THE UNIVERSITY of York

Keep it Simple? Predicting Primary Health Care Costs with Measures of Morbidity and Multimorbidity

CHE Research Paper 72

Keep it Simple? Predicting Primary Health Care Costs with Measures of Morbidity and Multimorbidity

[1]Samuel L Brilleman
[2]Hugh Gravelle
[1]Sandra Hollinghurst
[1]Sarah Purdy
[1]Chris Salisbury
[3]Frank Windmeijer

[1]Academic Unit of Primary Care, University of Bristol
[2]Centre for Health Economics, University of York
[3]Department of Economics, University of Bristol

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 presents independent research commissioned by the National Institute for Health Research School for Primary Care Research. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health. Helpful comments were received from Matt Sutton, and at the Society for Academic Primary Care 2011 Bristol conference, and the Health Economists' Study Group meeting in Bangor 2011. We are grateful to Tarita Murray-Thomas at GPRD for assistance and advice with the data. The project was covered by ethics approval for GPRD work granted by Trent Multicentre Research Ethics Committee, reference 05/MRE04/87.

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

In this paper we investigate the relationship between patients' primary care costs (consultations, tests, drugs) and their age, gender, deprivation and alternative measures of their morbidity and multimorbidity. Such information is required in order to set capitation fees or budgets for general practices to cover their expenditure on providing primary care services. It is also useful to examine whether practices' expenditure decisions vary equitably with patient characteristics.

Electronic practice record keeping systems mean that there is very rich information on patient diagnoses. But the diagnostic information (with over 9000 possible diagnoses) is too detailed to be practicable for setting capitation fees or practice budgets. Some method of summarizing such information into more manageable measures of morbidity is required. We therefore compared the ability of eight measures of patient morbidity and multimorbidity to predict future primary care costs using data on 86,100 individuals in 174 English practices. The measures were derived from four morbidity descriptive systems (17 chronic diseases in the Quality and Outcomes Framework (QOF), 17 chronic diseases in the Charlson scheme, 114 Expanded Diagnosis Clusters (EDCs), and 68 Adjusted Clinical Groups (ACGs)).

We found that, in general, for a given disease description system, counts of diseases and sets of disease dummy variables had similar explanatory power and that measures with more categories did better than those with fewer. The EDC measures performed best, followed by the QOF and ACG measures. The Charlson measures had the worst performance but still improved markedly on models containing only age, gender, deprivation and practice effects.

Allowing for individual patient morbidity greatly reduced the association of age and cost. There was a pro-deprived bias in expenditure: after allowing for morbidity, patients in areas in the highest deprivation decile had costs which were 22% higher than those in the lowest deprivation decile.

The predictive ability of the best performing morbidity and multimorbidity measures was very good for this type of individual level cross section data, with R^2 ranging from 0.31 to 0.46. The statistical method of estimating the relationship between patient characteristics and costs was less important than the type of morbidity measure. Rankings of the morbidity and multimorbidity measures were broadly similar for generalised linear models with log link and Poisson errors and for OLS estimation.

It would be currently feasible to combine the results from our study with the data on the number of patients with each QOF disease, which is available on all practices in England, to calculate budgets for general practices to cover their primary care costs

Keywords: multimorbidity; primary care; utilisation; costs; deprivation; budgets

JEL categories: I14, I18

1. Introduction

In this paper we investigate the relationship between patients' primary care costs (consultations, tests, drugs) and their age, gender, deprivation and alternative measures of their morbidity and multimorbidity. The relationship is of policy interest for two reasons. First, in many health care systems general practitioners (GPs) are paid prospectively via capitation fees for the patients on their lists. If policy makers are concerned either with equity, in the sense of ensuring that patients with greater needs for health care carry a larger capitation, or with reducing financial incentives for practices to cream skim or dump patients, then it is necessary to know how patients' expected expenditure varies with their characteristics (Schokkaert et al., 1998). Second, such information can also be used to investigate the equity of GPs' decisions about the allocation of primary healthcare resources amongst their patients. Horizontal equity requires that patients in equal need should receive equal amounts of health care. Vertical equity requires that there is greater expenditure on patients in greater need. Models of expenditure can be used to test whether general practices allocate primary care resources in accordance with vertical and horizontal equity: expenditure should be related to morbidity and should not vary with a patient's socio-economic status or the practice to which they belong (Gravelle et al., 2006).

Whether one wishes to inform the setting of prospective capitation fees, or to test for equity, the importance of morbidity as a determinant of costs implies that good measures of patient morbidity are essential for the empirical modelling of health care expenditure. With the development of electronic patient record keeping systems in general practice it has become possible to obtain information on detailed diagnoses for patients which can be used to construct such measures. The simplest approach with such data is to group diagnoses into a manageable number of morbidity categories which can then be included in regression models of patient costs as a set of dummy variables indicating the presence or absence of the different diagnoses.

But measuring morbidity as a set of morbidity dummies does not allow for the possibility that the effect of diagnoses is not additive. The cost of one multimorbid patient with both diabetes and depression may be more than the cost of two patients, one with diabetes and the other with depression, because it may be more difficult to control blood sugar levels for a patient who also has depression. Conversely, there could be cost savings with some multimorbid patients. For example, heart disease and diabetes are conditions where monitoring of cholesterol may be required but the associated costs need only be incurred once in a given period for a patient with both conditions. Allowing for the possible non-additive effects of multimorbidity is potentially important since the proportion of the population who are multimorbid is non-trivial (20% to 61% in our data set depending on the multimorbidity measure used) and has been growing over time (Hippisley-Cox and Pringle, 2007).

Multimorbidity measures range from simple counts of the number of morbidities to elaborate classification schemes such as the widely used John Hopkins Adjusted Clinical Groups (ACG) Case-Mix System (Johns Hopkins Bloomberg School of Public Health, 2008). More elaborate schemes should be better at predicting costs since they generally use more information and their construction has been guided in part by their predictive ability. But, particularly when used for setting capitation rates, simplicity is also a virtue. Simpler morbidity and multimorbidity schemes are easier for patients and GPs to understand. Setting capitation fees based on morbidity requires that patient morbidity be measured every budgetary period for every patient and more complex schemes have higher measurement and computation costs. Thus there may be a trade-off between simplicity and predictive power when alternative measures of morbidity and multimorbidity are considered.

The main aim of this paper is to compare the ability of a set of morbidity and multimorbidity measures of varying complexity to predict primary care costs. We believe it is the first to do so.[1]

The paper makes a number of other contributions. First, our research is relevant for recent and proposed policy changes in the English NHS which will encourage patients to shop around amongst general practices, with the lifting of restrictions on choice of practice (Department of Health, 2010) and websites such as NHS Choices providing information about practices. With greater mobility of patients

[1] Previous comparisons of the predictive power of alternative morbidity and multimorbidity schemes have focussed on hospital cost (Huntley et al., 2012; Winkelman and Mehmud, 2008; Perkins et al., 2004). Omar et al (2008) examined prescribing costs for English patients but did not compare morbidity measures.

across practices it will be more important to align capitation funding of practices with the expected costs patients impose on practices. Capitation fees attached to patients are currently set by a formula based on results from two models of the determinants of consultation rates and consultation length, as measures of GP workload. The study of consultation length did not include any morbidity variables and the consultation rate study was based on area level data and used area measures of mortality and of patient reported morbidity (Formula Review Group, 2007). Funding for general practice prescribing is allocated by a formula derived from a practice level model of prescribing costs which included three practice level disease prevalence measures (Department of Health, 2011). We investigate whether it is possible to get accurate estimates of cost to be used in setting capitation payments by using the data on individual patient morbidity which is readily available in practice records.

Second, there is some debate about estimation methods for health cost data, particularly about the usefulness of linear ordinary least squares models which do not take account of the skewed distribution of costs and the high proportion of patients with zero costs (Manning, 2006). This literature has focussed on modelling hospital costs whereas we estimate models of primary care cost. We examine whether the performance of alternative morbidity and multimorbidity measures is sensitive to the estimation method.

Third, studies of equity in primary care have been based on population surveys which use a limited range of self reported measures of health (Bago, D'Uva, 2005; Bago D'Uva et al., 2009; Morris et al., 2005; Sutton, 2002). As a by-product of our main analysis we are able to test for horizontal equity by examining whether expenditure varies with socio-economic status after controlling for a very much richer set of morbidity measures which are based on data recorded by health care professionals.

Section 2 describes the data, the construction of the morbidity and multimorbidity measures, and the estimation methods. The results are set out in Section 3 and discussed in Section 4.

2. Methods

2.1 Data

2.1.1 Sample

The General Practice Research Database (GPRD) contains primary care medical records for around 5 million patients currently registered throughout the United Kingdom (UK). An initial random sample of patients aged 18 years and over was drawn from the 182 practices included in the GPRD which had 'research standard' data continuously from 1st April 2005 to 31 March 2008, and which had given consent to link patient data to measures of area deprivation. The sample was stratified by age, gender and practice. The GPRD is considered broadly representative of the general population in the UK (Lawrenson et al., 1999). We dropped 8 practices with entirely missing deprivation data. In order to use the most up-to-date resource use data and the largest possible observation period for diagnoses, we included the 86,100 individuals from the original sample who were alive and registered at one of the remaining 174 practices on 1st April 2007.

2.1.2 Costing

We calculated the total cost of primary care resources used by each patient during the NHS financial year 1st April 2007 to 31st March 2008. This included consultations, prescription drugs, and tests initiated within primary care. All costs were valued in £ sterling at 2007/08 prices.

