Primary medical care – new workload formula for allocations to CCG areas

Authors: Lindsay Gardiner, Kath Everard

NHS England Analytical Services (Finance)

NHS England INFORMATION READER BOX

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

Publications Gateway Reference:	**05104**
Document Purpose	Resources
Document Name	Primary medical care – new workload formula for allocations to CCG areas
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 primary medical care formula 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Superseded Docs (if applicable)	N/A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	Lindsay Gardiner Analytical Services (Finance) Skipton House 80 London Road, London SE1 6LH 0113 824 9221

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.

Primary medical care – new workload formula for allocations to CCG areas

Version number: 1.0

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

1 Introduction

The previous formula for primary medical care (GP services) allocations to Area Teams and prior to that Primary Care Trusts (PCTs) was based on the contractual formula that is at the heart of the General Medical Services (GMS) contract, usually referred to as the Carr-Hill formula.

NHS England asked ACRA to advise on a new formula for primary medical care to be used to allocate budgets to CCG areas from 2016-17.

The key change to the formula is new estimates of workload per patient by age-sex group, which are used as the relative weights per head for allocations. Data were not available to update the relative costs, such as those related to rurality.

This paper summarises the modelling which was endorsed by ACRA. The results presented to ACRA have been pulled together into this single paper, along with details of the dataset used for the modelling.

2 Preferred model

2.1 Model outline

The analysis modelled workload as measured by the length of time patients' file are open (the dependent variable). The factors included in model were age and sex group, deprivation (measured by decile of the Index of Multiple Derivation (IMD)), the number of new registrations and rurality.

A number of other potential factors were considered but were either not available in the anonymised dataset, the data were not of sufficient quality, or data were not available for every GP practice in the country to permit implementation. This is set out in Annex A.

2.2 ACRA's recommendations

2.2.1 Attributed deprivation score

ACRA considered two options for the final workload models (Model 2 and Model 3), that used different methods to address the problem that we had a systematic group of patients that had no information concerning their IMD. ACRA felt both were improvements on the original model, with no attribution, and recommended that Model 2 be used for allocations to CCG areas.

For Model 2 the IMD data missing for new patients was imputed using a model of the likelihood that a new patient at a particular practice would have an IMD decile of 1, 2, 3….. The imputation was carried out multiple times to reflect that the model was probabilistic.

Model 3 dealt with the missing IMD scores by developing two regressions, one for patients who were in the practice at the start of the year which included IMD as a predictor and one for patients who joined during the year, which did not include IMD as a predictor and combining these together to calculate overall workload.

The recommended model is presented in section 3 below.

2.2.2 Rurality

ACRA considered whether rurality should be included as a factor in determining workload. They considered that it should be excluded because of the lack of certainty over whether it was reflective of additional workload or systematic behaviour in rural practices not arising from workload. This is discussed at 3.3 below.

2.2.3 Deprivation

ACRA looked at the impact on allocations at practice level of using model 2 or model 3 relative to Carr-Hill. They registered some unease that, under the new model the weightings for practices in deprived areas were lower, on average than under Carr-Hill. They reflected that the formula still had a positive gradient in respect of deprivation so was still progressive. They also considered that the narrower range of weightings under the new formula inevitably meant that practices that had the highest weightings previously (typically northern and deprived) would see a reduction. The information on impact of the new models is presented in 3.4 below.

2.2.4 New patients

For previous allocations the impact of new patients has been sterilised in the formula. ACRA advised that new patients should not be sterilised in the formula because the data shows a significantly higher workload associated with this group, they are not a random cross section of the population (being disproportionately very young or very old) and it cannot be assumed that they have moved from another practice within the same CCG.

2.2.5 Quality assurance

Quality assurance of our coding was undertaken by Deloitte who considered that the code was doing what we intended it to do. We undertook some more checking of the data extract and had some reassurance concerning apparent anomalies. However, we were not able to undertake a comprehensive data check in the time available.

3 Model 2

3.1 Methodology

3.1.1 Linear mixed effects model

We fitted a linear mixed effects model to the Clinical Practice Research Datalink (CPRD) data to estimate the effect of patient and practice characteristics on GP practice workload, the model is of the form:

$$\begin{aligned}&\text{Total weighted file opening time}\\&= \text{Age sex band} + \text{New registration} + \text{IMD decile}\\&+ \text{Rurality} + \text{Practice ID}\end{aligned}$$

Equation 3.1

The weights for the file opening times are by staff type, giving a lower weight to administrative staff than medical staff[1].

Registrations for part of the year are included in this particular model as though the person had been there for the whole year. Although the workload of such patients arises only in the part of the year that they are registered for, the data shows that their workload is the same or higher as full year patients with the same demographics. In applying this particular model we use an estimate of new registrations (based on the latest year of data on registrations) and their likely demography applied to each practice. We use the formula to estimate the workload for this subset of patients and add it to the workload estimate for the existing practice list.

3.1.2 Imputed IMD values

We imputed IMD values for individual patients that did not have IMD values associated with them (due new registration post IMD linkage[2]). To do this we used multiple imputation implemented in R. This uses a multinomial logit model to probabilistically assign the IMD of a patient with missing data to a decile multiple times. The model used for imputation is of the form:

$$\text{Odds that patient is in a particular decile} = \text{Age sex band} + \text{New Registration} + \text{IMD decile} + \text{Rurality} + \text{Morbidity count} + \text{Practice ID}$$

Equation 3.2

This model allows a random selection of IMD decile for each patient with a missing IMD decile based on the odds that they reside in each decile given their characteristics and practice.

Multiple imputation was used, which means that we imputed the IMD value for each patient with a missing IMD multiple times, so the variation between these will reflect the uncertainty in the imputation model. The 'file opening time' model is then fitted to each of the datasets in turn (with the first imputed values, then the second imputed values and so on)[3], and the results of these models are combined and jointly analysed to give combined coefficients and standard error estimates incorporating the additional uncertainty associated with the imputation process.

