ISER Working Paper Series

Baseline Health and Public Healthcare Costs Five Years On: A Predictive Analysis Using Biomarker Data in a Prospective Household Panel

Apostolos Davillas
Institute for Social and Economic Research
University of Essex

Stephen Pudney
School of Health and Related Research
University of Sheffield

No. 2019-01
February 2019

INSTITUTE FOR SOCIAL & ECONOMIC RESEARCH

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

University of Essex

www.iser.essex.ac.uk

Non-Technical Summary

There is a large epidemiological research literature examining the predictive power of personal characteristics, socio-economic status (SES) and current health state, on future health outcomes. Although this literature is driven by, and clearly focused on, a concern for welfare at the individual level, the associated implications for the monetary costs to wider society are often neglected or left implicit. In this paper, we add to a small but growing body of research on the individual-level determinants of healthcare demand and costs.

Health care costs have risen faster than economic growth in all OECD countries and this is projected to continue as a result of new medical technology, rising expectations and the increasing needs of the ageing population. In this policy setting, it is important for policymakers to be able to identify the sections of the population where costs are high and rising, to establish priorities for resource planning and preventative policy.

Using data from Understanding Society, our working sample is 2,314 adults who, at baseline in 2010/11, reported no history of diagnosed long-lasting health conditions and for whom a set of objective health measures (nurse-collected and blood-based biomarkers) were observed. Five years later, their utilisation of GP and hospital outpatient and inpatient services was observed. We develop econometric techniques appropriate for the purpose of our analysis and a statistical method of combining NHS episode cost data with Understanding Society data. This allows us to estimate the impact of differences in personal characteristics and socio-economic status (SES) on cost outcomes.

We find that a biomarkers summary measure, capturing several dimensions of physical health, is a powerful predictor of realised costs: among the group of individuals with excess allostatic load at baseline, we estimate that a reduction to achieve more normal biomarker levels reduces GP, outpatient and inpatient cost outcomes by around 18%. In addition to the expected strong effect of ageing on cost, we also find a large gender difference: on average women experience costs at least 20% higher than comparable men, because of their greater utilisation of GP and outpatient services. There is a strong SES gradient in healthcare costs: the average impact of moving from no educational qualifications to intermediate or from intermediate to degree level is approximately 16%. Income differences, on the other hand, have negligible impact on future costs.

The predictive power of personal characteristics and biomarker-based health measures gives a possible basis for sophisticated tailoring preventive interventions. A measure similar to our allostatic load proxy could be constructed from information gathered in the NHS England Health Check introduced in 2009, which offers quinquennial check-ups including blood tests. There are concerns about low take-up, which is a potential obstacle for any such preventive measure. However, the NHS Health Check is available to all adults aged 40-74 and thus targeted only on age in a simple way and our findings suggest that more tailored targeting could identify better the population groups with highest potential future healthcare needs and costs. Of course, for this to be worthwhile there needs to be an effective follow-up intervention that can be used to reduce those future health needs and costs.

Another potential policy application of our findings is in refining the design of capitation payment systems by reorienting the capitation formula to match more closely patient level morbidity data and other demographic and SES characteristics. This offers the prospect of improved allocation of resources as well as health outcomes by reducing incentives for health providers to “cream skim” the patient population by selecting patient groups with lower expected future healthcare costs.

Baseline health and public healthcare costs five years on: A predictive analysis using biomarker data in a prospective household panel

Apostolos Davillas
Institute for Social and Economic Research, University of Essex

Stephen Pudney
School of Health and Related Research, University of Sheffield

This version: February 12, 2019

Abstract: We investigate the utilisation of primary and secondary public healthcare services and the consequent public costs, using data from the British *Understanding Society* household panel. We use a sample of 2,314 adults who, at baseline in 2010/11, reported no history of diagnosed long-lasting health conditions and for whom a set of objective biomarkers were observed. Five years later, their utilisation of GP and hospital outpatient and inpatient services was observed. We develop an econometric technique for count data observed within ranges and a method of combining NHS episode cost data with *Understanding Society* data without exact individual-level matching. This allows us to estimate the impact of differences in personal characteristics and socio-economic status (SES) on cost outcomes. We find that a composite biomarker index approximating allostatic load is a powerful predictor of realised costs: among the group who are at least 1 standard deviation (SD) above mean allostatic load, we estimate that a reduction of 1 SD at baseline reduces GP, outpatient and inpatient cost outcomes by around 18%. In addition to the expected strong effect of ageing on cost, we also find a large gender difference: on average women experience costs at least 20% higher than comparable men, because of their greater utilisation of GP and outpatient services. There is a strong SES gradient in healthcare costs: the average impact of moving from no educational qualifications to intermediate or from intermediate to degree level is approximately 16%. Income differences, on the other hand, have negligible impact on future costs.

Keywords: Healthcare costs; Socioeconomic gradient; Biomarkers; Allostatic load, Understanding Society

JEL codes: C3, C8, I10, I18

Contact: Steve Pudney, ScHARR, University of Sheffield, 30 Regent Street, S1 4DA, UK; tel. +44(0)114 222 9187; email: steve.pudney@sheffield.ac.uk

Understanding Society is an initiative funded by the Economic and Social Research Council and various Government Departments, with scientific leadership by the Institute for Social and Economic Research, University of Essex, and survey delivery by NatCen Social Research and Kantar Public. The research data are distributed by the UK Data Service. Participants gave informed consent for their blood to be taken for future scientific analysis. Biomarker collection was approved by the National Research Ethics Service (10/H0604/2). We are grateful to the Economic and Social Research Council for financial support for this research via project *How can biomarkers and genetics improve our understanding of society and health?* (award no. ES/M008592/1). The funders, data creators and UK Data Service have no responsibility for the contents of this paper. We are grateful to members of the project team for many helpful comments. Any remaining errors are our sole responsibility.

1 Introduction

There is a large epidemiological research literature examining the predictive power of personal characteristics, socio-economic status (SES) and current health state, on future health outcomes. Although this literature is driven by, and clearly focused on, a concern for welfare at the individual level, the associated implications for the monetary costs to wider society are often neglected or left implicit. But costs are important and the link between the impacts of an individual's characteristics on his or her health outcome and on social cost is not necessarily simple. In this paper, we add to a small but growing body of research on the individual-level determinants of healthcare demand and costs (Brilleman et al., 2014; Doorslaer et al., 2004; Bago d'Uva and Jones, 2009; Sari, 2009; Sturm, 2002; Traczynski and Udalova, 2018).

Health care costs have risen faster than economic growth in all OECD countries and this is projected to continue as a result of new medical technology, rising expectations and the increasing needs of the ageing population (OECD, 2015). In Britain, the National Health Service (NHS) spends about 10% of UK GDP on health care, which is broadly in line with other European counties. That proportion has doubled since the establishment of the NHS in 1948 (Charlesworth and Bloor, 2018). However, the NHS has below the OECD average number of doctors, nurses and hospital beds per head and its performance on some health outcomes (for example, survival rates for breast and cervical cancer or health-care amendable mortality) is below the average among comparable countries (Barber et al., 2017). In this policy setting, it is important for policymakers to be able to identify the sections of the population where costs are high and rising, to establish priorities for resource planning and preventative policy.