Consultations included all face-to-face (including surgery consultation, home visit, clinic, out of hours) and telephone consultations. The unit cost of each consultation was based on a combination of consultation type and primary staff role (type of general practitioner, practice nurse or other health care professional leading the consultation). The unit costs for consultations are shown in Table 1. An additional cost was attached to cover administrative activities such as the recording of results or sending mail to a patient when this was performed by a receptionist, administrator, or secretary. Unit costs were taken from Curtis (2008) and from a report on GP earnings and expenses based on GP tax returns (Technical Steering Committee, 2010).

Table 1. Unit costs (£) per primary care encounter

Staff Type	Surgery consultation	Home visit	Clinic	Telephone consultation	Out of hours
GP: partner	24.47	81.37	35.98	14.85	36.97
GP: registrar/associate	15.92	52.92	23.40	9.66	24.05
GP: sole practitioner	27.55	91.61	40.51	16.72	41.63
Practice Nurse	9.00	-	9.00	5.46	-
Counsellor	64.00	-	-	-	-
Other Health Care Professional	15.00	-	15.00	9.10	-

Unit costs for prescription drugs were based on information provided by the GPRD which combined data from several sources, including the National Drug Tariff for generic products, and manufacturers for branded products. Each prescription drug in the patient level data was matched to unit cost using drug name, strength and formulation. Where there was more than one unit cost for a prescription drug we used the median unit cost.

To allow for possible ambiguities in the mapping of recorded drug quantities to costs we also computed a measure of prescription costs using data from the prescription cost analysis of the NHS Information Centre (2008a) which provides information on the net ingredient cost of all prescriptions dispensed in the community in England. We used the British National Formulary (BNF) code (British Medical Association and Royal Pharmaceutical Society, 2010) to attach these average costs to each prescription in the GPRD data, ignoring the quantity specified in the prescription.

With advice from a general practitioner member of the research team (SP) we determined which tests were performed within a standard surgery consultation and applied a zero unit cost, save for the cost associated with any consumables such as pregnancy test kits or urine dipsticks. Unit costs for the remaining tests were based on the National Health Service (NHS) Reference Costs (Department of Health, 2009). For laboratory tests the unit cost was based on pathology discipline. Hospital-based tests and investigations requested by the practice were costed using the NHS Reference Costs.

2.1.3 Measures of morbidity and multimorbidity

We use four methods of categorising diagnoses to create three measures of morbidity and five multimorbidity measures (Table 2).

Table 2. Morbidity and multimorbidity measures

Measure	Number diseases/ categories	Range of measure	Details
QOF disease dummy variables	17 Not mutually exclusive	0-1 dummies	17 chronic diseases in the clinical domain of the UK Quality and Outcomes Framework (QOF) pay for performance scheme. Not mutually exclusive
QOF disease count	17	0 to 17	Count of the QOF diseases.
Charlson disease dummy variables	17 Not mutually exclusive	0-1 dummies	17 chronic diseases (similar to but not identical to the QOF set of chronic diseases). Not mutually exclusive.
Charlson Index score	17	0 to 33	A weighted score, where weights of 1, 2, 3, or 6 are given to each of the 17 chronic diseases in the Charlson set depending on the strength of their relationship with patient mortality.
Expanded Diagnosis Clusters (EDCs) dummy variables	114 Not mutually exclusive	0-1 dummies	EDCs are clinically related groupings of diagnoses selected by the ACG System. We include the 114 (of the possible 264) EDCs which have been identified as chronic. Not mutually exclusive.
Count of EDCs	114	0 to 114	Count of EDCs
Adjusted Clinical Groups(ACGs)	68 mutually exclusive categories	0-1 dummies	ACGs are mutually exclusive categories. Classification into an ACG is based on an individual's age, gender, combination of morbidities and expected cost. The age range of our sample meant we used only 68 out of 82 possible ACG categories.
Resource Utilization Bands (RUBs)	6 mutually exclusive categories	0-1 dummies	The ACG software groups ACGs into 6 mutually exclusive Resource Utilization Bands on the basis of their expected costs: 0 - No or only invalid diagnoses, 1 - Healthy Users, 2 – Low, 3 – Moderate, 4 – High, 5, Very High. Patients assigned to higher bands are expected to have higher costs.

QOF disease categories. The electronic record systems in UK general practices use Read codes to record summary clinical information on patients. It is possible to use the Read codes to construct various sets of diagnostic categories.[2] We used the 17 chronic conditions included in the clinical domain of the 2006/7 version of the Quality and Outcomes Framework (QOF) which is a pay for performance scheme covering all practices in the UK. This set of morbidity markers is simple and likely to be reliably recorded because it forms the basis for payment under the QOF. It also has high face validity as the main business of general practices is dealing with chronic conditions, although it omits some chronic conditions such as skin disease and liver disease. The 17 conditions are asthma, atrial fibrillation, cancer, coronary heart disease (CHD), chronic kidney disease (CKD), chronic obstructive pulmonary disease (COPD), dementia, depression, diabetes, epilepsy, heart failure, hypertension, learning difficulties, mental health, obesity, stroke, and hyperthyroidism. We used the QOF Business Rules Version 16[3] which outlines the clinical Version 2 Read codes and any additional criteria required to include a patient on the relevant QOF disease register. The 17 QOF morbidities were included in the regression models as 17 dummy variables as description of patient morbidity.

In principle all possible types of multimorbidity can be represented by a suitable combination of the dummy variables for the separate morbidity categories. But with anything other than a very coarse set of categories, there will be too many combinations to be practical: with m different diagnostic categories there are 2^m possible morbidity states. Thus some means of aggregating the information on diagnostic categories into more manageable descriptions of multimorbidity is required. Our first multimorbidity measure is a count of the number of QOF morbidity categories into which a patient falls.

[2] It is possible to map from Read codes into ICD-10 which has over 9000 diagnostic categories in its finest classification and 20 in its coarsest (both counts excluding chapters XX and XXII). See http://www.who.int/classifications/icd/en/

[3] NHS Primary Care Contracting. QOF Implementation Business Rules v16. http://www.primarycarecontracting.nhs.uk/145.php

We drew on a review of the use of established indices of multimorbidity to consider which other measures might be appropriate to a UK primary care setting (Huntley et al, 2012). We selected the Charlson Index and measures derived from the John Hopkins ACG system because they are widely used and potentially straightforward to operationalise with routine data.

Charlson diseases. The Charlson Index is a diagnosis-based multimorbidity measure that weights 17 diseases on the basis of the strength of their association with mortality (Charlson et al., 1987). We use an adaptation by Khan et al (2010). The diseases included in the Charlson Index score, with associated weight in parentheses, were cerebrovascular disease (1), chronic pulmonary disease (1), congestive heart disease (1), dementia (1), diabetes (1), mild liver disease (1), myocardial infarction (1), peptic ulcer disease (1), peripheral vascular disease (1), rheumatological disease (1), cancer (2), diabetes with complications (2), hemiplegia and paraplegia (2), renal disease (2), moderate or severe liver disease (3), AIDS (6), and metastatic tumour (6). Khan et al. (2010) provides the clinical Version 2 Read codes for diagnosing each disease, using a translation from the widely used Deyo adaption of the Charlson Index for ICD-9 codes (Deyo et al., 1992). About half of the 17 conditions are similar to those in the set of QOF chronic conditions, though the precise definitions vary. We used the weighted sum of the Charlson diseases (the Charlson Index) as a multimorbidity measure and a vector of 17 dummy variables for the Charlson diseases to describe patient morbidity.

The John Hopkins ACG Case-Mix System is also diagnosis-based and was developed using administrative claims data in the United States (US) (Starfield et al. 1991; Weiner et al. 1991). The system has been validated in a number of studies in the US and has been studied in primary care in, for example, Sweden (Halling et al. 2006) and Canada (Reid et al., 2002). Further studies have used the ACG system to predict hospital referrals and prescribing rates in the UK (Sullivan et al., 2005; Omar et al., 2008). The ACG system has recently been expanded to allow the input of Version 2 Read codes. The software provides a variety of morbidity measures and multimorbidity measures.

Expanded Diagnosis Clusters (EDCs) are groupings of diagnostic codes which are clinically similar. An individual was assigned to an EDC if any of the Version 2 Read diagnostic codes relating to that EDC appeared in their clinical data. We designated 114 of the 264 EDCs as representing a chronic condition (Salisbury et al., 2011) and measured morbidity as a vector of 114 dummy variables. We measured multimorbidity by the number of chronic EDCs an individual was included in. We used the John Hopkins ACG System Version 8.2 (Johns Hopkins Bloomberg School of Public Health, 2008) to obtain the EDC classifications.

Adjusted Clinical Groups (ACGs) are mutually exclusive categories (unlike the QOF, Charlson, and EDC categorisations). Diagnoses are first grouped into 32 non-mutually exclusive Aggregated Diagnosis Groups (ADGs) by duration, severity of condition, diagnostic certainty, etiology (infectious, injury, or other), and specialty care involvement. (EDCs are more purely diagnosis based.) ADGs are in turn grouped into 12 collapsed ADGs (CADGs) based on persistence, severity, type of healthcare required. ACGs are mutually exclusive combinations of CADGs which group patients with similar healthcare needs. The grouping depends on ADGs and also on age and gender. The ACG classifications were based on all patient diagnoses recorded over the one-year period prior to resource use as suggested by the system software (Johns Hopkins Bloomberg School of Public Health, 2008). Every patient was placed into one of 68 mutually exclusive ACGs.[4] At least 35 of these ACGs are for multimorbid patients in that they contain patients with at least 2 ADGs.

Resource Utilization Bands (RUBs). The ACG software groups ACGs with similar expected expenditure into 6 Resource Utilization Bands. 0 - No or only invalid diagnoses, 1 - Healthy Users, 2 – Low, 3 – Moderate, 4 – High, 5 - Very High. Higher bands are expected to have higher costs and patients in them are more likely to be multimorbid.

