THE UNIVERSITY of York

CHE
Centre For Health Economics

Modelling Individual Patient Hospital Expenditure for General Practice Budgets

CHE Research Paper 73

Modelling Individual Patient Hospital Expenditure for General Practice Budgets

[a]Hugh Gravelle
[a]Mark Dusheiko
[b]Steve Martin
[c]Pete Smith
[a]Nigel Rice
[d]Jennifer Dixon

[a]Centre for Health Economics, University of York
[b]Department of Economics and Related Studies, University of York
[c]Imperial College, University of London
[d]Nuffield Trust, London

December 2011

Background to series

CHE Discussion Papers (DPs) began publication in 1983 as a means of making current research material more widely available to health economists and other potential users. So as to speed up the dissemination process, papers were originally published by CHE and distributed by post to a worldwide readership.

The CHE Research Paper series takes over that function and provides access to current research output via web-based publication, although hard copy will continue to be available (but subject to charge).

Acknowledgements

The paper is based on work done for the second round of the Person Based Resource Allocation project which was funded by the Department of Health (DoH). The views expressed are those of the authors and not necessarily those of the DoH. We are grateful to Paul Dixon for costing the HES data and to Rob Shaw for suggestions and comments. The PBRA team was - Health Dialog: Michael DeLorenzo; London School of Hygiene and Tropical Medicine: Colin Sanderson; New York University: John Billings; Nuffield Trust: Jennifer Dixon, Martin Bardsley, Theo Giorgiou; University of York: Paul Dixon, Mark Dushieko, Hugh Gravelle, Steve Martin, Nigel Rice; Imperial College: Pete Smith.

Disclaimer

Papers published in the CHE Research Paper (RP) series are intended as a contribution to current research. Work and ideas reported in RPs may not always represent the final position and as such may sometimes need to be treated as work in progress. The material and views expressed in RPs are solely those of the authors and should not be interpreted as representing the collective views of CHE research staff or their research funders.

Further copies

Copies of this paper are freely available to download from the CHE website www.york.ac.uk/che/publications/ Access to downloaded material is provided on the understanding that it is intended for personal use. Copies of downloaded papers may be distributed to third-parties subject to the proviso that the CHE publication source is properly acknowledged and that such distribution is not subject to any payment.

Printed copies are available on request at a charge of £5.00 per copy. Please contact the CHE Publications Office, email che-pub@york.ac.uk, telephone 01904 321458 for further details.

Centre for Health Economics
Alcuin College
University of York
York, UK
www.york.ac.uk/che

Abstract

The English NHS has introduced a system of budgets for general practices covering hospital expenditure for the patients on their lists. We model individual expenditure using diagnostic information from previous hospital spells, plus a large set of attributed variables measuring population, general practice, and local hospital characteristics. We show that, despite the large proportion of zero expenditures and the heavy right tail of expenditures, estimating models of untransformed expenditure via OLS yields better predictions at practice level than one or two part models using OLS with transformed expenditure or Generalised Linear Models. We describe a procedure for setting budgets for general practices which reduces the problem of the lags in the available data. We examine the distinction between need and non-need variables and the incentive implications of allowing past numbers of hospital encounters to determine practice budgets.

1. Introduction

The English National Health Service is funded almost entirely from general taxation and patients face no charges when consuming health care.[1] Patients register with a general practice which acts as a gatekeeper for access to elective hospital care. The average practice has a patient list of 6500 and 3.5 general practitioners (GPs), operating as a partnership, not as state employees. The practice receives revenue from a mixture of capitation fees for the patients on the list, lump sums (related, for example, to experience and location), and quality incentives. The practice meets the cost of providing primary care (practice nurses, premises etc.) and shares the practice profit amongst the GP partners.

Hospital care for NHS patients is provided mainly by publicly owned but independent hospitals with a small proportion of NHS elective patients treated in privately owned hospitals. Prospective pricing was introduced from 2003/4 and now covers most hospital care. Until recently all hospitals providing care to NHS patients were paid by local health authorities (Primary Care Trusts) which received a formula determined grant from the Department of Health (see Figure 1). In 2005/6, under a new policy known as Practice Based Commissioning (PBC) practices could choose to hold indicative budgets carved out from the PCT hospital care budget. The practice budget was intended to cover most elective and emergency hospital care. The Conservative Liberal coalition government elected in May 2010 decided to extend PBC and to abolish Primary Care Trusts (DH, 2010) replacing them with consortia of general practices known as Clinical Commissioning Groups (CCGs). The CCGs will, like PCTs, hold hard budgets for hospital care and PBC will be extended to all practices (Dixon et al, 2011).

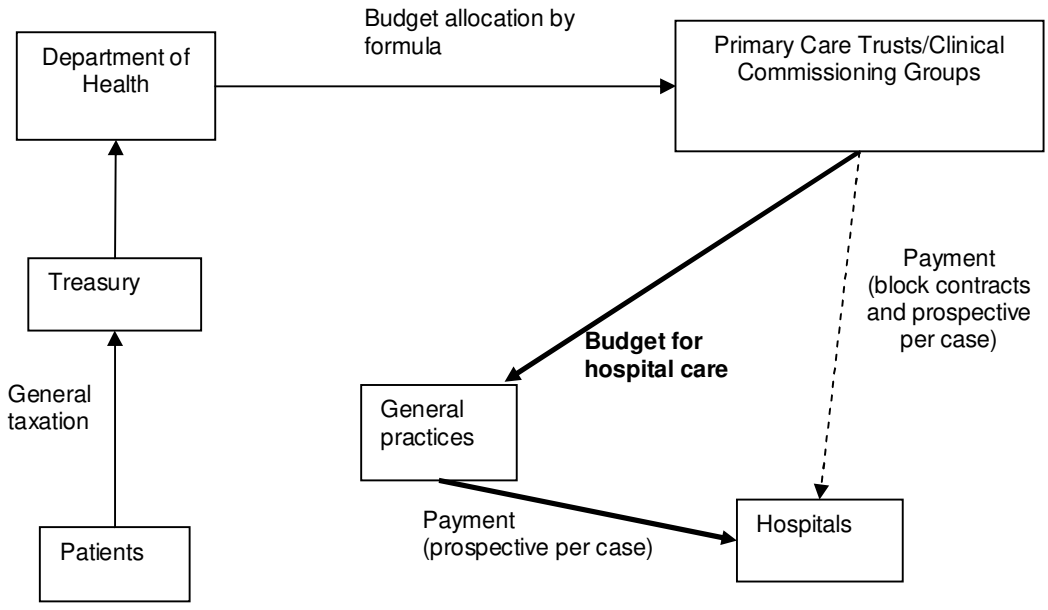


Figure 1. Financial flows in English NHS before and after introduction of Practice Based Commissioning with practice budgets for hospital care.

PBC is an enlarged version of the fundholding system which operated between 1990/1 and 1998/9. Under fundholding, practices could opt to hold a budget for a subset of elective treatments and about 50% of them did so. Fundholding reduced practice admissions by 3%-5% (Dusheiko et al, 2006) and

[1] There is a charge for drugs prescribed by general practitioners but because of wide set of exemptions around 90% of prescriptions are dispensed without charge to patients.

induced hospitals to compete for fundholder business by reducing the waiting times of patients in fundholding practices (Propper et al, 2002; Dusheiko et al, 2004).

Fundholding budgets were set by reference to practice expenditure in the year before the practice opted to become a fundholder and some practices responded by increasing expenditure before becoming a fundholder (Croxson et al, 2001). The Department of Health decided that PBC budgets should be related to the healthcare needs of the practice population.

Need is a normative concept: it is the set of factors that it is believed ought to determine access. It is assumed that an individual's need depends on their current health status which, together with the state of medical technology, determines their capacity to benefit from health care. Need may also depend on socio-economic circumstances. For example, one may wish to deal differently with patients with a given health state depending on their age, gender, home circumstances, or health affecting behaviour. Need or capacity to benefit from care does not depend on supply factors (bed capacity, distances, waiting times etc) which affect the availability of care.

The basic assumption underlying the NHS formulae for allocating funds to PCTs is that the lower level decision makers and front line staff in NHS have better information about the needs of individual patients and take, on average, appropriate decisions about their utilisation of services. But utilisation is not a direct measure of need because in deciding on appropriate use, account is also taken of supply conditions since these also affect the net social benefit from use of the NHS by the individual patient. Thus by examining the relationship between utilisation in different small areas and the health and socio economic characteristics of the area populations it is possible to determine which area population characteristics are good measures of the need for health care and hence to derive a resource allocation formula which is related to those characteristics and thus reflects need. Because supply factors can affect use it is necessary to include these in the empirical modelling and then to sterilize their effects when calculating the allocations.

This paper describes an analogous procedure to calculating a needs-based target allocation for practices. However, our approach is different from previous utilisation based modelling approaches. First, we model the expenditure of individuals, not the mean expenditure of individuals in small areas as in previous modelling for NHS resource allocation (Morris et al, 2007; Sutton et al, 2002; Carr-Hill et al, 1994). Second, we link individuals to practices. Third, we use diagnostic information from past hospital utilisation of individuals to construct individual level morbidity measures. Previous modelling has had to use fairly crude morbidity measures (self assessed health or limiting long term illness) averaged at small area level.

Our approach is therefore similar to that adopted in US and European literature on risk adjustment (Van de Ven and Ellis, 2000). Setting target allocations for general practices using the characteristics of their patients, particularly their past morbidity, has some similarities to developing an insurance premium based on patient risk characteristics. Indeed one could regard general practices as insurers in that they are allocated a risk adjusted capitated sum for the patients on their list with which to commission (purchase) hospital care for the patients' registered with them.[2] The actual budget allocated to practices is somewhere between a needs based target allocation (generated through a resource allocation formula described in this paper) and the historic expenditure on the practice's patients. PCTs will decide, according to local circumstances, on the speed at which actual budget will move toward the needs based target allocation for the practice. At present the budgets set by PCTs for practices are likely to be soft ("notional") with practices intended to break even over several years.

There are three key issues in all types of risk adjustment modelling: the choice of estimation method for modelling expenditure, the set of explanatory variables used, and the distinction between explanatories which should affect premia or allocations and those which should not.

[2] Practices are very small when viewed as insurance pools and PCTs/CCGs will adopt various risk sharing strategies and will encourage practices to pool budgets and commission care jointly with other practices.

Hospital cost data for individuals are non-negative, typically exhibiting a non-normal distribution with a heavy right-hand tail, and a mass point at zero. They also tend to be heteroskedastic. See Figures 2 and 3. These characteristics pose challenges for estimation and ordinary least squares (OLS) may be inappropriate. Alternatives to OLS include generalized linear models (GLM) and transformed ordinary least squares where the dependent variable is transformed by, for example, taking its square-root to help deal with skewness. Two-part estimation separately models the probability of positive cost and the level of cost for individuals with positive cost. This latter stage can be modelled using all the estimators available for one part expenditure models, including OLS, GLM, and transformations of the cost variable.

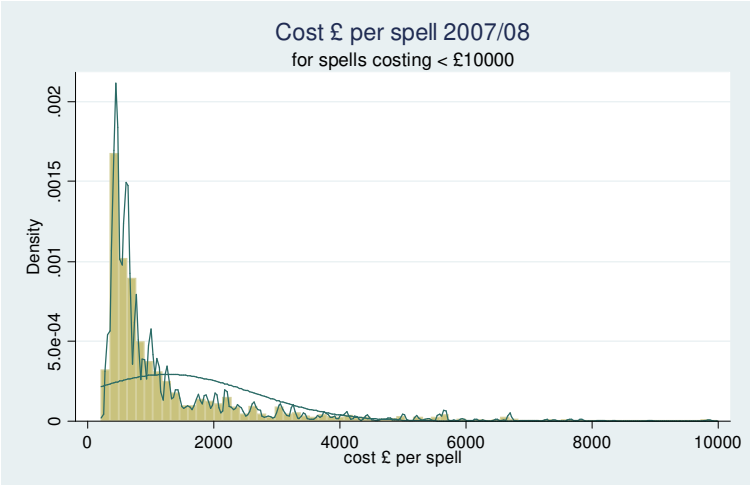


Figure 2 Cost £ per spell 2007/08 for spells costing less than £10,000

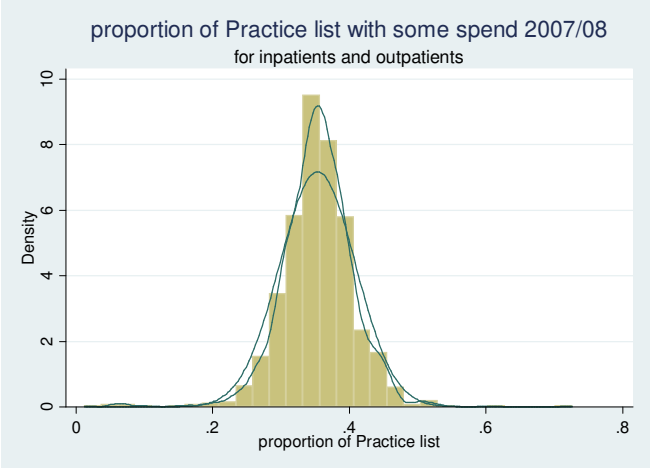


Figure 3. Proportion of Practice list incurring some inpatient or outpatient cost in 2007/08

Our objective is to model individual hospital expenditure within general practices in PCTs to provide predictions of expenditures at per capita practice level to inform the setting of budgets for practices for CCGs. Accordingly, we focus on the predictive ability of alternative estimators when comparing performance.

Given the number of possible sets of explanatory variables (see section 2) it is not feasible to re-examine the choice of estimation method every time the set of explanatories is changed. We therefore first compare estimation procedures using a fixed intuitively plausible set of variables. We then use the preferred estimation method to compare alternative sets of regressors.

Having selected models with good predictive powers we then consider how the explanatories should affect practice budgets. Some variables may make meaningful contribution to explaining the variation in expenditure across individuals but should not affect budgets. Such “non-solidarity” variables (Van de Ven and Ellis, 2000) may be characteristics of individuals, such as ethnicity, which it is felt should not affect allocations, or they may be supply side factors, or they generate perverse incentives.

Section 2 describes the data sets. Section 3 discusses and compares alternative estimators. Section 4 explains how we use the chosen estimation method to select preferred sets of explanatory variables and compares results from two of them. Section 5 describes a procedure for setting practice budgets which minimises the impact of the data lags, discusses potential incentive effects, and the implications for the set of explanatories which should determine practice budgets.

2. Data

2.1 Linking of data sets

We had information from Hospital Episode Statistics on inpatient and outpatient hospital use at individual level for all NHS hospitals in England from 2002/3 to 2007/8. We also had information on general practice patient lists for all patients in England for the same period. We linked the two data sets at individual level using an encrypted NHS identifier. Around 5% of HES records could not be linked at individual level to patients in practices, and the probability of link failure was higher for younger patients and for emergency admissions. We were able to link about half the records with missing NHS numbers to practices when calculating practice-age-gender means of the explanatories (see section 5). Around 0.6% of patients were registered in more than one practice at the start of the financial year. We assigned patients with duplicate registrations to practice they had joined most recently.