Administrative data on healthcare utilisation have been used to describe the associations of demographic characteristics and morbidity with utilisation (Cawley and Meyerhoefer, 2012; Brilleman et al., 2014; Carreras et al., 2018). But research aiming to inform the

development of forward-looking policy to control costs cannot be done solely on the basis of records accumulated by the health care system, since they do not contain information on the full range of personal and SES characteristics, and do not have good coverage of individuals with latent health conditions that have not yet reached the stage of diagnosis.[1]

An ideal approach would be to use data from a nationally representative longitudinal survey with administrative data from NHS records matched to survey respondents to add detail on the amounts and types of treatment received by each respondent. However, there are also difficulties here. First, the UK has no national-scale individual-level administrative datasets on general practitioner (GP) and community health service delivery. Second, although comprehensive hospital episodes (HES) data exist in the UK, there are barriers to data linkage which currently prevent matching to suitable longitudinal surveys. Those barriers arise both from legal and ethical data security restrictions and also practical difficulties of data quality.[2] A third disadvantage is that matching to survey data requires informed consent from respondents, which is achievable for fewer than 70% of respondents in surveys like the *Understanding Society* household panel (also known as the UK Household Longitudinal Study; UKHLS) used in this study.[3] Conditioning the analysis on informed consent thus introduces a further possible source of bias so even when (if) a matched UKHLS-HES dataset becomes available, there will still be a case for alternative approaches to contribute to a robust research consensus. A final difficulty is that there exists no individual-specific data on treatment costs, so even with a matched dataset, it would be necessary to estimate cost rather than observe it directly.

The alternative to matched HES data is self-reported information on utilisation of healthcare resources, in response to survey questions asking for counts of numbers of medical con-

[1] The extensive population databases assembled for some Scandinavian countries come closest to this ideal, but still lack some critical information, especially pre-diagnostic states of ill-health.

[2] At present, HES data for England have been matched to the Biobank dataset, which is unrepresentative and lacking important socioeconomic contextual data, and to the Millenium Cohort Study, which is limited to a single birth cohort.

[3] Consent rates fall much lower than this when interviewing is conducted online with no interviewer present to establish a relationship of trust, see Jäckle et al. (2018).

sultations within a reference period. In the case of the UKHLS, these counts are reported as the numbers of GP, outpatient and day patient (OP) consultations within the preceding year, and the number of days spent as a hospital inpatient (IP) over the same period. There is a potential problem of recall error, mainly in the form of under-reporting, but comparisons with macro-level administrative data (section 2.2 and Appendix 1) suggest that such error is moderate in size. In the UKHLS and many other surveys, consultation counts are reported in grouped rather than exact form. In section 3, we develop a new method of modelling such data econometrically using an interval negative binomial specification. Our econometric modelling operates in a 5-year-ahead predictive framework using waves 2 and 7 of the UKHLS, and gives us estimates of the impacts on resource utilisation five years later of baseline indicators of health risks (measured by biomarkers), demographic characteristics, and SES, which are discussed in section 4.

A key feature of our empirical modelling is the use of biomarkers to measure the baseline health state, which is a core predictor of subsequent healthcare utilisation and costs. Biomarkers are more objective than conventional self-assessed health and have the potential to act as indicators of health problems prior to the symptomatic and diagnostic stages of disease (Goldman et al., 2006; Geronimus et al., 2006; Turner et al., 2016). We use a measure of cumulative biological risk factors, often called allostatic load, which combines biomarkers relevant to different biological systems (Davillas and Pudney, 2017; Howard and Sparks, 2016; Seeman et al., 2004). Section 2.1 of the paper details its construction. We focus on individuals who appeared, from a clinical point of view, to be healthy at baseline, so we excluded from our analysis those who reported any past or recent diagnosed long-lasting health condition.

A final step is needed to generate estimates of resource costs, in the absence of survey information on the type of disorder and treatment involved in reported contacts with the health service (section 5). We attach average unit costs to the GP and OP consultation counts, but develop a more sophisticated method of statistical assignment of treatment types

and corresponding costs to the predictions of the IP utilisation model. This approach exploits reported information on hospital stay duration, age and gender to tailor the probabilities of each treatment type to the survey individual. We find large predictive impacts of age, gender, SES and health measured at baseline.

There are potential policy implications of our results. They can be used to indicate priority areas for interventions with such as screening programmes and health education initiatives to control future treatment costs among individuals who have not yet reached the stage of diagnosis. They are also relevant to the design of capitation fee systems, used in the UK and elsewhere to pay providers prospectively for treatment of patients to whom they agree to provide health care. There is a need to tailor capitation payments closely to expected future healthcare costs to reduce incentives for providers to engage in "cream-skimming" behaviour. Currently, most capitation payments are not based on patient-level data, apart from age and gender, neglecting other potentially important patient-level characteristics (Brilleman et al., 2014; Shepherd, 2017).

2 Data: the *Understanding Society* panel (UKHLS)

The UKHLS is a longitudinal, nationally representative study of the UK, designed as a two-stage stratified random sample of the general population. We use the Great Britain (GB) subsample, excluding the Northern Ireland component of the UKHLS which does not provide biomarker data. As part of wave 2 (2010-2011), nurse-measured and non-fasted blood-based biomarkers were collected, giving a potential pool of 6,337 survey respondents with valid data on all the nurse-collected and blood-based biomarkers used in our analysis. From those, 4,759 individuals had non-missing data on SES and demographic covariates at baseline (wave 2) and were successfully followed up at wave 7, when healthcare utilisation measures are collected. Our focus is on individuals who appeared (from the viewpoint of clinical diagnosis) to be healthy at baseline, so we further excluded from our analysis those who reported any

past diagnosis of a long-lasting health condition (asthma, chronic bronchitis, congestive heart failure, coronary heart disease, heart attack or myocardial infarction, stroke, cancer or malignancy, diabetes, high blood pressure, arthritis and liver condition), or a hospital inpatient stay with a newly diagnosed health condition (UKHLS wave 2). This allows us to follow a set of 2,314 respondents in apparently good health at baseline, up to five years later (UKHLS wave 7).

2.1 A multi-system measure of health risks at baseline

Allostatic load was developed as a measure of biological risk arising from the cumulated effects of chronic exposure to physical, psychosocial and environmental stressors that may lead to physiological dysregulation and increased risk of manifest diseases (Howard and Sparks, 2016; Seeman et al., 2004). Allostatic load is a multisystem risk score, sensitive to morbidities that may be yet undiagnosed (Geronimus et al., 2006; Turner et al., 2016).

A large set of physical measurements and blood-based biomarkers, spanning multiple dimensions of health, were collected by trained nurses at UKHLS wave 2. Our index combines markers for adiposity, blood pressure, heart rate, lung function, inflammation, blood sugar levels, cholesterol levels, liver function and steroid hormone.[4]

We use waist-to-height ratio to measure adiposity and resting heart rate (HR), systolic blood pressure (SBP) and high-density lipoprotein cholesterol (HDL) to measure cardiovascular health.[5] Lung function is measured, using a spirometer, as forced vital capacity (FVC), the total amount of air forcibly blown out after a full inspiration; higher FVC values indicate better lung functioning. C-reactive protein (CRP) is our inflammatory biomarker,

[4]Some authors include cortisol, in addition to the stress-related hormone DHEAS, to capture primary responses to stress. However, cortisol is excluded here because of time-of-day and other measurement difficulties in the UKHLS context. Similar constructions to ours have been used extensively in previous studies (Davillas and Pudney, 2017; Howard and Sparks, 2016; Vie et al., 2014).