Table 2 summarises the three morbidity measures (vectors of QOF, Charlson, EDC morbidity markers) and the five measures of multimorbidity (counts of the QOF, Charlson, EDCs plus ACGs and RUBs) used in our analysis. The QOF, Charlson, and EDC morbidity categories were constructed using all historic diagnoses on patients' general practice records up to 31st March 2007. The ACG and RUB measures use diagnoses over the one-year period 1st April 2006 to 31st March 2007.

[4] Though the ACG System Version 8.2 identifies up to 82 default ACGs (or up to 93 if optional branching is turned on), the age range of our sample meant that some of these categories were not populated.

2.1.4 Covariates

For each gender, age at 1st April 2007 was categorised into ten-year age bands with the exception of 90+ years as the upper category. Deprivation, measured by deciles of the Index of Multiple Deprivation (IMD) 2007 (Department for Communities and Local Government, 2008) which is based on seven dimensions of deprivation, was attributed to the individuals by GPRD using patients' Lower Layer Super Output Area (LSOA) of residence. There are 32,482 LSOAs in England with a mean population of 1500.

2.2 Modelling

We estimated separate regression models of individual cost using the eight morbidity and multimorbidity measures. The three numerical multimorbidity measures (QOF count, Charlson Index score, EDC count) and the ordered RUB multimorbidity measure were included as dummy variable categories to estimate the most flexible relationships between multimorbidity and cost. In our dataset, the maximum QOF count was 10, the maximum Charlson score was 13, and the maximum EDC count was 28. We used 6 categories for the QOF count (1,2,..,6 or more), 7 for the Charlson (1,2,…,7 or more) and 18 for the EDC count (1,2,…,18 or more) as there were few patients with larger numerical scores. We used dummy variables for the 68 mutually exclusive ACG categories. For the models with non-mutually exclusive QOF, Charlson, and EDC morbidity categories, we used dummy variables for each of the categories.

The distribution of healthcare costs for individual patients usually has a long right hand tail and a spike at zero cost reflecting non-use by a non trivial proportion of the population. This has led to some debate about the appropriate estimation method for models of individual cost with suggestions including transformation of the cost variable, two part models, and Generalised Linear Models (GLMs) (Blough et al., 1999; Buntin and Zaslavsky, 2004; Manning, 2006; Manning and Mullahy, 2001; Manning et al., 2005; Mullahy, 1998). Our primary care cost data are right skewed (skewness 7.43, with the mean cost being 2.5 times the median), although the proportion of patients with no cost in 2007/8 is smaller (12.3%) than in typical distributions of hospital costs.

We report results from estimated GLMs in which a link function of the conditional expected 2007/8 cost of the *i*'th patient is linear in the explanatory variables:

$$g\left(Ec_{iadmp}\right) = \beta_0 + \sum_{a'} \beta_{a'} D_{ia'} + \sum_{d'} \beta_{d'} D_{id'} + \sum_{m'} \beta_{m'} D_{im'} + \sum_{p'} \beta_{p'} D_{ip'} \qquad (1)$$

and the error distributions were assumed to be Poisson (variance proportional to the mean), or Guassian (normal). The D_{ia} are 15 age/gender group dummies, the D_{id} are 9 deprivation decile dummies, the D_{im} are morbidity or multimorbidity category dummies, and the D_{ip} are 173 practice dummies. All the explanatories are measured at the start of the expenditure year 2007/8 with morbidity and multimorbidity variables based on patient morbidity records up to 31st March 2007.

We used both log link functions ($g(Ec) = \ln Ec$) and linear link functions ($g(Ec) = Ec$) The log form allows for the right skewness of the patient cost data and use of the GLM specification means that we do not have to correct for retransformation bias (Manning 1998) or adjust the dependent variable because a proportion of patients have zero cost. With a linear link function and a normal error distribution the GLM is equivalent to Ordinary Least Squares (OLS).

We estimated models with practice fixed effects D_{ip} to allow for lack of information on practice characteristics (such as the number of GPs per patient) and for intangible "practice style" which might affect patient costs. Robust standard errors were used to allow for clustering of errors within practices. Models were estimated using STATA Version 11.2 (StataCorp, 2009).

We summarise model performance with four goodness of fit measures. The Bayesian Information Criterion (BIC) (Schwarz, 1978) allows for differing numbers of explanatory variables in the models being compared and is calculated as

$$\text{BIC} = -2\ln L^{*} + K\ln N \tag{2}$$

where L^{*} is the maximised likelihood from the model, N is the number of observations and K the number of parameters estimated.

The deviance-based R_D^2 for a model is

$$R_D^2 = 1 - \left[\ln L^S - \ln L^*\right] / \left[\ln L^S - \ln L^0\right] = \left[\ln L^* - \ln L^0\right] / \left[\ln L^S - \ln L^0\right] \tag{3}$$

where L^0 is the likelihood from a model with only a constant, L^{*} is the maximised likelihood for the estimated model, and L^S is the likelihood from a saturated model with as many parameters as observations. R_D^2 can be interpreted as the fraction of empirical uncertainty in total patient cost which has been explained by the model (Cameron and Windmeijer, 1997). It is equal to the usual R^2 in an OLS model (ie GLM with linear link and normal error distribution). Like the BIC its value depends on the assumed error distribution.

The Efron R^2 (Efron, 1978) is the squared correlation coefficient between the estimated cost from a model and actual cost and is readily obtained from an OLS regression of estimated on actual costs. It is comparable across different underlying models for estimating cost and for OLS regression models it is equal to the model R^2 and to R_D^2. We also computed the mean absolute error over all observations (MAE), which is the average absolute difference in £'s between observed and estimated cost. Like the Efron R^2 it does not depend on the assumed error distribution and so can be used to compare models with the same set of explanatories but different error distributions.

3. Results

3.1 Summary statistics

There were 86,100 individuals alive and registered at the beginning of the 2007/8 resource use year. Table 3 gives the frequency distribution by age and gender alongside the mean, standard deviation and median number of consultations, prescriptions, and tests per individual for 2007/8. Also shown is the percentage of individuals who had at least one consultation, prescription, or test. At most ages women have higher rates of resource use but resource use increased more rapidly with age for men and was higher for men aged 80-89. The mean number of consultations and prescriptions is similar to those in other UK datasets for the same period (Hippisley-Cox and Vinogradova, 2009; NHS Information Centre 2008b).

Table 3. Consultations, prescriptions and tests 2007/8: mean, (standard deviation), [median].

	Age category	N	Number of consultations	% at least one consultation	Number of prescriptions	% at least one prescription	Number of tests	% at least one test
Male	20-29	6021	1.7 (3.0) [1]	54	2.6 (10.0) [0]	44	0.7 (2.2) [0]	19
	30-39	7204	1.9 (3.5) [1]	54	4.1 (13.4) [0]	47	1.0 (2.9) [0]	22
	40-49	8902	2.6 (4.4) [1]	60	6.6 (18.3) [1]	53	1.8 (4.1) [0]	30
	50-59	7486	3.7 (5.4) [2]	69	13.4 (27.1) [2]	64	3.2 (5.7) [0]	44
	60-69	6481	5.7 (6.9) [4]	83	26.5 (37.7) [15]	81	5.7 (7.9) [3]	64
	70-79	4112	8.0 (7.8) [6]	92	42.9 (45.2) [33]	92	8.0 (8.9) [6]	77
	80-89	1878	9.9 (9.7) [7]	93	54.4 (55.9) [44]	94	9.2 (10.2) [7]	79
	90+	253	7.3 (6.6) [6]	87	46.1 (52.0) [32]	88	6.0 (7.1) [4]	68
Female	20-29	5551	4.4 (4.8) [3]	84	5.8 (9.8) [3]	81	2.6 (4.6) [0]	47
	30-39	6930	4.6 (5.5) [3]	81	7.9 (20.4) [3]	76	2.9 (5.3) [1]	51
	40-49	8447	4.5 (5.7) [3]	81	10.0 (23.6) [3]	74	3.1 (5.7) [1]	51
	50-59	7525	5.0 (5.8) [3]	82	15.5 (28.2) [6]	78	4.2 (6.7) [1]	65
	60-69	6563	6.3 (6.7) [5]	88	28.0 (40.5) [16]	87	5.9 (8.2) [3]	73
	70-79	4954	8.3 (8.1) [6]	93	44.7 (50.8) [33]	93	7.8 (8.9) [6]	77
	80-89	3057	9.1 (8.7) [7]	93	59.2 (68.0) [44]	95	7.9 (9.0) [6]	77
	90+	736	7.9 (7.8) [6]	90	64.6 (75.5) [47]	94	6.6 (7.8) [4]	72
Overall		86100	4.8 (6.3) [3]	77	18.5 (36.3) [4]	72	3.9 (6.8) [1]	52

Notes. Consultations are face to face or telephone consultations. 18.3% of patients had no consultation, prescription or test.

Table 4 gives the total cost per patient, stratified by age and gender, and categorised by consultations, prescription drugs, and tests and investigations. Prescription drugs account for just over half (57%) of the total cost.

The mean total patient cost for men ranged from £90 at 20-29 years to £822 at 80-89 years. For women the mean total patient cost ranged from £176 at 20-29 years to £709 at 80-89 years. The share of prescription drugs in total cost increased with age and was higher for men than for women. The share of consultations in total cost decreased with age, whereas the share of tests was fairly stable across age bands.

Deprivation has a weaker unconditional association with cost than age. The mean total patient cost for the most deprived decile is 1.34 times as large as mean total cost for the least deprived (£394 (CI: £379 to £408) compared to £289 (CI: £278 to £300)).