2.2 Dependent variable

The dependent variable is an individual's NHS hospital expenditure in financial year *t* for hospital spell finishing in year *t*.[3] We include expenditure on both inpatient care and outpatient attendances. We had data on several years but costed all components of care using the same (2008/9) unit cost for all years.

We exclude expenditure on both mental health and maternity spells. The data on the former is poor and budgets for both were constructed by other means. Since the aim is to allocate budgets to cover the cost of NHS care we excluded the costs of care provided to private patients in NHS hospitals. Since practices will be charged for care when it is completed we excluded the costs of spells which started in the financial year but which were unfinished at the end of the year. We include spells which finish in the year even if they started in the previous year

2.3 Attributed needs

In addition to 38 age/gender categories we had available a large set of potential determinants of hospital expenditure which we could attribute to individuals via their lower super output area (LSOA)[4] of residence. These included 77 variables derived from the 2001 Census, the Indices of Multiple Deprivation, and modelled estimates of air quality, obesity etc. We also had a 53 category classification of LSOAs in socio-economic types eg "Multicultural suburbia", "Urban terracing" based on a cluster analysis of the 2001 Census by the Office of National Statistics. We also had the prevalence rate for 11 diseases as reported by each general practice.

2.4 Individual morbidity measures

HES records ICD10 diagnoses for inpatients in up to 14 fields. We used the HES records for each individual covering the two years before the start of the expenditure year (2006/7) to construct 6 sets of morbidity indicators:

- 22 ICD10 chapters
- 152 ICD10 categories – as used by NHS Information Centre to group HES admissions for reporting purposes
- 281 ICD10 groups. Our project colleagues (Health Dialog) grouped diagnosis codes into clinically meaningful categories and selected an aggregation level based on the relationship of the codes to cost. Of the 281 categories, 10 categories are at chapter level, 131 categories are at grouping level 2 as defined by WHO, and 140 categories are at grouping level 3 as defined by three-digit ICD 10 codes.
- 70 Hierarchical Conditions Classification groups. The HCC grouper was developed as part of a model to explain expenditure of patients in the US Medicare scheme which is primarily for patients age 65 and over. The HCC researchers developed a hierarchy of diagnoses, so that for

[3] Financial years run from 1 April to 31 March.
[4] There are 34,378 LSOAs each containing an average of 1500 people, with a minimum of 1000.

related diagnoses a patient was assigned to only the most serious category (so avoiding double counting). On the basis of previous research, the researchers first collapsed over 15,000 health care intervention codes into 804 diagnostic groups. These were then further aggregated into 70 HCCs that reflected clinically meaningful categories of diagnosis (Pope et al, 2004; Smith, 2007).

- 185 Augmented HCC groups. Not all diagnosis codes map to a HCC70 group. To map these other diagnosis codes to a group, our project colleagues Health Dialog identified a further set of 185 ICD10 groups. These 185 groups are based primarily on the first three digits of the diagnosis code
- 260 Clinical Classification Software groups. This grouper was developed by the Agency for Healthcare Research and Quality.

We also used four encounter variables derived from the HES records for the previous two years:

- number of inpatient episodes
- number of outpatient episodes
- whether any outpatient attendance was coded urgent
- whether treatment was received at any outpatient attendance

2.5 Supply

We constructed 70 supply variables which were attributed to individuals by their LSOA or their practice. They included median waiting times for outpatient and inpatient care for patients in the LSOA, distances to providers, and measures of local provided capacity (beds, staffing, equipment) weighted by distance and competing populations.

For practices we had 36 measures of quality derived from the Quality and Outcomes Framework (for example the practice's achievement in controlling Hba1c levels for their diabetic patients) and 30 practice characteristics such as the total list size, the number GPs, and the type of practice contract.

We also used PCT dummies and the shares of practice admissions at each hospital to capture unobserved factors (for example, PCT clinical governance, past levels of PCT resourcing, hospital admission threshold, hospital data recording conventions).

3. Alternative estimation procedures

3.1 Explanatories for the estimation procedure comparison

We compare alternative estimators using the same set of explanatory variables. Assuming expenditure model is linear we specify:

$$c_{ijpt} = \beta_0 + x'_{ijpt}\beta_x + m'_{ijpt-1,2}\beta_m + n'_{ijpt-1}\beta_n + v'_{ijpt-1,2}\beta_v + s'_{ijpt}\beta_s + PCT'_{pt}\beta_{PCT} + \varepsilon_{ijpt}, \quad (1)$$

where c_{ijpt} is expenditure in year *t* on patient *i* who is in practice *j* located in PCT *p* on the first day (1 April) of the financial year *t*. x'_{ijpt} is a vector of age and gender dummies, $m'_{ijpt-1,2}$ is a vector of 152 morbidity characteristic indicators based on ICD10 diagnoses in HES finished consultant episodes in the previous two years.

n'_{ijpt-1} is a vector of six attributed needs variables: proportion of LSOA population who are Disability Living Allowance claimants aged 60 and over, proportion non-white, proportion of lone parent households with dependents, standardized limiting long-term illness, proportion of individuals living alone aged 75 and over, and proportion of all pension credit claimants. $v'_{ijpt-1,2}$ is a vector of encounter variables derived from HES records for the previous two years: number of inpatient episodes, number of outpatient attendances, whether any outpatient attendance was an urgent referral, whether there was treatment at any outpatient attendance.

s'_{ijpt} is a vector of six supply characteristics: accessibility score for acute provider capacity, proportion of inpatients waiting less than 3 months, average distance to five nearest acute providers, average outpatient wait for first attendance, average distance to outpatients used, and road distance to nearest GP main surgery. PCT'_{pt} is a vector of 151 PCT dummy variables. ε_{ijpt} is an idiosyncratic error term assumed to be independently and identically distributed.

We estimated (1), or non-linear versions of it, using the routines outlined below and assess the predictive ability of the various estimators using both within-sample and out-of-sample tests. We supplement the tests by considering the predictive ability of the models across the full distribution of cost. Note that since the models are intended to be used to set prospective practice allocations we are primarily interested in their predictive ability and are less concerned about the individual coefficients beyond their relevance for assessing model plausibility.

3.2 One step estimation

3.2.1 Ordinary least squares

OLS is attractive due to its simplicity and ease of computation. Estimation operates directly on the cost scale (levels of cost) and marginal effects (the impact of a change in the value of a regressor on cost) are readily estimated. Heteroskedasticity renders OLS regression inefficient. OLS estimation of (1) assumes linearity between the set of regressors and cost which can lead to out-of-range (negative) predictions. It may also fail to predict well over the full range of the distribution of costs.

3.2.2 Transformed cost

An alternative to modelling cost on its natural scale is to apply a concave transform to produce a dependent variable which has less right hand skewness. Popular choices include the logarithmic (log) and the square-root transformation (Ettner et al, 1998). Cost data with zeros render the use of the log transformation problematic. Adding a small positive constant to the data prior to transformation is unsatisfactory because the results from a log transformed regression can be sensitive to changes in the

left-hand tail of the distribution of cost (Buntin and Zaslavsky, 2004). It also fails to deal with the large mass point in the distribution.

There is a difficulty with predictions from models which use non-linear transformations of the cost dependent variable. As Manning (1998) points out "Congress does not appropriate log dollars". Similarly, PCTs seek to allocate budget pounds to practices, not the log or square-root of pounds. Accordingly, where models are estimated on a transformed outcome, predictions must be retransformed back to the original scale. But we cannot simply take the inverse of the transformation: the conditional mean of a non-linearly transformed variable is not the inverse transform of the conditional mean, for example, for the log model $E[y \mid x] \neq \exp\{E[\ln(y) \mid x]\}$. Instead, for a log transformation, assuming normally distributed errors, the expected value of cost is given by:

$$E[y \mid x] = \exp\left(x'\beta + 0.5\sigma_\varepsilon^2\right) = \exp\left(x'\beta\right)\exp\left(0.5\sigma_\varepsilon^2\right) \tag{2}$$

where *x* is a vector of model regressors, β is a vector of parameter estimates and σ_ε^2 is the variance of the distribution of residuals.[5]

Where the distribution of the error is not normally distributed, but is homoskedastic, Duan's (Duan, 1983) smearing estimator can be applied for log transformations, where

$$E\left[y \mid x\right] = \varphi \exp\left(x'\beta\right) \tag{3}$$

where φ is the smearing factor, estimated as $\hat{\varphi} = N^{-1}\sum_i \exp\left(\hat{\varepsilon}_i\right)$, with $\hat{\varepsilon}_i = \ln y_i - x_i'\hat{\beta}$.

In practical applications the homoskedastic errors assumption is unlikely to be tenable for individual cost data. If the errors are heteroskedastic then Duan's smearing estimator will be biased. If the form of heteroskedasticity, as a function *p*(*x*) of regressors x, is known then unbiased predictions of cost are given as:

$$E\left[y \mid x\right] = p\left(x\right)\exp\left(x'\beta\right) \tag{4}$$

Where the variance is a function of multiple regressors, and where regressors are continuous rather than discrete, specifying the form of heteroskedasticity is problematic (as the exact from is often unknown). In such cases, it can be useful to compute a smearing estimator for percentiles of the range of fitted values from the estimated model.[6] In the following we split the distribution of fitted values into five-percentile ranges and compute the resulting 20 smearing factors.

Similar retransformation corrections can be applied to square-root transformed OLS models of expenditures. In this case, however, the correction term is additive rather than multiplicative, so that for the homoskedastic case

$$E\left[y \mid x\right] = \left(x'\beta\right)^2 + \phi, \tag{5}$$

where the smearing factor, ϕ is estimated as $\hat{\phi} = N^{-1}\sum_i \hat{\varepsilon}_i^2$.

[5] If the errors are heteroskedastic it is also the case that OLS estimation produces biased estimates of the coefficients (Santos Silva and Tenreyo, 2006)

[6] Based on percentiles of $\exp\left(x'\hat{\beta}\right)$.

3.2.3 Generalised linear models

Generalised linear models (GLMs) offer a flexible way to estimate expenditure models on the original scale (£s) (Blough et al, 1999; Buntin and Zaslavsky, 2004; Mullahy, 1998; Manning and Mullahy, 2001). GLM models can be applied directly to the entire distribution of costs including zero expenditures. Both the mean and variance function of expenditures are specified by the analyst. The mean function specifies how the linear index $(x'\beta)$ relates to the expected outcome, $g(E[y \mid x]) = x'\beta$, where g(.) is the link function, or $E[y \mid x] = \mu(x'\beta)$ where μ is the inverse of the link function. For expenditure data, a log link function is often chosen such that, $g\left(E[y \mid x]\right) = \ln(E[y \mid x]) = x'\beta$. Alternatively, the identity link, where the conditional expectation of y is linearly related to the vector of explanatory variables, can be used. Whichever link function is chosen, the results from a GLM regression can be interpreted directly on the original cost scale without retransformation. Thus in the case of the log link $g\left(E[y \mid x]\right) = \ln(E[y \mid x]) = x'\beta$ and $E[y \mid x] = \exp(x'\beta)$.

Specification of the relationship between the conditional variance and the conditional mean of y is also required in the GLM framework. Typically the variance is modelled as being proportional to a power function of the mean: $\mathrm{var}[y \mid x] \propto \mu(x'\beta)^{\lambda}$ which belong to linear exponential family of distributions. When $\lambda = 0$, the variance is constant and the conditional distribution of y_i is normal. With $\lambda = 1$, the variance is proportional to the mean and the conditional distribution is of the Poisson type . With λ = 2, the distribution is a gamma type and with λ = 3 the distribution is an inverse-normal type (Blough et al, 1999). The choice of distributional family allows heteroskedasticity to be modelled directly as a functions of the conditional mean. The distribution and link functions can be combined freely, although there are canonical links for each distribution. Perhaps the most commonly specified GLM for health care costs has combined the log link with a gamma type error.

If the mean function[7] of a GLM model is correctly specified, misspecification of the variance function will lead to inefficient parameter estimates and, in the extreme, might cause the estimation routine to fail to converge. If the mean function is misspecified, then the model will fail to fit the observed data well across its full range. In such circumstances the choice of variance function will affect both the efficiency of the estimator and model fit (predictions versus observed data) (Buntin and Zaslavsky, 2004).

Other estimation approaches to expenditure data are available including exponential conditional mean models (for example the Poisson model), hazard type models and finite mixture models and a useful summary of these is provided by Jones (2011). We restrict attention to estimators commonly encountered in the applied literature modelling health care costs.

3.3 Two-part models

The estimators described above may fail to accommodate the large mass point at zero expenditure and as Figure 3 shows most patients on a practice list do not generate any hospital expenditure during a year. Accordingly two-part models seem potentially suitable. The first part of a two-part model estimates the probability of zero versus positive expenditure. For example using a probit model:

$$\Pr(y_i > 0) = \Phi(x'\beta) \tag{6}$$

where $\Phi(\cdot)$ is the standard normal cumulative distribution function.

[7] Both the link function and the specification of the linear predictors.

The second part of the model estimates expenditure for individuals who incur hospital costs: $E[y \mid y > 0, x]$. All of the estimators considered above can be used to model costs conditional on use. Transformations can include logs since we are considering positive expenditures only. Transformed outcomes again require retransformation to obtain predictions on the original cost scale.

Total predicted expenditure can be obtained from a two-part model by multiplying the predicted probability of experiencing positive health care costs with the expected level of cost derived from the second part of the model:

$$E[y_i \mid x_i] = \Pr(y_i > 0 \mid x_i) E[y_i \mid y_i > 0, x_i] \tag{7}$$

3.4 Assessing model performance and model validation

3.4.1 Estimation and validation samples

It is well known that within-sample tests of model fit produce overly optimistic results caused by overfitting of, for example, rare but costly procedures (Copas, 1987). To avoid this we estimate the model on one sample and assess its predictive accuracy on two other samples.

(i) Estimation sample. A 10% random sample of individuals registered within each general practice at 1 April 2006. Practices with populations less than 1,000 were removed from the data prior to sampling. The estimation sample has over 5M individuals registered across all practices within 152 PCTs. This large sample is computationally burdensome, particularly for certain estimation routines such as GLM.
(ii) Individual level validation sample. A second 10% random sample of individuals registered in general practices at April 2006. Practices with populations less than 1,000 were removed from the data prior to sampling.
(iii) Practice level validation sample. The second validation sample is *all* patients registered with a random sample of 10% of all practices, stratified by PCT. The sample has 812 practices, with over 5M patients. The sample is used to provide an assessment of the ability of the estimation routines to predict both total practice expenditure and per patient practice expenditure.

3.4.2 Model evaluation criteria

We assess and compare the various approaches to model estimation by their predictive performance, measured by R^2, root-mean-squared error (RMSE), and mean absolute prediction error (MAPE).

The R^2 is less straightforward to compute for non-linear models, GLM models and for models requiring a retransformation of the predicted outcome. To compute an R^2 value comparable across all estimators, we first compute predicted expenditure, $\hat{y}_i$, on the original cost scale following OLS, GLM or transformed OLS regression. Predictions were then regressed against actual expenditure, y_i. The R^2 from these auxiliary regressions are then compared across the different estimation routines.