[5]SBP is the maximum pressure in an artery at the moment when the heart is pumping blood; it is generally considered more relevant to health risks than diastolic blood pressure (Haider et al., 2003). Low HDL cholesterol levels are associated with increased cardiovascular risks, while low HR and SBP indicate lower risks.

which rises as part of the immune response to infection and is associated with general chronic or systemic inflammation.[6] Glycated haemoglobin (HbA1c) is our blood sugar biomarker, and is a validated diagnostic test for diabetes. Albumin is used to proxy liver functioning, with low albumin levels suggesting impaired liver function. We also use dihydroepiandrosterone suphate (DHEAS), a steroid hormone in the body, in our composite index of health. DHEAS is one of the primary mechanisms through which psychosocial stressors may affect health, with low levels associated with cardiovascular risk and all-cause mortality (Vie et al., 2014). We calculated a composite risk score measure to proxy allostatic load after converting HDL, Albumin and DHEAS to negative values to reflect ill-health rather than good health, and then transforming each biomarker into a z-score and summing to produce the composite measure.[7]

2.2 Health care utilisation measures

Retrospective information on the number of GP consultations, attendance at a hospital or clinic as an out-patient or day patient (OP), and in-patient (IP) days in the preceding 12 months were also collected at UKHLS wave 7. The numbers of GP and OP consultations were collected as five-category variables: 0, 1-2, 3-5, 6-10, more than 10. Respondents were asked how many days they spent in a hospital or clinic as an IP in the preceding 12 months. To ensure that our health care utilisation measures are not contaminated by any pregnancy-related visits, we excluded women who reported any in-patient days for childbirth during this period (about 0.5 per cent of our sample), so our cost analysis excludes services related to childbirth.

There are clear distributional differences between age groups, with GP and OP consul-

[6] We exclude CRP values over 10mg/L because such values may reflect acute rather than systemic inflammatory processes (Pearson et al., 2003).

[7] When used singly in econometric models, each of these biomarkers has a statistically significant coefficient, but their strong intercorrelations make it impossible to estimate robust models involving all nine biomarkers jointly as covariates.

tations being more evident for those at older ages (Appendix Figures A1 and A2). Figure 1 shows gender differences in the distributions of GP and OP consultations, indicating that women tend to seek care from GP or OP consultations more frequently than men.

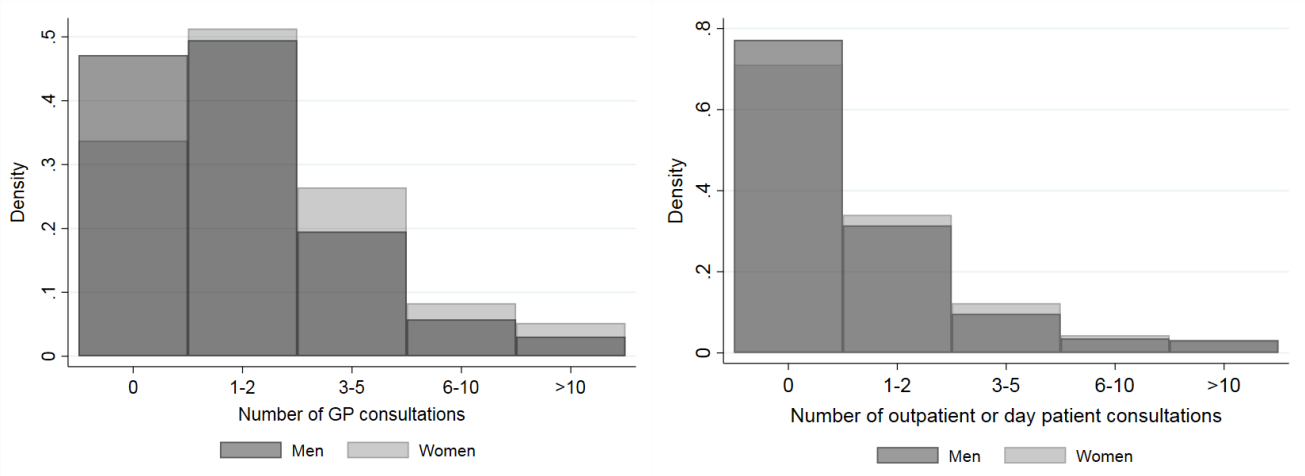

Figure 1: Distribution of the numbers of GP and OP consultations in the preceding 12 months by gender

The GP, IP, and OP utilisation counts are retrospective self-reports of utilisation of health services over the past year, so they are potentially subject to well-known biases in long-term recall (Bound et al., 2001). To check this, we can compare the full wave 7 UKHLS data with external sources of information (Hobbs et al., 2016; ISD Scotland, 2017; NHS Digital, 2017; NHS Improvement, 2017). Appendix Tables A1 and A2 give comparisons of GP and OP consultation data for England and Scotland and IP days for England only. These comparisons are not straightforward, since the UKHLS GP and OP data are interpolated, there are minor differences in timing, and the administrative data relate to the whole population whereas the UKHLS is a sample from the household population only, subject to variations in response rates across population groups.[8]

Overall, we find that the administrative GP consultation rates for England and Scotland

[8]The comparisons presented in the appendix use unweighted UKHLS data. Since the UKHLS response weights are built up sequentially over waves, missing data causes progressive loss of information and, by wave 7 almost a quarter of individuals have zero weights. If we use the official weights, results are not changed in any important way from those presented in Tables A1 and A2.

are reasonably close to mean counts interpolated[9] from the UKHLS interval data (Table A1). There is some evidence of moderate under-reporting in the UKHLS, with discrepancies larger for women than men, for older than younger respondents and for the English rather than Scottish subsample. For OP consultations (Table A2) we have no demographic breakdown of the administrative data; the overall mean counts are reasonably close to the ratio of aggregate consultations to relevant population size, for both England and Scotland (UKHLS rates lower by 4% and 12% respectively). For IP utilization, we only have administrative data for England. Unlike most of the comparisons for GP and OP consultations, the UKHLS mean IP count is larger (by almost 10%) than the corresponding administrative estimate (Table A2), but this is largely due to definitional differences – an IP episode completed within one day is recorded as a zero-days duration in the HES data, but would generally be reported by UKHLS respondents as a one-day episode. When linking costs to durations (section 5) we allow for this by adding 1 to durations in the HES data.

These differences should be borne in mind when interpreting our results, but do not seem large enough to greatly distort econometric results. The main cause for concern is the possible under-reporting of GP consultations by older women, which would suggest that the large demographic effects reported in Tables 2-4 may be underestimates.

2.3 Costs

Financial cost is the natural metric for distilling the three categories of resource use into a single measure of burden on public healthcare resources. However, this is not straightforward because the UKHLS interview gives no details of the types of treatment involved, nor is it possible to match survey respondents to records of the public healthcare system.[10] Instead,

[9] Interpolation is done here by fitting negative binomial distributions (with zero-inflation where appropriate) to the interval data, then calculating the expected value of the count conditional on the observed interval.

[10] Consents for matching of UKHLS data to hospital episodes administrative data were obtained for a subset of UKHLS respondents, but a usable matched dataset is not expected to be available for a considerable time. Moreover, such a dataset would not cover GP consultations and would raise significant issues of non-consent

we pursue a data combination strategy, exploiting average cost data published in varying detail for the GP, OP and IP resource classes.[11]

For simplicity, we use reference costs and caseload composition figures from NHS England for the whole of the UKHLS sample, including the relatively small Scottish and Welsh subsamples (making up 5% and 2% of the analysis sample, respectively). Robustness checks reported in section 6 confirm that results are not materially affected by restricting the sample to respondents resident in England.

GPs are the gatekeepers to NHS healthcare services but they are self-employed contractors rather than employees of the NHS and, consequently, financial data relating to GP services are not available on the same detailed basis as for the rest of the NHS. We use the mean unit cost per consultation estimated by Curtis and Burns (2017) as £66.20 per consultation, comprising £37 for GP costs and £29.20 for associated prescription costs (on a net ingredient cost basis).