Table 4. Patient primary care costs (£) 2007/8: mean, (standard deviation), [median].

	Age category	Consultation cost	Prescription drug cost	Test and investigations cost	Total patient cost
Male	20-29	41 (69) [19]	43 (341) [0]	6 (32) [0]	90 (367) [24]
	30-39	48 (84) [21]	57 (248) [0]	9 (40) [0]	113 (300) [26]
	40-49	61 (98) [26]	94 (477) [2]	13 (50) [0]	168 (524) [40]
	50-59	86 (117) [49]	174 (443) [12]	22 (67) [0]	282 (530) [92]
	60-69	130 (151) [93]	294 (503) [111]	34 (87) [8]	458 (622) [264]
	70-79	184 (177) [139]	449 (602) [268]	49 (102) [14]	681 (722) [484]
	80-89	237 (218) [178]	524 (643) [326]	60 (124) [17]	822 (783) [608]
	90+	213 (208) [160]	320 (381) [192]	34 (73) [10]	567 (510) [428]
Female	20-29	102 (110) [72]	57 (160) [17]	17 (50) [0]	176 (239) [106]
	30-39	106 (126) [72]	85 (330) [14]	20 (58) [0]	210 (407) [110]
	40-49	105 (135) [68]	111 (324) [16]	24 (68) [0]	240 (429) [111]
	50-59	119 (134) [80]	175 (418) [34]	40 (84) [11]	334 (515) [168]
	60-69	148 (150) [109]	288 (533) [109]	47 (89) [16]	483 (636) [293]
	70-79	195 (180) [150]	396 (541) [236]	48 (101) [14]	639 (662) [463]
	80-89	234 (217) [179]	431 (665) [287]	45 (95) [13]	709 (772) [555]
	90+	245 (237) [189]	395 (434) [245]	32 (70) [10]	673 (579) [528]
Overall		114 (147) [69]	189 (458) [24]	27 (74) [0]	330 (563) [134]

Notes. Consultation costs are the sum of the costs of face to face or telephone consultations plus the costs of administration for repeat prescriptions and other administration not requiring face to face or phone contact. 12.3% of individuals had no cost.

Tables 5 and 6 have distributions of patients for each of the 8 morbidity and multimorbidity measures. The proportion of patients in the lowest multimorbidity categories is much smaller for the three measures (EDC count, ACG categories, RUB categories) derived by applying the ACG software which uses a much finer classification of morbidity than the 17 diseases in the QOF and Charlson schemes. Although there are 68 mutually exclusive ACG categories, the eight largest categories accounted for over 70% of the sample patients.

The distribution of the EDC count has a larger range than the QOF chronic disease count and Charlson Index score because of the greater number of relatively minor diseases that the EDC count includes. According to the QOF chronic disease count 20% of patients were multimorbid (had a count of two or more) whereas 61% were multimorbid according to the EDC count. Women had slightly higher scores than men on the three count multimorbidity measures (QOF count, Charlson Index score, EDC count). There were significant positive Spearman rank correlations amongst these three count measures (for the top censored counts used in the models) – QOF and Charlson: 0.63; QOF and EDC: 0.72; Charlson and EDC: 0.59.

There are differences across the QOF, Charlson and EDC morbidity categorisations in proportions of patients with some of the diseases. For example, 14.3% of patients have asthma in the EDC scheme but only 6.5% in the QOF scheme. The QOF payments for asthma patients relate mainly to the monitoring of patients and therefore patients require a recent inhaler prescription to be classified as asthmatic, whereas EDC requires only a diagnosis of asthma and therefore may capture all patients who had childhood asthma. The QOF distinguishes between asthma and chronic obstructive pulmonary disease and so only records 2.1% of patients as having chronic obstructive pulmonary disease, whereas the Charlson scheme records 16.6% as having chronic pulmonary disease because its definition includes asthmatics.

Table 5. Distribution of patients by QOF disease categories, QOF disease count, Charlson diseases, Charlson Index score, selected EDC categories, EDC count

QOF indicators	%	QOF count	%	Charlson diseases	%	Charlson score	%	Selected EDC categories	%	EDC count	%
None	55.05	0	55.05	None	55.05	0	68.84	None	19.39	0	19.39
Asthma	6.5	1	24.81	AIDS	0.01	1	19.27	Low back pain	25.80	1	19.56
Atrial fibrillation	2.1	2	11.14	Cancer	3.83	2	6.62	Dermatitis and eczema	19.57	2	16.36
Cancer	1.6	3	5.20	Cerebrovascular disease	2.31	3	2.98	Hypertension	18.36	3	12.54
CHD	5.6	4	2.39	Chronic pulmonary disease	16.63	4	1.30	Anxiety, neuroses	16.50	4	9.20
CKD	3.8	5	0.92	Congestive heart disease	1.21	5	0.56	Depression	16.21	5	6.75
COPD	2.1	6+	0.49	Dementia	0.41	6	0.23	Asthma	14.26	6	4.90
Dementia	0.5			Diabetes	4.28	7+	0.19	Cervical pain syndromes	13.20	7	3.51
Depression	14.9			Diabetes with complications	0.91			Arthritis	11.17	8	2.41
Diabetes	5			Hemiplegia/paraplegia	0.18			Irritable bowel syndrome	6.78	9	1.74
Epilepsy	0.9			Metastatic tumour	0.12			Gastroesophageal syndrome	6.70	10	1.25
Heart failure	1.1			Mild liver disease	0.17			Acute myocardial infarction	5.84	11	0.82
Hypertension	18.2			Moderate or severe liver disease	0.04			Malignant neoplasms of the skin	2.53	12	0.57
Learning diff.	0.4			Myocardial infarction	1.70			Malignant neoplasms, breast	1.09	13	0.36
Mental health	0.9			Peptic ulcer	2.15			Emphysema, chronic bronchitis, COPD	2.38	14	0.27
Obesity	9.9			Peripheral vascular disease	1.43					15	0.15
Stroke	2.4			Renal disease	4.68					16	0.09
Hyperthyroidism	3.9			Rheumatological disease	2.05					17	0.07
										18+	0.08

Notes. CHD: coronary heart disease; CKD: chronic kidney disease; COPD: chronic obstructive pulmonary disease; AIDS: acquired immune deficiency syndrome.

Table 6. Distribution of patients by selected ACG categories and RUBs

Selected ACG categories	%	RUB	%
Non-Users	9.44	Non-user	9.44
No Diagnosis or Only Unclassified Diagnosis	20.05	Healthy user	36.36
Preventive/Administrative	6.47	Low morbidity	25.03
Acute Minor, Age 6+	9.63	Moderate	27.26
Chronic Medical: Stable	2.49	High	1.70
2-3 Other ADG Combinations, Age 35+	9.98	Very high	0.20
4-5 Other ADG Combinations, Age 45+, no Major ADGs	1.83		
4-5 Other ADG Combinations, Age 45+, 1 Major ADGs	2.28		
6-9 Other ADG Combinations, Age 35+, 0-1 Major ADGs	1.52		
10+ Other ADG Combinations, Age 18+, 2 Major ADGs	0.05		
6-9 Other ADG Combinations, Males, Age 18 to 34, 1 Major ADGs	0.01		

Note. ACGs: Adjusted Clinical Groups; ADGs: Aggregated Diagnosis Groups; used to construct ACGs; RUB: Resource Utilization Band constructed by aggregation of ACGs.

3.2 Comparison of GLM specifications

For the regression modelling we dropped 154 individuals with missing deprivation data to leave an estimation sample of 85,946.

Table 7. Goodness of fit of alternative specifications of model for total patient cost

	Log, Poission				Linear, Gaussian (OLS)		
Model specification	BIC	R_D^2	Efron R^2	MAE	BIC	Efron R^2 = R_D^2 = R^2	MAE
Age, gender (Model 1)	38446223	0.21	0.13	285	1320939	0.13	285
Age, gender, and deprivation (Model 2)	38041309	0.22	0.13	283	1320555	0.13	284
Age, gender, and practice (Model 3)	37725187	0.22	0.14	282	1322027	0.14	283
Age, gender, deprivation, and practice (Model 4)	37522639	0.23	0.14	281	1321883	0.14	282
(Model 4) + QOF disease indicators	29339460	0.40	0.25	244	1302791	0.31	234
(Model 4) + QOF chronic disease count	28523441	0.42	0.29	239	1305547	0.29	240
(Model 4) + Charlson indicators	32546370	0.33	0.22	259	1310235	0.25	255
(Model 4) + Charlson Index score	32274547	0.34	0.23	258	1311982	0.23	260
(Model 4) + EDC indicators	26255449	0.46	0.29	231	1295660	0.37	222
(Model 4) + EDC count	26196861	0.46	0.32	229	1302546	0.31	236
(Model 4) + ACG	28694630	0.41	0.27	242	1308944	0.27	248
(Model 4) + RUB	30367472	0.38	0.24	250	1312522	0.23	259

Notes. BIC: Bayesian Information Criterion. Smaller BIC indicates better fit and is comparable for different models with same error distribution. R_D^2 : deviance based R^2, which is comparable for different models with same error distribution. Efron R^2: correlation coefficient from OLS regression of estimated cost on actual cost. For OLS models the Efron R^2, the deviance based R^2, and the model R^2 are equal. MAE: mean absolute error

Table 7 has goodness of fit statistics (BIC, deviance-based R-squared values R_D^2 , Efron R^2, MAE) for GLM models with log link and Poisson errors and OLS models (linear link, normal distribution) with various sets of covariates. For the models without any morbidity or multimorbidity measures the log link Poisson models and OLS models have very similar performance in terms of MAE and Efron R^2 for any given set of covariates. The log link Poisson GLM has lower MAE and higher Efron R^2 than OLS for 5 of the 8 models with morbidity or multimorbidity measures. OLS does better when morbidity is measured by dummies for the Charlson diseases, QOF diseases and the EDC indicators.