---

[1] The file opening times have been weighted based on the relative salaries of the type of staff member the opening was associated with. The weights used are:
i) GP =1 (HSCIC Income and Expenses 2011-12)
ii) Practice Nurse = 0.25 (NHS Staff Earnings Estimates for qualified nurses AfC band 5)
iii) Practice Administrator = 0.21 (HSCIC Income and Expenses 2011-12)

[2] We have not imputed data for patients that were missing due to individual opt out, change of LSOA boundaries or where the practice refused to allow linkage for any of its patients. We prioritised understanding the impact of new patients given the computational capacity and timeframe available, and further work would be required to consider the appropriateness of this methodology for the other missing values given the nature of the absence of these data.

[3] We tested different numbers of iterations and found that the results were sufficiently stable after 20 iterations.

3.2 Results

The fitted coefficients are in Table 3.1 below together with t-statistics and p-values.

Table 3.1: Imputed data analysis with 20 imputations

	Coefficients	Std error	t-stat	p-value
Intercept	**34.7**	0.9	37.7	0.000
Male 5 - 14	**-22.4**	0.4	-63.1	0.000
Male 15 - 44	**-17.2**	0.3	-56.0	0.000
Male 45 - 64	**6.7**	0.3	20.8	0.000
Male 65 - 74	**41.1**	0.4	109.9	0.000
Male 75 - 84	**80.5**	0.4	187.8	0.000
Male 85+	**116.7**	0.6	195.7	0.000
Female 0-4	**-3.2**	0.4	-7.9	0.000
Female 5 - 14	**-20.9**	0.4	-58.6	0.000
Female 15 - 44	**9.1**	0.3	29.5	0.000
Female 45 - 64	**25.7**	0.3	80.0	0.000
female 65 - 74	**48.1**	0.4	130.2	0.000
Female 75 - 84	**89.4**	0.4	220.5	0.000
Female 85+	**123.5**	0.5	259.4	0.000
IMD decile 2	**1.3**	0.2	5.3	0.000
IMD decile 3	**2.5**	0.2	10.6	0.000
IMD decile 4	**4.0**	0.2	16.2	0.000
IMD decile 5	**5.2**	0.3	19.5	0.000
IMD decile 6	**5.6**	0.3	20.0	0.000
IMD decile 7	**7.8**	0.3	28.0	0.000
IMD decile 8	**9.3**	0.3	32.7	0.000
IMD decile 9	**10.3**	0.3	33.7	0.000
IMD decile 10	**13.7**	0.4	35.4	0.000
Rural	**4.5**	2.1	2.1	0.033
New Registration	**5.1**	0.2	25.5	0.000

The intercept represents the estimated average number of weighted contact minutes per year that a patient on the registration list at the start of the year with baseline characteristics has with their GP surgery. In this model that is a male patient aged 0-5, in IMD decile 1, in a practice in an urban area.

Each coefficient represents the estimated average number of additional weighted contact minutes per year that a patient with those demographic characteristics has with their GP surgery compared to a patient with the baseline characteristics. So for example, a patient who was in IMD[4] decile 10 would have 13.7 additional weighted contact minutes compared to a patient with the same demographic characteristics in IMD decile 1.

The coefficients follow the expected pattern and are highly significant.

We compared the weightings of each GP practice under model 2 and model 3 and these are very similar. The correlation coefficient is 0.9986.

[4] We have used IMD2010 as this was the data that was linked to CPRD at the time of our data extraction (and CPRD would not have performed a new linkage since IMD 2015 was published).

3.3 Rurality

3.3.1 Positive coefficient

In our models we have included rurality as an explanatory factor of workload, based on the 2011 Census definition of rurality. We have found that rurality has a positive and significant impact on workload, equal to around four minutes per patient per year.

3.3.2 Possible explanations

There are two possible explanations of the link between longer aggregate file openings and rurality (once demography and deprivation have been taken into account):

a) in rural areas patients consult with their GPs over conditions they would take to a different provider of healthcare services if one were available locally; or

b) in rural areas practices are less rushed and are able to devote more time to consultations, or recommend additional appointments.

3.3.3 Average duration and average number of file openings

We used our CPRD sample to estimate the average duration of an appointment and average number of file openings in urban and rural areas by age. These are shown in Table 3.2.

The ratio of rural duration to urban duration was above 1 for some age groups and below for other suggesting that *file openings in rural areas are not systematically longer*. However, for 15-44 year old women, file openings are 10% longer in rural areas.

We also looked at the average number of file openings. Here the ratio of rural to urban was consistently greater than 1, indicating that *there were more file openings in rural areas.* However, we cannot say whether these additional openings related to supply or demand side factors.

Table 3.2: Rural / urban ratio of duration and number of file openings

	Ratio of average duration of opening rural / urban	Ratio of average number of openings rural / urban
Male 0-4	1.015903	1.004811
Male 5-14	1.061438	1.013476
Male 15-44	1.066692	1.049934
Male 45-64	0.99287	1.044042
Male 65-74	0.976925	1.042205
Male 75-84	1.014924	1.048767
Male 85+	1.056232	1.032668
Female 0-4	1.002325	1.007326
Female 5-4	1.040613	1.024199
Female 15-44	1.101448	1.028055
Female 45-64	0.955983	1.033963
Female 65-74	0.953966	1.050248
Female 75-84	0.983393	1.036265
Female 85+	0.987621	1.020401

3.3.4 Carr-Hill adjustments

It should be noted that Carr-Hill does not currently adjust workload for rurality, and that allocations have not taken account of rurality in the cost adjustment either.