NHS reference cost data for OP and IP activity in England give unit costs broken down in great detail by type of treatment and compiled according to standard measurement conventions (DOH, 2016). We use the national schedules of reference costs (NHS Improvement, 2017), providing data on average unit cost for each service submitted by the NHS providers in 2016/17, a similar period to that covered by UKHLS wave 7. For OP cases, the average unit cost and aggregate number of attendances in each treatment category relate to both outpatient and day-case visits. After excluding paediatric categories which are not relevant to UKHLS adult respondents, and any category with fewer than 50 cases in the year, we are left with 1,355 treatment categories with an average caseload of 55,704 attendances per category and a mean (caseload-weighted) unit cost of £163.32 per OP consultation.

For IP cases, reference costs relate to episodes of care, defined as "the time spent under

bias.

[11] GP, OP and IP costs are only part of the cost picture. The UKHLS questions do not cover resources like community nursing, ambulance services, etc. Moreover published hospital reference costs exclude some activities such as screening (DOH (2016), section 15).

the care of one consultant", are available as average unit cost by groups of patient events that have been judged to consume a similar level of resource, known as Healthcare Resource Groups (HRG), along with the aggregate number of HRG episodes (NHS Improvement, 2017). Elective and non-elective IP treatment are separated in the official activity and cost data, and we treat them as distinct treatment types. Treatment categories are further separated into elective (E), non-elective long stay (NEL) and non-elective short stay defined as 2 days or less (NES). In our analysis, we treat these types as separate categories, exploiting the fact that caseload, unit cost and mean length of stay (but not other episode characteristics) are broken down by type of episode (NHS Digital, 2017; NHS Improvement, 2017). After excluding categories which cover paediatric cases, or involve fewer than 50 cases, or have missing or invalid unit cost, mean stay data, we are left with 3,827 IP categories, with a (caseload-weighted) mean stay length of 3.5 days, and a mean total of case-days of 9,184 per category. The overall mean daily unit cost defined as the ratio of aggregate cost to aggregate number of days IP treatment is £542 per day.

3 Grouped count data models of healthcare utilisation

Let $Y_i \geq 0$ be the ith observation on a dependent variable (the GP, OP or IP utilisation count), which takes non-negative integer values, and $\boldsymbol{X}_i$ a vector containing the explanatory covariates. We allow for the possibility of zero-inflation: a double hurdle or mixture process, where some individuals have a degenerate zero count with probability 1, while others have a count drawn from a standard distribution. The probability of a degenerate zero is specified as probit:[12]

$$Pr(\text{degenerate } 0|\boldsymbol{X}_i) = \Phi(\boldsymbol{X}_{i1}\boldsymbol{\gamma}) \tag{1}$$

where $\boldsymbol{X}_{i1}$ is a subvector of $\boldsymbol{X}_i$.

[12]We also estimated logit specifications which gave almost identical estimates.

The distribution of Y among the non-degenerate population is $g(y|\boldsymbol{X}_{i2})$, where $\boldsymbol{X}_{i2}$ is another subvector of $\boldsymbol{X}_i$. The mixture distribution of Y is:

$$f(y|\boldsymbol{X}_i) = \begin{cases} \pi(\boldsymbol{X}_{i1}\boldsymbol{\gamma}) + (1 - \pi(\boldsymbol{X}_{i1}\boldsymbol{\gamma}))g(0|\boldsymbol{X}_{i2}) & \text{if } y = 0 \\ (1 - \pi(\boldsymbol{X}_{i1}\boldsymbol{\gamma}))g(y|\boldsymbol{X}_{i2}) & \text{if } y > 0 \end{cases} \tag{2}$$

Our observations are not necessarily on Y_i itself but rather an interval within which Y_i lies. Consequently, we have a pair of observed dependent variables, $[L_i, U_i]$ with the property that $L_i \leq Y_i \leq U_i$. For the GP and OP consultation counts, the observable limit pairs are in the set $\{(0,0),(1,2),(3,5),(6,10),(11,\infty)\}$; for the IP count we have exact observation, so $L_i = Y_i = U_i$. The likelihood for individual i is the conditional probability of observing the event $L_i \leq Y_i \leq U_i$:

$$Pr(L_i \leq Y_i \leq U_i|\boldsymbol{X}_i) = \sum_{y=L_i}^{U_i} f(y|\boldsymbol{X}_i) \tag{3}$$

The model is completed by a specifying a parameterized functional form for the non-degenerate component distribution $g(.|\boldsymbol{X}_i)$. We initially considered three alternative base models, binomial, Poisson and negative binomial (NB). The NB specification gave by far the best fit in every case (Pudney, 2018). It is derivable as a Poisson$(\lambda_i\nu)$-gamma(α^{-1},α) mixture, where $\lambda_i = e^{\boldsymbol{X}_{i2}\boldsymbol{\beta}}$ and $\ln\alpha$ is treated as an unrestricted constant parameter. This gives a distribution for y with mean λ_i and variance $1 + \alpha\lambda_i$.[13] The ML estimator for is implemented in a new Stata command `intcount`, documented in Pudney (2018).

4 Parameter estimates

The explanatory covariates $\boldsymbol{X}$ used in our healthcare utilization model represent individual characteristics that been shown to affect health outcomes either directly or indirectly

[13] In the terminology of Cameron and Trivedi (2013), this is the NB2 parameterization of the regression model.

(Davillas and Pudney, 2018; Carrieri and Jones, 2017; Doorslaer et al., 2004). They were collected as part of the UKHLS wave 2 main survey, along with our biomarker measures. We use two indicators of SES. Educational attainment is captured as a 3-category classification: degree-equivalent, intermediate, and no/basic qualification. Household income is the sum of the gross incomes of all household members but, to avoid spurious correlation arising from the fact that disability resulting from ill-health creates eligibility for disability benefits (Morciano et al., 2015), income from those sources is excluded. We allow for differences in household composition by equivalising household income using the modified OECD equivalence scale before using a log transformation to allow for the concavity of the health-income association. A flexible quadratic function of age and gender is used to capture demographic differences. Finally, we also allow for differences between the three nations of Great Britain (England, Scotland and Wales), since NHS policy is determined on a national basis.

In implementing the NB models, we embed an important feature of the healthcare system in the UK. GPs normally act as gatekeepers to the hospital system, so OP or IP episodes are mostly preceded by GP consultations (Doorslaer et al., 2004; Brilleman et al., 2014). For that reason, we model OP and IP utilisation counts conditional on the number of GP consultations, with $\boldsymbol{X}$ extended to include categorical indicators of the number of GP consultations. Parameter estimates for our preferred models are shown in Table 1 (columns 2, 4 and 6). Marginal OP and IP models, estimated without conditioning on the GP consultation count, are also shown for comparison (columns 3 and 5). For the OP and IP counts, the best-fitting model involves zero-inflation, distinguishing between individuals with zero and non-zero GP consultation counts. The estimated impact of a zero GP count on the OP and IP counts is almost completely sharp, with large negative intercept and large positive coefficient. That implies negligible zero-inflation for the OP and IP counts if the GP count is positive, and large probabilities of a degenerate zero (0.69 for the OP count and 0.98 for the IP count) if the GP count is zero.[14]

[14]In practice the gatekeeper role of GPs is not completely sharp, since GP consultations leading to an OP consultation or IP episode may not fall in the same 12-month recall period; also some emergency IP

We used initially a larger set of covariates than that shown in Table 1, including additionally urban/rural area type, marital status, housing tenure and household size, but coefficients of those variables were statistically insignificant in all of the models, so the smaller set of covariates listed in Table 1 was adopted. Inclusion of smoking and physical activity produced no significant effects in any model after accounting for allostatic load, indicating that information on unhealthy lifestyles at baseline has no additional predictive power for subsequent health care utilization beyond what can be achieved using biomarkers.

cases may reach hospital without GP involvement. Consequently we have chosen to leave the model fully parameterised rather than imposing a zero probability of zero-inflation when the GP count is zero. For zero-inflated models of the GP count, the ML optimisation always led to corner solutions where the probability of a degenerate zero count was essentially zero.