3.3 Comparison of morbidity and multimorbidity measures

Table 7 shows that the inclusion of any measure of morbidity or multimorbidity boosts the performance of the regression models considerably. For example, with age and gender groups, deprivation deciles, and practice effects the R_D^2 for the log link Poisson GLM specification is 0.23. Adding the set of Charlson indicators, the worst performing of the eight morbidity and multimorbidity measures, to the model increases the R_D^2 to 0.34. Similar increases in performance are seen with the OLS model.

For any given estimation method, the rankings of the eight morbidity and multimorbidity measures by the BIC, R_D^2, Efron R^2 and MAE criteria are very similar. In the log link Poisson GLM specification, the EDC count has the best performance on all goodness of fit statistics (provided we use the third decimal place for the R_D^2) closely followed by the set of 114 EDC indicators. These two EDC based measures are noticeably better than the QOF count, 68 ACG indicators, 17 QOF disease categories, and the 6 RUBs, which in turn are markedly better than the Charlson Index score and 17 Charlson disease categories.

Under OLS estimation, the EDC indicators have the best performance followed by the QOF indicators and then the EDC count. The three sets of morbidity category dummies (EDC, QOF, Charlson) performed better than the corresponding count multimorbidity measures.

Tables 8 and 9 give cost ratios for the morbidity and multimorbidity measures from the log link Poisson models. (OLS models yield similar results.) The log link GLM model estimates $\ln Ec = x'\beta$ or $Ec = \exp(x'\beta)$. Since all explanatory variables are categorical, the cost or risk ratio for a variable x_k is the ratio of expected cost when x_k = 1 to the expected cost when x_k = 0 and is $\exp(\beta_k * 1 + \sum_{j \neq k} \beta_j x_j) / \exp(\beta_k * 0 + \sum_{j \neq k} \beta_j x_j) = \exp \beta_k$.

For the three count multimorbidity measures (QOF, Charlson, and EDC) and the two mutually exclusive multimorbidity categorisations (ACGs, RUBs) the cost ratio is relative to a zero count or to non-users. For the three sets of non-mutually exclusive morbidity dummies (QOF diseases, Charlson diseases, EDCs) the cost ratio is the cost of the disease relative to not having that disease, rather to not having any disease. For example, a patient with an EDC count of 4 has an expected cost which is 6.13 times as large as a patient with a zero EDC count. A patient diagnosed as having cancer under the QOF scheme has an expected cost 1.74 times as large as a patient without cancer.

Amongst the QOF morbidity categories epilepsy is the chronic disease with the biggest relative effect (2.32 compared with no epilepsy), though only 0.9% of the sample have the condition. The most common QOF condition is hypertension (18.2% of the sample) with a cost ratio of 1.42. All the QOF disease cost ratios are significantly greater than 1 but their range is limited (1.09 to 2.32). The 17 Charlson diseases also have a similarly limited range of cost ratios (0.98 to 2.43).[5] Of the 114 EDCs cost ratios, 82 are significant. They also have a limited range: all but 8 are less than 2.0 and the largest (Transplant status) has a cost ratio of 3.6 but only 0.05% of patients are in this category).

The baseline ACG category is non-users and all other categories have a cost ratio in excess of 1. The results for selected ACGs in Table 9 suggest that in general patients with more morbidities have higher costs. The fact that cost ratios increase with RUB levels also suggests that multimorbid patients are more costly as patients in higher RUBs are more likely to be multimorbid.

[5] Patients with AIDS do not increase general practice costs significantly because in the UK their care and the prescribing of AIDS drugs is generally managed in secondary care.

Table 8. Cost ratios for QOF disease categories, QOF disease count, Charlson disease categories, Charlson Index score, EDC count

QOF disease categories*			QOF disease count**			Charlson disease categories*			Charlson Index score**			EDC count**		
Disease	Cost ratio	95% CI	Count	Cost ratio	95% CI	Disease	Cost ratio	95% CI	Score	Cost ratio	95% CI	Count	Cost ratio	95% CI
Asthma	1.80	(1.74,1.86)	0	1.00	-	AIDS	0.98	(0.51,1.88)	0	1.00	-	0	1.00	-
Atrial fibrillation	1.11	(1.06,1.17)	1	2.41	(2.34,2.48)	Cancer	1.54	(1.48,1.60)	1	2.05	(2.00,2.10)	1	2.11	(2.01,2.21)
Cancer	1.74	(1.65,1.84)	2	3.80	(3.67,3.93)	Cerebrovascular dis.	1.32	(1.25,1.39)	2	2.59	(2.50,2.68)	2	3.34	(3.20,3.48)
CHD	1.45	(1.41,1.50)	3	4.86	(4.68,5.04)	Chronic pulmonary dis.	1.64	(1.61,1.68)	3	3.16	(3.02,3.31)	3	4.64	(4.44,4.85)
CKD	1.15	(1.10,1.20)	4	5.81	(5.55,6.07)	Congestive heart dis.	1.24	(1.17,1.32)	4	3.60	(3.42,3.80)	4	6.13	(5.85,6.42)
COPD	1.64	(1.56,1.72)	5	7.27	(6.85,7.73)	Dementia	1.29	(1.16,1.44)	5	3.86	(3.58,4.16)	5	7.73	(7.37,8.11)
Dementia	1.27	(1.15,1.41)	6+	7.56	(7.04,8.11)	Diabetes	1.98	(1.92,2.05)	6	4.13	(3.70,4.62)	6	9.08	(8.65,9.53)
Depression	1.53	(1.49,1.57)				Diabetes with comp.	2.43	(2.27,2.61)	7+	4.54	(3.98,5.18)	7	11.01	(10.40,11.66)
Diabetes	1.73	(1.67,1.79)				Hemiplegia/paraplegia	1.86	(1.46,2.36)				8	12.16	(11.49,12.86)
Epilepsy	2.32	(2.13,2.53)				Metastatic tumour	1.36	(1.07,1.71)				9	13.74	(12.92,14.61)
Heart failure	1.09	(1.02,1.17)				Mild liver disease	1.39	(1.16,1.67)				10	15.02	(14.13,15.97)
Hypertension	1.42	(1.38,1.46)				Mod/sev liver disease	2.11	(1.32,3.38)				11	16.96	(15.81,18.20)
Learning	1.70	(1.41,2.05)				Myocardial infarction	1.42	(1.35,1.50)				12	17.18	(15.89,18.58)
Mental health	2.06	(1.88,2.26)				Peptic ulcer	1.31	(1.25,1.38)				13	19.52	(17.87,21.32)
Obesity	1.30	(1.26,1.34)				Periph vascular disease	1.26	(1.20,1.33)				14	19.05	(17.17,21.13)
Stroke	1.22	(1.16,1.27)				Renal disease	1.34	(1.28,1.40)				15	21.04	(18.65,23.73)
Hyperthyroidism	1.21	(1.16,1.26)				Rheumatological dis.	1.46	(1.39,1.53)				16	24.69	(21.43,28.46)
												17	24.00	(19.98,28.83)
												18+	27.60	(24.08,31.63)

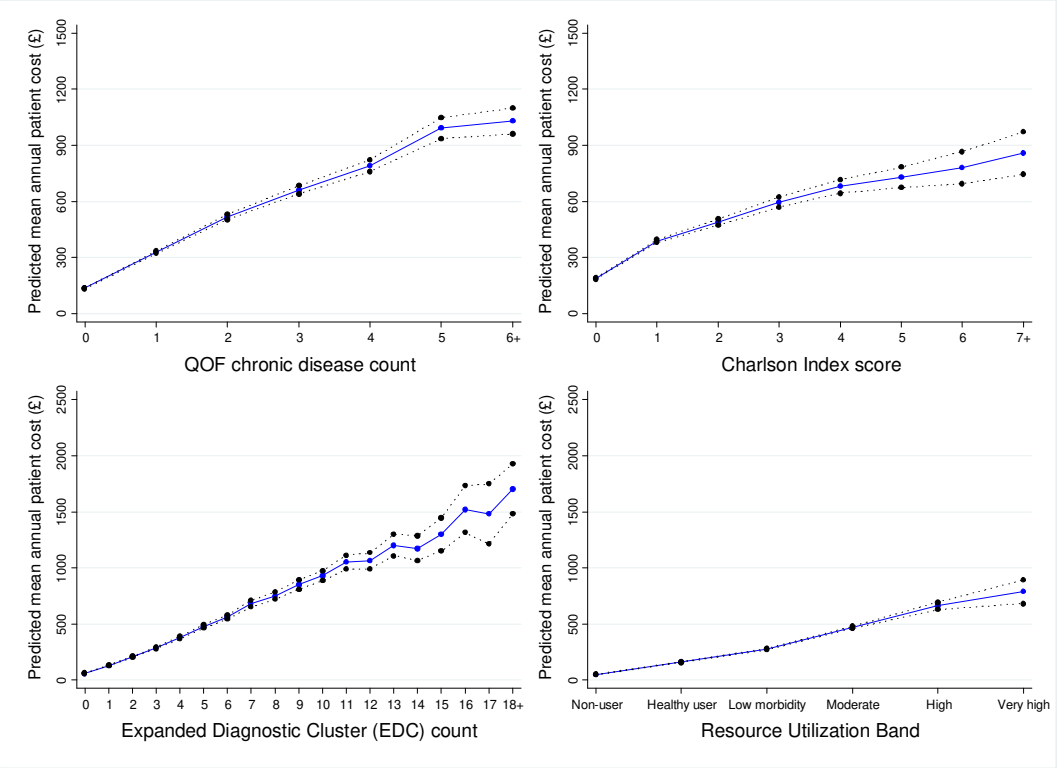
Notes. *Cost ratios are the estimated costs for a patient with the relevant disease divided by the estimated cost for a patient without that disease. **Cost ratios are the estimated costs for a patient with the relevant count divided by the estimated cost for a patient with no disease (zero count). Estimates from GLM log Poisson model with age/gender, deprivation and practice effects. EDC: Expanded Diagnosis Cluster. CHD: coronary heart disease; CKD: chronic kidney disease; COPD: chronic obstructive pulmonary disease; AIDS: acquired immune deficiency syndrome.

Table 9. Cost ratios for selected EDC categories and for selected ACG categories and Resource Utilization Bands

Selected EDC categories*			Selected ACG categories**			Resource Utilization Bands**		
Category	Cost ratio	95% CI	Category	Cost ratio	95% CI	Band	Cost ratio	95% CI
Low back pain	1.10	(1.08,1.13)	Non-Users	1.00	-	Non-user	1.00	-
Dermatitis and eczema	1.12	(1.09,1.15)	No Diagnosis or Only Unclassified Diagnosis	2.58	(2.40,2.77)	Healthy user	3.28	(3.06,3.51)
Hypertension	1.38	(1.35,1.42)	Preventive/Administrative	4.04	(3.72,4.39)	Low morbid	5.60	(5.22,6.01)
Anxiety, neuroses	1.21	(1.18,1.24)	Acute Minor, Age 6+	4.25	(3.95,4.57)	Moderate	9.54	(8.90,10.23)
Depression	1.29	(1.26,1.33)	Chronic Medical: Stable	7.06	(6.49,7.68)	High	13.44	(12.36,14.62)
Asthma	1.50	(1.46,1.54)	2-3 Other ADG Combinations, Age 35+	8.93	(8.31,9.60)	Very high	16.02	(13.81,18.58)
Cervical pain syndromes	1.10	(1.07,1.12)	4-5 Other ADG Combinations, Age 45+, no Major ADGs	10.90	(10.05,11.82)			
Arthritis	1.13	(1.09,1.16)	4-5 Other ADG Combinations, Age 45+, 1 Major ADG	12.50	(11.56,13.51)			
Irritable bowel syndrome	1.19	(1.14,1.23)	6-9 Other ADG Combinations, Age 35+, 0-1 Major ADG	16.32	(15.03,17.71)			
Gastroesophageal reflux	1.27	(1.23,1.31)	10+ Other ADG Combinations, Age 18+, 2 Major ADGs	21.75	(16.90,27.99)			
Acute myocardial infarction	1.26	(1.22,1.30)	6-9 Other ADG Combinations, Male, Age 18 to 34, 1 Major ADG	46.78	(20.99,104.22)			
Malignant neoplasm of the skin	1.07	(1.02,1.12)						
Malignant neoplasms, breast	1.56	(1.46,1.68)						
Emphysema, chron bronchitis, COPD	1.30	(1.24,1.36)						

Notes. * Cost ratios for EDCs are the estimated costs for a patient with the relevant disease divided by the estimated cost for a patient without that disease. ** Cost ratios for mutually exclusive ACG categories and RUBs are the estimated costs for a patient in the ACG or RUB with the relevant count divided by the estimated cost for a patient with no use. Estimates from GLM log Poisson model with age, gender, deprivation and practice effects. EDC: Expanded Diagnosis Cluster; ADGs: Aggregated Diagnosis Groups.

The cost ratios for the three count measures increase with the counts (except for EDC counts 14 and 17, which are slightly smaller, respectively, than EDC counts 13 and 16) implying, like the ACG and RUB results, that patients with more diseases have higher costs. We re-estimated the QOF count and EDC count log Poisson models with the counts and their squares rather than with categories for the counts. The cost ratio for the squared QOF count was 0.921 (CI: 0.918-0.925) and for the squared EDC count was 0.985 (CI: 0.985-0.986). This means that the proportionate effect of an additional disease on the cost ratio declines with the number of diseases.[6] This is confirmed by inspection of the cost ratios on the counts. For the EDC count, having 2 diseases increases costs by 3.34/2.11 = 1.58 times compared with having one disease, having 3 diseases increases costs by 1.39 times compared to having two diseases, and so on until having 6 diseases increases costs by 1.17 times compared to having 5 diseases. Since the cost ratios for the counts increase less than proportionately with the counts, the *level* of cost increases less than exponentially with the number of diseases.[7] Figure 1 plots predicted costs against the EDC count, QOF count, Charlson score and RUB level and suggests a roughly linear, rather than exponential, effect of increasing multimorbidity on the level of costs.



Note: Estimated costs from log link, Poisson model with age/gender bands, deprivation and practice effects.

Figure 1. Estimated mean annual patient cost (with 95% confidence interval) for multimorbidity counts and Resource Utilization Bands

3.4 Effects of age, gender, deprivation and practice

Table 7 suggests that including fixed practice effects in the models did not contribute greatly to model performance. Adding practice effects to models with only age and gender or only age, gender and deprivation increases model R_D^2 by at most 0.01. Dropping all the practice fixed effects from the full EDC count model reduces the R_D^2 by 0.02.

[6] With n equal to the count of diseases, the estimated model is $\ln(\mathrm{E}c) = \beta_0 + \beta_1 n + \beta_2 n^2$ so that the cost (risk) ratio on the n^2 term is $\exp(\beta_2)$, and $d\ln(\mathrm{E}c)/dn = \beta_2 < 0$ if $\exp(\beta_2) < 1$.

[7] Let the cost with n diseases be $C(n)$. Then if the cost ratios increase linearly with the count $C(n)/C(0) = \exp(\beta n)$ so that $C(n)/C(n-1) = \exp(\beta)$ and $C(n) = C(0)\exp(\beta n)$.

Table 10 has the cost ratios for age/gender categories and deprivation deciles from the log link Poisson model with EDC counts. (Results with other morbidity or multimorbidity measures are similar.) The effects of age and gender are qualitatively similar to those in Table 5 but with smaller differences between men and women at any given age and with much less of an effect of age on cost. This suggests that older patients are more expensive mainly because they are sicker.

Table 10. Conditional effects of age, gender and deprivation.

	Age category	Cost ratio	95% CI	Deprivation decile	Cost ratio	95% CI
Male	20-29	1.00	-	(least deprived) 1	1.00	-
	30-39	1.17	(1.05,1.32)	2	1.04	(0.99,1.09)
	40-49	1.48	(1.32,1.67)	3	1.07	(1.02,1.12)
	50-59	1.90	(1.71,2.12)	4	1.05	(1.00,1.10)
	60-69	2.22	(2.00,2.48)	5	1.05	(1.00,1.11)
	70-79	2.39	(2.14,2.67)	6	1.07	(1.01,1.13)
	80-89	2.38	(2.12,2.67)	7	1.13	(1.07,1.19)
	90+	1.81	(1.56,2.10)	8	1.08	(1.02,1.14)
Female	20-29	1.54	(1.39,1.72)	9	1.14	(1.08,1.20)
	30-39	1.57	(1.40,1.75)	(most deprived) 10	1.22	(1.15,1.29)
	40-49	1.60	(1.44,1.79)			
	50-59	1.85	(1.66,2.06)			
	60-69	2.11	(1.89,2.35)			
	70-79	2.17	(1.94,2.42)			
	80-89	2.06	(1.84,2.30)			
	90+	2.01	(1.77,2.27)			

Notes. Cost ratios for age bands are the estimated costs for a patient in the relevant age category divided by the estimated cost for a male patient in the lowest age band. Cost ratios for deprivation deciles are estimated costs for a patient in the relevant decile divided by the estimated cost for a patient in the lowest decile. Estimates are from GLM log Poisson model with EDC count, age/gender, deprivation and practice effects.

Table 7 shows that including deprivation deciles in the regression models in addition to age and gender improves their fit modestly: the R_D^2 increases by 0.01, which is about the same as the improvement due to adding practice effects. Once morbidity measures are included in the model the contribution of deprivation is very small: dropping the deprivation deciles from the full log link Poisson model with EDC counts does not change the R_D^2 (to 2 decimal places) and increases the BIC slightly by 45711. By contrast dropping all the age/gender bands reduces the R_D^2 slightly to 0.45 and increases the BIC by 851817.

The gradient of cost with respect to deprivation is flatter than with respect to age and after inclusion of morbidity measures the cost ratio between the 10th and 1st deciles is reduced to 1.22 (CI: 1.15, 1.29).

3.5 Sensitivity analyses

There were some patients with extremely high costs: the highest four costs ranged from £15,128 to £27,810 compared to the median cost of £134. The use of log link in the GLM models reduced this discrepancy considerably (the log of the highest cost was 10.233 compared to the log of the median of 4.895). Results were not sensitive to dropping the patients above the 99th cost centile (£2471 or 7.812 in logs). The R_D^2 of the log Poisson model with EDC counts increased to 0.48 compared 0.46 for the model with a full set of observations. There was a slight reduction in the cost ratios for higher EDC counts.

We also estimated GLM models with a log link and gamma error distribution (error variance proportional to the square of the mean). In all cases, the log link gamma models had higher MAE and

smaller Efron R^2 than the log link Poisson models. The Efron R^2 was much smaller for the models with EDC indicators (0.03 vs 0.29) and QOF indicators (0.11 vs 0.25).

Our results were not sensitive to the method for calculating prescribing costs which are the largest part of the costs. Using national average amounts per prescription rather than recorded amounts to calculate total primary care costs reduced model performance somewhat but estimated costs ratios were similar and the model rankings were unchanged.