We also report the root-mean squared error calculated as the square-root of the mean of the square of the difference between predicted and actual expenditures

$$RMSE = \sqrt{\sum_i (y_i - \hat{y}_i)^2 N^{-1}} \tag{8}$$

and the mean absolute prediction error

$$MAPE = N^{-1} \sum_n \text{abs}(y_i - \hat{y}_i) \tag{9}$$

Better model performance is associated with higher R^2, and lower RMSE and MAPE.[8]

[8] Note that $R^2 = 1 - (RMSE)^2 / Var(y)$ so that the statistics are negatively monotonically related.

3.4.3 GLM models

To assess the choice of variance function for a GLM specification, deviance residuals[9] from a GLM model can be assessed using normal plots, which plot actual residuals against the values they would take if they were normally distributed. If the variance function is chosen correctly then deviance residuals should be approximately normally distributed and follow the 45° line on a normal plot (McCullagh and Nelder, 1989; Davison and Gigli, 1989).

A further test for the adequacy of the variance function, suggested by Manning and Mullahy (2001), is the Park test (Park, 1966). Since the variance function is assumed to be a power function of the conditional mean: $\text{var}[y \mid x] = \kappa(E[y \mid x])^{\lambda}$, the test suggests regressing $\ln(y_i - \hat{y}_i)^2$ on $\ln(\hat{y}_i)$ and a constant, where $\hat{y}_i$ is obtained from a preliminary GLM regression[10]. The estimated value of the parameter λ attached to $\ln(\hat{y}_i)$ provides an indication of the appropriate power of the variance function.

We also report the Akaike information criterion (AIC): $AIC = 2k - 2\ln(L)$, where $\ln(L)$ is the model likelihood and k the number of model parameters (constant across the set of GLM models estimated here) to further aid comparison across model. A smaller AIC indicates better model fit.

3.4.4 Model validation

The predictive or forecast accuracy of the alternative estimation techniques are assessed on the validation samples using R^2, RMSE and MAPE. The first validation sample is a random sample of 10% of practice patients from all practices. This replicates the estimation sample and provides a direct measure of the predictive accuracy of the model at the individual level.

The second validation dataset is a 10% sample of practices which is used to assess predictive ability at practice level. Actual and predicted expenditures are aggregated to practice level before computing the R^2 from a regression of aggregated actual costs on aggregated predicted costs. We also compute for each practice the per patient actual and predicted expenditure and report the R^2, RMSE, and MAPE from the regression of actual per capita expenditure on predicted per capita expenditure.

Given the skewness of the cost data and the spike at zero cost, we might expect predictions to be less accurate at the extremes of the distribution. Hence we further assess the prediction properties of the alternative estimation routines across the full range of cost by plotting the mean of actual and mean of predicted costs by percentiles of the distribution of predicted costs. This allows the assessment of the extent of the range of costs for which the estimation approach performs well. Plots are generated for predictions from both the individual and practice level validation samples.

3.5 Results from estimation procedures

Tables 1 and 2 present statistics for the alternative estimation routines used to predict expenditure in 2006/7 for individuals on practice lists at 1 April 2006 using morbidity measures calculated from HES data for 2004/5 and 2005/6.[11]

---

[9] The deviance residual for an observation is $[\text{sgn}(y_i - \hat{\mu}_i)]\sqrt{d_i^2}$ where $\hat{\mu}_i = \mu\left(x_i'\hat{\beta}\right)$ is the estimated conditional mean, and $d_i = \left[2\left\{\ln L(\hat{\beta}, x_i) - \ln L(\beta^*, x_i)\right\}\right]^{\frac{1}{2}}$ is the square root of twice the difference between the logs of the maximised likelihood and the likelihood from a saturated model with as many explanatories as observations.

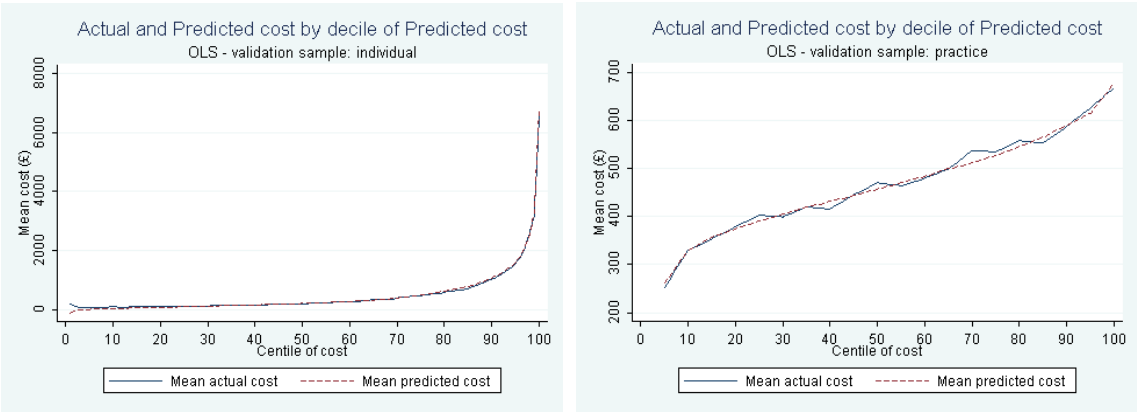
[10] For a log link function, $\ln\left(\text{var}[y \mid x]\right) = \ln(k) + \lambda \ln\left[\exp\left(x'\beta\right)\right]$ which can be approximated via the regression: $\ln\left(y - \hat{y}\right)^2 = \alpha + \lambda \ln\left(\hat{y}\right) + e$.

[11] Results for models to explain 2006/7 expenditure for individuals on practice lists at 1 April 2005 using the same set of explanatories but with morbidity data from HES records for 2003/4 and 2004/5 yielded identical conclusions about the relative performance of the different estimators, though, because of the longer lags, overall performance was worse.

3.5.1 One-part results

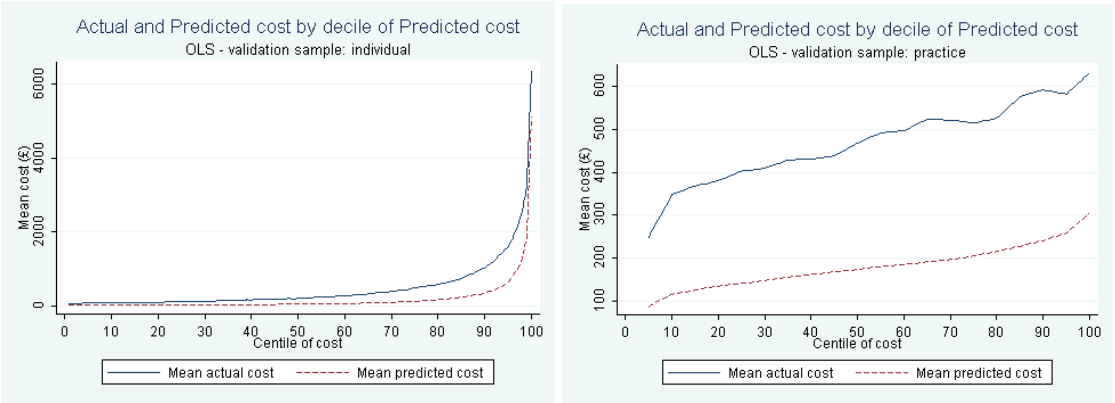
OLS. Table 1 shows that OLS has better R^2 (and hence RMSE) than any of the other one part models on the estimation and the two validation samples. For the practice validation sample OLS is best on all criteria. Figure 4 plots, using the individual and practice validation samples, the mean actual expenditure against mean predicted expenditure, for individuals and per capita for practices, within each percentile. The plots show that, for both individual and total per capita practice cost, predicted and actual expenditures are closely aligned across the distribution of cost.[12]

Transformed OLS. While the square-root transformation without smearing records the lowest value for the MAPE for both the estimation sample and the individual level validation sample, its performance is otherwise poor compared to the other transformed OLS models with smearing (Figure 5). This is particularly the case for the practice level validation sample. An inspection of the plots of actual versus predicted expenditure shows that retransformation with smearing improves performance and that heteroskedastic retransformation (Figure 6) does best. These improvements are particularly evident using the practice level validation sample. Across the three square-root models, the heteroskedastic retransformed model has the best performance on all samples except the MAPE on estimation and individual validation samples. This is unsurprising, given that separate retransformations are applied to each of the centiles of the distribution of predicted cost.



(i) individual **(ii) practice per capita**

Figure 4: One-part OLS. Validation – predicted vs actual

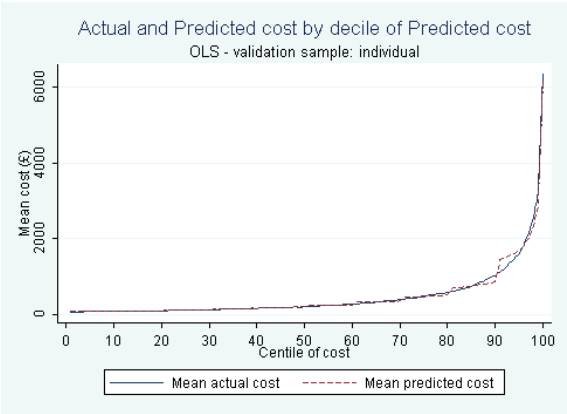


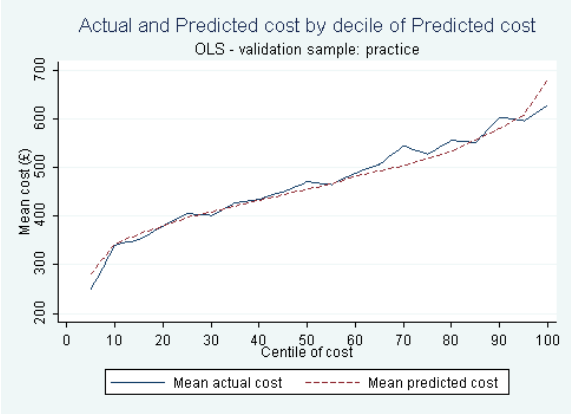
(i) individual **(ii) practice per capita**

Figure 5. One part OLS square root. Validation – prediction versus actual.

Square-root without retransformation:

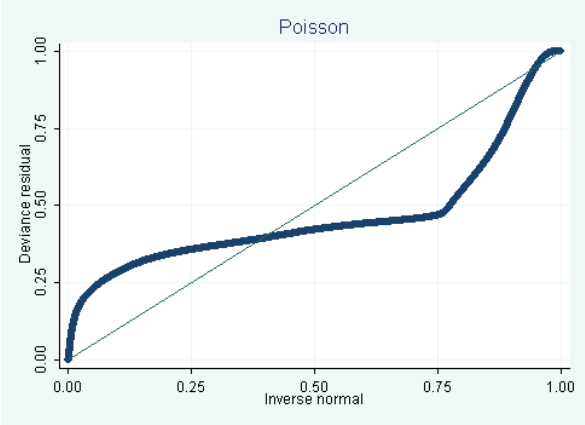
[12] Since the samples contain over 5M individuals, centile means are based on approximately 50,000 individual observations.

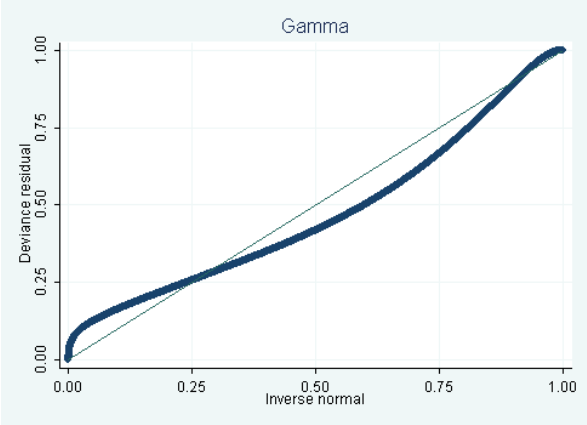




(i) Individual level **(ii) Practice level per capita**

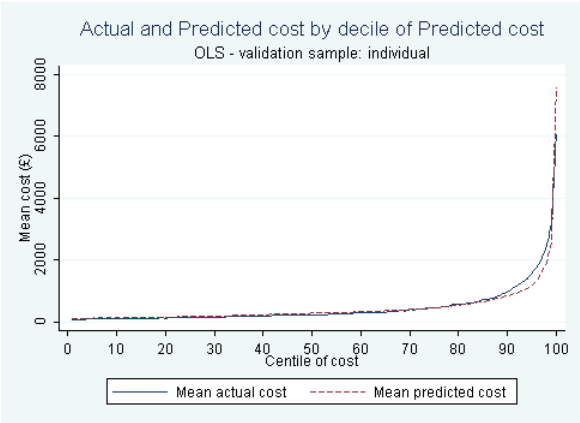
Figure 6. One part OLS square root models with heteroscedastic smearing transformation. Validation – prediction versus actual.





(i) mean (Poisson) **(ii) mean squared (Gamma)**

Figure 7. One part GLM. Normal plots of deviance residuals for log link models with variance proportional to (i) mean (Poisson) (ii) mean squared (Gamma)



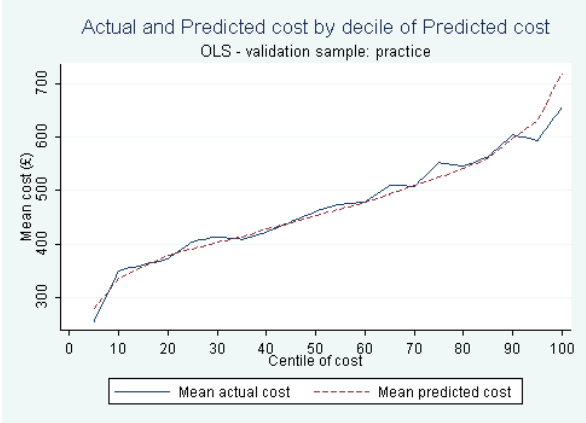


Figure 8 One part GLM, log link, variance proportional to mean: validation – prediction versus actual.

Generalised linear models. In comparison to OLS, GLM models have disappointing R^2, RMSE and MAPE statistics. GLM with linear link functions failed to estimate except for the model with constant variance which is equivalent to OLS. All models with log link functions, except where the variance function was

proportional to the mean cubed, were estimable. Results of the Park test suggests that a model with a variance function somewhere between the Poisson family (variance proportional to the mean) and a gamma family (variance proportional to the mean squared) is optimal (λ = 1.67). The AIC favours the gamma family, as do normal plots of deviance residuals (Figure 7). However, the prediction tests at practice level suggest that the model with variance proportional to the mean is the best of the GLM log link models (Figure 8).

The contrast between prediction tests and the Park test and AIC suggest that these latter diagnostic tests might not always be helpful in choosing between different GLM variance functions, in large samples where predictive performance is of primary interest. Similar findings have been reported elsewhere (Buntin and Zaslavsky, 2004).

3.5.2 Two-part model results

The two-part models reported in Table 2 were estimated using the same explanatories for both the probit model of the probability of incurring positive expenditures and for models of the level of expenditure, conditional on incurring positive expenditures. As with one-part models, OLS does best on the prediction criteria. For the practice validation sample, probit plus OLS returns the highest R^2 and lowest forecast prediction errors (RMSE and MAPE) compared to all other model specifications. With the exception of MAPE, this is also true for the validation at an individual level and for the estimation sample. Plots show good performance across the entire cost range.