Table 1: Estimated parameters for grouped negative binomial models of GP, OP and IP utilisation: baseline demographic, SES and health state coefficients

		OP model		IP model	
Parameter	GP	marginal	conditional	marginal	conditional
Allostatic load	0.210***	0.129**	-0.030	0.555***	0.219
	(0.036)	(0.053)	(0.055)	(0.162)	(0.175)
§Age	-0.053*	0.033	0.068	0.241	0.177
	(0.030)	(0.046)	(0.045)	(0.154)	(0.137)
§Age^2	0.017	0.018	0.002	-0.024	-0.051
	(0.016)	(0.026)	(0.026)	(0.081)	(0.072)
Male	-0.198***	-0.327***	-0.319***	-0.285	-0.493
	(0.076)	(0.121)	(0.110)	(0.397)	(0.370)
§Age × male	0.146***	0.187***	0.119*	-0.367*	-0.165
	(0.042)	(0.069)	(0.067)	(0.209)	(0.201)
§Age^2× male	-0.025	0.024	0.061*	0.182	0.247**
	(0.025)	(0.038)	(0.037)	(0.117)	(0.111)
No qualifications	0.092	0.033	-0.016	0.744	1.481*
	(0.092)	(0.142)	(0.137)	(0.566)	(0.718)
Intermediate qualifications	0.016	0.104	0.112	0.691**	0.745**
	(0.065)	(0.103)	(0.096)	(0.321)	(0.290)
ln(income)	-0.140***	-0.106*	-0.002	-0.209	0.101
	(0.045)	(0.062)	(0.059)	(0.237)	(0.206)
Wales	0.555***	0.140	-0.365	-0.061	-1.062*
	(0.192)	(0.332)	(0.249)	(0.694)	(0.590)
Scotland	-0.018	-0.246	-0.311	0.278	-0.405
	(0.135)	(0.231)	(0.198)	(0.776)	(0.680)
Intercept	1.793***	1.001**	0.165	-0.270	-1.573
	(0.351)	(0.483)	(0.481)	(1.883)	(1.670)
Impact of conditioning on GP consultation count					
1-2			-0.237		-1.384*
			(0.206)		(0.794)
3-5			0.534***		-1.248
			(0.206)		(0.817)
6-10			0.952***		-0.203
			(0.214)		(0.860)
more than 10			1.977***		1.640**
			(0.270)		(0.832)
$ln(\alpha)$	0.012	1.099***	0.683***	3.895***	3.307***
	(0.056)	(0.058)	(0.080)	(0.121)	(0.150)
Zero-inflation parameters					
Zero GP consultation count			16.677***		13.424***
			(0.320)		(0.724)
Intercept			-16.168***		-11.319***
			(0.230)		(0.560)
AIC	6067.9	4913.3	4480.0	1514.5	1445.4
BIC	6142.6	4988.0	4589.2	1589.2	1554.6

§ Age measured in decades from an origin of 50. Standard errors in parentheses. Sample size $N = 2314$.
Statistical significance: * = 10%, ** = 5%, *** = 1%.

Table 1 shows a strong predictive role of allostatic load for GP consultations, implying

an expected increase of $e^{0.21} - 1 = 23\%$ in GP consultations five years after a 1-standard deviation increase in allostatic load. In models for OP and IP that condition on the GP count, there is no further statistically significant direct impact of allostatic load, so the effect of allostatic load is primarily channeled through the increased engagement with primary healthcare. The marginal models of OP and IP that do not condition on the GP count have highly significant coefficients of 0.129 and 0.555, implying total five-year impacts of a standard deviation increase in allostatic load as 14% for OP consultations and 74% for IP days. The statistical dependence between the GP count and the OP and IP counts is confirmed by the large significant coefficients for the GP variables in the conditional OP and IP models, and the much higher AIC and BIC statistics for the models that do not condition on GP visits.

The highly significant coefficients for the male gender dummy and its interaction with age is consistent with the differences shown in Figures 1, A1 and A2, and they imply that men tend to use health resources less than women, but that their use increases faster with age. These results accord with recent research on demographic variations in primary health care costs (Brilleman et al., 2014).

With respect to our SES variables, there is a highly significant gradient of future GP consultations with respect to baseline household income, with an elasticity of -0.14. This strong gradient in primary care demand explains why the negative income gradient in OP consultation counts (with an elasticity of -0.11) is completely attenuated when we condition the OP model on GP consultations. Educational attainment has no significant estimated influence on the demand for primary and OP care, but it exerts a very large influence on the demand for IP care, with expected IP days doubling as we go from degree-level to intermediate attainment and doubling again if there are no qualifications.[15] Since IP treatment generally results from more serious conditions, this indicates that education, rather than income, is the key SES influence on severity.

[15]Using the conditional model, $e^{.745}$ and $e^{1.481-.745}$ are both approximately 2.1).

The nations of Great Britain have separate NHS organisation and funding arrangements, and we also find a significant divergence between Wales and the rest of Great Britain. Respondents resident in Wales have significantly higher engagement with the primary care system but lower probability of passing onto the hospital system through OP or IP consultations, with a net effect of no significant difference in overall access to OP or IP care in the marginal models. This is intriguing but cannot be pursued further owing to the small size of the Welsh biomarker subsample (2% of the whole GB sample). We find no significant difference between England and Scotland in marginal or conditional models.

Statistical significance does not necessarily mean that estimated effects are large enough to matter in practice. To quantify more fully the impacts of differences in baseline demographics, SES and allostatic load, we focus on a set of sub-groups defined by their baseline characteristics, evaluate the predicted outcome five years later, and then calculate the impact on that outcome that would be generated by a hypothetical change in the baseline characteristics defining each group. We use the following sequential Monte Carlo simulation, conditional on the baseline covariates, where $r = 1 \dots 250$ indexes pseudo-random replication sequences:

(*i*) Draw pseudo-random Y_{ir}^{GP} from the conditional distribution $Pr(Y_i^{GP}|\boldsymbol{X}_i^{GP})$ derivable from the fitted model for GP consultations.

(*ii*) Use Y_{ir}^{GP} to construct the extended covariate vectors $\boldsymbol{X}_i^{OP}$ and $\boldsymbol{X}_i^{IP}$. Then evaluate $Pr(Y_i^{OP}|\boldsymbol{X}_i^{OP})$ and $Pr(Y_i^{IP}|\boldsymbol{X}_i^{IP})$ for each individual, using the fitted models for OP and IP consultation counts, and draw pseudo-random Y_{ir}^{OP} and Y_{ir}^{IP} from those distributions.

(*iii*) Compute any desired summary measures (means, probabilities of positive counts, etc.) of the distribution of Y^{GP}, Y^{OP}, Y^{IP} from the R replications.