4. Conclusion

4.1 Limitations

The GPRD practices included in our sample were all of "research standard". One implication is that the recording of patient diagnoses may be less thorough at other UK general practices, particularly for diagnoses which are not included in the QOF. Measures of morbidity or multimorbidity constructed for practices with lower quality data may therefore be underestimates. Basing a capitation scheme on results from high data quality practices could lead to lower capitation payments for practices with less good quality data. However, recording of QOF morbidity variables is less likely to show this kind of variation because they are used as part of a national pay for performance scheme.

It is possible that part of the observed association between morbidity and cost is because patients who have, for unobserved reasons, a propensity to consult more frequently are more likely to acquire more diagnoses and to be given more prescriptions. This effect is likely to be most marked for measures which include a large number of diagnostic categories and may partly explain why EDCs have the strongest association with expenditure.

We have estimated drug costs using data on prescriptions issued by the practice. This will overestimate costs because not all prescriptions issued to patients result in the dispensing of medicines. Patients must pay for dispensed prescriptions but because of exemptions from payment on grounds of age, income, and health, over 90% of NHS prescriptions are dispensed without charge to the patient. Thus there is little financial incentive for patients not to have prescriptions dispensed. One study estimated that 5.2% of prescriptions written in a general practice were not dispensed (Beardon et al., 1993). It therefore seems unlikely that there is any large upward bias in our estimates of drug costs.

We included cost data for patients who were alive and registered with our sample practices on 1st April 2007 at the start of the financial year 2007/8. We made no adjustment for patients who deregistered from the practice list during the resource use year. Thus if patients moved to another practice during the year their total primary care costs including those incurred in other practices would be under recorded in our data. If the aim is to estimate future costs which will be incurred by the practice in which the patient is initially registered this does not present a problem. On average, setting capitation fees based on our results would ensure that practices would be paid for costs actually incurred by patients on their list at the start of the year. A retrospective adjustment could be made at the end of the financial year for costs attributable to newly registered patients. We also make no adjustment for patients who die. Such patients may have higher costs but to the extent that the morbidity measures are correlated with mortality in the coming year the estimated effects of different types of morbidity will include some of the differentially higher costs of patients at higher risk of death.

We had limited information on patient socio-economic circumstances because the data suppliers were concerned to ensure that there was no risk of identifying patients. Thus we had only a single measure of deprivation (IMD) which was attributed to patients by their small area of residence. The IMD has been widely used in other studies and is intended to reflect a number of dimensions of deprivation (including income, education, and housing). But nevertheless it would have been useful to have individual level data on these patient characteristics and also on the ethnicity of patients which is not included in the IMD measure.

4.2 Discussion

4.2.1 Practice effects

The practice dummy variables pick up the effects of characteristics of practices such as the GP to patient ratio, idiosyncratic practice treatment styles, and differences in the practice means of both observed and unobserved patient characteristics. The small impact of practice dummies implies that the practice characteristics and mean patient characteristics are either offsetting, or that there is little cross practice variation in these variables, or that they have little effect. An alternative possibility is that 'practice style' manifests itself in differences in case finding and thus the number and type of diagnoses that are recorded for patients. Hence models of cost which include measures of morbidity will not find any difference across practices. But we found that practice effects are not important in explaining costs even when no account is taken of morbidity, suggesting that case finding does not explain the small impact of practice effects on patient costs.

4.2.2 Age and gender

Age is one of the factors affecting capitation payments to general practices and our results show that age does have a marked effect on expenditure, with a roughly five fold variation in expenditure between the lowest age band (20-29) and those aged 80-89 if no allowance is made for morbidity or multimorbidity. However, once morbidity or multimorbidity is allowed for the variation in cost with age is much less marked: those aged 80-89 have costs which are only about twice as large as those aged 20-29, rather than five times as large. Thus a major reason that the elderly have higher costs is that they have greater morbidity.

4.2.3 Deprivation

Deprivation is often suggested as having a strong relationship with health status and general practice workload and until 2004 the capitation payments received by general practices depended on the deprivation level of the areas of residence of their patients. The results here suggest there is positive relationship between deprivation and primary care expenditure. The positive deprivation gradient is flatter than the age gradient and is reduced by allowing for patient morbidity, but it is still statistically and economically significant. Previous survey based studies have reported mixed results about the relationship between deprivation and general practice workload. Bago D'Uva (2005), using the British Household Panel Survey, found that patients with higher income had more consultations. Morris et al (2005) used data from the Health Survey for England which has more health measures and found a negative but insignificant association of income and higher social class with consultations. Dixon et al (2007) reviewed the evidence and suggest that the NHS is pro-poor in terms of the distribution of GP consultations once allowance is made for morbidity, age and gender. Our results, based on much more detailed health data than earlier studies support this broad conclusion. We cannot however rule out the possibility that poorer health status in more deprived groups may be relatively under-recorded.

4.2.4 Estimation methods

With any of the morbidity or multimorbidity measures, both the log Poisson GLM and OLS estimation methods had good explanatory power, especially for individual level cross-section regressions. OLS performed a little worse with five of the eight morbidity and multimorbidity measures. This was unexpected because in large samples of patients it has been found that OLS performs as well or better than GLM models in explaining hospital costs (Gravelle et al., 2011; Van de Ven and Ellis, 2000) and our sample of primary care patients had a smaller proportion of patients with no expenditure than in these studies. However, the reduction in explanatory power was not large and OLS estimation with the 114 EDC categories had the lowest MAE and highest Efron R^2 over all sets of explanatories and estimation methods. The estimation method was less important than the choice of morbidity or multimorbidity measure.

4.2.5 Morbidity and multimorbidity measures

Having finer categories of morbidity unsurprisingly improved the ability to predict patient costs. Thus measures using the 114 EDC categories were better than those using the 17 category QOF scheme, whether the categories were used to measure morbidity as a set of dummy variables or used to construct counts of disease categories to measure multimorbidity. The QOF based measures of morbidity and multimorbidity performed considerably better than the Charlson based measures which had the same number of morbidity categories. The poor performance of the widely used Charlson Index score and of the Charlson disease dummy variables may be a reflection of the fact that the

Charlson scheme was originally intended to predict mortality rather than the cost of general practice activities. The two QOF based measures had about the same predictive power as the 68 mutually exclusive ACG categories derived using purpose built case-mix software. This may be because the 17 QOF categories were selected for a primary care pay for performance scheme targeted at care for chronic patients who are the main business of general practices. The ACG categories included non-chronic diagnoses which were grouped in part by their anticipated effect on all patient costs, including hospital costs.

Measuring morbidity as a set of disease dummy variables and measuring multimorbidity as a count of diseases are both simple ways of using the rich information available in electronic patient health records. Their simplicity requires possibly incorrect assumptions about the relationship between morbidity and cost. With linear models of cost, using morbidity categories allows the effect of costs to vary across diseases but assumes that there are no interactions between diseases in their effect on cost.[8] If there are non-additive effects of multimorbidity then these will be partially picked up in the estimated coefficients on the morbidity category dummies since they estimate the average cost of patients with only that disease plus some of the effect on costs of patients with that and other diseases.[9] On the other hand, using a count of diseases with dummies for each count value allows flexibly for non-additivity but assumes that all diseases are essentially identical in their effect on cost. For example, the coefficient on the first count is the average cost of all patients with a single disease, the coefficient on the second count is the average cost of all patients with two diseases, and so on. Thus the choice between using disease dummies to measure morbidity or a count of the same diseases to measure multimorbidity reflects a trade-off between two possible sources of bias in estimated cost function coefficients. The fact that the Charlson disease, QOF disease, and EDC indicator dummies have better MAE and Efron R^2 in the linear OLS models than the corresponding count multimorbidity measures suggests that it is more important to allow for different diseases to have different cost impacts than to allow for possible non-additivity. However, the difference between the multimorbidity counts and the corresponding morbidity categories in terms of ability to explain patient costs is not great. This may be due to the fact that costs increased roughly linearly with the number of diseases, and there was relatively little variation in costs of particular diseases.

4.2.6 Practice budget setting

We included measures based on the 17 chronic QOF conditions because such information is readily available and reliably coded in all UK general practices. The measures performed nearly as well as measures requiring considerably more information. As the QOF includes major chronic diseases it is likely that similar measures could be readily constructed using routine data in other countries. Moreover, it would currently be feasible to use the results from the linear OLS model with a simple set of 17 QOF disease dummies to allocate budgets to practices to cover their primary care costs. With an additive linear model, practice budgets can be set without information on the QOF diseases of all individual patients. For a practice level allocation the coefficients on the QOF diseases can be applied to the number of a practice's patients with each of the 17 QOF diseases. Data on the numbers with each disease in each practice is already produced by the QOF process. Similarly applying the deprivation and age/gender coefficients to numbers on practice lists in each deprivation and age/gender category yields an estimate of total practice primary care needs.[10] Such a procedure seems likely to improve on the current methods of allocating funds to practices to cover their costs of providing primary care.

---

[8] The discussion is in terms of linear cost models but with suitable reinterpretation it applies to log models. For example additivity in the log model means that the proportional effects on cost (rather than absolute effects) of having a disease do not depend on whether the patient has other diseases.

[9] Suppose there are two diseases with prevalence π_i (unconditional probability), with $\pi_1 + \pi_2 < 1$ so that there are some non-morbid patients, and that the patient cost is $c = \alpha_0 + \beta_1 D_1 + \beta_2 D_2 + \beta_{12} D_1 D_2 + \varepsilon$ where $D_i = 1$ if and only if the individual has disease i. If the diseases are statistically independent and we estimate a model $c = a_0 + b_1 D_1 + b_2 D_2 + e$, then plim $\hat{b}_i = \beta_i + \pi_j \beta_{12}$, where π_j is the probability of disease j. Even in this simple case we also get an inconsistent estimate of the cost of multimorbid patient with both diseases by adding the separate coefficients: plim $(\hat{b}_1 + \hat{b}_2) = \beta_1 + \beta_2 + (\pi_1 + \pi_2)\,\beta_{12} \neq \beta_1 + \beta_2 + \beta_{12}$.

[10] Assumes that variations in practice effects reflect differences in practice behaviour rather than patient need and should not therefore determine the budget, the expenditure which is associated with needs variables in practice p is

$$n_p \hat{\beta}_0 + \sum_{a'} \hat{\beta}_{a'} n_{pa'} + \sum_{d'} \hat{\beta}_{d'} n_{pd'} + \sum_{m'} \hat{\beta}_{m'} n_{pm'}$$

where n_p is the total list, and n_{pa}, n_{pd}, n_{pm} are the number of patients in the practice in age/gender band a, deprivation decile d, and with QOF disease m.

4.2.7 Simpler is better?

Our study is the first to compare across a range of estimation methods and measures of morbidity and multimorbidity with respect to their ability to explain primary health care resource use in the UK. The best performing multimorbidity measures were simple counts of the number of chronic conditions patients suffered from or simple sets of disease dummies. In future work it would be useful to test whether other reasonably straightforward specifications would improve the ability to predict patient costs. One interesting, and simple, possibility is to add indicators for particular combinations of diseases to models.

References

Bago d'Uva, T. 2005. Latent class models for use of primary care: evidence from a British panel. *Health Economics*. 14, 873-892.

Bago d'Uva, T., Jones, A.M., van Doorslaer, E., 2009. Measurement of horizontal inequity in health care utilisation using European panel data, *Journal of Health Economics*, 28, 280-289.

Beardon P, McGilchrist M, McKendrick A, McDevitt D, MacDonald T. 1993. Primary non-compliance with prescribed medication in primary care. *British Medical Journal*. 307: 846-848.

British Medical Association and Royal Pharmaceutical Society. 2010. http://bnf.org. Accessed 12 May 2011.

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.

Cameron, A.C., Windmeijer, F.A.G. 1997. An R-squared measure of goodness of fit for some common nonlinear regression models. *Journal of Econometrics*, 77, (2) 329-342.

Charlson, M.E., Pompei, P., Ales, K.L., & Mackenzie, C.R. 1987. A new method of classifying prognostic co-morbidity in longitudinal-studies - development and validation. *Journal of Chronic Diseases*, 40, (5) 373-383.

Curtis, L. 2008. *Unit costs of health and social care 2007* PSSRU, University of Kent.
Department for Communities and Local Government. *The English Indices of Deprivation* 2007. 28-3-2008.

Department of Health. 2009. "NHS Reference Costs 2007/08". [online] http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsLibrary/index.htm. Accessed 10 May 2011.

Department of Health. 2010. *Equity and excellence*: Liberating the NHS. Cm 7881.

Department of Health. 2011. Resource Allocation: Weighted Capitation Formula. 7th Edition. http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/documents/digitalasset/dh_124947.pdf. Accessed 5 September 2011.

Deyo, R.A., Cherkin, D.C., Ciol, M.A. 1992. Adapting a clinical comorbidity index for use with ICD-9-CM administrative databases. *Journal of Clinical Epidemiology*, 45, (6) 613-619.

Dixon, A., Le Grand, J., Henderson, J., Murray, R., Poteliakhoff, E. 2007. Is the British National Health Service equitable? The evidence on socioeconomic differences in utilization. *Journal of Health Services Research and Policy*. 12, 2, 104-109.

Efron, B. 1978. Regression and ANOVA with zero-one data: measures of residual variation. *Journal of the American Statistical Association*. 73, 113-121.

Formula Review Group. 2007. *Review of the General Medical Services global sum formula*. British Medical Association and NHS Employers. http://www.nhsemployers.org /SiteCollectionDocuments/frg_report_final_cd_090207.pdf Accessed 5 September 2011.

Gravelle, H., Dusheiko, M., Martin, S., Rice, N., Smith, P., Dixon, J. 2011. Modelling individual patient hospital expenditure for general practice budgets. *CHE Research Paper* 72. Centre for Health Economics, University of York.

Gravelle, H., Morris, S., Sutton, M. 2006. Economic studies of equity in the consumption of health care. In Jones, A.M. (ed.), *Elgar Companion to Health Economics*, Edward Elgar,193-204.

Halling, A., Fridh, G., & Ovhed, I. 2006. Validating the Johns Hopkins ACG case-mix system of the elderly in Swedish primary health care. *Bmc Public Health*, 6., available from: ISI:000239803600001

Hippisley-Cox J, Pringle M. 2007. *Comorbidity of diseases in the new General Medical Services Contract for General Practitioners: analysis of QRESEARCH data*. http://www.qresearch.org/Public_Documents/DataValidation/Co-morbidity%20of%20diseases%20in%20the%20new%20GMS%20contract%20for%20GPs.pdf

Hippisley-Cox, J. & Vinogradova, Y. 2009. Trends in consultation rates in general practice 1995 to 2008: Analysis of the QResearch database. Final Report to the NHS Information Centre and Department of Health. National Health Service (NHS) Information Centre and QResearch.

Huntley, A., Johnson, R., Purdy, S., Valderas, J., Salisbury, C. 2012. Measures of multimorbidity and morbidity burden for use in primary care settings: a systematic review and guide. *Annals of Family Medicine*. (In press)

Johns Hopkins Bloomberg School of Public Health. 2008. The Johns Hopkins ACG® Case-mix System Version 8.2. Baltimore,

Khan, N.F., Perera, R., Harper, S., & Rose, P.W. 2010. Adaptation and validation of the Charlson Index for Read/OXMIS coded databases. *BMC Family Practice*, 11, available from: ISI:000274624100001

Lawrenson, R., Williams, T., & Farmer, R. 1999. Clinical information for research; the use of general practice databases. *Journal of Public Health Medicine*, 21, (3) 299-304.

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

Manning, W.G. 2006. Dealing with skewed data on costs and expenditures. In Jones, A.M. (ed.), *Elgar Companion to Health Economics*, Edward Elgar, 439-446.

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

Manning, W.G., Basu, A., Mullahy, J. 2005. Generalised modelling approaches to risk adjustment of skewed outcomes data. *Journal of Health Economics*. 24. 465-488.

Morris, S., Sutton, M., Gravelle, H. 2005. Inequity and inequality in the use of health care in England: an empirical investigation. *Social Science and Medicine*, 60, 1251-1266.

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

National Health Service (NHS) Information Centre. 2008a. Prescription cost analysis 2007.

National Health Service (NHS) Information Centre. 2008b. Prescriptions dispensed in the community, statistics for 1997 to 2007: England.

Omar, R.Z., O'Sullivan, C., Petersen, I., Islam, A., & Majeed, A. 2008. A model based on age, sex, and morbidity to explain variation in UK general practice prescribing: cohort study. *British Medical Journal*, 337, doi:10.1136/bmj.a238

Perkins, A.J., Kroenke, K., Unutzer, J., Katon, W., Williams, J.W., Hope, C., & Callahan, C.M. 2004. Common comorbidity scales were similar in their ability to predict health care costs and mortality. *Journal of Clinical Epidemiology*, 57, (10) 1040-1048.

Reid, R.J., Roos, N.P., MacWilliam, L., Frohlich, N., & Black, C. 2002. Assessing population health care need using a claims-based ACG morbidity measure: A validation analysis in the province of Manitoba. *Health Services Research*, 37, (5) 1345-1364.

Salisbury, C., Johnson, L., Purdy, S., Valderas, J.M., & Montgomery, A.A. 2011. Epidemiology and impact of multimorbidity in primary care: a retrospective cohort study. *British Journal of General Practice*, 61, (582) 18-24.

Schokkaert, E., Dhaene, G., and Van de Voorde, C. 1998. Risk adjustment and the trade-off between efficiency and risk selection: an application of the theory of fair compensation. *Health Economics* 7:465-480.

Schwarz, G. 1978. Estimating dimension of a model. *Annals of Statistics*, 6, (2) 461-464.

Starfield, B., Weiner, J., Mumford, L., & Steinwachs, D. 1991. Ambulatory care groups - A categorization of diagnoses for research and management. *Health Services Research*, 26, (1) 53-74.

StataCorp. 2009. *Stata Statistical Software: Release 11*. College Station, TX: StataCorp LP.

Sullivan, C.O., Omar, R.Z., Ambler, G., & Majeed, A. 2005. Case-mix and variation in specialist referrals in general practice. *British Journal of General Practice*, 55, (516) 529-533.

Sutton, M., 2002. Vertical and horizontal aspects of socio-economic inequity in general practitioner contacts in Scotland. *Health Economics*. 11, 537–549.

Technical Steering Committee 2010, *GP Earnings and expenses 2007/08.* NHS Information Centre.

van de Ven, W. P. and R. P. Ellis (2000): Risk adjustment in competitive health plan markets, Elsevier, vol. 1 of *Handbook of Health Economics*, 755–845.

Weiner, J.P., Starfield, B.H., Steinwachs, D.M., & Mumford, L.M. 1991. Development and application of a population-oriented measure of ambulatory care case-mix. *Medical Care*, 29, (5) 452-472.

Winkelman, R., Mehmud, S. 2007. A comparative analysis of claims based tools for health risk assessment. Society of Actuaries. www.soa.org/files/pdf/risk-assessmentc.pdf. Accessed 18 October 2011.