Probit plus transformed OLS. For models with a square-root transformation for the second-part of the two-part model, improvement in predictive performance is achieved through the use of smearing and, in particular, heteroskedastic smearing. This is the case across both estimation and validation samples, with the exception of the MAPE which favours an estimator without a smeared retransformation. Plots of predicted versus actual expenditure show good predictive properties across the range of expenditures.

Smearing estimators appear to improve the predicted performance of log models when applied to the practice level validation sample, but fails to improve performance when applied to the individual level validation sample. An inspection of plots of average predicted expenditure against average actual expenditure reveals that for the latter, predictions using smearing estimators (homoskedastic and heteroskedastic) at the upper end of the cost distribution hugely over predict expenditures. While, performance across the majority of the range of predicted values appears reasonable, differences at the end of the distribution account for the poor R^2.

Probit plus generalised linear models. The Park test undertaken on GLM models with a log link function suggest a gamma family of variance functions (λ = 1.87) This is supported by the AIC and normal plots of residual deviance However, on the basis of predictive performance (R^2, RMSE and MAPE), the estimation sample and individual validation sample favour a constant variance model while at practice level a variance function proportional to the mean (Poisson family) does best.

Plots of predicted expenditure versus actual expenditures are similar to those from the one-step GLM models. When the variance function is assumed to be proportional to the mean squared (gamma type family) and mean cubed (inverse normal type family), predictions at the upper end of the distribution of predicted costs far exceed actual costs. This would appear to explain the poor performance of these models, even though they appear to predict expenditures well over the majority of the range of cost.

3.6 Conclusions

Our results suggest that one stage OLS out-performs alternative estimators when the aim is to predict health care expenditures in order to set practice budgets. OLS is the simplest to implement, is widely used and understood, and avoids the problem of having to make adjustments for heteroskedasticity which are inherent in models using a transformed dependent variable. The approach also avoids the necessity to estimate separate models for the probability of incurring expenditure and the level of conditional expenditure. A potential disadvantage of applying OLS is that it can produce negative predictions of cost for some individuals, a problem avoided with, for example, GLM with a log mean link function. However,

when individual predictions are aggregated to practice level negative predictions are not a problem – no practices had negative predicted expenditure. While the two part probit plus OLS model marginally outperforms one step OLS for the estimation sample and on the individual level validation sample, one step OLS is marginally better at predicting expenditure at practice level. We find that for our purposes the one part OLS model is a robust estimator of hospital expenditure at the general practice level.

4. Modelling expenditure with OLS

4.1 Morbidity markers

Having decided on the one part OLS estimator, we next chose a set of explanatories for predicting hospital expenditure for individuals on practice lists. We first compared models with age/gender categories and PCT dummies and alternative sets of morbidity indicators and then alternative sets of attributed need and supply variables. Table 3 reports results from models using variants of the 6 sets of morbidity indicators listed in section 2.[13] Morbidity sets with more categories predict better on the estimation and individual validation sample. However, the best models in terms of per capita predictions on the practice sample use the 22 and 152 ICD10 category sets which have larger R^2 and a larger proportion of practices with actual per capita expenditure within 10% of predicted. The simple 22 category set performed as well as the 152 category set but the larger set was preferred on grounds of greater appeal to GPs.

Basing the morbidity markers on two rather than one year of data improved led to a small improvement in explanatory power and predictive power.

4.2 Attributed need and supply variables

Having decided to use the 152 category set of morbidity indicators we next selected from amongst the large set of attributed need and supply variables.[14] We started with the full model (152 morbidity markers, 4 encounter variables, 2 variables measuring past use of NHS hospitals as a private patient, 151 PCT dummies, 37 age/gender dummies and 169 attributed needs and 106 attributed supply variables.[15] We then generated a parsimonious model in stages:

(i) re-estimate the full model retaining only those attributed need and supply variables whose absolute t-ratio was greater than 0.20

(ii) re-estimate the model estimated in (i) retaining only those attributed need and supply variables whose absolute t-ratio was greater than 0.40

(iii) re-estimate the model estimated in (ii) retaining only those attributed need and supply variables whose absolute t-ratio was greater than 0.60

(iv) continue this process until only those variables with an absolute t-ratio greater than 2.00 remain

(v) inspect the coefficients on the remaining need and supply variables and drop those variables with 'incorrect/unexpected' signs

(vi) re-estimate the model with the remaining attributed need and supply variables

(vii) re-estimate the model estimated in (vi) retaining only those attributed need and supply variables whose absolute t-ratio is greater than 2.20

(viii) continue the process outlined in steps (v) - (vii) until only those variables with an absolute t-ratio of greater than 2.58 remain.

We then repeated the process starting from a full model excluding the four encounter variables and derived a restricted parsimonious model.

[13] The models also contained age/gender cateogories and PCT dummies. The models estimated expenditure in 2006/7 for individuals on practice lists at 1 April 2005.

[14] The project timetable and data availability meant that the morbidity variable sets were chosen on the basis of models to explain 2006/7 expenditure for individuals on practice lists at 1 April 2005 with morbidity data from 2003/4 and 2004/5, whereas the modelling of the non-morbidity explanatories used models to explain 2007/8 expenditure for individuals on lists at 1 April 2007 with morbidity explanatories from 2005/6 and 2006/7. The reduction in the data lag boosts model performance considerably. The performance of the latter models was also boosted because a change in coding of intensive care cases produced a distribution of costs with fewer extreme outliers, though the median cost per spell was unaffected.

[15] We also investigated including the proportions of practice expenditure accounted for by each of 165 hospitals but found that they had no effect on model performance.

4.3 Preferred parsimonious models compared

Table 4 has summary statistics on alternative models. Using only 38 age and gender produces low individual R^2 of 0.037 and the per capita practice performance is also not impressive with 58% of practices having a discrepancy between actual and predicted per capita expenditure of over 10%. Figure 9 has the age/gender cost curves. Adding the 152 morbidity markers increases performance dramatically with an R^2 on the individual validation sample of 0.261. PCT dummies have little effect on individual R^2 but do improve practice per capita R^2. This is probably because variations in the total budget allocated to PCTs do not lead to variations in expenditure across patients in the same practice in the same PCT but do lead to variations in average levels of practice expenditure across PCTs. Finally, adding the full set of attributed needs and supply variables and two dummies for private outpatient and inpatient use of NHS hospitals, leads to a further small improvement. Dropping insignificant attributed need supply variables to produce the parsimonious model has a negligible effect on overall model performance.

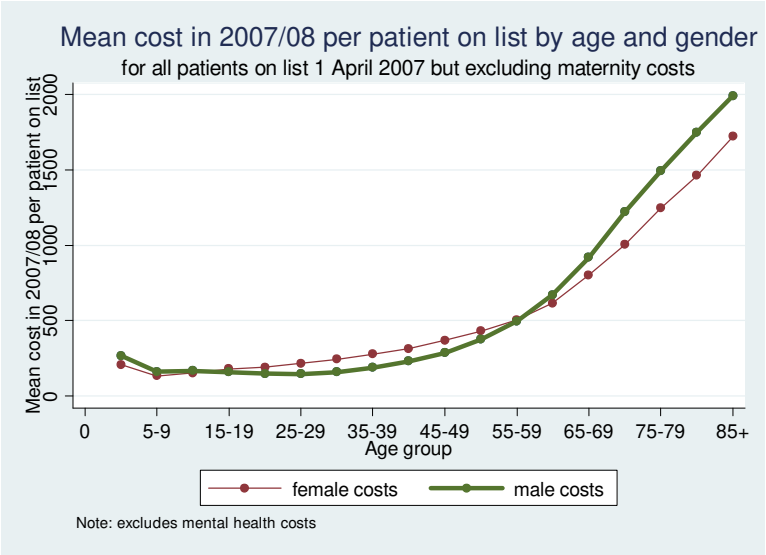


Figure 9. Mean cost per patient excluding maternity by age and gender for 2007/08.

We next compare, in Tables 5 and 6 and Figures 10 and 11 the preferred parsimonious model with encounter variables and the re-estimated preferred restricted parsimonious model with the encounter variables excluded. The age and gender patterns (Figures 10, 11) are very similar for the two models, with slightly higher coefficients for older people in the restricted model with no encounters compared with the unrestricted model. This is probably because the propensity to have more encounters is increasing in age. The age pattern is very similar for males and females. The effect of age on costs increases for adults, despite the rich morbidity information in the models, presumably because older individuals have more severe instances of any given diagnosis. However, comparison with the unconditional age/gender cost means in Figure 9, shows that the effect of age is smaller in the models with morbidity variables.

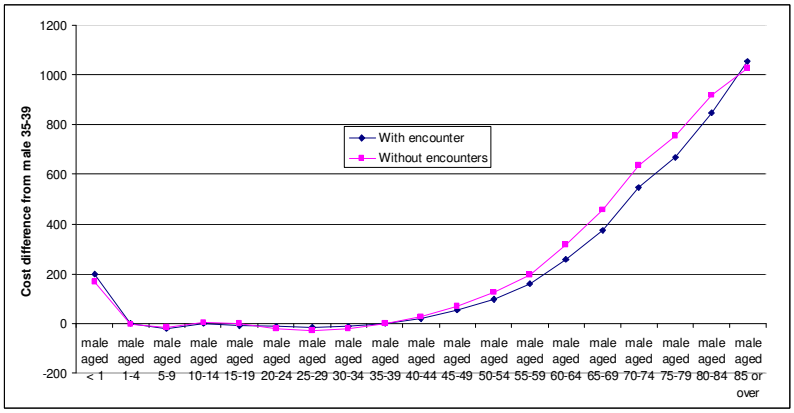


Figure 10. Estimated male age effects relative to male 35-39 for alternative models

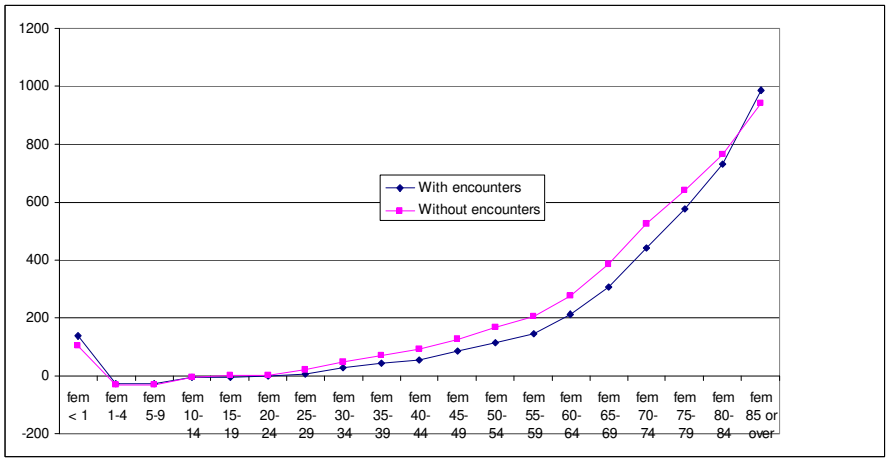


Figure 11. Estimated female age effects relative to male 35-39 for alternative models

The two count variables (number of outpatient attendances and number of inpatient episodes) have remarkably high t-stats and a considerable effect on the individual level R^2, though a much smaller one on the practice per capita R^2 between predicted and actual expenditure.[16] These variables can be interpreted as capturing severity (since additional episodes with the same ICD10 code after the first will not affect the ICD10 morbidity markers). They may also reflect the non-additive effects of having several episodes with different ICD10 codes. Models with hospital effects which allow for differences in admission threshold or inter-consultant referral thresholds produce very similar results, suggesting that if the encounter variables are reflecting supply side factors they arise from unobserved differences in practice characteristics (idiosyncratic practice or GP style).

Because of the large number of past HES based morbidity markers (152 ICD10 groupings) included in the models only a few of the large number of attributed need variables are significant in the two models. These attributed needs variables make relatively difference to model fit.

Persons in social rented housing, disability allowance claimants, and lack of qualifications have plausible positive coefficients in both models. The ONS area classification Mature City Professionals has a negative coefficient in both models. We interpret this as picking up the greater likelihood that people in these areas are more likely to use private health care and so incur less NHS expenditure. Similarly we

[16] In the models to explain 2006/7 expenditure for individuals on practice lists at 1 April 2005 with explanatories from 2003/4 and 2004/5 we found that dropping the encounter variables also had a marked effect on R^2 at individual level but less of an effect at practice per capital level (see Table 3).

suggest that the student population variable has a negative sign in both models because it is picking up private health care use.

The coefficient on the dummy for having any private episodes is negative in both models. Having a private inpatient episode in an NHS hospital in the previous two years reduces NHS expenditure in 2007/8 by £491 or £556 depending on the model. This is also plausible since individuals who have been private patients in the past are more likely to go private in the future if ill and thus to generate less NHS expenditure. Since the aim of the resource allocation process is to allocate resources for NHS care we treat the private care variable as measure of need. Practices with more patients who go private in NHS providers will receive, ceteris paribus, a smaller budget.

When the four encounter variables are forced out of the model, asthma prevalence (calculated from QOF data) become positive and significant. This is probably because asthma patients are more likely to have repeated attendances at outpatients.

There few significant attributed supply variables, possibly because of the inclusion of PCT dummies. Practice stroke care quality (as measured by QOF stroke achievement) has a negative effect on expenditure in both models. Since we treat stroke care quality as a supply variable in the allocation calculations (see section 5) there is no disincentive effect for QOF quality via the practice budget. The supply of MRI scanners also has a positive effect of expenditure in both models. Access to residential beds reduces costs in the model with encounters because easier access will reduce length of stay in NHS hospitals.

Table 5 reports the coefficients on the 152 ICD10 category dummies for HES admission during 2005/5 or 2006/7. Consider the unrestricted model coefficients. It might at first sight seem surprising that 80 of the 150 non null coefficients are negative. The coefficient on a morbidity category is the effect on costs of an admission in that category, given the value of the other explanatories. The unrestricted model also contains a count of admissions and outpatient attendances in 2005/6 and 2006/7. Anyone with a diagnostic code from 2005/6 or 2006/7 must have had a past admission in 2005/6 or 2006/7 and a high proportion of them will also have had an outpatient attendance. Thus the coefficient on the code dummy is the additional expected cost of having that type of diagnosis. Thus the full effect of an admission in 2005/6 or 2006/7 on costs in 2007/8 is the sum of the coefficient on the diagnostic category of that past admission, the coefficient on the number of past admissions, and (a proportion of) the coefficient on the number of past outpatient attendances. The coefficient on the number of past admission is £299 and on outpatient attendances is £46. When the coefficient on the number of past admissions is added to the diagnostic coefficients the sum the number of negative codes falls to 16 and if the coefficient on outpatient attendances is added as well the number falls to 11.

The restricted model omits the count variables. Hence the cost effect of a past diagnosis will now pick up the cost element attributed to past attendance and episode number in the unrestricted model. Forcing out the encounter variables increases 143 of the coefficients in the restricted model and the average proportionate increase[17] in coefficients is 1.75 compared to the unrestricted model. In the restricted model only 22 of the coefficients on past diagnostic categories are negative.