This procedure is repeated (reusing the same pseudo-random number sequences), after perturbing the covariate values appropriately. Table 2 shows the simulated impacts of nine types: a 1-standard deviation (1σ) reduction in allostatic load for all those who are more than

1σ above the mean; increasing age by 10 years for all members of each of five baseline age groups; changing gender for the two gender groups in turn; increasing educational attainment by one category for each of the unqualified and intermediate groups; and a universal 10% increase in equivalised income.

First, note that allostatic load is a strong predictor of future healthcare demand, with potential for a substantial reduction in resource usage if effective interventions could be targeted on those with high allostatic load. A 1σ deviation reduction among that group is predicted to reduce GP and OP consultations by 19% and 12% respectively, and the more costly IP resource by over 40%. This indicates that allostatic load is a hgihly effective predictor, particularly for relatively serious conditions requiring hospital stays.

Demographic impacts are also large. Gender differences have implications ranging from 23% to 33% for GP and OP consultations. Smaller differences of 1-7% are evident for IP resources, confirming that the main source of gender difference is the greater engagement with the primary care and hospital outpatient system by women than men. Age is an extremely important factor, particularly in the oldest over-75 group for IP treatments, where a uniform 10-year increase in age raises IP utilisation by over 250%.

After controlling for allostatic load and demographics, there remains a SES gradient in healthcare utilisation. As measured by educational attainment, the gradient is only statistically significant for IP resources, but there the effect is large: a 1-category increase in educational attainment reduces the expected IP day count by over 50%. Although the income coefficient is significant at the 5% level in the GP count model, the magnitude of the implied impact is small – a 10% increase in equivalised income is estimated to cut utilisation only by about 1%.

Table 2: Estimated impacts of personal characteristics on expected resource utilization counts

Base sample and variation	Mean base count	% change in mean count	Base proportion positive	% change in positive count
GP consultations				
High allostatic load - 1σ	3.12	-18.9	0.75	-5.4
Age 16-29 + 10 years	1.94	-5.1	0.62	0.5
Age 30-44 + 10 years	2.01	-2.4	0.65	-0.1
Age 45-59 + 10 years	2.12	1.4	0.67	0.7
Age 60-74 + 10 years	2.51	4.1	0.70	1.1
75 and over + 10 years	2.97	5.2	0.74	1.1
Males → female	1.80	26.9	0.62	10.0
Females → male	2.45	-23.2	0.72	-9.1
No qualifications → intermediate	2.70	-7.4	0.72	-2.1
Intermediate → degree	2.13	-1.6	0.66	-0.5
All incomes + 10%	2.16	-1.3	0.67	-0.4
OP consultations				
High allostatic load - 1σ	1.80	-12.4	0.46	-4.9
Age 16-29 + 10 years	1.03	-3.6	0.36	-1.0
Age 30-44 + 10 years	1.05	5.1	0.37	2.8
Age 45-59 + 10 years	1.18	16.7	0.39	6.9
Age 60-74 + 10 years	1.63	37.4	0.44	10.2
75 and over + 10 years	2.58	62.0	0.51	9.7
Males → female	1.04	32.3	0.35	17.5
Females → male	1.42	-28.5	0.42	-16.5
No qualifications → intermediate	1.53	7.1	0.43	2.5
Intermediate → degree	1.28	-11.6	0.40	-4.9
All incomes +10%	1.25	-1.0	0.39	-0.4
IP days				
High allostatic load - 1σ	0.94	-43.5	0.07	-15.4
Age 16-29 + 10 years	0.21	-22.6	0.04	-7.7
Age 30-44 + 10 years	0.18	8.9	0.04	2.9
Age 45-59 + 10 years	0.27	17.8	0.04	9.0
Age 60-74 + 10 years	0.56	70.8	0.06	16.2
75 and over + 10 years	1.94	253.5	0.08	17.4
Males → female	0.32	-1.1	0.04	10.5
Females → male	0.35	-7.1	0.05	-10.8
No qualifications → intermediate	1.10	-57.8	0.07	-24.0
Intermediate → degree	0.29	-54.5	0.05	-28.4
All incomes +10%	0.34	-1.5	0.05	-0.3

5 Impacts on costs

Our procedure for inferring costs necessarily differs between the GP, OP and IP resource types because of differences in the detail available from NHS reference cost statistics. In Britain, GPs are independent contractors to the NHS and there is consequently less detailed administrative data relating to the treatment profile of their caseloads and the corresponding costs than there is for hospital treatments. For GP consultations we have used a single average unit cost figure of $\bar{c}^{GP} = £66.20$ per consultation (Curtis and Burns, 2017). To exploit this unit cost figure, we assume that the unobserved true individual-specific average cost of a GP consultation may vary between individuals, but is uncorrelated with the number of consultations, conditional on personal characteristics, implying that the conditional expected cost incurred for individual i is:

$$E\left(C_i^{GP}|\boldsymbol{X}_i^{GP}\right) = \bar{c}^{GP} E\left(Y_i^{GP}|\boldsymbol{X}_i^{GP}\right) \tag{4}$$

For OP cases in each treatment category j, there is a unit cost $\bar{c}^{OP}(j)$ and aggregate number of treatment episodes $n^{OP}(j)$, from which category proportions can be constructed as $\pi^{OP}(j) = n^{OP}(j)/\sum n^{OP}$. By the same reasoning as before, we arrive at a conclusion that $E\left(C_i^{OP}|\boldsymbol{X}_{ir}^{OP}\right) = \bar{c}^{OP} E\left(Y_i^{OP}|\boldsymbol{X}_i^{OP}\right)$, where $\bar{c}^{OP} = \sum_j \pi^{OP}(j)\bar{c}^{OP}(j)$ and $\boldsymbol{X}_{ir}^{OP}$ is the covariate vector in the OP consultation model, constructed at replication r using Y_{ir}^{GP}.

Table 3 gives the results for variations in personal characteristics. Expected GP costs are particularly high for the group with high allostatic load, the over-75 age group and the group with no educational qualifications. The largest proportional impacts are for gender, where women incur roughly 25% higher costs than men, after controlling for other characteristics; and for baseline allostatic load, where a 1σ reduction for the high-allostatic load group would reduce GP costs by almost one-fifth. Mean OP costs are uniformly higher than GP costs across the set of baseline population groups and are proportionately much more responsive

to variations in most personal characteristics, especially gender and ageing within the older population.

Table 3: Predicted mean impacts of personal characteristics on annual GP and OP costs

	GP consultations		OP consultations	
Base sample and hypothetical variation	Mean base cost	% change in cost	Mean base cost	% change in cost
High allostatic load - 1σ	£207	-18.9	£295	-12.3
Age 16-29 + 10 years	£128	-5.1	£168	-3.5
Age 30-44 + 10 years	£133	-2.4	£173	5.2
Age 45-59 + 10 years	£140	1.4	£194	16.8
Age 60-74 + 10 years	£166	4.1	£266	37.2
75 and over + 10 years	£197	5.2	£420	62.4
Males → female	£119	26.9	£171	32.6
Females → male	£162	-23.2	£231	-28.5
No qualifications → intermediate	£179	-7.4	£251	7.1
Intermediate → degree	£141	-1.6	£208	-11.6
All incomes + 10%	£143	-1.3	£205	-1.0

For IP cases, we have much richer cost and caseload information (section 2.3). For each treatment category, we observe caseload broken down by age group and (separately) by gender. We also observe average unit cost and upper and upper and lower quartiles of unit cost for normal length episodes. Treatment categories are further separated into elective (E), non-elective long stay (NEL) and non-elective short stay defined as 2 days or less (NES). We treat these types as separate categories, exploiting the fact that caseload, unit cost and mean length of stay (but not other episode characteristics) are broken down by type of episode.