3.3.5 ACRA recommendation

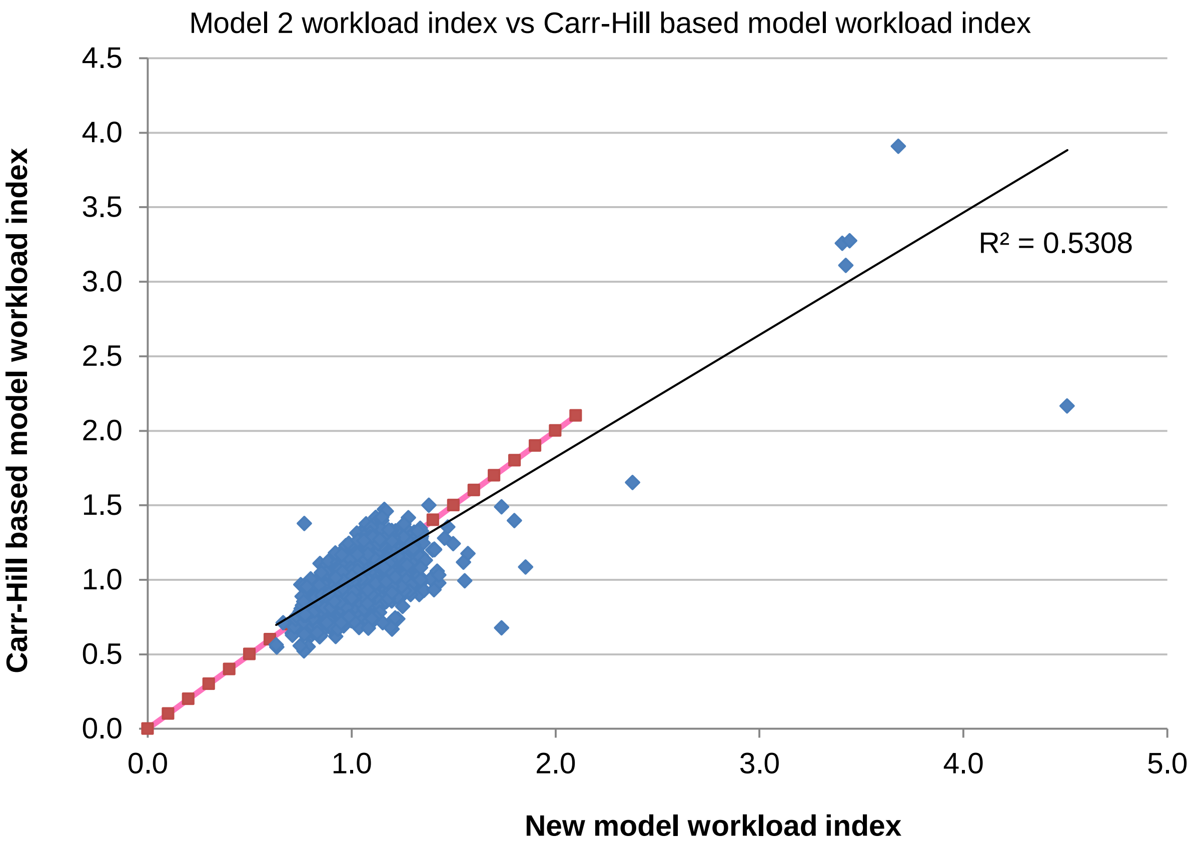
ACRA advised that rurality should be excluded from allocations, therefore the models were rerun excluding rurality. This gives a very small change to the intercept, the new value is 35.4. The remaining coefficients are unchanged, as they are measured and modelled at the patient level.

3.4 Impact on allocations relative to Carr-Hill

3.4.1 Model 2 - results

We have compared the weighting of each GP practice under Model 2 with the weightings under Carr-Hill. Figure 3.1 shows the two weighting values.[5]

Figure 3.1: Scatterplot showing the two weighting values



3.4.2 Distribution of weightings

The general distribution of weightings under the new formula is slightly narrower than under Carr-Hill. The central 90% of practices have indices between 0.87 and 1.16 under the new model (excluding rurality) compared with 0.83 and 1.20 under Carr-Hill.

[5] Practices without comparable Carr-Hill weights have been excluded (10 of 7,711 practices).

3.4.3 Change to the weightings

There is notable change to the weightings of individual practices. The correlation between a practice's weighting under Carr-Hill and under Model 1 is 0.53 without rurality included in the model and 0.47 if rurality is included.

3.5 More detail on gainers and losers under Model 2

3.5.1 Overall

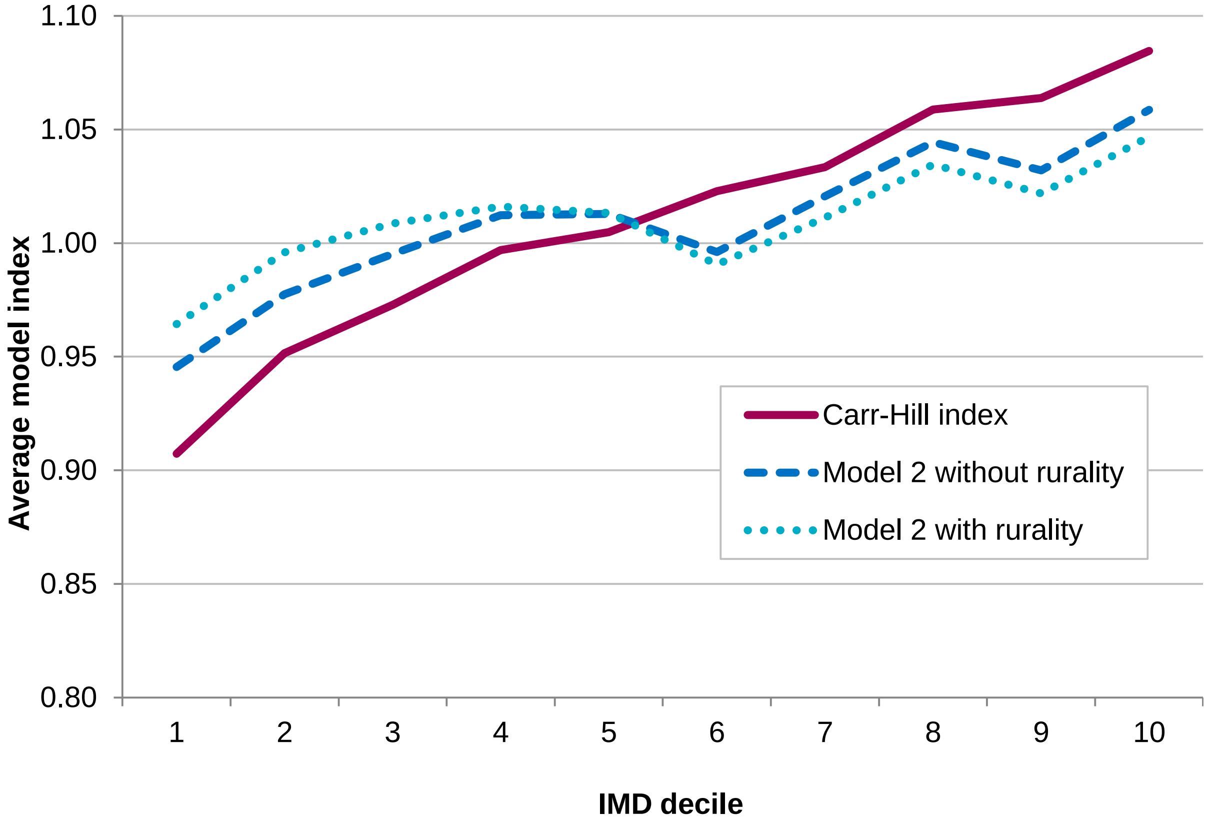
3,878 practices see an increase in their weighting, 3,832 see a decrease.