None of the negative coefficients in the restricted model are large except A90-A99: Arthropod-borne viral fevers & viral haemorrhagic fevers which is extremely rare. Some negative coefficients are because past successful treatment reduced the likelihood of future admission. For example, with K35-K38 (Diseases of the appendix), if the patient survives the operation they can never get appendicitis again. Alternatively, a past inpatient episode with dementia (F00 - F03) diagnosis may reduce 2007/8 cost because this was an initial occurrence and the patient was then transferred to a care home.

[17] Calculated as (restricted coefficient – unrestricted)/(absolute value of unrestricted coefficient).

5. Setting needs based practice budgets

The target allocation to a practice in a CCG/PCT is the practice's share of the total need in the CCG/PCT multiplied by the total budget the CCG/PCT decides to give to its practices for commissioning. Practice need is calculated by applying a set of regression coefficients from a model to explain individual expenditure to a corresponding set of explanatory variables for the practice. In this section we first describe a procedure for calculating practice needs which makes use of the latest data which is available for the estimation of regression coefficients and the calculation of practice explanatories. We then discuss two reasons why variables which explain expenditure should not influence practice target allocations: (a) they do not reflect need for NHS resources[18] or (b) they will create perverse incentives if they are allowed to affect the practice budget. In the light of these considerations we then compare three candidate calculations of practice need and suggest that the choice amongst them will depend on the relative importance of a better estimate of practice need versus the wish to avoid perverse incentives.

5.1 Calculating practice allocations

Budgets for say 2010/11 are set in the autumn of 2009 and because of lags in the availability of data on hospital use and expenditure, they have to be based on models to explain 2007/8 expenditure using data from 2005/6 and 2006/7. The following procedure makes use of the most recently available data at each step:

(i) estimate a model for 2007/8 expenditure for individuals in a practice at 1 April 2007 using explanatories from 2005/6 and 2006/7.

(ii) calculate the means of the explanatory variables for 2006/7, 2007/8 for each age/gender group in each practice where the means are based on the patients registered in the practice at 1 April 2008. (For supply variables use the national age/gender means for all practices.) This uses the most up to date information on explanatories.

(iii) apply the coefficients from the step 1 model to the practice age/gender explanatory means from step (ii) to calculate per capita need in £s in each practice/age/gender group.

(iv) multiply the practice age/gender per capita needs by the number of patients in each practice age/gender group at budget setting date in autumn 2009 and sum over all age/gender groups within a practice. This makes use of the most recent information on practice lists and produces a total practice need measured in £s.

(v) share out the total CCG/PCT budget in proportion to the practices shares of total need as calculated at step (iv).

This approach does not require individual level predictions for those moving practices, and therefore removes the necessity to track specific individuals at the time they move. It also reduces the impact on model performance of data lags because the age-gender group average needs index are more stable over time than individual level needs variables (and therefore expenditures).

The procedure for calculating allocations rests on the assumption that applying coefficients from the regression model which uses 2005/6 and 2006/7 data to explain 2007/8 expenditure to the age/gender/practice means calculated from data for 2006/7 and 2007/8 gives a reasonably accurate estimate of average need per person in the age/gender/practice group in 2010/11. We are thus not attempting to predict need in 2010/11 at *individual* level for individuals on the practice list at budget setting date in autumn 2009. Individual level explanatories are likely to vary over time and accordingly the relationship between explanatories and expenditure on the individual will therefore depend on the chosen lag length. We are using age/gender/group means, not individuals.

[18] In the terminology of Van de Ven and Ellis (2000) need variables are solidarity variables.

The procedure assumes that the relationship between explanatories (at *t*-1, *t*-2) and expenditure at *t* is reasonably stable over time (ie we would get similar coefficients using *t+s*-1, *t+s*-2 explanatories to explain *t+s* expenditure. It also requires that practice/age/gender mean explanatories are reasonably stable over time (unlike individual explanatories) so that the calculation of practice need is not affected by the lag between the date at which the practice/age/gender means of explanatories are calculated and the budget year.

We investigated the robustness of the procedure in two ways. First, we applied the coefficients from a regression model which uses 2004/5 and 2005/6 data to explain 2006/7 expenditure to the age/gender/practice means calculated from data for 2006/7 and 2007/8. We then calculated practice shares of total PCT need from this procedure which used lagged coefficients with the practice shares of total PCT need when the coefficients were from the most up to date model. The correlation of practice shares was 0.999

Second, we tested for stability with respect to lags in the data used to calculate the age/gender/practice means of the explanatories. We applied coefficients from the regression model which uses 2005/6 and 2006/7 data to explain 2007/8 expenditure to the age/gender/practice means calculated from data for 2005/6 and 2006/7. We calculated practice shares of total PCT need with this lagged explanatory data and compared the allocations with those using the most recent data for 2006/7 and 2007/8. The correlation of practice shares of total PCT need was 0.995.

Whilst the most recent data should be used to estimate the underlying models and to calculate the age/gender/practice means of the explanatories, the calculation of practice allocations seems robust to lags in timing. The reasons are that shares of allocations are driven by practices shares of the PCT population and the procedure uses age/gender/practice means rather than individual data which would be much more volatile.

5.2 Do encounter variables measure need or supply?

There are three types of variables which should not affect allocations: (a) measures of supply; (b) socio-economic variables which have inappropriate or counter intuitively signed coefficients, for example negative coefficients on ethnicity or unemployment measures; (c) variables which would generate perverse incentives. In the current study, because of the richness of our individual level morbidity measures, there were no inappropriately signed socio-economic variables. Both the count of inpatient episodes and the count of outpatient attendances have large, highly significant, positive coefficients in our preferred parsimonious model. It is clear that these explanatory variables reflect something which makes a substantial contribution to the variation in costs. The question is whether they reflect need or supply; and, even if they have reflected need in the past, whether letting them influence future allocations will create seriously perverse incentive.

The rationale for their interpretation as need variables is that the number of times an individual has been admitted or attended in the past conveys something about their morbidity over and above the ICD10 dummy morbidity category variables. The morbidity category dummies, whilst very powerful as explanatories, do not reflect repeated admissions falling in the same ICD10 category. The count variables will therefore partially reflect severity. The use of morbidity categories also means that different morbidities have an additive effect: they do not reflect the possible effects of comorbidity. The count data measures will also pick up some part of the non-additive effects of comorbidity.

But it can also be argued that the count variables are also picking up supply factors. For example, better resourced areas may have providers with shorter waiting times or different providers may have different admission thresholds or provide better quality of care which leads to increased referrals from GPs, or secondary care may be a substitute or complement to GP services. Our models have included a large number of supply variables, PCT dummies, and QOF practice quality scores. We have also estimated models with additional provider effects which should pick up differences in provider outpatient and admission policies. Including these variables has little effect on the coefficients on the encounter

variables. It is however possible that the count variables reflect purely idiosyncratic GP behaviour ("practice style" or practice quality not correlated with QOF quality).

5.3 Perverse incentives?

5.3.1 GP incentives

Under PBC practices have a nominal budget with the intention, inter alia, of making them consider the costs of their decisions affecting hospital use, as well as their benefit to patients. Budget surpluses are not intended to be treated as personal income: they are meant to be spent on patients, for example by commissioning non-hospital care. Thus, unless GPs find illicit means of appropriating budget surpluses, their incentives as regards costs falling on their hospital budget depends on the marginal value they place on using surpluses to fund other types of care. The more altruistic the GP the greater the incentive to reduce hospital costs.

Admissions and referrals. If the outpatient and inpatient count variables are allowed to affect the budget, additional referrals or admissions will increase the budget two and three years later. The coefficient on the outpatient count is £46. Thus an additional outpatient referral would increase the practice budget two and three years later by £46 in each year (multiplied by the ratio of PCT total practice budget to total need). An adult first referral has a price of (2008/9) £160. The implication of treating the outpatient count as a need variable is that the net impact of a GP outpatient referral on the practice budget is greatly reduced.

GPs can clearly influence patient decisions to seek elective inpatient treatment. They can also affect emergency admissions via the quality of care they provide for Ambulatory Care Sensitive Conditions which can be managed in primary care (Bindman et al, 1995; Dusheiko et al, 2011a; Dusheiko, et al., 2011b). A practice budget will impose the costs of admissions for ACSCs on the practice and therefore incentivise prevention. But this incentive will be reduced because the additional admission will increase the future practice budget. However, to the extent that the effects of preventive care on admissions do not arise immediately but may be delayed, the future effects of changes in prevention will also be delayed. Moreover, around 8% of patients move between practices in a year, and the future financial effect of current preventive care provided to them will not be borne by the practice making the prevention decisions. There are also direct and immediate financial incentives for practice prevention from the Quality and Outcomes Framework (Campbell et al, 2009). Thus the dilution of incentives for prevention if the count variables affect budgets may not be great.

Cream-skimming. It is possible that budget holding practices will also consider patient hospital costs when deciding whether to accept or retain patients on their lists. If practices have information about patients' characteristics, and therefore their expected future costs, in addition to that used in setting practice budgets, then they may cream-skim potential patients or dump existing patients.

The budget formula is based primarily on hospital records. Practices will have better information on existing patients. In principle this would give them the ability to identify patients whose expected costs greatly exceed their contribution to the practice budget and attempt to remove them from practice lists. However, it is more difficult for practices to remove existing patients from their list than to refuse to accept new patients, and it would be easier to regulate by monitoring patient lists. Cream skimming of new patients is less easy to monitor but GPs' ability to predict patient costs may be no better than the budget model and possibly worse since they do not have access to potential patients past hospital records.

5.3.2 Incentives for hospitals

Hospitals can affect practice budgets by admitting a patient with new ICD10 diagnosis, and if the outpatient appointment and inpatient episode variables affect budgets, by additional admissions and outpatient appointments. The resulting increased practice budget may lead to increased demand for hospital care. However, the perverse incentive effects for providers are weak. First, because of lags in data availability, an increase in the admissions or outpatient referrals in year t will not affect the practice budget until years t+2 and t+3. Second, the practice need not spend its increased budget at the provider:

it will use a mix of providers (Dusheiko et al, 2008) and it may not spend its budget entirely on hospital care.

5.4 Comparison of practice allocations

Since there are arguments for and against letting practice budgets be affected by their encounter variables we compare three sets of allocations. The first treats encounter variables as predominantly measuring need and places a low weight on any consequent perverse incentives and so permits variations in encounters across practices that affect their allocations.

The second and third calculations of allocations do not let variations in encounters affect practice allocations, either because they reflect supply or because of their incentive effects. They differ in the way in which the encounter variables are sterilised. The second method (frozen parsimonious model) uses the coefficients from the parsimonious model but sterilises encounters by calculating budgets using the national age/gender means values of the encounters, rather than practice/age/gender specific values. The third method (restricted parsimonious model) uses the coefficients from the restricted parsimonious model which does not use encounters to explain expenditure. As we show in the Appendix, the choice between these two methods of sterilising non need variables rests on untestable assumptions about the underlying structural model determining expenditure (Gravelle et al, 2003; Smith, Rice and Carr-Hill, 2001). The earliest English utilisation based allocation formulae used the third method to sterilise supply variables (Carr-Hill et al, 1994). Later English formulae used the second method of leaving the non-needs variables in the model but applying their coefficients to their mean values (Sutton et al, 2002; Morris et al, 2007).

The frozen parsimonious model produces allocations which vary less across practices than the other two methods. The reason is that by retaining the positive coefficients on the encounter variables but multiplying by the national means of the variables, a large positive constant is added to the calculated need (£s) of all practices and thus there is less variation across practices in their shares of total calculated need.

Table 7 shows the correlations in the practice budgets amongst the three calculations. We have also calculated the practice allocation which would result from applying to practices the previous resource allocation formula (CARAN) based on small area data which was used to allocate funds to PCTs. The four allocations are reasonably highly correlated. Table 8 shows the correlations of the difference between each individual based allocation and the CARAN allocation with the mean of the individual based and CARAN allocations. If we interpret the mean of the individual and CARAN based allocations as an estimate of the true need based allocation a positive correlation implies that the individual based allocation is more responsive to need than the CARAN allocation. Table 8 suggests that parsimonious model, with no freezing of encounters, produces allocations which are more responsive to need than the new parsimonious model. The frozen parsimonious model based allocations are less responsive to need than CARAN based allocations.

Table 9 shows how far the practice budget shares calculated by the three methods are from practice shares of total expenditure in 2007/8. The budgets from the parsimonious model would be closest to actual expenditure shares and those from the frozen parsimonious model the furthest away. From Table 10 we see that distances between practice budget shares and expenditure shares are only very weakly correlated with practice need

6. Conclusions

Despite the problematic distribution of individuals' hospital expenditure (a high proportion of zero expenditures, a long right tail) we found that, because of the large (5M) sample sizes we were able to use a simple one part OLS which did better in predicting expenditure than the more elaborate one and two part OLS models with transformed expenditure or GLM. This is in line with other work in health care systems with different institutional and financial arrangements (Van de Ven and Ellis, 2000). The R^2 on our models were also in line with studies in other countries.

Our dependent variable was total expenditure with mental health and maternity excluded. In future work it will be worth investigating whether disaggregating expenditure into more homogenous types of hospital care improves model fit. One obvious disaggregation would be into expenditure on elective and emergency care since different factors are likely to influence the need for these groups of care. Another possibility is to decompose expenditure into functional programme budget categories.

The key explanatory variables were the morbidity variables derived from ICD10 diagnoses reported for hospital use in the two previous years. We found that a relatively straightforward partition with 152 categories, used by the Information Centre to report volume of activity, did as well in terms of per capita practice predictions as systems with more categories. Further refinements (for example age/gender interactions with some of the diagnostic categories, adding more years of history, allowing for the time since diagnosis) are likely to boost predictive performance.

Encounter variables, especially the number of outpatient appointments and the number of inpatient episodes in the previous two years, were very powerful explanatory variables, yielding considerable increases in predictive power at individual level and worthwhile, though smaller increases in R^2 at per capita practice level. However, the interpretation of these variables as reflecting past morbidity or supply factors and the possibility that using them to set allocations could generate perverse incentives for practices, makes it highly debatable whether they should be allowed to influence target allocations.

The choice between the three methods for determining target allocations for practices depends on factual judgements about the underlying unobservable structural model determining expenditure and the magnitude of perverse incentives. It also depends on value judgements about the importance of differences in need across practice populations. It is clear, irrespective of one's preferred procedure, that actual allocations differ considerably from those based on measures of need derived from the expenditure models.

References

Bindman, A., K. Grumbach, D. Osmond, A. Komaromy, K. Vranizan, and N. Lurie. 1995. Preventable hospitalizations and access to health care. *Journal of the American Medical Association,* 274:305–11.

Blough, D.K., Madden, C.W., Hornbrook, M.C. 1999. Modeling risk using generalized linear models. *Journal of Health Economics*, 18:153-171.

Buntin, M.B., Zaslavsky, A.M. 2004. Too much ado about two-part models and transformation? Comparing methods of modelling Medicare expenditures. *Journal of Health Economics*, 23:525-542.

Campbell, S., Reeves, D., Kontopantelis, E., Sibbald, B. and Roland, M. 2009. Effects of pay for performance on the quality of care in England. *New England Journal of Medicine*, vol. 361:pp.368-378.

Carr-Hill R, Hardman G, Martin S, Peacock S, Sheldon TA, Smith PC. 1994. *A formula for distributing NHS revenues based on small area use of hospital beds*. Centre for Health Economics, University of York, September.

Copas, J.B. Cross-validation shrinkage of regression predictors. 1987. *Journal of the Royal Statistical Society, Series B*, 54:175-183.

Croxson, B., Propper, C., Perkins, A., 2001. Do doctors respond to financial incentives? UK family doctors and the GP fundholder scheme. *Journal of Public Economics* 79:375–398.

Davison, A.C., Gigli, A. 1989. Deviance residuals and normal scores plots. *Biometrika*. 1989, 76:2:211-221.

Department of Health. 2010. *Equity and excellence: liberating the NHS*. Cm 7881. www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_117353

Dixon, J., Smith, P.C., Gravelle, H., Martin, S., Bardsley, M., Rice, N., Georghiou, T., Steventon, A., Dusheiko, M., Billings, J., De Lorenzo, M., and Sanderson, C. 2011. Calculating budgets for commissioning general practices in England. *British Medical Journal*. 343:d6608 doi: 10.1136/bmj.d6608. 22 November 2011.

Duan, N. 1983. Smearing estimate: a nonparametric retransformation method. *Journal of the American Statistical Association*, 78:605-610.

Dusheiko, M., Doran, T., Gravelle, H., Fullwood, C., Roland, M. 2011a. Does higher quality of care in general practice reduce unplanned hospital admissions? *Health Services Research*, 46:1:27-46.

Dusheiko, M., Gravelle, H., Martin, S., Rice, N., Smith, P. 2011b. Does disease management reduce hospital costs? Evidence from English primary care. *Journal of Health Economics*. 30:919-932.

Dusheiko, M., Goddard, M., Gravelle, H. and Jacobs, R. 2008. Explaining trends in concentration of health care commissioning in the English NHS. *Health Economics*, 27:907-927.

Dusheiko, M., Gravelle, H., Jacobs, R., 2004. The effect of practice budgets on patient waiting times: allowing for selection bias. *Health Economics* 13:941–958.

Dusheiko, M., Gravelle, H., Jacobs, R., Smith, P.C. 2006. The effect of budgets on doctor behaviour: evidence from a natural experiment. *Journal of Health Economics*, 25:449-478.

Ettner, S.L., Frank, R.G., McGuire, T.G., Newhouse, J.P., Notman, E.H.1998. Risk adjustment of mental health and substance abuse payments. *Inquiry*, 35:223-239.

Gravelle, H., Sutton, M., Morris, S., Windmeijer, F., Leyland, A., Dibben, C. and Muirhead, M. 2003. Modelling supply and demand influences on the use of health care: implications for deriving a needs-based capitation formula. *Health Economics*, 12:985-1004.
Jones, A.M. 2011. Models for health care. In Hendry, D., Clements, M. (eds.) *Oxford handbook of economic forecasting*, Oxford University Press.

Manning, W.G. The logged dependent variable, heteroskedasticity, and the retransformation problem. *Journal of Health Economics*, 1998;17:283-295.

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

McCullagh, P., Nelder, J.A. 1989. *Generalized linear models* (Second Edition). London, Chapman & Hall.

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*. Final Report, 30 November 2007. ACRA(2007)22. http://www.dh.gov.uk/prod_consum_dh /groups/dh_digitalassets/documents/digitalasset/dh_093169.pdf

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

Park, R. 1966. Estimation with heteroskedastic error terms. *Econometrica*, 34:888.

Pope, G., J. Kautter, R. Ellis, A. Ash, J. Ayanian, L. Iezzoni, M. Ingber, J. Levy and J. Robst (2004). Risk Adjustment of Medicare Capitation Payments Using the CMS-HCC Model. *Health Care Financing Review* 25(4):119-141.

Propper, C., Croxson, B., Shearer, A., 2002.Waiting times for hospital admissions: the impact of GP fundholding. *Journal of Health Economics* 21:227–252.

Santos Silva, J., Tenreyo, S. 2006. The log of gravity. *Review of Economics and Statistics*. 88(4):641–658.

Smith PC. 2007. *Formula funding of public services*. Routledge. London.

Smith PC, Rice N, Carr-Hill R. 2001. Capitation funding in the public sector. *JRSS Ser A* , 164:217–257.

Sutton, M., Gravelle, H., Morris, S., Leyland, A., Windmeijer, F., Dibbin, C., Muirhead, M. 2002. *Allocation of Resources to English Areas: Individual and Small Area Determinants of Morbidity and Use of Health Care*. Report for Department of Health, December 2002.
http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/@dh/@en/documents/digitalasset/dh_4137768.pdf

Van de Ven, W., Ellis, R. 2000. Risk adjustment in competitive health plans, in Culyer, A.J. and Newhouse, J. eds., *Handbook of Health Economics*, Vol 1A, North Holland.

Appendix: Structural model

Consider a simple model of need, utilisation (encounters) and expenditure:

$$N_i = \alpha_0 + \alpha_1 x_{1i} + \alpha_2 x_{2i} + \varepsilon_i^N \tag{10}$$

$$U_i = \delta_0 + \delta_1 z_{1i} + \delta_2 z_{2i} + \delta_3 N_i + \varepsilon_i^U \tag{11}$$

$$C_i = \beta_0 + \beta_1 z_{1i} + \beta_2 z_{2i} + \beta_3 N_i + \beta_4 U_i + \varepsilon_i^C \tag{12}$$

where N is need, U is utilisation, C is expenditure. The x are need variables, z are supply variables. We observe x_1 but not x_2, and z_1 but not z_2. We do not observe N. Figure A1 illustrates. The three errors in the model are conditionally mean independent of the explanatories and mutually uncorrelated. The need and supply variables may be correlated with each other but the errors in three equations are uncorrelated, conditional on the need and supply variables.

Suppose we estimate two models of expenditure by OLS in an attempt to calculate unobserved need. The first model (corresponding to our parsimonious model) includes observed need variables, observed supply variables and utilisation. The second estimated model (corresponding to our restricted parsimonious model) drops utilisation.

$$C_i = a_0 + a_1 x_{1i} + a_2 z_{1i} + a_3 U_i + \varepsilon_i^1 \tag{13}$$

$$C_i = c_0 + c_1 x_{1i} + c_2 z_{1i} + \varepsilon_i^2 \tag{14}$$

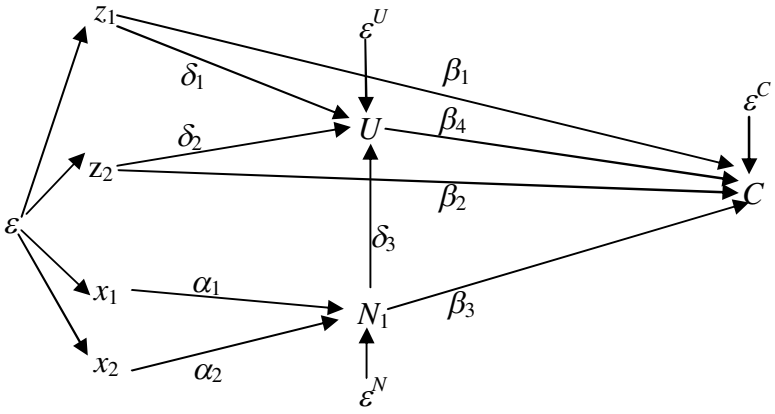


Figure A1. Model of need, utilisation, and cost

From (8) the true cost model corresponding to estimation model (9) which includes the utilisation variable is

$$C_i = \beta_0 + \beta_3\delta_0 + \beta_1 z_{1i} + \beta_2 z_{2i} + \beta_3\alpha_1 x_{1i} + \beta_3\alpha_2 x_{2i} + \beta_4 U_i + \varepsilon_i^C + \beta_3\varepsilon_i^N \tag{15}$$

and so the coefficients on the variables included in (9) pick up their effects plus the effects of the omitted need and supply variables (x_2, z_2) with which they are correlated:

$$\text{plim}\ \hat{a}_0 = \beta_0 + \beta_3\delta_0 + b_{x_2 0.x_1 z_1 U}\beta_3\alpha_2 + b_{z_2 0.x_1 z_1 U}\beta_2$$

$$\text{plim}\ \hat{a}_1 = \beta_3\alpha_1 + b_{x_2 x_1 . z_1 U}\beta_3\alpha_2 + b_{z_2 x_1 . z_1 U}\beta_2 \tag{16}$$

$$\text{plim}\ \hat{a}_2 = \beta_1 + b_{x_2 z_1 . x_1 U}\beta_3\alpha_2 + b_{z_2 z_1 . z_1 U}\beta_2 \tag{17}$$

$$\text{plim}\ \hat{a}_3 = \beta_4 + b_{x_2 U . x_1 z_1}\beta_3\alpha_2 + b_{z_2 U . x_1 z_1}\beta_2 \tag{18}$$

where, for example, $b_{x_2 0.x_1 z_1 U}$ is the constant term from the linear multiple regression of x_2 on x_1, z_1 and U, and $b_{x_2 x_1 . z_1 U}$, $b_{x_2 z_1 . x_1 U}$ are the coefficients on x_1 and z_1.

Similarly the true cost model corresponding to estimation model (10) which drops the utilisation variable is

$$C_i = \beta_0 + (\beta_3 + \beta_4\delta_3)\alpha_0 + \beta_4\delta_0 + (\beta_1 + \beta_4\delta_1) z_{1i} + (\beta_2 + \beta_4\delta_1) z_{2i} + (\beta_3 + \beta_4\delta_3)\alpha_1 x_{1i} + (\beta_3 + \beta_4\delta_3)\alpha_2 x_{2i} + \varepsilon_i^C + (\beta_3 + \beta_4\delta_3)\varepsilon_i^N + \beta_4\varepsilon_i^U \quad (19)$$

and the estimated coefficients for (10) have probability limits

$$\text{plim}\ \hat{c}_0 = \beta_0 + (\beta_3 + \delta_3\beta_4)\alpha_0 + \beta_4\delta_0 + b_{x_2 0.x_1 z_1}(\beta_3 + \delta_3\beta_4)\alpha_2 + b_{z_2 0.x_1 z_1}(\beta_2 + \beta_4\delta_1) \quad (20)$$

$$\text{plim}\ \hat{c}_1 = (\beta_3 + \delta_3\beta_4)\alpha_1 + b_{x_2 x_1 . z_1}(\beta_3 + \delta_3\beta_4)\alpha_2 + b_{z_2 x_1 . z_1}(\beta_2 + \beta_4\delta_1) \quad (21)$$

$$\text{plim}\ \hat{c}_2 = (\beta_1 + \delta_3\beta_4) + b_{x_2 z_1 . x_1}(\beta_3 + \delta_3\beta_4)\alpha_2 + b_{z_2 z_1 . x_1}(\beta_2 + \beta_4\delta_1) \quad (22)$$

where, for example, $b_{x_2 0.x_1 z_1}$ is the constant term from the linear multiple regression of x_2 on x_1, z_1 and $b_{x_2 x_1 . z_1}$, $b_{x_2 z_1 . x_1}$ are the coefficients on x_1 and z_1.

We can calculate three measures of need using the results and then use them to calculate practice allocations according to the procedure outlined in section 5.1. The first assumes that utilisation is a need variable which should affect the allocation, but freezes supply at its mean

$$\hat{N}_i^1 = \hat{a}_0 + \hat{a}_1 x_{1i} + \hat{a}_2 \bar{z}_1 + \hat{a}_3 U_i \quad (23)$$

The second assumes that utilisation is not a need variable and freezes it and supply at the national means:

$$\hat{N}_i^2 = \hat{a}_0 + \hat{a}_1 x_{1i} + \hat{a}_2 \bar{z}_1 + \hat{a}_3 \bar{U} \quad (24)$$

The first two approaches use the estimated cost model (18) which includes the utilisation variable.

The third approach, like the second approach (24), also assumes that utilisation should not affect allocations but achieves this by dropping utilisation when estimating the expenditure model (15) and basing the allocation on the estimated coefficient on the need variable:

$$\bar{N}_i^3 = \hat{c}_0 + \hat{c}_1 x_{1i} + \hat{c}_2 \bar{z}_1 \quad (25)$$

First we compare $\hat{N}_i^1$ and $\hat{N}_i^2$. The difference between these measures of need is that there is less variation across practices for $\hat{N}_i^2$ than for $\hat{N}_i^1$ because more the encounter variable U does not vary across practices for $\hat{N}_i^2$. The effect of the variation in U across practices in $\hat{N}_i^1$ depends on the estimated coefficient on U ($\hat{a}_3$). This coefficient will pick a direct effect of U on costs via β_4. Since this does not reflect need this part of $\hat{a}_3$ should not influence the practice allocation. Similarly neither should the part of $\hat{a}_3$ which reflects supply variation ($b_{z_2 U . x_1 z_1}\beta_2$. But $\hat{a}_3$ also picks up an indirect effect of the unobserved variable x_2 which determines need ($b_{x_2 U . x_1 z_1}\beta_3\alpha_2$) and we would want to have this affect the allocation. Unfortunately there is no means of decomposing $\hat{a}_3$ into its constituent parts and just using the part related to need. The choice between $\hat{N}_i^1$ and $\hat{N}_i^2$ as a measure of need

depends on a judgement on whether including the need component of $\hat{a}_3$ improves the measure of need more than the other two components worsen it.

Comparing $\hat{N}_i^3$ and $\hat{N}_i^2$, there will again be less variation in $\hat{N}_i^2$ because U is frozen at its mean. And, as a result of dropping the utilisation variable U in estimating the expenditure model, the coefficient on the observed need variable is changed from $\hat{a}_1$ to $\hat{c}_1$. The difference in the probability limits is

$$\begin{aligned}\text{plim}\,(\hat{c}_1 - \hat{a}_1) = \delta_3\beta_4\alpha_1 + \left[\left(b_{x_2x_1.z_1} - b_{x_2x_1.z_1U}\right)\beta_3 + b_{x_2x_1.z_1}\delta_3\beta_4\right]\alpha_2 \\ + \left(b_{z_2x_1.z_1} - b_{z_2x_1.z_1U}\right)\beta_2 + b_{z_2x_1.z_1}\beta_4\delta_1 \end{aligned} \quad (26)$$

The first term $\delta_3\beta_4\alpha_1$ reflects the observed need variable x_1 via its effect on utilisation which is frozen out of $\hat{N}_i^2$. The second term is the change in the extent to which the unobserved need variable x_2 is picked up by the observed need variable because it is now also picking the effects of x_2 via the correlation of U and x_1. The third term is the change in the extent to which the coefficient on x_1 is contaminated by the unobserved supply variable z_2. The last term is an additional contamination of the coefficient on x_1 via the effect of the unobserved supply variable on utilisation. Again a judgement about the relative importance of including more of the effect of need in $\hat{N}_i^3$ against picking more of the supply effects is required to choose between $\hat{N}_i^3$ and $\hat{N}_i^2$.

The comparison of $\hat{N}_i^1$ and $\hat{N}_i^3$ is even more complicated since in addition to the comparison of $\hat{a}_1$ to $\hat{c}_1$ we have to take account of the fact $\hat{N}_i^1$ allows for U to affect the allocation.

If one believes that incentive effects should carry less weight than having a good measure of need and that encounter variables reflect need more than supply, then the best measure of need is $\hat{N}_i^1$. If one believes that the encounter variables are either reflecting supply more than need or one places more weight on perverse incentives than on a better measure of need, then the choice is between $\hat{N}_i^3$ and $\hat{N}_i^2$ which removed the effects of encounter variables in different ways.

The above procedure freezes supply (or other non-need variables which should not affect allocations) its national mean. This convention is essentially arbitrary but has been used in previous allocation formulae. One obvious alternative would be to set all non-need variables to zero. This would have the effect of making calculated practice need more variable across practices. As a consequence the second method of sterilising the effect of encounter variables by freezing them ($\hat{N}_i^2$) would then lead to greater variation in calculated need than the dropping them from the estimated model ($\hat{N}_i^3$).

Table 1. Comparison of one part models of 2006/7 patient expenditure

One-Part Models	Estimation Sample N = 5,208,225				Validation Sample 1 N = 5,207,347			Validation Sample 2 Practices: N = 812			
	R^2	RMSE	MAPE	AIC	R^2	RMSE	MAPE	R^2 total practice	RMSE practice per capita	MAPE practice per capita	R^2 practice per capita
OLS	.1672	2462.6	599.5	96,122,909	.1685	2475.1	601.0	.9632	60.5	41.2	.7441
Square-root models											
SQRT(cost) – No smearing	.1303	2533.5	456.8	44,254,235	.1306	2548.7	458.5	.9184	301.1	288.4	.5724
SQRT(cost) – Smearing	.1303	2517.2	633.5	44,254,235	.1306	2532.5	635.2	.9088	87.2	62.8	.5724
SQRT(cost) – Heteroskedastic (by decile)	.1437	2500.8	598.6	44,254,235	.1444	2516.0	600.2	.9446	71.8	49.3	.6395
GLM models with identity link function†											
GLM: Constant variance	.1672	2462.6	599.5	96,122,909	.1685	2475.1	601.0	.9632	60.5	41.2	.7441
GLM: Variance = Mean	*	*	*	*	*	*	*	*	*	*	*
GLM: Variance = Mean^2	*	*	*	*	*	*	*	*	*	*	*
GLM: Variance = Mean^3	*	*	*	*	*	*	*	*	*	*	*
GLM models with Log link function											
GLM: Constant variance	.1367	2509.1	704.8	96,317,753	.0962	2609.0	708.2	.8432	157.3	122.5	.3059
GLM: Variance = Mean	.0609	2878.0	629.3	8.164e+9	.0439	3292.5	633.3	.9442	73.5	49.4	.6297
GLM: Variance = Mean^2	.00000005	3.0e+15	2.5e+25	68,062,912	2.2e-8	9.8e+14	4.2e+24	.0003	1.2e+14	1.2e+24	.0010
GLM: Variance = Mean^3	*	*	*	*	*	*	*	*	*	*	*

* Model failed to estimate. † A Park test yields $\lambda = 1.67$, suggesting that the GLM variance function most closely resembles the gamma (variance proportional to the mean squared) family. The AIC for the log models also favours the gamma family.

Models are for 2006/7 expenditure of patients on practice list at 1 April 2006 using 38 age/sex categories, 152 ICD10 morbidity categories, 4 encounter variables, 6 attributed need, 6 attributed supply variables, 152 PCT effects, with morbidity and encounters measured for 2004/5, 2005/6.

Table 2. Comparison of two part models of 2006/7 patient expenditure

	Estimation Sample N = 5,208,225				Validation Sample 1 N = 5,207,347			Validation Sample 2 Practices: N = 812			
	R^2	RMSE	MAPE	AIC	R^2	RMSE	MAPE	R^2 total practice	RMSE practice per capita	MAPE practice per capita	R^2 practice per capita
Probit + OLS	.1680	2461.5	591.7	32,176,474	.1697	2473.6	593.1	.9631	61.0	41.5	.7434
Square-root models											
Probit + SQRT(cost) – No smearing	.1514	2493.2	508.9	15005327	.1513	2507.1	510.3	.9516	189.7	173.9	.6811
Probit + SQRT(cost) – Smearing	.1557	2480.0	603.5	15005327	.1560	2493.8	604.9	.9422	71.2	48.1	.6660
Probit + SQRT(cost) – Heteroskedastic (by decile)	.1585	2475.5	593.6	15005327	.1594	2489.3	595.0	.9520	67.6	45.5	.6832
Log models											
Probit + Log(cost) – No smearing	.00045	46707	521.4	5395689.9	.0003	63082	532.4	.0584	598.2	280.1	.0024
Probit + Log(cost) – Smearing	.00045	111812	713.6	5395689.9	.0003	151063	739.1	.0584	1309.1	152.9	.0024
Probit + Log(cost) – Heteroskedastic (by decile)	.00045	129441	743.0	5395689.9	.0003	174883	772.3	.0496	1517.1	180.6	.0024
GLM models with identity link function†											
Probit + GLM: Constant variance	.1680	2461.5	591.7	32,176,474	.1697	2473.6	593.1	.9631	61.0	41.5	.7434
Probit + GLM: Variance = Mean	*	*	*	*	*	*	*	*	*	*	*
Probit + GLM: Variance = Mean^2	.1535	2490.5	592.0	26,647,127	.1558	2502.1	593.4	.9523	70.1	46.3	.6813
Probit + GLM: Variance = Mean^3	.1441	2506.3	590.7	33,867,407	.1464	2158.5	592.2	.9483	74.3	49.1	.6588
GLM models with Log link function											
Probit + GLM: Constant variance	.1526	2484.1	618.4	32,206,841	.1302	2539.6	620.9	.9174	88.9	65.6	.5552
Probit + GLM: Variance = Mean	.1183	2572.7	602.4	4.5e+9	.1048	2651.1	604.9	.9594	63.7	43.5	.7168
Probit + GLM: Variance = Mean^2	.0290	173363	1463.9	26,655,960	.0115	336518.8	1691.2	.0815	3353.9	966.7	.0283
Probit + GLM: Variance = Mean^3	.0012	9.0e+13	8.2e+10	33,867,405	.0002	4.2e+14	2.2e+11	.0002	2.1e+12	1.0e+11	.0002

* Model failed to estimate. ‡ AIC applies to the second stage of the two-part model. † A Park test gives $\lambda = 1.87$, suggesting that the GLM variance function most closely resembles the gamma (variance proportional to the mean squared) family. The AIC for the log models also favours the gamma family.

Models are for 2006/7 expenditure of patients on practice list at 1 April 2006 using 38 age/sex categories, 152 ICD10 morbidity categories, 4 encounter variables, 6 attributed need, 6 attributed supply variables, 152 PCT effects, with morbidity and encounters measured for 2004/5, 2005/6.

Table 3. Comparison of morbidity marker sets

	Sample					
	Estimation		Individual validation		Practice validation	
Morbidity marker set	$\bar{R}^2$		R^2		Per capita R^2	Per capita actual/predicted < +/-10%[1]
152 ICD10 markers, 4 encounter variables, 2003/4, 2004/5	0.0897		0.0917		0.6968	64
152 ICD10 markers, 4 encounter variables, 2004/5	0.0875		0.0909		0.6843	62
22 ICD10 markers, 4 encounter variables, 2003/4. 2004/5	0.0865		0.0884		0.6968	64
22 ICD10 markers, 4 encounter variables, 2004/5	0.0843		0.0878		0.6801	61
281 ICD10 markers, 4 encounter variables, 2003/4. 2004/5	0.0937		0.0963		0.6946	63
281 ICD10 markers, 4 encounter variables,2004/5	0.0920		0.0959		0.6821	61
70 HCCs, 185 ICD10 markers, 4 encounter variables, 2004/5	0.0919		0.0958		0.6818	61
70 HCCs, 152 ICD10 markers, 4 encounter variables, 2004/5	0.0916		0.0958		0.6819	60
260 CCS markers, 4 encounter variables, 2004/5	0.0913		0.0945		0.6841	61
152 ICD10 markers, 2003/4, 2004/5	0.0501		0.0502		0.6462	58

All models contain 36 age/gender groups, 152 PCT groups. Dependent variable: expenditure in 2006/7 on individuals on the practice list at 1 April 2005.
[1] % difference between per capita actual and per capita predicted expenditure in practice is less than 10% in absolute value.

Table 4. Comparison of models for 2007/8 expenditure

	Est indiv R^2	Individ validation R^2	Pract per capita R^2	% Practices with less than 10% diff between per capita predicted & actual
Model 1: age and gender	0.0373	0.0366	0.3444	42
Model 2: age and gender, morbidity markers set 1	0.2656	0.2610	0.7394	62
Baseline model: age and gender, morbidity markers set 1, PCT dummies	0.2659	0.2612	0.8046	71
Full model: age and gender, morbidity markers set 1, PCT dummies, 2 private use variables, 135 attributed needs, 63 supply	0. 2662	0.2615	0.8254	77
Parsimonious model: age and gender, morbidity markers set 1, PCT dummies, 2 private use variables. 5 attributed needs & 3 supply	0.2662	0.2615	0.8224	76
Restricted parsimonious model: age and gender, PCT dummies, 1 private use variable, 6 attributed needs & 3 supply, 152 IC ICD10 categories	0.1272	0.1229	0.7735	68

Models to explain 2007/8 expenditure for individuals on practice lists at 1 April 2007 with morbidity variables from 2005/6, 2006/7. Morbidity markers set 1: 152 ICD10 categories used by Information Centre, 4 encounter variables.

Table 5. Comparison of parsimonious and restricted parsimonious models

	Parsimonious model		Restricted Parsimonious model	
	coef	t	coef	t
number of episodes, 2005/06 & 2006/07	299.07	945.99		
number of attendances, 2005/06 & 2006/07	45.69	232.31		
any priority referral to outpatients, 2005/06 & 2006/07	70.99	25.38		
any treatment received at outpatients, 2005/06 & 2006/07	55.68	14.05		
Persons in social rented housing	0.28	3.77	0.35	4.38
All disability living allowance claimants	337.04	5.42	422.42	6.25
Persons 16-74 with no qualifications - age standardised	23.97	5.68	23.97	5.20
Asthma prevalence rate, 2006			3.17	3.94
ONS area: Mature City Professionals	-23.82	-2.83	-28.25	-3.08
Proportion of students in population	-1319.46	-9.34	-1571.83	-10.28
Whether had a private episode in 2005/06 or 2006/07	-490.59	-29.45	-555.74	-31.11
Whether had a private attendance in 2005/06 or 2006/07	-167.14	-10.16		
MRI machine accessibility score	5436.26	3.76	7781.37	5.26
Residential home beds accessibility score	-7.85	-4.39		
Stroke QOF population achievement 2005	-0.44	-3.11	-0.74	-4.72
Practice list size, 2006			-0.00004	-3.63
Age/gender categories	Yes		Yes	
147 ICD10 morbidity categories	Yes		Yes	
PCT dummies	Yes		Yes	
Model R^2: individual level	0.2662		0.1272	
Model R^2: practice level	0.8224		0.7735	
% practices actual cost per capita within +/- 10% of predicted	75%		68%	

Table 6 Alternative models for resource allocation: morbidity coefficients

		Parsimonious model		Restricted parsimonious model		Diagnoses as % of total diagnoses (2004/5, 2005/6)
		Coef	t	Coef	t	%
A00-A09	*Intestinal infectious diseases*	-84.21	-4.97	90.06	4.88	0.3717
A15-A19	*Tuberculosis*	-90.48	-1.39	856.50	12.04	0.0224
A20-A49	*Certain bacterial diseases*	255.36	11.86	606.18	25.82	0.3705
A50-A64	*Infections with predominantly sexual mode of transmission*	92.16	1.58	288.83	4.53	0.0260
A65-A79	*Other infectious and parasitic disorders*	-412.69	-2.89	-187.79	-1.21	0.0037
A80-A89	*Viral infections of the central nervous system*	-116.33	-1.46	253.94	2.92	0.0125
A90-A99	*Arthropod-borne viral fevers & viral haemorrhagic fevers*	-1189.98	-1.90	-1146.79	-1.68	0.0003
B00-B09	*Viral infections characterized by skin & mucous mem. lesns.*	-109.80	-3.36	331.28	9.30	0.0870
B15-B19	*Viral hepatitis*	488.00	11.42	934.23	20.06	0.0537
B20-B24	*Human imrnunodeficiency virus [HIV] disease*	152.47	2.04	465.51	5.72	0.0166
B25-B34	*Other viral diseases*	-228.78	-12.97	91.65	4.77	0.2676
B35-B49	*Mycoses*	104.29	3.47	354.94	10.82	0.1163
B50-B64	*Protozoal diseases*	-651.65	-6.29	-262.67	-2.32	0.0085
B65-B83	*Helminthiases*	-172.24	-1.40	-70.71	-0.53	0.0055
B85-B99	*Other infectious and parasitic diseases*	231.06	18.90	288.75	21.66	0.8036
C00-C14	*Malignant neoplasm of liporal cavity and pharynx*	478.76	7.96	1239.01	18.89	0.0321
C15-C26	*Malignant neoplasm of digestive organs*	168.45	7.53	1179.78	48.41	0.3424
C30-C39	*Malignant neoplasms of respiratory & intrathoracic organs*	273.44	7.68	978.99	25.24	0.1974
C40-C41	*Malignant neoplasm of bone and articular cartilage*	1192.52	8.77	2349.29	15.84	0.0070
C43-C44	*Malignant neoplasms of skin*	-253.75	-14.77	243.84	13.05	0.2888
C45-C49	*Malignant neoplasms of mesothelial and soft tissue*	536.67	7.26	1493.27	18.52	0.0289
C50	*Malignant neoplasm of breast*	-494.04	-23.19	561.58	24.28	0.2430
C51-C58	*Malignant neoplasms of female genital organs*	369.50	10.26	1043.23	26.57	0.0959
C60-C63	*Malignant neoplasms of male genital organs*	64.74	2.84	380.92	15.31	0.2146
C64-C68	*Malignant neoplasms of urinary tract*	347.95	14.03	1022.19	37.82	0.1862
C69-C72	*Malignant neoplasms of eye, brain & other parts of CNS*	1255.31	18.30	2372.25	31.72	0.0347
C73-C80, C97 Malignant neoplsm. of thyroid and oth. endo. Glands etc.		-10.99	-0.50	1001.88	42.14	0.5386
C81-C96	*Malignant neoplasms of lymphoid, haematopoietic & rel. tiss.*	973.31	37.78	3217.12	115.14	0.1919
D00-D48	*In situ & benign neoplasms and others of uncertainty*	-187.39	-21.88	153.26	16.45	1.2906
D50-D64	*Anaemias*	291.44	27.85	598.87	52.50	1.0715
D65-D89	*Diseases of the blood and blood-forming organs*	256.23	13.35	1396.20	66.88	0.2859
E00-E07	*Disorders of thyroid gland*	0.27	0.02	42.42	3.30	0.6698
E10-E14	*Diabetes Mellitus*	371.41	45.33	540.63	60.62	1.6892
E15-E90	*Endocrine nutritional and metabolic diseases*	37.78	4.87	239.50	28.35	1.9132
F00-F03	*Dementia*	-289.96	-14.48	-435.67	-19.96	0.4390
F04-F09	*Other organic including symptomatic mental disorders*	20.62	0.47	-15.79	-0.33	0.0622
F10-F19	*Mental and behavioural disorders due to psychoactive subst.*	74.60	5.82	226.30	16.19	0.6117
F20-F29	*Schizophrenia, schizotypal and delusional disorders*	-577.99	-24.17	-48.75	-1.87	0.1738
F30-F39	*Mood [affective] disorders*	-140.28	-9.92	92.62	6.01	0.5105
F40-F69	*Neurotic, bahavioural & personality disorders*	-203.89	-10.63	189.43	9.06	0.2463
F70-F79	*Mental retardation*	-1201.03	-19.19	131.43	1.93	0.0250
F80-F99	*Other mental and behavioural disorders*	37.73	1.54	382.93	14.35	0.1384
G00-G09	*Inflammatory diseases of the central nervous system*	598.74	10.95	731.83	12.27	0.0303
G10-G13, G30-G32 Other degenerative diseases (incl. Alzheimer).		18.27	0.70	99.68	3.52	0.2087