We follow NHS reporting practices which report episode unit costs for durations within a specified limit ("trim point") and a lower unit cost for "excess stays" – the part of any episode beyond the trim point. So, for the jth treatment category, the episode-specific cost function is:

$$c_j(Y) = \theta_{1j} \min(Y, T_j) + \theta_{2j} \max(0, T_j - Y) \tag{5}$$

where T_j is the trim point, θ_{1j} is the per diem unit cost for "inlier" episodes completed within the normal time and θ_{2j} is the per diem unit cost for excess days.

To incorporate the unit cost information, in each replication of the Monte Carlo simulation outlined in section 4, we construct an individual-specific probability of each treatment type, conditional on the simulated treatment duration and observed characteristics of each individual. Those probabilities are then used to calculate the conditional expected treatment cost, which is then averaged over the 250 Monte Carlo replications. The procedure is necessarily complex and is set out in detail in Appendix 2.

The first panel of Table 4 summarises two alternative cost estimates. The first uses only the simulated IP duration to tailor treatment type probabilities to individuals; the second uses duration, age group and gender to tailor the treatment probabilities. Perhaps surprisingly, the use of demographic information changes the simulated costs rather little.

The predictive power of allostatic load is again clear. The subgroup with allostatic load more than 1σ above the mean are predicted to generate a mean total cost of just over £320 five years later (compared to a mean prediction of approximately £175 for the whole sample). If allostatic load were hypothetically reduced by 1σ for each member of this group, the implications would be a reduction of almost a quarter in their future IP costs, so there is clear scope for a hypothetical effectively-targeted intervention to reduce health care costs significantly.

As anticipated, the influence of age on cost is very strong: simulated average IP cost for people aged over 75 is approximately £450, over three times the per capita cost generated by the 16-29 age group. Moreover, these costs increase steeply with further ageing: adding 10 years to the age of each person in the older group is predicted to increase their average IP costs by over 50%. Gender differences in IP costs are modest – less than half the size of corresponding differences in GP and OP costs (Table 3).

The SES gradient in IP costs differs from the gradients evident in GP and OP costs. Although we again find no significant evidence of an income gradient, the cost differences between classes defined by educational attainment are much larger. The simulated increase in

IP costs caused by hypothetically raising attainment from the unqualified to the intermediate level or from intermediate to degree level is around 35%.

The second panel of Table 4 combines the results for GP, OP and IP costs to give a picture of the overall influences on total direct treatment costs and confirms the general picture of major impacts of allostatic load, demographic differences and education-related SES.

Table 4: Predicted mean impacts of personal characteristics on annual expected IP costs and on total costs (GP+OP+IP)

	Statistical allocation on ...			
	duration only		duration+age+gender	
Base group and hypothetical variation	Mean base cost	% change in cost	Mean base cost	% change in cost
Conditional mean IP costs				
High allostatic load - 1σ	£321	-23.2	£323	-23.4
Age 16-29 + 10 years	£142	-17.4	£136	-15.9
Age 30-44 + 10 years	£134	2.4	£125	7.9
Age 45-59 + 10 years	£163	11.7	£162	16.3
Age 60-74 + 10 years	£243	30.4	£258	28.3
75 and over + 10 years	£450	55.6	£442	52.4
Males → female	£160	10.0	£163	6.1
Females → male	£187	-11.3	£182	-8.2
No qualifications → intermediate	£354	-37.1	£358	-35.3
Intermediate → degree	£179	-36.6	£175	-34.5
All incomes + 10%	£175	-0.8	£174	-0.1
Conditional mean total (GP+OP+IP) costs				
High allostatic load - 1σ	£823	-18.2	£825	-18.3
Age 16-29 + 10 years	£438	-8.5	£432	-7.9
Age 30-44 + 10 years	£439	2.1	£430	3.6
Age 45-59 + 10 years	£496	10.8	£495	12.2
Age 60-74 + 10 years	£675	26.6	£690	25.9
75 and over + 10 years	£1,067	49.0	£1,059	47.6
Males → female	£450	23.1	£453	21.6
Females → male	£581	-21.5	£576	-20.6
No qualifications → intermediate	£784	-16.2	£788	-15.5
Intermediate → degree	£529	-17.4	£524	-16.5
All incomes + 10%	£523	-1.0	£522	-0.8

6 Limitations and robustness

Our analysis has significant limitations, some of which are inherent in any research in this area. Like any survey-based analysis, our estimates are subject to possible distortion from various types of general and item non-response, particularly related to the biomarker data used to measure baseline health objectively. Moreover, any attempt to attach costs to health-care utilisation involves accounting and recording errors inherent in the available reference cost data, which are in any case averages across groups of cases rather than true individual-specific costs and exclude some elements of medical costs (such as most community health services).

Our methodology of statistical cost allocation rests on assumptions that seem strong at first sight, although we would argue that they are more innocent than they appear. In estimating IP costs, we assume each individual's reported number of days in hospital stem from a single episode. This is certainly true for the majority of individuals. The expected total cost over multiple episodes is the sum of the expected cost of each so multiple episodes have no inherent impact on expected total cost. However, we use duration in our calculation of the individual-specific probabilities of alternative disease/treatment types, so multiple episodes do have a modest indirect impact via the quality of cost allocation. We also assume that different patient characteristics (age, gender etc) are distributed independently across patients conditional on condition/treatment type, which is is far less stringent than assuming full independence. We have found that incorporating age and gender information into the cost simulation makes rather little difference to the cost results in Table 4, which suggests that any breaches of the conditional independence assumption may be of little importance.

The outcomes that we study are limited. We look at healthcare utilisation five years after the baseline as a single snapshot, rather than a long-term sequence of outcomes, and we can say nothing directly about the duration of those impacts on the public healthcare system. Our cost analysis is a distributional analysis in the sense that it assigns cost to the

individual whose ill-health generates the need for treatment. That analysis contributes to our understanding of the processes leading to escalating health costs, but it does not tell us about the distribution of the financial burden of those costs across the population, which depends on the redistributive nature of the tax system used to fund public healthcare costs.

Two further limitations are amenable to simple robustness checks. First, we have used English NHS reference costs to the whole sample, which includes individuals resident in Wales and Scotland. We now investigate robustness by repeating the model estimation and cost analysis on a sample restricted to English residents (93% of the full GB sample). Second, in constructing our allostatic load health measure, we ignored the fact that someone with biomarkers maintained at clinically low levels using medication may have an underlying health state different to someone with similar biomarker levels maintained naturally. We now repeat the analysis, excluding individuals who, at baseline, were taking statins or medication for high blood pressure, cardiovascular conditions, diabetes or respiratory conditions. This effectively tightens the criterion we used for limiting the sample to those apparently in good health, and reduces the sample size by 9%. Table 5 summarises the result of re-running the simulation of GP, OP and IP costs using the full and restricted samples in each case (parameter estimates of the count data models are given in Appendix Tables A3 and A4).

The England-only results are very close to those obtained for the full GB sample, so our earlier results are geographically robust for England – and also for Scotland and Wales if the reference treatment unit costs for England are representative of NHS Scotland and NHS Wales also.

The differences are larger when we restrict the sample to those not taking medications at baseline. There are two aspects of these changes: the impact of restricting the sample on the estimated parameters of the count data models; and the change in membership of the population groups. Comparison of Tables 1 and A4 confirms that the impact on the estimated model parameters is very slight, so the differences in Table 5 are primarily the

result of changes in the size and composition of the population groups considered. The presence of respondents who report no diagnosed conditions despite taking medications is itself the result of two factors: under-reporting of diagnosed conditions and the prescribing of medication without a formal diagnosis.[16]

As Table 5 shows, for most population groups the effect of the exclusion is to reduce the expected average cost, by a large amount in some cases: 38% for people with no educational qualifications, 17% for those with high baseline allostatic load and 13% for the oldest group. In two important cases, the sample exclusion makes a major difference to the estimated impact of changing the group's defining characteristic. For those with no qualifications, the impact of adding an intermediate-level qualification changes from -15% to +28%, which indicates that taking prescribed medication can act as a strong indicator of the SES gradient at low educational levels (see also Powdthavee (2010)). For the older group, the sample exclusion greatly increases the estimated impact of ageing, with the impact of a 10-year increase in age changing from 48% to 71%.[17]

[16]The latter is particularly common for statins which are often prescribed as a prophylactic measure rather than treatment for a specific condition – 41% of those excluded because of medication were taking statins only.

[17]The no-qualification, age 60-74 and over-75 groups had the greatest proportions of people on medication at baseline: 21%, 22% and 37% respectively, compared to 9% for the sample as a whole.

Table 5: Robustness checks: alternative estimates of total costs (GP+OP+IP)

	Full sample		England only		No medication	
Base sample and variation	Mean base cost	change in mean cost	Mean base cost	change in mean cost	Mean base cost	change in mean cost
High allostatic load - 1σ	£825	-18.3%	£832	-14.6%	£689	-15.7%
Age 16-29 + 10 years	£432	-7.9%	£446	-8.5%	£445	-12.5%
Age 30-44 + 10 years	£430	3.6%	£432	3.4%	£399	3.4%
Age 45-59 + 10 years	£495	12.2%	£496	11.5%	£444	17.6%
Age 60-74 + 10 years	£690	25.9%	£684	25.9%	£606	39.9%
75 and over + 10 years	£1,059	47.6%	£1,079	44.2%	£942	71.2%
Males → female	£453	21.6%	£399	25.2%	£399	25.2%
Females → male	£576	-20.6%	£516	-23.4%	£516	-23.4%
No qualifications → intermediate	£788	-15.5%	£765	-13.1%	£487	28.2%
Intermediate → degree	£524	-16.5%	£528	-14.8%	£518	-19.4%
All incomes +10%	£522	-0.8%	£524	-0.9%	£465	-0.5%

7 Discussion and conclusions

In this paper we have adopted a forward-looking approach to explore the predictive power of biomarkers and other personal characteristics on the utilisation of health services and associated costs, five years later. To the best of our knowledge, it is the first analysis of its kind. Using data from UKHLS on a group of individuals with no history of diagnosed health conditions, we find that a biomarker-based approximation to allostatic load that reflects pre-diagnosed and pre-symptomatic pathways (Goldman et al., 2006; Geronimus et al., 2006; Turner et al., 2016; Seeman et al., 2004) is a powerful predictor of future burden for the GP and hospital healthcare systems, both in terms of service utilisation and healthcare costs.

Our analysis also indicates the demographic and SES characteristics most strongly associated with future treatment costs and this has highlighted some interesting aspects. There is a large gender difference for GP and OP utilisation levels and costs of around 25% between women and men. The finding of gender differences in primary healthcare utilisation is not new (Wang et al., 2013). However, we have also found that the gender difference is negligible

for IP costs, suggesting that GP and OP resources are driven more by demand and less by clinicians' assessment of need than is the case for more serious IP treatments. We offer no judgements about whether men's demand for primary care is too low and women's too high in any sense but, if attempts to encourage men to use primary care services to the same extent as women (as recommended by WHO (2018)) were successful, our results suggest that there would be a very large increase in healthcare costs as a consequence. That would need to be incorporated into the resourcing of interventions targeted at men.

Another interesting feature of our estimates is the combination of a large education-related gradient and quantitatively negligible gradient with respect to current income. This suggests that the SES gradient in treatment costs might have its basis in long-run human capital accumulation or social norms linked to social class, rather than current access to economic resources. That would in turn suggest that redistributive policy would have quite limited impact on treatment costs in the short to medium term, and that more far-reaching social reform than simple income redistribution might be required to address the SES gradient in healthcare need. This is consistent with some recent evidence relating to some specific healthcare services (Terraneo, 2015; Labeit and Peinemann, 2017), but our finding of a large difference between the impact of education on primary care and on hospital inpatient care appears to be new.

The predictive power of personal characteristics and biomarker-based health measures gives a possible basis for sophisticated tailoring preventive interventions. A measure similar to our allostatic load proxy could be constructed from information gathered in the NHS England Health Check introduced in 2009, which offers quinquennial check-ups including blood tests. There are concerns about low low take-up (Robson et al., 2016), which is a potential obstacle for any such preventive measure. However, the NHS Health Check is available to all adults aged 40-74 and thus targeted only on age in a simple way and our findings suggest that more tailored targeting could identify better the population groups with highest potential future healthcare needs and costs. Of course, for this to be worthwhile,

there needs to be an effective follow-up intervention that can be used to reduce those future costs to the individuals themselves and wider society. The evidence so far on the ability of the NHS Health Check to prompt change to reduce disease risks (Chang et al., 2016; Hinde et al., 2017) is controversial, but it is possible that a more assertive and better-targeted intervention might be cost-effective.

There is a continuing debate on capitation-based payments that are currently used to allocate budgets to GPs, and its possible extension as the future payment system for two models of care by NHS England (Brilleman et al., 2014; Shepherd, 2017). A potential policy application of our findings is in refining the design of capitation payment systems by re-orienting the capitation formula to match more closely patient level morbidity data and other demographic and SES characteristics (Shepherd, 2017). This offers the prospect of improved allocation of resources as well as health outcomes by reducing incentives for health providers to "cream skim" the patient population by selecting patient groups with lower expected future healthcare costs.

In addition to these substantive contributions, we have also made some new methodological developments. Our econometric procedure has extended the standard zero-inflated negative binomial regression model to allow estimation from count data reported by survey respondents in interval form, rather than as exact counts. The accompanying software (Pudney, 2018) makes this technique widely available.

We have also developed a simple simulation-based method of assigning costs to the service utilisation levels predicted by our count data models. The matching of administrative records such as HES data to survey data is often seen as an ideal solution to the difficult problem of analysing utilisation and costs at the individual level. However, suitable matched datasets are not currently available and, in any case, low consent rates mean that matching introduces additional assumptions which may be questionable.[18] We argue that

[18] Alternative assumption-free partial identification methods of dealing with non-consent and missing data typically yield rather uninformative results (Manski, 2003).

other approaches resting on alternative assumptions should be used in parallel as part of a robust research picture. For inpatient treatment episodes, we have developed a method of statistical allocation of costs that assigns multiple treatment types, with corresponding individual-specific probabilities to each survey respondent, using reported duration and demographic characteristics to presonalise the assignment. This can be done for the whole sample rather than the subgroup of respondents who consent to data matching.

References

Bago d'Uva, T. and Jones, A. M. (2009). Health care utilisation in europe: new evidence from the echp. *Journal of Health Economics*, 28(2):265–279.

Barber, R. M., Fullman, N., Sorensen, R. J., Bollyky, T., McKee, M., Nolte, E., Abajobir, A. A., Abate, K. H., Abbafati, C., Abbas, K. M., et al. (2017). Healthcare Access and Quality Index based on mortality from causes amenable to personal health care in 195 countries and territories, 1990–2015: a novel analysis from the Global Burden of Disease Study 2015. *The Lancet*, 390(10091):231–266.

Bound, J., Brown, C., and Mathiowetz, N. (2001). