(9.623)
N17-N19 Renal failure	-190.2	(-0.486)	2063.1	(19.501)	801.1	(19.584)
N20-N23 Urolithiasis	545.0	(1.113)	147.7	(3.712)	-13.80	(-0.197)
N25-N29 Other disorders of kidney & ureter	-112.1	(-0.587)	320.4	(2.310)	443.8	(4.196)
N30-N39 Other diseases of the urinary system	226.3	(4.737)	357.8	(12.646)	263.2	(8.240)

	Age Group 0-14		Age Group 15-64		Age Group 65+	
	Coefficient	T-Statistic	Coefficient	T-Statistic	Coefficient	T-Statistic
N40-N51 Diseases of male genital organs	38.25	(1.594)	37.38	(1.478)	-49.08	(-1.416)
N60-N64 Disorders of breast	-92.83	(-0.512)	152.8	(3.149)	-46.34	(-0.388)
N70-N77 Inflammatory diseases of female pelvic organs	263.9	(1.251)	100.0	(3.820)	-89.33	(-0.791)
N80-N98 Noninflammatory disorders of female genital tract	200.0	(2.186)	152.5	(9.665)	8.611	(0.219)
N99 Other disorders of the genitourinary system	-317.9	(-1.438)	52.51	(0.345)	-194.5	(-1.224)
O00-O08 Pregnancy with abortive outcome	-15.01	(-0.155)	95.18	(14.108)	2632.6	(1.453)
O10-O75, O85-O92, O95-O99 Complications of labour and delivery	41.95	(0.197)	134.0	(14.198)	-113.6	(-0.165)
O80-O84 Delivery	140.1	(1.552)	61.54	(6.249)	.	.
P00-P04 Complications of foetus/neonate affected by maternal	73.03	(0.168)	.	.	.	.
P05-P96 Other conditions originating in the perinatal period	468.0	(2.745)	239.2	(0.639)	-1287.8	(-1.424)
Q00-Q89 Congenital malformations	282.9	(12.637)	340.8	(8.529)	-87.40	(-0.716)
Q90-Q99 Chromosomal abnormalities nec.	601.3	(4.845)	528.0	(3.204)	11.47	(0.014)
R00-R09 Symptoms & signs inv. the circulatory/respiratory system	247.1	(8.607)	186.7	(11.284)	226.0	(9.126)
R10-R19 Symptoms & signs inv. the digestive system & abdomen	171.1	(7.044)	228.0	(18.455)	194.0	(7.538)
R20-R23 Symptoms & signs inv. the skin & subcutaneous tissue	182.4	(4.045)	171.8	(3.648)	235.7	(2.859)
R25-R29 Symptoms & signs inv. the nervous & musculoskeletal sys.	287.5	(2.911)	438.7	(6.406)	490.4	(11.637)
R30-R39 Symptoms & signs involving the urinary system	222.7	(3.552)	166.0	(6.535)	5.120	(0.170)
R40-R46 Symptoms & signs inv. Cognition, perception etc.	387.9	(3.664)	189.7	(5.047)	93.94	(2.543)
R47-R49 Symptoms & signs inv. speech & voice	115.0	(0.740)	-9.595	(-0.118)	187.9	(2.117)
R50-R68 General symptoms & signs	279.4	(12.532)	237.4	(12.576)	179.7	(6.764)
R69 Unknown & unspecified causes of morbidity	-608.6	(-3.649)	73.34	(0.331)	-144.8	(-0.274)
R70-R89 Abnormal findings of bodily fluids or samples without diag.	104.9	(0.646)	439.6	(7.136)	224.1	(4.131)
R90-R94 Abnormal findings on diagnostic imaging/function studies	500.8	(2.302)	249.9	(4.529)	63.98	(1.217)
R95-R99 Ill-defined & unknown causes of mortality	.	.	580.1	(4.584)	-499.1	(-3.475)
S00-S09 Injuries to the head	129.8	(4.045)	175.6	(6.218)	187.3	(3.788)
S10-S19 Injuries to the neck	423.7	(1.886)	166.5	(1.755)	356.2	(1.620)
S20-S29 Injuries to the thorax	331.6	(1.797)	131.5	(1.912)	212.0	(2.056)
S30-S39 Injuries to abdomen, lower back, lumbar spine & pelvis	60.18	(1.568)	268.3	(4.069)	-1.605	(-0.020)
S40-S49 Injuries to the shoulder & upper arm	82.18	(1.835)	292.1	(6.134)	64.09	(0.921)
S50-S59 Injuries to the elbow & forearm	112.8	(3.847)	203.0	(5.380)	-39.37	(-0.663)
S60-S69 Injuries to the wrist & hand	116.1	(2.360)	164.0	(6.192)	-74.53	(-0.982)
S70-S79 Injuries to the hip & thigh	284.5	(4.013)	453.8	(5.306)	-317.2	(-6.003)
S80-S89 Injuries to the knee & lower leg	200.2	(4.072)	185.4	(5.066)	37.81	(0.582)
S90-S99 Injuries to the ankle & foot	265.0	(2.082)	158.5	(2.864)	175.0	(1.494)
T00-T07 Injuries involving multiple body regions	63.59	(0.627)	352.0	(2.014)	-3.468	(-0.015)

	Age Group 0-14		Age Group 15-64		Age Group 65+	
	Coefficient	T-Statistic	Coefficient	T-Statistic	Coefficient	T-Statistic
T08-T14 Injuries to unspecified part of trunk limb or body	44.02	(0.293)	219.2	(1.364)	-191.8	(-0.825)
T15-T19 Effects of foreign body entering through natural orifice	49.13	(1.519)	239.6	(2.730)	149.4	(0.729)
T20-T32 Burns and corrosions	139.9	(2.998)	66.09	(0.867)	667.9	(1.740)
T33-T35 Frostbite	.	.	3606.4	(1.347)	-2309.2	(-2.814)
T36-T50 Poisonings by drugs medicaments & biological substances	181.8	(2.831)	255.7	(5.869)	-49.23	(-0.362)
T51-T65 Toxic effects of substances chiefly non-medicinal as to source	82.45	(1.303)	121.7	(2.912)	-77.32	(-0.358)
T66-T78 Other and unspecified effects of external causes	103.5	(2.021)	207.1	(2.168)	486.3	(2.884)
T79 Certain early complications of trauma	-23.43	(-0.092)	199.7	(1.072)	-486.0	(-2.263)
T80-T88 Complications of surgical & medical care nec.	573.3	(3.617)	483.9	(12.053)	637.6	(13.307)
T90-T98 Sequelae of injuries of poisoning & other consequences	88.72	(0.893)	60.01	(1.259)	300.4	(1.853)
VVV	187.0	(4.287)	81.24	(2.058)	-19.89	(-0.199)
WWW	183.6	(5.443)	102.8	(3.657)	143.1	(3.199)
XXX	255.2	(5.642)	105.0	(2.808)	200.8	(2.152)
YYY	-59.68	(-0.599)	244.9	(8.157)	142.0	(3.423)
Z00-Z13 Examination and investigation	148.3	(3.841)	112.9	(6.635)	32.33	(1.182)
Z20-Z29 Potential health hazards related to communicable diseases	522.4	(2.513)	324.9	(7.714)	331.0	(2.955)
Z30-Z39 Health services in circumstances related to reproduction	28.15	(0.098)	75.13	(5.912)	289.6	(0.371)
Z40-Z54 Persons encountering health services for specific care	581.7	(12.429)	745.2	(42.780)	810.4	(32.990)
Z55-Z65 Potential health hazards related to socioeconomic & psychosoc.l	187.2	(1.750)	175.8	(3.600)	213.6	(3.461)
Z70-Z76 Persons encountering health services in other circs.	954.9	(4.798)	163.8	(6.451)	65.05	(1.476)
Z80-Z99 Persons with potential health hazards related to family	290.7	(8.405)	252.2	(24.373)	159.8	(8.800)
U Unclassified	1002.0	(2.851)	1461.2	(6.433)	888.5	(5.117)
Morbidity Interactions						
A00B99-G00G99	518.4	(2.933)	234.6	(1.720)	89.64	(0.751)
A00B99-H00H59	194.1	(0.995)	701.9	(2.968)	220.6	(1.601)
A00B99-O00O99	2439.3	(4.098)	-533.2	(-7.121)	0	(.)
A00B99-Q00Q99	368.6	(2.850)	440.9	(1.150)	676.4	(1.219)
A00B99-Z00Z99	-16.49	(-0.308)	298.0	(4.548)	111.0	(1.447)
C00D48-H00H59	278.3	(0.445)	-186.0	(-1.203)	-184.1	(-2.824)
C00D48-L00L99	34.44	(0.069)	-177.5	(-1.661)	-209.0	(-2.727)
C00D48-N00N99	110.2	(0.134)	-90.64	(-1.722)	-124.5	(-2.281)
C00D48-P00P96	116.5	(0.261)	-1616.8	(-2.485)	-93.33	(-0.099)
C00D48-Z00Z99	-65.64	(-0.339)	102.4	(2.961)	129.9	(3.359)
D50D89-K00K93	-276.2	(-0.853)	-134.0	(-1.920)	-371.2	(-6.197)
D50D89-O00O99	437.0	(0.703)	-634.2	(-8.222)	-5455.2	(-11.310)

	Age Group 0-14		Age Group 15-64		Age Group 65+	
	Coefficient	T-Statistic	Coefficient	T-Statistic	Coefficient	T-Statistic
D50D89-Z00Z99	948.9	(3.097)	482.6	(7.185)	274.8	(4.891)
E00E90-G00G99	164.2	(0.592)	82.71	(1.792)	52.16	(1.094)
E00E90-H00H59	1100.1	(2.086)	308.0	(5.204)	46.03	(1.372)
E00E90-I00I99	-91.62	(-0.184)	-156.5	(-5.389)	-133.1	(-4.657)
E00E90-L00L99	103.7	(0.396)	322.3	(4.763)	166.8	(2.615)
E00E90-O00O99	523.5	(1.309)	-149.1	(-5.955)	11434.9	(27.238)
E00E90-R00R99	331.8	(3.510)	11.73	(0.449)	28.33	(0.992)
E00E90-Z00Z99	346.3	(2.904)	59.74	(2.449)	-42.39	(-1.619)
F00F99-I00I99	-53.37	(-0.119)	-12.55	(-0.485)	-61.85	(-1.834)
F00F99-J00J99	212.6	(1.922)	-31.04	(-1.358)	-28.44	(-0.682)
F00F99-O00O99	235.7	(1.397)	-65.39	(-5.939)	3827.9	(4.943)
F00F99-R00R99	-20.88	(-0.244)	-13.28	(-0.730)	14.63	(0.391)
H00H59-L00L99	404.6	(1.477)	564.0	(2.985)	117.1	(1.230)
I00I99-K00K93	427.8	(1.431)	-60.55	(-2.617)	-48.69	(-2.164)
I00I99-L00L99	102.9	(0.195)	282.1	(4.401)	131.7	(2.315)
J00J99-O00O99	293.2	(1.058)	-159.6	(-10.191)	-5688.3	(-7.085)
K00K93-N00N99	181.3	(1.474)	38.45	(1.097)	55.48	(1.453)
K00K93-Q00Q99	310.9	(2.706)	324.8	(2.601)	111.6	(0.519)
L00L99-M00M99	454.1	(1.617)	142.2	(2.153)	86.62	(1.365)
L00L99-R00R99	43.04	(0.890)	201.1	(3.441)	159.4	(2.524)
M00M99-N00N99	180.5	(0.492)	16.25	(0.351)	-248.0	(-6.475)
M00M99-O00O99	-401.8	(-0.877)	-243.0	(-10.727)	-434.1	(-0.567)
N00N99-Q00Q99	90.88	(0.818)	136.4	(1.060)	803.7	(3.503)
N00N99-S00T98	115.4	(0.587)	392.3	(5.676)	50.56	(0.942)
N00N99-Z00Z99	1.115	(0.016)	-15.26	(-0.742)	113.6	(3.367)
O00O99-R00R99	-2928.3	(-6.181)	-132.3	(-10.900)	3555.2	(8.255)
O00O99-Z00Z99	123.2	(0.628)	-0.142	(-0.010)	-6772.7	(-15.553)
V01Y98-Z00Z99	-102.4	(-1.748)	-2.651	(-0.107)	-33.75	(-0.884)
Morbidity Count 9						
No morbidities	.	.	.	.	.	.
2 morbidities	-98.34	(-5.067)	-51.81	(-4.621)	181.2	(5.811)
3 morbidities	-273.5	(-10.831)	-209.7	(-20.649)	34.92	(1.946)
4 morbidities	-390.0	(-10.682)	-248.0	(-15.721)	120.7	(3.150)
5 morbidities	-288.7	(-6.721)	-239.8	(-16.789)	37.52	(1.076)
6 morbidities	-508.1	(-12.227)	-400.7	(-26.824)	-101.1	(-4.479)

	Age Group 0-14		Age Group 15-64		Age Group 65+	
	Coefficient	T-Statistic	Coefficient	T-Statistic	Coefficient	T-Statistic
7 morbidities	-541.6	(-9.641)	-362.1	(-18.442)	25.70	(0.672)
8 morbidities	-512.7	(-6.698)	-398.9	(-17.653)	-52.20	(-1.329)
9 morbidities	-691.8	(-10.264)	-604.2	(-26.795)	-208.0	(-7.490)
CCG Dummies						
NHS Darlington CCG	40.43	(2.673)	-7.452	(-0.459)	-230.7	(-4.018)
NHS Durham Dales, Easington and Sedgefield CCG	11.00	(0.852)	-15.43	(-1.153)	-278.4	(-5.981)
NHS Gateshead CCG	-10.32	(-0.657)	54.84	(3.277)	-37.93	(-0.590)
NHS Newcastle North and East CCG	-55.65	(-4.094)	-26.75	(-1.938)	-163.7	(-2.119)
NHS Newcastle West CCG	-61.39	(-4.849)	-29.33	(-1.998)	-144.6	(-1.882)
NHS North Durham CCG	-46.52	(-4.366)	-40.98	(-3.172)	-237.1	(-4.740)
NHS Hartlepool and Stockton-On-Tees CCG	-28.86	(-2.355)	-11.15	(-0.827)	-290.8	(-5.976)
NHS Northumberland CCG	-56.87	(-5.380)	-28.15	(-2.152)	-229.4	(-5.077)
NHS South Tees CCG	-23.70	(-1.874)	-14.98	(-1.057)	-163.8	(-2.976)
NHS South Tyneside CCG	-24.16	(-2.066)	81.64	(4.662)	-72.35	(-1.103)
NHS Sunderland CCG	37.20	(2.699)	27.14	(1.974)	-94.92	(-1.755)
NHS Blackburn With Darwen CCG	-14.10	(-0.977)	-0.759	(-0.041)	-135.2	(-2.007)
NHS Blackpool CCG	-45.02	(-3.182)	-70.97	(-4.541)	-334.4	(-6.493)
NHS Bolton CCG	-74.52	(-7.372)	-18.27	(-1.394)	-195.1	(-4.113)
NHS Bury CCG	-69.62	(-6.058)	-60.77	(-4.566)	-333.3	(-6.603)
NHS Central Manchester CCG	-103.4	(-9.218)	-24.58	(-1.887)	-48.66	(-0.612)
NHS Chorley and South Ribble CCG	-37.90	(-2.912)	3.861	(0.245)	-125.3	(-2.341)
NHS Oldham CCG	-72.55	(-6.533)	-10.54	(-0.764)	-132.6	(-2.501)
NHS East Lancashire CCG	-6.990	(-0.582)	-21.37	(-1.662)	-21.29	(-0.407)
NHS Eastern Cheshire CCG	-47.20	(-2.701)	-21.76	(-1.501)	-204.4	(-4.422)
NHS Heywood, Middleton and Rochdale CCG	-74.44	(-7.118)	-27.92	(-1.911)	-230.1	(-4.425)
NHS Greater Preston CCG	-36.42	(-3.275)	-16.98	(-1.269)	-13.09	(-0.213)
NHS Halton CCG	21.39	(1.138)	-1.904	(-0.110)	7.168	(0.102)
NHS Salford CCG	-81.32	(-7.677)	-17.66	(-1.321)	-188.1	(-3.299)
NHS Cumbria CCG	-21.85	(-2.154)	8.261	(0.694)	-51.66	(-1.284)
NHS Knowsley CCG	-14.50	(-0.740)	15.42	(0.878)	-193.0	(-3.047)
NHS Lancashire North CCG	-33.00	(-2.057)	8.339	(0.547)	-145.3	(-2.743)
NHS North Manchester CCG	-83.89	(-7.377)	-24.11	(-1.672)	-154.8	(-2.301)
NHS South Manchester CCG	-91.10	(-8.145)	-23.70	(-1.721)	-86.12	(-1.178)
NHS South Cheshire CCG	-15.46	(-0.967)	-24.46	(-1.773)	-242.9	(-4.882)
NHS South Sefton CCG	-46.53	(-2.758)	12.48	(0.790)	-64.58	(-1.091)

	Age Group 0-14		Age Group 15-64		Age Group 65+	
	Coefficient	T-Statistic	Coefficient	T-Statistic	Coefficient	T-Statistic
NHS Southport And Formby CCG	-40.30	(-2.468)	15.88	(0.919)	-150.5	(-2.768)
NHS Stockport CCG	-48.98	(-4.488)	-12.18	(-0.941)	-86.32	(-1.776)
NHS St Helens CCG	23.16	(1.304)	-11.47	(-0.732)	-149.3	(-2.832)
NHS Tameside And Glossop CCG	-51.34	(-4.501)	-24.35	(-1.811)	-28.21	(-0.525)
NHS Trafford CCG	-74.65	(-6.846)	9.471	(0.707)	-76.19	(-1.402)
NHS Vale Royal CCG	-12.15	(-0.656)	-42.70	(-2.590)	-160.8	(-2.329)
NHS Warrington CCG	-4.036	(-0.332)	-28.47	(-2.063)	-115.7	(-2.070)
NHS West Cheshire CCG	6.510	(0.497)	1.283	(0.079)	120.1	(1.931)
NHS West Lancashire CCG	-16.69	(-1.278)	-0.0887	(-0.005)	-115.0	(-1.942)
NHS Wigan Borough CCG	-36.07	(-3.527)	-14.67	(-1.176)	-129.9	(-2.818)
NHS Fylde and Wyre CCG	-56.13	(-4.706)	-57.86	(-3.741)	-243.5	(-4.588)
NHS Airedale, Wharfedale and Craven CCG	-13.83	(-0.650)	-4.896	(-0.300)	-66.47	(-1.203)
NHS Barnsley CCG	-36.98	(-3.086)	-11.18	(-0.856)	-93.97	(-1.870)
NHS Bassetlaw CCG	-33.92	(-2.101)	-36.08	(-2.050)	-137.7	(-1.951)
NHS Bradford Districts CCG	-37.68	(-3.741)	-4.848	(-0.384)	-200.0	(-4.304)
NHS Calderdale CCG	-47.09	(-4.033)	-44.87	(-3.522)	-194.6	(-3.791)
NHS Leeds North CCG	-49.26	(-3.184)	16.31	(1.004)	-67.60	(-1.105)
NHS Bradford City CCG	-25.59	(-1.696)	26.08	(1.570)	-253.4	(-2.754)
NHS Doncaster CCG	-41.07	(-3.660)	-24.35	(-1.917)	-53.90	(-1.071)
NHS East Riding of Yorkshire CCG	-46.41	(-4.536)	-37.64	(-3.158)	-200.1	(-4.778)
NHS Greater Huddersfield CCG	-56.46	(-5.541)	-18.35	(-1.350)	-304.8	(-6.394)
NHS Leeds West CCG	-53.67	(-5.048)	14.99	(1.156)	-62.38	(-1.127)
NHS Hambleton, Richmondshire and Whitby CCG	1.208	(0.064)	-29.61	(-2.010)	-12.92	(-0.233)
NHS Harrogate and Rural District CCG	-24.47	(-2.034)	22.40	(1.521)	-65.22	(-1.267)
NHS Hull CCG	-45.92	(-3.887)	-29.97	(-2.462)	-173.4	(-3.301)
NHS Leeds South and East CCG	-61.91	(-5.773)	25.51	(1.626)	-114.1	(-2.049)
NHS North East Lincolnshire CCG	-29.06	(-2.308)	-24.14	(-1.629)	-145.1	(-2.667)
NHS North Kirklees CCG	-41.90	(-2.774)	-9.597	(-0.709)	-145.4	(-2.603)
NHS North Lincolnshire CCG	-20.56	(-1.701)	-30.85	(-2.146)	-186.3	(-3.721)
NHS Rotherham CCG	-49.71	(-4.649)	-31.90	(-2.492)	-182.4	(-3.861)
NHS Scarborough and Ryedale CCG	-16.67	(-1.030)	-42.58	(-2.773)	-55.64	(-0.837)
NHS Sheffield CCG	-21.58	(-1.638)	-4.071	(-0.356)	-42.61	(-0.943)
NHS Vale of York CCG	-26.81	(-2.502)	4.426	(0.361)	85.95	(1.762)
NHS Wakefield CCG	-50.33	(-3.318)	-31.95	(-2.559)	-204.7	(-4.401)
NHS Lincolnshire East CCG	-16.52	(-1.378)	-12.87	(-0.920)	-143.1	(-3.200)

	Age Group 0-14		Age Group 15-64		Age Group 65+	
	Coefficient	T-Statistic	Coefficient	T-Statistic	Coefficient	T-Statistic
NHS Corby CCG	-96.88	(-7.090)	-75.27	(-4.186)	80.81	(0.755)
NHS East Leicestershire and Rutland CCG	-67.37	(-7.034)	-37.70	(-3.144)	-108.2	(-2.377)
NHS Erewash CCG	-68.13	(-4.258)	-34.44	(-2.244)	-155.9	(-2.413)
NHS Hardwick CCG	-7.956	(-0.371)	-37.92	(-2.018)	-75.75	(-1.055)
NHS Leicester City CCG	-83.55	(-8.922)	-46.27	(-3.869)	-205.3	(-3.905)
NHS Lincolnshire West CCG	-15.75	(-1.376)	-12.62	(-0.926)	-214.3	(-4.672)
NHS Mansfield and Ashfield CCG	-46.14	(-4.285)	-23.33	(-1.524)	-171.7	(-3.250)
NHS Milton Keynes CCG	-19.27	(-1.407)	-18.18	(-1.446)	-127.4	(-2.581)
NHS Nene CCG	-45.98	(-4.722)	-28.18	(-2.552)	-59.19	(-1.464)
NHS Newark and Sherwood CCG	-30.01	(-2.453)	-40.28	(-2.617)	-238.8	(-4.543)
NHS North Derbyshire CCG	-49.57	(-4.450)	-21.33	(-1.607)	-71.53	(-1.525)
NHS Nottingham City CCG	-66.34	(-5.921)	-51.76	(-4.600)	-199.9	(-3.853)
NHS Nottingham North and East CCG	-61.96	(-4.160)	-60.02	(-4.337)	-240.8	(-4.708)
NHS Nottingham West CCG	-56.56	(-3.959)	-55.72	(-3.561)	-223.3	(-3.631)
NHS Rushcliffe CCG	-68.62	(-4.435)	-21.53	(-1.370)	-199.8	(-3.604)
NHS South West Lincolnshire CCG	-22.64	(-1.612)	-17.53	(-1.142)	-83.72	(-1.371)
NHS Southern Derbyshire CCG	-42.62	(-4.440)	-2.102	(-0.184)	-114.5	(-2.705)
NHS West Leicestershire CCG	-44.75	(-4.492)	-43.04	(-3.667)	-129.8	(-2.921)
NHS Birmingham South and Central CCG	-31.95	(-1.932)	2.215	(0.167)	-96.60	(-1.550)
NHS Cannock Chase CCG	-7.925	(-0.509)	31.83	(1.610)	-36.25	(-0.506)
NHS Coventry and Rugby CCG	-23.51	(-1.924)	-40.34	(-3.522)	-150.8	(-3.450)
NHS Dudley CCG	-62.61	(-6.162)	-32.80	(-2.676)	-143.8	(-3.198)
NHS East Staffordshire CCG	-48.26	(-4.019)	-6.682	(-0.484)	-4.210	(-0.074)
NHS Herefordshire CCG	4.004	(0.250)	-25.35	(-1.795)	-130.3	(-2.791)
NHS North Staffordshire CCG	-35.31	(-2.951)	-25.23	(-1.749)	-136.5	(-2.781)
NHS Warwickshire North CCG	-6.687	(-0.477)	6.203	(0.372)	47.79	(0.788)
NHS Redditch and Bromsgrove CCG	0.834	(0.049)	14.94	(0.973)	37.92	(0.642)
NHS Sandwell and West Birmingham CCG	-37.48	(-3.573)	-26.17	(-2.342)	-137.4	(-2.915)
NHS Shropshire CCG	12.32	(0.700)	-6.469	(-0.525)	-31.84	(-0.738)
NHS Solihull CCG	-28.41	(-1.744)	-14.87	(-1.194)	-150.3	(-3.219)
NHS South East Staffordshire and Seisdon Peninsula CCG	-22.16	(-1.825)	10.47	(0.802)	47.45	(0.851)
NHS South Warwickshire CCG	-17.05	(-1.415)	-5.109	(-0.412)	47.40	(0.973)
NHS South Worcestershire CCG	-27.75	(-2.265)	-12.17	(-0.959)	-137.6	(-3.135)
NHS Stafford And Surrounds CCG	-12.94	(-0.687)	-25.28	(-1.826)	-160.9	(-3.214)
NHS Stoke on Trent CCG	-35.92	(-2.772)	-45.79	(-3.508)	-130.5	(-2.542)

	Age Group 0-14		Age Group 15-64		Age Group 65+	
	Coefficient	T-Statistic	Coefficient	T-Statistic	Coefficient	T-Statistic
NHS Telford and Wrekin CCG	-19.66	(-1.577)	11.85	(0.833)	-43.49	(-0.777)
NHS Walsall CCG	-22.53	(-1.740)	0.655	(0.043)	8.035	(0.152)
NHS Wolverhampton CCG	-39.85	(-3.639)	-0.440	(-0.030)	-42.17	(-0.671)
NHS Wyre Forest CCG	-29.07	(-2.003)	-40.85	(-2.662)	-214.6	(-3.712)
NHS Bedfordshire CCG	-54.18	(-5.604)	-40.01	(-3.443)	-180.4	(-4.383)
NHS Cambridgeshire and Peterborough CCG	-52.66	(-5.841)	-27.81	(-2.628)	-138.0	(-3.597)
NHS East and North Hertfordshire CCG	-25.56	(-2.594)	-9.634	(-0.856)	-129.7	(-3.135)
NHS Ipswich and East Suffolk CCG	16.04	(1.370)	-16.24	(-1.348)	218.5	(4.630)
NHS Great Yarmouth and Waveney CCG	-22.63	(-1.583)	3.206	(0.192)	-135.5	(-2.846)
NHS Herts Valleys CCG	-52.73	(-5.696)	-28.51	(-2.639)	-132.2	(-3.141)
NHS Luton CCG	-56.70	(-3.772)	-49.49	(-3.968)	-266.5	(-4.951)
NHS Mid Essex CCG	-18.91	(-1.839)	19.59	(1.610)	68.67	(1.449)
NHS North East Essex CCG	54.92	(3.958)	9.491	(0.683)	-60.04	(-1.392)
NHS North Norfolk CCG	-67.81	(-4.667)	-39.85	(-2.727)	-80.38	(-1.696)
NHS Norwich CCG	-80.69	(-6.114)	-38.04	(-2.980)	-225.2	(-4.667)
NHS South Norfolk CCG	-66.15	(-4.333)	-44.37	(-3.600)	-132.8	(-2.844)
NHS Thurrock CCG	-20.91	(-1.432)	-21.11	(-1.316)	-157.0	(-2.492)
NHS West Essex CCG	-56.61	(-5.213)	-13.61	(-1.083)	-162.7	(-3.495)
NHS West Norfolk CCG	-29.68	(-2.108)	-54.43	(-3.732)	-203.0	(-3.966)
NHS West Suffolk CCG	4.654	(0.337)	-9.481	(-0.698)	228.0	(4.406)
NHS Barking and Dagenham CCG	-64.93	(-5.593)	3.309	(0.221)	-129.0	(-1.722)
NHS Barnet CCG	-49.78	(-5.028)	13.78	(1.084)	89.11	(1.262)
NHS Bexley CCG	-13.67	(-0.908)	20.77	(1.277)	-79.45	(-1.339)
NHS Brent CCG	-42.46	(-2.903)	8.457	(0.711)	-47.54	(-0.769)
NHS Bromley CCG	-42.05	(-3.947)	-2.194	(-0.159)	-118.6	(-2.344)
NHS Camden CCG	-46.40	(-4.375)	22.55	(1.622)	144.4	(1.404)
NHS City and Hackney CCG	-44.99	(-4.078)	11.09	(0.702)	74.68	(0.695)
NHS Croydon CCG	-51.86	(-4.731)	-3.981	(-0.328)	-118.0	(-2.267)
NHS Ealing CCG	-55.51	(-5.492)	8.365	(0.707)	-48.19	(-0.865)
NHS Enfield CCG	-44.98	(-4.162)	20.59	(1.282)	155.8	(1.972)
NHS Hounslow CCG	-38.12	(-3.384)	5.025	(0.405)	42.35	(0.627)
NHS Greenwich CCG	-30.29	(-2.402)	-2.303	(-0.164)	-43.55	(-0.655)
NHS Hammersmith and Fulham CCG	-2.246	(-0.156)	14.19	(1.020)	-78.27	(-1.028)
NHS Haringey CCG	-40.12	(-3.050)	22.85	(1.446)	173.3	(1.811)
NHS Harrow CCG	-44.79	(-3.826)	11.80	(0.931)	-109.5	(-1.883)

	Age Group 0-14		Age Group 15-64		Age Group 65+	
	Coefficient	T-Statistic	Coefficient	T-Statistic	Coefficient	T-Statistic
NHS Havering CCG	-59.70	(-4.789)	-7.422	(-0.551)	-105.8	(-1.998)
NHS Hillingdon CCG	-25.67	(-2.304)	-4.581	(-0.369)	-177.7	(-3.439)
NHS Islington CCG	-36.32	(-2.708)	11.16	(0.831)	297.5	(2.397)
NHS Kingston CCG	108.9	(8.128)	69.08	(4.325)	571.6	(5.933)
NHS Lambeth CCG	13.14	(0.872)	43.63	(3.050)	154.0	(1.642)
NHS Lewisham CCG	17.45	(1.176)	50.42	(3.254)	254.1	(2.786)
NHS Newham CCG	-52.74	(-4.426)	15.41	(1.191)	-62.07	(-0.815)
NHS Redbridge CCG	-66.62	(-5.907)	15.78	(1.162)	-0.884	(-0.014)
NHS Richmond CCG	-45.71	(-3.936)	27.05	(1.737)	-42.76	(-0.656)
NHS Southwark CCG	25.43	(1.611)	40.10	(2.651)	353.9	(3.107)
NHS Merton CCG	-28.07	(-2.242)	41.04	(2.754)	86.79	(1.083)
NHS Sutton CCG	11.41	(0.899)	24.10	(1.742)	297.8	(3.575)
NHS Tower Hamlets CCG	-57.37	(-5.514)	12.22	(0.896)	298.0	(2.376)
NHS Waltham Forest CCG	-56.77	(-5.363)	15.30	(1.048)	-54.41	(-0.795)
NHS Wandsworth CCG	-13.39	(-0.950)	23.06	(1.930)	124.6	(1.530)
NHS West London CCG	-45.64	(-3.555)	31.46	(2.197)	-32.63	(-0.438)
NHS Central London (Westminster) CCG	-50.76	(-4.045)	-9.962	(-0.760)	-120.4	(-1.388)
NHS Ashford CCG	3.503	(0.281)	-4.912	(-0.299)	-147.3	(-2.584)
NHS Brighton and Hove CCG	-61.23	(-5.645)	-26.05	(-2.122)	-150.5	(-2.925)
NHS Canterbury and Coastal CCG	-21.04	(-1.701)	-41.38	(-3.270)	-188.7	(-3.608)
NHS Eastbourne, Hailsham and Seaford CCG	-21.04	(-1.634)	-0.648	(-0.043)	-171.4	(-3.762)
NHS Coastal West Sussex CCG	-21.46	(-1.835)	-42.61	(-3.478)	-259.4	(-6.915)
NHS Crawley CCG	-74.73	(-5.337)	-46.54	(-2.866)	-171.1	(-2.740)
NHS Dartford, Gravesham and Swanley CCG	-24.30	(-2.031)	56.94	(3.978)	-133.9	(-2.794)
NHS East Surrey CCG	45.46	(1.814)	59.31	(3.331)	515.7	(6.629)
NHS Guildford and Waverley CCG	8.226	(0.687)	10.90	(0.806)	324.0	(5.253)
NHS Hastings and Rother CCG	-13.42	(-1.091)	-6.641	(-0.459)	-148.8	(-3.207)
NHS Medway CCG	-22.41	(-1.972)	-34.85	(-2.672)	-287.3	(-6.028)
NHS Horsham And Mid Sussex CCG	-55.88	(-5.405)	-21.11	(-1.594)	-220.6	(-4.718)
NHS North West Surrey CCG	38.68	(2.308)	48.20	(3.737)	645.7	(10.388)
NHS South Kent Coast CCG	23.73	(1.638)	-19.36	(-1.403)	-177.7	(-3.776)
NHS Surrey Heath CCG	16.87	(1.062)	69.41	(2.732)	476.5	(4.478)
NHS Swale CCG	-33.14	(-2.173)	-55.99	(-3.416)	-287.3	(-4.681)
NHS Thanet CCG	-8.627	(-0.519)	-10.68	(-0.655)	-54.75	(-0.864)
NHS Bracknell and Ascot CCG	-18.30	(-1.383)	13.34	(0.770)	-34.55	(-0.461)

	Age Group 0-14		Age Group 15-64		Age Group 65+	
	Coefficient	T-Statistic	Coefficient	T-Statistic	Coefficient	T-Statistic
NHS Chiltern CCG	-15.52	(-1.387)	-33.66	(-2.634)	-73.74	(-1.461)
NHS North Hampshire CCG	-8.671	(-0.686)	25.01	(1.532)	1.753	(0.029)
NHS Fareham and Gosport CCG	-63.38	(-5.685)	-50.55	(-3.836)	-195.1	(-4.149)
NHS Isle of Wight CCG	-33.92	(-2.035)	-32.85	(-1.898)	-92.71	(-1.727)
NHS Newbury and District CCG	-55.19	(-4.964)	-30.50	(-2.000)	-164.4	(-2.766)
NHS North and West Reading CCG	-62.38	(-5.563)	16.89	(0.834)	-143.8	(-2.218)
NHS Oxfordshire CCG	-66.78	(-7.337)	-22.09	(-2.028)	-160.9	(-3.981)
NHS Portsmouth CCG	-60.56	(-5.237)	-52.99	(-4.130)	-220.9	(-4.074)
NHS Slough CCG	-24.28	(-1.930)	13.77	(0.957)	-145.5	(-2.062)
NHS South Eastern Hampshire CCG	-37.12	(-2.503)	-48.13	(-3.302)	-198.6	(-4.200)
NHS South Reading CCG	-48.23	(-3.978)	10.90	(0.719)	-187.0	(-2.678)
NHS Southampton CCG	-23.87	(-1.082)	-10.05	(-0.741)	17.94	(0.299)
NHS Aylesbury Vale CCG	-27.48	(-2.418)	-29.99	(-2.257)	-96.60	(-1.814)
NHS West Hampshire CCG	-25.80	(-2.471)	-3.834	(-0.320)	-77.45	(-1.925)
NHS Windsor, Ascot and Maidenhead CCG	-6.267	(-0.391)	-9.324	(-0.536)	-85.85	(-1.287)
NHS Wokingham CCG	-38.68	(-2.752)	6.291	(0.387)	14.48	(0.215)
NHS Bath and North East Somerset CCG	-45.17	(-3.922)	-44.33	(-3.350)	-239.1	(-4.934)
NHS Bristol CCG	-55.68	(-5.714)	-39.03	(-3.443)	-241.7	(-4.815)
NHS Dorset CCG	-7.061	(-0.641)	-7.984	(-0.722)	-32.52	(-0.863)
NHS Gloucestershire CCG	-14.28	(-1.308)	-29.23	(-2.703)	-180.5	(-4.738)
NHS Kernow CCG	-21.11	(-1.721)	-26.79	(-2.192)	-179.8	(-4.509)
NHS North Somerset CCG	-20.26	(-0.936)	-31.43	(-2.397)	-210.3	(-4.638)
NHS Somerset CCG	-14.41	(-1.166)	-29.16	(-2.523)	-152.3	(-3.861)
NHS South Gloucestershire CCG	-58.03	(-5.760)	-30.11	(-2.205)	-101.8	(-1.830)
NHS Swindon CCG	-27.45	(-2.268)	-27.11	(-2.111)	-188.2	(-3.669)
NHS Wirral CCG	-32.57	(-3.064)	-15.80	(-1.226)	-147.0	(-3.228)
NHS Birmingham Crosscity CCG	-20.78	(-1.717)	-2.574	(-0.235)	-157.2	(-3.801)
NHS Liverpool CCG	-75.26	(-7.451)	14.92	(1.179)	-81.20	(-1.648)
NHS North Tyneside CCG	-28.99	(-2.359)	-16.29	(-1.164)	-153.5	(-2.378)
NHS South Lincolnshire CCG	-10.03	(-0.779)	-24.32	(-1.748)	-201.9	(-4.100)
NHS Basildon and Brentwood CCG	-45.98	(-4.392)	-40.34	(-3.061)	-132.8	(-2.604)
NHS Castle Point and Rochford CCG	-34.44	(-3.017)	-38.16	(-2.690)	-287.8	(-5.952)
NHS Southend CCG	-36.43	(-2.815)	-32.77	(-2.313)	-275.4	(-4.954)
NHS Surrey Downs CCG	11.63	(0.965)	35.71	(2.692)	194.0	(3.776)
NHS West Kent CCG	-32.01	(-2.385)	36.86	(2.554)	50.70	(0.945)

	Age Group 0-14		Age Group 15-64		Age Group 65+	
	Coefficient	T-Statistic	Coefficient	T-Statistic	Coefficient	T-Statistic
NHS High Weald Lewes Havens CCG	-44.57	(-3.165)	9.793	(0.565)	25.88	(0.397)
NHS North East Hampshire and Farnham CCG	4.463	(0.230)	49.98	(2.935)	373.5	(5.116)
NHS Wiltshire CCG	-31.63	(-2.970)	-15.26	(-1.304)	-121.1	(-2.899)
NHS Northern, Eastern and Western Devon CCG	-7.108	(-0.720)	-18.14	(-1.612)	-120.8	(-3.243)
NHS South Devon and Torbay CCG	.	.	.	.	.	.
New GP practice	28.62	(10.340)	45.98	(21.526)	-43.83	(-2.584)
Private care	-77.98	(-1.444)	-35.63	(-0.861)	-182.0	(-2.890)
Attributed Needs						
Log population variance			-60.48	(-11.059)		
All Usual Residents Aged 16+					-0.659	(-7.681)
All Usual Residents Aged 16 to 74					0.721	(8.096)
Resident Population					-0.856	(-4.384)
Proportion Single Pensioner Households					10.20	(8.809)
Proportion aged 16-74 people never worked			1.840	(8.003)		
Proportion Single (never married)			0.459	(5.435)	3.105	(5.345)
Proportion Divorced			1.627	(4.598)		
Rented from private landlord or letting agency	-0.430	(-5.853)	-1.088	(-13.758)	-2.352	(-4.973)
Proportion (un standardised) with not good health (NGH)			1.651	(2.811)	15.80	(8.424)
All people living in the area					0.851	(4.412)
Average with (long term) medical condition for those with at least one			9.755	(2.524)	54.45	(2.936)
2012-13 QOF KD Total Exceptions			0.0400	(2.610)	0.110	(2.076)
2012-13 QOF Epilepsy Prevalence			32.00	(7.204)		
2012-13 QOF Mental Health Prevalence			7.240	(3.161)		
Health Deprivation and Disability Score	15.40	(15.681)	13.43	(7.393)		
Attributed Supply						
Adult critical beds Jan 13					13.41	(3.367)
2012-13 Median waiting times (weeks) for Dermatology Patients					-28.25	(-2.502)
2012-13 Median waiting times (weeks) of the 95th percentile for Neurosurgery Patients			-3.911	(-4.001)	-22.65	(-4.429)
Constant	369.8	(34.047)	59.16	(4.517)	290.8	(5.596)
R-Squared	0.0984		0.117		0.0974	
Adjusted R-Squared	0.0981		0.117		0.0972	

15 Appendix D: Goodness of fit metrics for G&A model

In this appendix, the goodness of fit statistics are explained with mathematical equations to help aid the replicability of the work. Individual level goodness of fit statistics and practice level goodness of fit statistics are used.

Costs and predicted costs are used to generate goodness of fit metrics at both the individual level and practice level.

For each individual in the practice sample we observe c_{ip}, the actual cost of individual i registered with practice p. For each model, we can create a predicted cost for the individual, $\hat{c}_{ip}$. The mean observed cost for individuals registered with each practice is given by Equation 15.1:

$$\bar{c}_p = \frac{\sum_{i \in p} c_{ip}}{N_p}$$

Equation 15.1

The mean predicted cost for individuals registered with each practice is given by Equation 15.2:

$$\bar{\hat{c}}_p = \frac{\sum_{i \in p} \hat{c}_{ip}}{N_p}$$

Equation 15.2

Where N_p is the number of individuals registered with practice p.

Within the practice sample, we also include statistics which can be used to compare against a base model; these statistics help to identify the impact of including specific variables on the each practice’s estimated share.

15.1.1 Adjusted R^2 (R-squared)

The R-squared statistic measures the amount of variation in the dependent variable that is explained by the model. The calculation of this metric is given by Equation 15.3 and the calculation of the adjusted R-squared is given by Equation 15.4:

$$R^2 = \frac{1/(N-1)\sum_{i=1}^{N}(\hat{y}_i - \bar{y})^2}{1/(N-1)\sum_{i=1}^{N}(y_i - \bar{y})^2}$$

Equation 15.3

Where $y = \text{total cost}, \hat{y} = \text{predicted cost}, \bar{y} = \text{mean cost}, N = \text{sample size}$

$$\text{adjusted } R^2 = 1 - \frac{(1 - R^2)(N - 1)}{N - k - 1}$$

Equation 15.4

Where N = sample size and k = number of parameters

Model predictive accuracy is measured by the adjusted R^2statistic, where the larger the value, the better the model is at predictions. The adjusted R^2statistic is used instead of the R^2statistic. This is because the adjusted R^2statistic accounts for the number of parameters within the model. If variables which do not improve the model fit are added, the value estimated using the R^2statistic will always increase. As the adjusted R^2 punishes the use of insignificant parameters, the goodness of fit is a more accurate representation of a model's predictive power.

Goodness of fit at practice level is the adjusted R^2 statistic from a regression of $\bar{c}_p$ on $\bar{\hat{c}}_p$.

15.1.2 Mean absolute error (MAE)

This metric measures the average, absolute error in the model's prediction of total costs compared with observed costs. This statistic is in pounds sterling (£). Therefore the higher the value the MAE, the higher monetary value of the difference between predicted and observed costs. The calculation of the MAE is given by Equation 15.5 :

$$MAE = \frac{|(y - \hat{y})|}{N}$$

Equation 15.5

Where $y =$ actual costs, $\hat{y}$ = predicted costs and N = number of observations

MAE is used on the estimation sample to calculate the mean absolute error on an individual level. MAE is used on the validation sample to calculate the mean absolute error on a practice level using Equation 15.1, Equation 15.2 and Equation 15.5 in combination.

15.1.3 Proportion Not Within 10%

Measures the proportion of predictions which are not within 10% of observed costs. For this percentage value, the higher the value the worse the goodness of fit as a larger percentage of predictions are incorrect by over 10%. The equations to calculate the proportion of predictions not within 10% are given by Equation 15.6 and Equation 15.7:

$$w_i = 1 \text{ if } \frac{|\hat{y}_i - y_i|}{y_i} > 0.1$$

Equation 15.6

$$\text{Within } 10\% = \frac{\sum w_i}{N}$$

Equation 15.7

Where $y_i =$ actual costs, $\hat{y}_i$ = predicted costs, N = observations and i = individual

This metric is calculated only on an individual level and is therefore conducted on estimation sample only.

15.1.4 Redistribution Index (RI)

The redistribution index compares the level of redistribution for each GP practice when allocations are based on two models. Two base categories were used. The first base category is actual observed costs and the second base category is the Nuffield PBRA specification model.

The Redistribution Index (RI) is the proportion of the total budget that is redistributed from ‘losing’ practices to ‘gaining’ practices when comparing model 1 with a reference model, Model 0. It is calculated by summing the predicted costs at practice level from the two models, taking the absolute values of the differences, summing across all practices, and dividing the sum by two (since the sum of the losses must equal the sum of the gains). The RI falls within the range 0 to 0.5. This is calculated using Equation 15.8:

$$RI^1 = \frac{\sum_p \left| \left(\sum_{i \in p} \hat{c}_{ip}^1 - \sum_{i \in p} \hat{c}_{ip}^0 \right) \right|}{2}$$

Equation 15.8

A low redistribution index value is indicative of a low redistribution of predicted costs between GP practices.

15.1.5 Mean Absolute Percentage Change In Share (MAPCIS)

The Mean Absolute Percentage Change In Share (MAPCIS) summarises the extent to which practice indicative shares are affected by a change in the model from a reference model, Model 0, to an alternative model, Model 1. It is calculated by summing the predicted costs at practice level from the two models, dividing the predicted costs by the national total to obtain the indicative practice shares (s_p), dividing the absolute changes in indicative shares by the shares from the reference model, multiplying by 100, and taking the mean of the practice values. This is calculated using Equation 15.9, Equation 15.10 and Equation 15.11:

$$s_p^0 = \sum_{i \in p} \hat{c}_{ip}^0 / \sum_p \sum_{i \in p} \hat{c}_{ip}^0$$

Equation 15.9

$$s_p^1 = \sum_{i \in p} \hat{c}_{ip}^1 / \sum_p \sum_{i \in p} \hat{c}_{ip}^1$$

Equation 15.10

$$MAPCIS^1 = \left(100 \times \frac{|s_p^1 - s_p^0|}{s_p^0} \right) / P$$

Equation 15.11

This metric is only conducted on the GP validation sample as we are interested in the distribution of shares affected for each GP practice and not each individual.

15.1.6 Percentage of Practice Shares Substantially Affected (PoPShaSA)

PoPShaSA summarises the proportion of practices whose MAPCIS has changed by 5%. This is calculated using Equation 15.12 and Equation 15.13:

$$\rho_p = 1 \; if \; MAPCIS > 5$$

Equation 15.12

$$PopShaSA = \frac{\sum \rho_p}{N_p}$$

Equation 15.13

15.1.7 Information Criterion

Information criteria were not used when assessing the goodness of fit for the final models. These metrics were not used as these statistics are not interpretable and are not bounded by a range of values. Due to these two reasons, the values from the information criterion cannot be used to measure the level of goodness of fit in isolation.

- When metrics are not bounded between two numbers, a change in the sample size will change the goodness of fit of a model, regardless of how well the model fits.

Information criterion offer a larger penalty for including variables which do not aid in the goodness of fit and are therefore suitable metrics when comparing across different models. The lower the value of the information criterion metrics, the better the goodness of fit.

Information criterions were used to assess the goodness of fit when testing for the optimal number of diagnostic fields to use to generate the set of morbidity flags. These metrics were used alongside the adjusted R-squared and Mean Absolute Error.

- When looking at the number of diagnostic positions to account for, the models that tested this has the same number of observations and variables, therefore the goodness of fit using Akaike and Bayesian Information Criterion was appropriate.

15.1.8 Akaike Information Criterion

The Akaike Information Criterion is calculated using Equation 15.14 and Equation 15.15:

$$AIC = N \times \ln\left(\frac{RSS}{N}\right) + 2 \times K$$

Equation 15.14

Where N = number of observations, K = number of parameters, RSS = residual sum of squares

$$RSS = \sum_{ip=1}^{N} (y_{ip} - f(x_{ip}))^2 = \sum_{ip=1}^{N} (e_{ip})^2$$

Equation 15.15

Where x_{ip} = all covariates, individual i, in practice p, e is the error term and y denotes total cost.

15.1.9 Bayesian Information Criterion

The Bayesian Information Criterion is calculated using Equation 15.14 and Equation 15.16:

$$BIC = N \times \ln\left(\frac{RSS}{N}\right) + K \times \ln(N)$$

Equation 15.16

Where N = number of observations, K = number of parameters, RSS = residual sum of squares