Taking (Model 2 weight) / (Carr-Hill weight) for each practice, we see an interquartile range from 0.95 to 1.05. This indicates that 50% of practices see a relatively small change in their weighting.

3.5.2 Deprivation decile (IMD 2010)

Figure 3.2 below shows the average index for Carr-Hill (solid purple), Model 2 excluding rurality (dashed dark blue) and Model 2 including rurality (dotted light blue) by deprivation decile. This shows that whilst there is an increase in the average index of the least deprived and a decrease in the average index of the most deprived deciles, this is part of a general narrowing of the range of indices. The general relationship of increasing index as deprivation increases is preserved. Note that including rurality in Model 2 further moves the average index of each decile towards 1.

Figure 3.2: Average model index by deprivation decile



3.5.3 List size

There appears to be a slight negative correlation between Model 2 weighting and list size (a few very small practices still have high weights but had even higher weights under Carr-Hill, though these tend to have lots of older patients which increases the estimated workload). Our ratio of (Model 2 weight) / (Carr-Hill weight) is closer to 1 as practices get bigger.

3.5.4 Region

Table 3.3 below shows the average change by region. Regions that have an average Carr-Hill index of below 1 having an increase in index under Model 2 and those with an average Carr-Hill index of above 1 have a decrease in index under Model 2. The final column shows the comparable figures for Model 2 including rurality.

Table 3.3: Average change by region

Region	No of practices	Average list size	Average Carr-Hill index	Average Model 2 index (excluding rurality)	Change Carr-Hill - Model 2 (excluding rurality)	*Average Model 2 index* (including rurality)
Wessex	312	9,096	0.988	1.023	▲	*1.027*
London	1,388	6,750	0.922	0.984	▲	*0.973*
Yorkshire and the Humber	762	7,514	1.046	1.021	▼	*1.021*
Lancashire and Greater Manchester	715	6,261	1.095	1.017	▼	*1.009*
Cumbria and North East	467	6,950	1.126	1.038	▼	*1.047*
Cheshire and Merseyside	398	6,460	1.129	1.049	▼	*1.042*
North Midlands	503	7,305	1.054	1.015	▼	*1.019*
West Midlands	664	6,621	1.034	1.031	▼	*1.026*
Central Midlands	559	8,563	0.951	0.963	▲	*0.972*
East	542	8,174	0.979	1.013	▲	*1.028*
South West	396	8,187	1.025	1.059	▲	*1.076*
South East	581	8,117	0.974	1.006	▲	*1.012*
South Central	423	8,949	0.914	0.949	▲	*0.960*
Total	**7,710**	**7,426**	**1.009**	**1.009**		***1.010***

3.5.5 Rurality

Including rurality in the model slightly increases the spread of the weightings by practice. The central 90% of practices have weightings between 0.86 -1.18, compared with 0.87-1.16 under the model excluding rurality.

4 Annex A - Data for workload modelling

4.1 Data requirements

4.1.1 Factors influencing workload

We needed to measure general practice workload and consider how the attributes of practices and the patients in them influenced that workload. We therefore needed a dataset that contained a proxy for workload and data on patient and practice characteristics associated with that workload.

4.1.2 Data required across all patient groups

In order to capture the variation experienced by the majority of practices we required information on the workload associated with all patient groups (where groups are defined by age, sex, morbidity and any other factor that may drive practice workload) and all sizes and locations of practice.

4.1.3 Sufficient depth and sample size

We needed to have data relating to all patient groups in sufficient depth to analyse the impacts on workload of these groups. However, the number of patients in each group in the sample did not need to be proportionate to the numbers in each group nationally in order to do this. This means that the sample could contain smaller numbers of patients in some groups without undermining the analysis, provided there were enough patients in each group. This is discussed in more detail later.

4.1.4 Possible factors

There are a lot of factors that could potentially influence workload. Our analysis focuses on the main ones identified by our Technical Group and previous work on general practice formulae, namely:

- *Registration*
  - Age (estimated on year of birth only)
  - Gender
  - Whether patient is new to the practice
  - Temporary & permanent patients
  - Ethnicity (as exists in HES and primary care record)
  - Practice list size (estimated at midpoint of the calendar year)
- *Consultation*
  - Staff group & number
  - Consultation type and number
  - Number of home visits
- *Morbidity*
  - QOF indicator condition
- *Deprivation*
  - IMD – index only quintile, deciles or twentiles
- *Geography*
  - Rurality at the practice level (2001Census)
  - Geographical region (SHA only)

4.1.5 Other influences

Arguments have been made in support of a wider array of influences that could have a significant impact on some practices. We have not been able to incorporate this full array into the analysis for a variety of reasons:

a) it may be that our sample data set does not have enough practices with a particular circumstance to measure the impact of that circumstance robustly;

b) it may be that there are no national data available on the circumstance that would enable it to be calculated for all practices in the country for payment purposes; and

c) we have had to trade-off between the comprehensiveness of the formula and its simplicity.