G20-G26	*Extrapyramidal & movement disorders (incl. Parkinsonism).*	816.90	31.34	916.53	32.24	0.1714
G35-G37	*Demyelinating diseases (incl Multiple Sclerosis) of the CNS.*	479.51	14.66	904.52	25.37	0.0837
G40-G59	*Epilepsymigraine & other episodic disorders*	13.13	1.41	360.56	35.47	1.0738
G50-G73, G90-G99	*Other diseases & disorders of the nervous syst.*	767.86	36.79	1229.11	54.02	0.2343
G80-G83	*Cerebral palsy & other paralytic syndromes*	414.97	19.19	436.59	18.51	0.2601
H00-H06,H15-H22, H30-H36, H43-H59	*Other disorders of the eye etc.*	-112.49	-11.18	297.99	27.23	0.9350
H10-H13	*Disorders of conjunctiva (including conjunctivitis)*	-136.46	-3.40	108.22	2.48	0.0554
H25-H28	*Disorders of lens (including cataracts)*	-329.22	-35.39	92.20	9.11	1.1944
H40-H42	*Glaucoma*	-249.97	-10.93	-7.51	-0.30	0.1803
H60-H95	*Diseases of the ear and mastoid process*	-160.57	-12.49	178.33	12.74	0.5438
I00-I09	*Rheumatic heart disease*	307.24	8.71	624.97	16.24	0.0841
I10-I15	*Hypertensive diseases*	59.31	10.00	121.24	18.75	4.0337
I20-I25	*Ischaemic heart diseases*	87.47	11.59	247.18	30.04	2.5176
I26-I28	*Pulmonary heart disease & diseases of pulmonary circulation*	-23.25	-0.91	294.00	10.58	0.1773
I30-I52	*Other forms of heart disease*	180.63	22.44	312.42	35.61	2.3034
I60-I69	*Cerebrovascular diseases*	-28.08	-1.93	26.84	1.69	0.7501
I70-I79	*Diseases of arteries, arterioles & capillaries*	575.66	38.37	908.43	55.55	0.5096
I80-I89	*Diseases of veins & lymphatic system nec.*	-159.71	-15.29	162.38	14.27	0.8687
I95-I99	*Other & unspecified disorders of the circulatory system*	102.09	4.88	176.01	7.71	0.2546
J00-J06	*Acute upper respiratory infections*	-215.68	-17.49	105.31	7.83	0.5226
J10-J18	*Influenza & pneumonia*	-1.02	-0.07	133.51	8.64	0.8090
J20-J22	*Other acute lower respiratory infections*	100.96	8.39	359.66	27.43	0.8291
J30-J39	*Other diseases of upper respiratory tract*	-254.89	-21.15	69.70	5.31	0.6698
J40-J47	*Chronic lower respiratory diseases*	138.94	21.01	289.93	40.22	2.3834
J60-J70	*Lung diseases due to external agents*	385.84	8.70	263.30	5.44	0.1125
J80-J99	*Other diseases of the respiratory system*	395.11	26.65	564.52	34.92	0.6597
K00-K14	*Diseases of oral cavity, salivary glands & jaws*	-299.20	-30.99	121.33	11.55	0.9008
K20-K31	*Diseases of oesophagusstomach & duodenum*	-74.19	-9.19	182.78	20.78	1.6714
K35-K38	*Diseases of appendix*	-326.29	-15.17	-93.49	-3.99	0.1884
K40-K46	*Hernia*	-166.77	-19.20	10.23	1.08	1.3122
K50-K52	*Noninfective enteritis & colitis*	7.86	0.79	355.00	32.77	0.9693
K55-K63	*Other diseases of intestines*	-41.29	-5.68	185.93	23.48	1.8992
K65-K67	*Diseases of peritoneum*	462.27	15.14	720.93	21.65	0.1114
K70-K77	*Diseases of liver*	846.51	37.53	1050.54	42.71	0.2321
K80-K87	*Disorders of gall bladder, biliary tract & pancreas*	-134.94	-10.57	193.07	13.87	0.5887
K90-K93	*Other diseases of the digestive system*	29.49	1.86	326.26	18.86	0.4486
L00-L14, L55-L99	*Other infections and disorders of the skin*	-63.03	-7.65	221.39	24.69	1.5199
L20-L30	*Dermatitis and eczema*	-207.62	-8.95	60.20	2.38	0.1530
L40-L45	*Papulosquamous disorders (including Psoriasis)*	12.40	0.37	653.99	17.86	0.0736
L50-L54	*Urticaria and erythems*	-233.82	-5.29	176.99	3.67	0.0434
M00-M25	*Arthropathies*	148.62	23.38	438.69	63.58	2.4448
M30-M36	*Systemic connective tissue disorders*	383.01	14.96	924.55	33.13	0.1422
M40-M54	*Dorsopathies*	71.95	7.55	406.44	39.12	0.9845
M60-M79	*Soft tissue disorders*	-137.75	-14.39	223.71	21.47	0.9597
M80-M94	*Osteopathies and chondropathies*	232.89	16.94	456.05	30.43	0.5298
M95-M99	*Other disorders of the musculoskeletal system & conn. tiss.*	-123.87	-2.38	133.57	2.35	0.0316
N00-N08, N10-N16	*Diseases of the kidney*	830.44	42.10	2145.18	99.94	0.2580
N17-N19	*Renal failure*	1835.29	114.35	4112.19	237.45	0.6264
N20-N23	*Urolithiasis*	-295.39	-14.87	5.25	0.24	0.2138
N25-N29	*Other disorders of kidney & ureter*	953.76	29.40	1555.41	43.98	0.0928
N30-N39	*Other diseases of the urinary system*	-44.05	-4.87	51.83	5.25	1.4596
N40-N51	*Diseases of male genital organs*	-279.83	-24.50	-93.26	-7.49	0.7275

N60-N64	*Disorders of breast*	-284.72	-11.25	80.22	2.91	0.1372
N70-N77	*Inflammatory diseases of female pelvic organs*	-191.35	-9.56	-35.46	-1.62	0.2340
N80-N98	*Noninflammatory disorders of female genital tract*	-279.83	-33.60	43.93	4.85	1.3536
N99	*Other disorders of the genitourinary system*	-290.81	-5.08	-349.03	-5.59	0.0300
O00-O08	*Pregnancy with abortive outcome*	-364.46	-29.73	45.84	3.44	0.5753
O10-O75, O85-O92, O95-O99 Complications of labour and delivery		-536.41	-63.55	-34.34	-3.74	2.8151
O80-O84	*Delivery*	-205.01	-15.29	13.11	0.90	0.5584
P00-P04	*Complications of fetus/neonate affected by maternal*	-6.31	-0.18	36.30	0.93	0.0772
P05-P96	*Other conditions originating in the perinatal period*	-7.52	-0.61	107.39	7.99	0.7717
Q00-Q89	*Congenital malformations*	284.11	22.81	722.35	53.27	0.5771
Q90-Q99	*Chromosomal abnormalities nec.*	676.43	12.96	1080.78	18.99	0.0347
R00-R09	*Symptoms & signs inv. the circulatory/respiratory system*	-98.59	-14.55	208.10	28.19	2.2415
R10-R19	*Symptoms & signs inv. the digestive system & abdomen*	-144.01	-23.00	166.50	24.42	2.6585
R20-R23	*Symptoms & signs inv. the skin & subcutaneous tissue*	-145.45	-8.99	72.82	4.13	0.3381
R25-R29	*Symptoms & signs inv. the nervous & musculoskeletal sys.*	118.09	5.73	99.34	4.42	0.2423
R30-R39	*Symptoms & signs involving the urinary system*	-59.20	-5.72	100.36	8.90	0.9711
R40-R46	*Symptoms & signs inv. Cognition, perception etc.*	38.34	2.72	63.26	4.12	0.5745
R47-R49	*Symptoms & signs inv. speech & voice*	-28.90	-0.93	138.49	4.10	0.1082
R50-R68	*General symptoms & signs*	38.25	5.17	271.35	33.67	2.0899
R69	*Unknown & unspecified causes of morbidity*	36.09	4.08	354.44	36.74	1.5654
R70-R89	*Abnormal findings of bodily fluids or samples without diag.*	-31.58	-1.85	235.03	12.66	0.3178
R90-R94	*Abnormal findings on diagnostic imaging/function studies*	103.03	5.58	218.00	10.84	0.3062
R95-R99	*Ill-defined & unknown causes of mortality*	dropped		dropped		0.0003
S00-S09	*Injuries to the head*	-77.66	-5.89	66.40	4.62	0.7506
S10-S19	*Injuries to the neck*	287.67	6.24	421.96	8.39	0.0423
S20-S29	*Injuries to the thorax*	85.28	2.82	86.36	2.62	0.1099
S30-S39	*Injuries to abdomen, lower back, lumbar spine & pelvis*	-88.36	-3.59	-27.05	-1.01	0.1693
S40-S49	*Injuries to the shoulder & upper arm*	-49.89	-2.24	101.32	4.17	0.2033
S50-S59	*Injuries to the elbow & forearm*	-257.15	-14.65	-10.21	-0.53	0.3510
S60-S69	*Injuries to the wrist & hand*	-302.64	-17.69	-6.23	-0.33	0.3701
S70-S79	*Injuries to the hip & thigh*	93.60	4.92	-72.41	-3.49	0.3954
S80-S89	*Injuries to the knee & lower leg*	-113.36	-6.48	93.82	4.92	0.3632
S90-S99	*Injuries to the ankle & foot*	-135.80	-4.31	32.31	0.94	0.0885
T00-T07	*Injuries involving multiple body regions*	138.75	1.90	163.52	2.05	0.0157
T08-T14	*Injuries to unspecified part of trunk limb or body*	-157.43	-3.43	5.14	0.10	0.0467
T15-T19	*Effects of foreign body entering through natural orifice*	-207.05	-5.26	39.41	0.92	0.0584
T20-T32	*Burns and corrosions*	-207.86	-4.39	-26.72	-0.52	0.0428
T33-T35	*Frostbite*	833.87	1.24	1063.15	1.46	0.0002
T36-T50	*Poisonings by drugs medicaments & biological substances*	-243.57	-10.85	-49.88	-2.04	0.4012
T51-T65	*Tox. effcts. of substances. chiefly non-medicinal as to source*	-143.08	-4.96	-12.22	-0.39	0.1278
T66-T78	*Other and unspecified effects of external causes*	21.85	0.62	197.35	5.17	0.0691
T79	*Certain early complications of trauma*	-347.78	-5.35	-131.15	-1.85	0.0228
T80-T88	*Complications of surgical & medical care nec.*	296.52	23.96	1035.76	76.90	0.8436
T90-T98	*Sequelae of injuries of poisoning & other consequences*	-49.54	-1.68	7.84	0.24	0.1073
Z00-Z13	*Examination and investigation*	-230.59	-28.36	82.49	9.31	1.4851
Z20-Z29	*Potential health hazards related to communicable diseases*	69.12	3.89	276.72	14.27	0.2522
Z30-Z39	*Health services in circumstances related to reproduction*	-327.30	-47.16	-86.42	-11.43	6.0527
Z40-Z54	*Persons encountering health services for specific care*	-123.09	-17.36	858.62	112.14	2.4993
Z55-Z65	*Potential health hazards reltd. to socioeconomic & psychosoc.l*	99.72	6.68	61.65	3.79	0.5344
Z70-Z76	*Persons encountering health services in other circs.*	-58.49	-7.31	63.26	7.25	1.5769
Z80-Z99	*Persons with potential health hazards related to family*	45.76	9.13	248.08	45.48	5.6386
U00-U99	*New diseases, bacterial agents*	dropped		dropped		0.0000

V00-V99	*Transport accidents*	-133.64	-6.39	3.07	0.13	0.2776
W00-W99	*Falls, submersion, electric current, extreme temperatures*	-28.60	-2.53	26.44	2.15	1.9121
X00-X99	*Fire, venomous animals, self-harm, assault*	-79.00	-4.56	109.67	5.81	0.7474
Y00-Y99	*Undetermined intent, war, complications of medical care*	-97.59	-9.00	64.01	5.41	1.1323

Table 7. Correlations between alternative practice needs indices (n=8222; 48 practices with fewer than 500 patients have been excluded)

	Original Parsimonious	Frozen Parsimonious	Restricted Parsimonious	CARAN
Original Parsimonious	1.000			
Frozen Parsimonious	0.824	1.000		
Restricted Parsimonious	0.949	0.879	1.000	
CARAN	0.841	0.941	0.887	1.000

Note: All variables are (practice share of predicted PCT need/practice share of PCT population.

Table 8. Correlations of difference between PBRA practice-based needs indices and CARAN indices with average of PBRA practice-based needs indices (PCT-based) and CARAN indices.

	Mean of parsimonious and CARAN	Mean of frozen parsimonious and CARAN	Mean of restricted parsimonious and CARAN
Difference : Parsimonious and CARAN	**0.273**		
Difference: Frozen parsimonious and CARAN		**-0.795**	
Difference: Restricted parsimonious and CARAN			**0.151**

Note: All variables relate to the PCT based need index per person (that is, the practice share of predicted PCT need/the practice share of PCT population). n=8,222 and excludes 48 practices with fewer than 500 patients.

Table 9. Distribution of practices by distance from target (DFT)

	Percentage of practices more than x% away from target		
	> +/- 5%	**> +/- 10%**	**> +/- 20%**
DFT relative to PCT mean			
Parsimonious model	53.2	26.1	11.1
Frozen parsimonious	69.5	45.7	19.4
Restricted parsimonious model	61.1	34.6	14.0

Table 10. Correlation between practice need and distance from target (DFT)

	Practice need per person		
DFT relative to PCT mean	**Parsimonious model**	**Frozen parsimonious model**	**Restricted model**
Parsimonious model	0.0378		
Frozen parsimonious model		0.0867	
Restricted parsimonious model			0.0400