4.2 Possible data sources

There were two main sources of data that could have met the requirements set out above:

a) a diary study; and
b) extraction of information from GP records.

4.2.1 Diary study

A diary study would involve a number of GP practices recording the activities of all staff over a period of time, usually a week. This information would be linked to summary information about patients in the practice and the practice itself in order to assemble a set of data on which to test how patient and practice characteristics drove practice activity.

Although a diary study would have given more tailored information on which to conduct analysis, there were a number of challenges that made this route less attractive, namely:

a) need for an extremely large study to get enough responses for modelling, as only a proportion of practices will respond;

b) completing the study would have been onerous on respondents;

c) a very long lead time would have been needed to set up the study and obtain the approvals for it, making delivery in time uncertain;

d) data would have required extensive validation to check for bias including selection bias of those responding, response bias within results and impact of selecting one week in the year over another;

e) even if tailored information was available, in order to use it in the resulting formula it would have to be available for all practices in the country; and

f) high cost and less certainty about how usable the results would be.

4.2.2 Extraction of information from GP records

To obtain data from GP records there are a number of organisations that routinely extract data for research purposes. As this approach is used for research, and was used for the original Carr-Hill formula and in the work by the Formula Review Group in 2007[6], we have more information at the outset on the quality of data that will be available. We have therefore been able to assure ourselves that sufficient data are available, and to anticipate their limitations. Although data specification and processing still takes time, the timescales are a few months rather than in excess of a year.

There are some high level limitations of this data, and more detail is examined below:

a) the system only records numbers of file openings and duration as file openings as a proxy for workload, rather than details of all practice workload; and
b) we are limited to the fields of information recorded on the system. In particular, we do not have practice ID so we cannot readily link the practice to local information relating to the practice.

4.2.3 Best approach

Given the time constraints with this work, and the risk of unforeseeable problems from using a diary study as its basis, we considered that using an extraction of data from GP records was the best approach.

We worked with our Technical Group and procurement partners to develop criteria to assess the best supplier of this type of data. We assessed a number of suppliers and selected CPRD, one of the organisations that extracts data from Vision Clinical software.

4.3 CPRD dataset

The Clinical Practice Research Datalink (CPRD) is an ongoing primary care database of anonymised medical records from general practitioners, with coverage of over 11.3 million patients from 674 practices in the UK. With 4.4 million active (alive, currently registered) patients meeting quality criteria, approximately 6.9% of the UK population are included and patients are broadly representative of the UK general population in terms of age, sex and ethnicity.

Figure 4.1 and Figure 4.2 show the percentage representation of age groups in the CPRD population and the UK 2011 Census [7]. There is some difference between the populations[8], but this is relatively small.

[6] http://www.nhsemployers.org/your-workforce/primary-care-contacts/general-medical-services/gms-finance/global-sum-formula

[7] Oxford Journals. 2015. *Data Resource Profile: Clinical Practice Research Datalink (CPRD)*. Available from: http://ije.oxfordjournals.org/content/early/2015/06/06/ije.dyv098.full

[8] What matters for the accuracy of the workload formula is not how much the samples differ but whether they differ in respect of unobserved variables which affect workload. But similarity in respect of observed variables is encouraging since it is plausible that the samples are similar in respect of unobserved variables.

Figure 4.1: Age distribution of the CPRD primary care data (March 2011) compared with UK Census data (2011), men

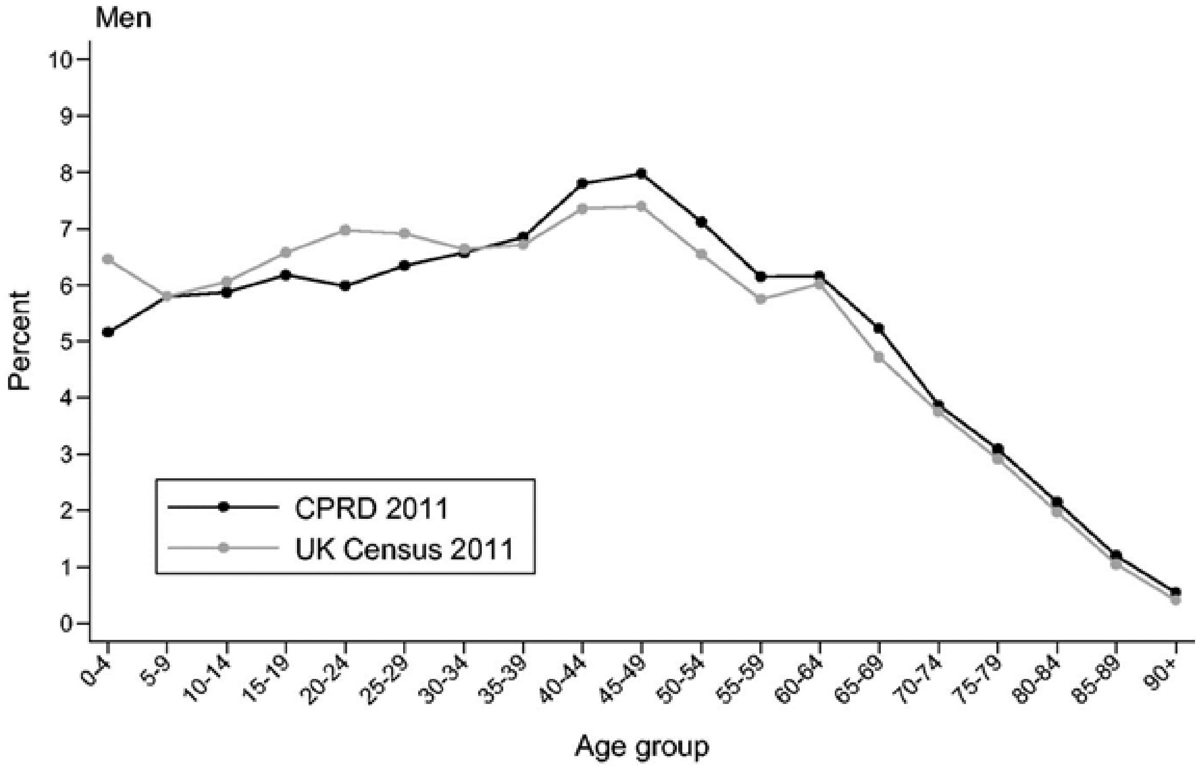


Figure 4.2: Age distribution of the CPRD primary care data (March 2011) compared with UK Census data (2011), women

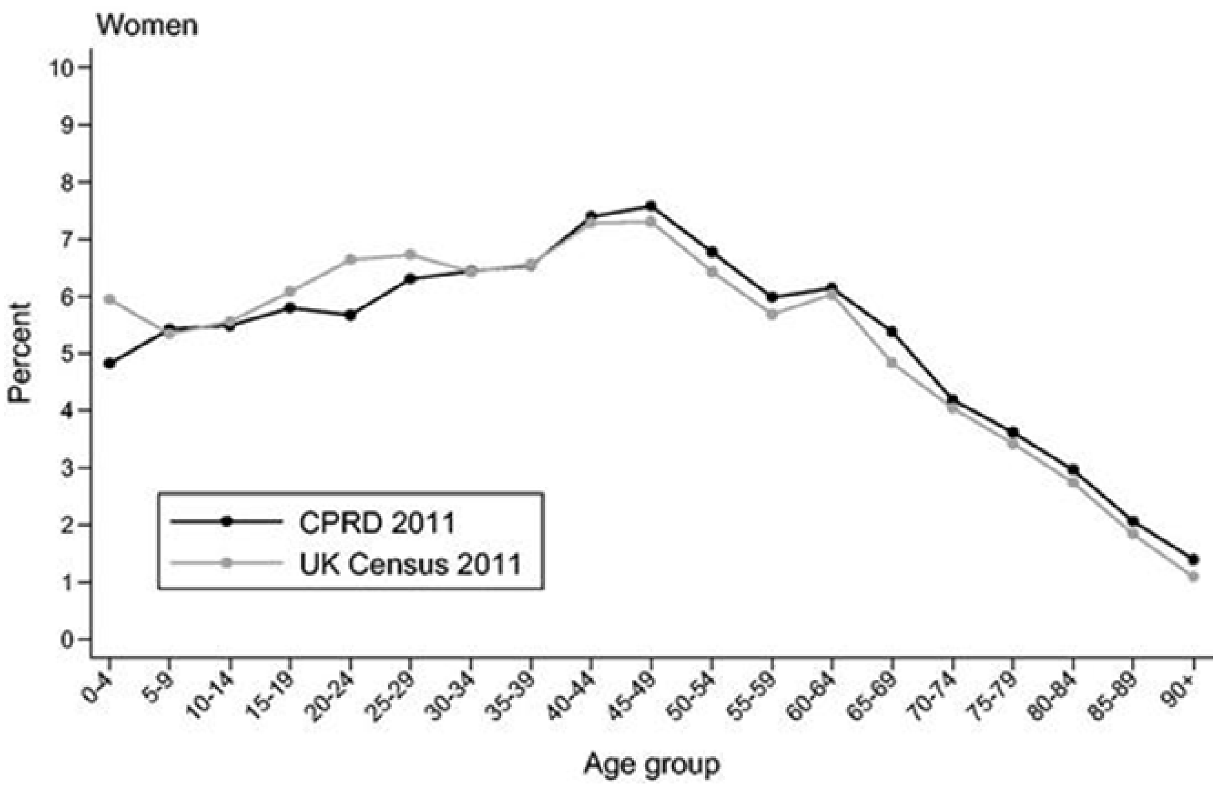


Table 4.1 shows the regional distribution of active patients in CPRD. It contains numbers in each region to support robust analysis, though clearly there are fewer patients in some regions.

Table 4.1: CPRD by age and region (2013 extract)

	Active patients	
No. patients	4,425,016	
Sex	**Number in CPRD**	**% of CPRD**
Men, n	2,183,161	49.3
Women, n	2,241,855	50.7
Age in 2013	**Number in CPRD**	**% of CPRD**
<18	742,765	20.2
18-64	4,402,926	61.8
65+	1,728,514	18.1
Region	**Number in CPRD**	**% of CPRD**
North East	67,639	1.5
North West	52,3356	11.8
Yorkshire & The Humber	48,480	1.1
East Midlands	29,954	0.7
West Midlands	39,4115	8.9
East of England	306,538	6.9
South West	377,821	8.5
South Central	544,979	12.3
London	600,824	13.6
South East Coast	474,593	10.7
Northern Ireland	153,576	3.5
Scotland	499,969	11.3
Wales	403,172	9.1

4.3.1 Size of dataset used

Of the total set of practices, we utilised data on 272 that are in England and have "research quality" information.

4.4 CPRD specification

Our specification to CPRD included the following significant aspects of data cleaning:

a) short and long file openings;

b) openings relating to work not paid for through the global sum;

c) handling of long term conditions from which recovery is possible; and

d) ethnicity.

These are discussed in turn.

4.4.1 Short and long file openings

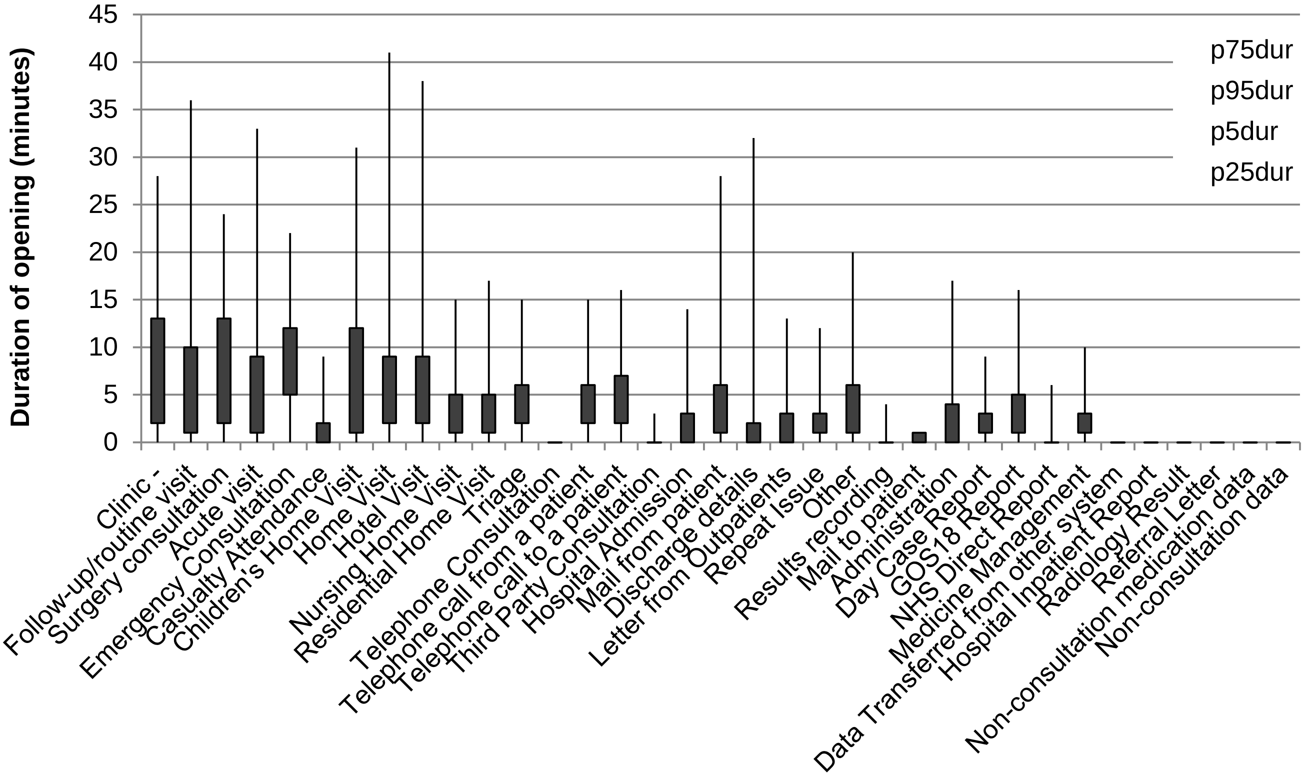
Vision Clinical Software rounds file openings down to the nearest whole minute. As a result there are file openings of 0 length with activity associated with them.

Conversely, if a patient record is left open then Vision will record the whole time that the file is open, even if there is no activity. Thus there are some file opening times of 23 hours 59 minutes in the data.

4.4.2 Range of durations of file openings

Figure 4.3 shows the range of durations of file openings. The solid bar shows the inter quartile range of file openings (e.g. for clinics, 25% of file openings are less than two minutes, and 25% are above 12 minutes with 50% of openings between this range). The line shows the range for 90% of openings with only 5% being longer than this and 5% being shorter. At least 5% of openings for all consultation types are recorded as zero minutes.

Figure 4.3: Range of durations of file openings



4.4.3 Adjustment for short file openings

We therefore specified that each file opening should have 30 seconds added to it to, which matches the data specification used in the original Carr-Hill work. This means that all activity is shown as significant but the difference between short activity and longer activity is retained.

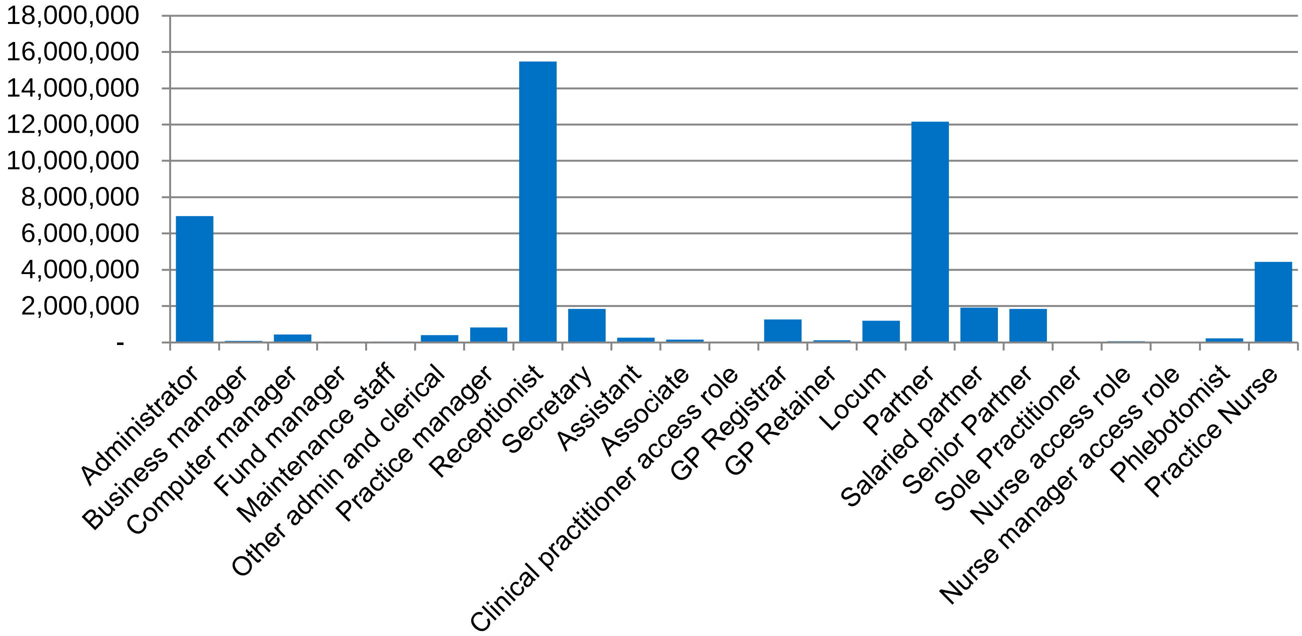
4.4.4 Adjustment for long file openings

The data also showed that the majority of file openings were less than 30 minutes. For openings by staff in all clinical roles, at least 95% were less than 29 minutes. 95% of openings were less than 19 minutes for partners. For administrative staff more than 95% of files were open for less than 15 minutes (except for secretaries, 95% less than 24 minutes). We therefore set an upper band of 30 minutes for a file opening and file openings longer than this were truncated. Although some file openings in excess of 30 minutes might reflect patient work, it was considered more likely that long openings related to staff leaving the file open whilst undertaking other work.

4.4.5 Openings for activities not paid through the global sum

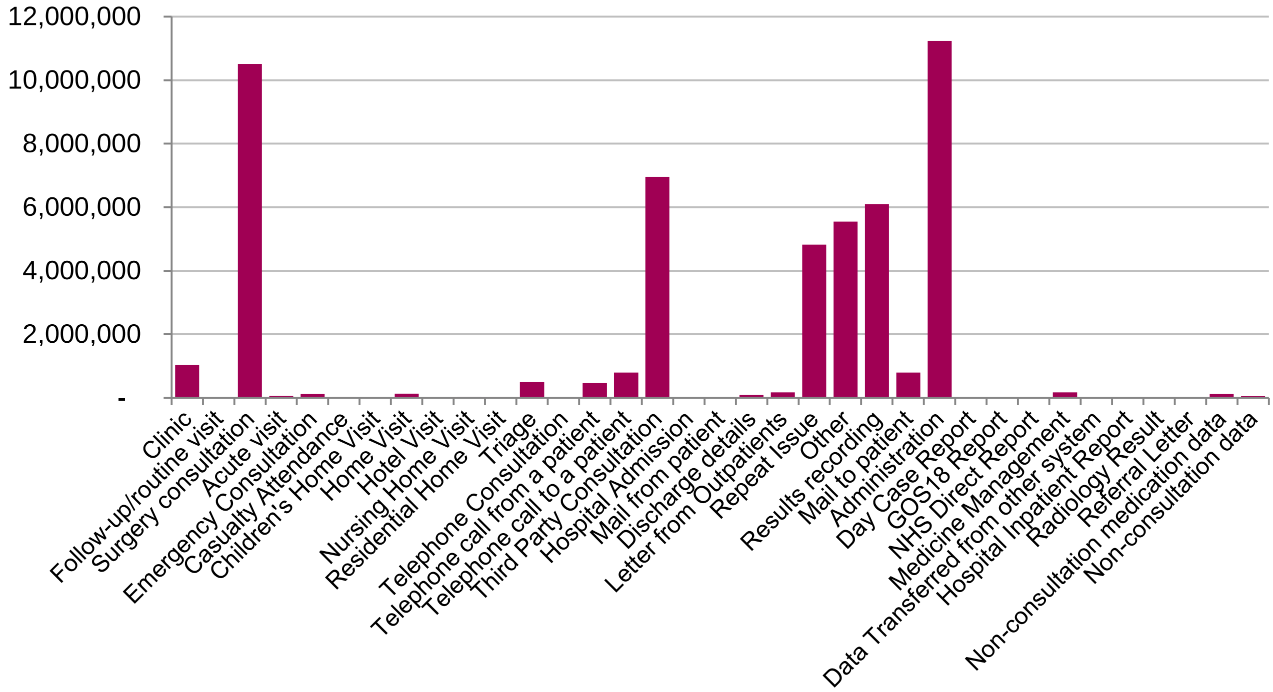
In a practice there are a variety of staff that might have access to patient records. Some will be the key staff paid by the global sum, such as the admin staff, nurses and GPs. However, others may be paid via different contractual arrangements, such as health visitors. We excluded the latter type from the data set. Figure 4.4 gives the number of file openings by staff type.

Figure 4.4: Number of file openings by staff type



There are also a range of consultation types, some of which are out of scope. Figure 4.5 shows the number of file openings by consultation type recorded on the system.

Figure 4.5: Number of file openings by consultation type



These values should be considered alongside the information above when considering workload. There are a large number of file openings associated with the recording of results but these tend to be very short openings. For example, there are a large number of openings for the purposes of results recording but generally these are very short.

4.4.6 Handling of long term conditions

We considered using Quality and Outcomes Framework (QOF) prevalence data in the CPRD data as an indicator of long term conditions, as these codes are already used for GP payments and are available for all practices in the country. Once a GP records in Vision that a patient has a condition, then this will always be associated with that patient except in the case of COPD and depression, in which case a GP can then record COPD or depression resolved, if that is the case. In practice, the patient notes record relatively few cases of the condition being resolved. It was out with the scope of this project to investigate why that might be, but to address the possibility that this was a product of under-recording of cases being resolved we included the time period of diagnosis of both of these conditions in the data set. Hence, we could test whether there was lower workload associated with a patient with depression diagnosed in 2007 relative to one diagnosed in 2015.

We were not able to use the QOF data in the end, as the QOF definitions in the CPRD data were different to the national definitions, and so the coefficients form any models could not be applied to all GP practices in the country.

4.4.7 Ethnicity

Ethnicity is available in CPRD through links to the Hospital Episode Statistics. This means ethnicity information is available for patients that have had a hospital episode whilst registered with their current GP and that GP using the Vision system. This means that there is ethnicity coding for approximately 70% of patients. However, on inspecting the data we have found that that the recorded ethnicity could not be related to even a simple grouping of “white” and “non-white”.

4.4.8 Temporary residents

We obtained data on temporary residents from CPRD. However, we found that the quality of coding for such registrations was poor (no consistency in giving end dates for the registration and missing data on patient characteristics) so we could not include them in the model. Furthermore, there is no national data available on numbers of temporary residents attending practices, making it impossible to apply a “temporary resident” coefficient.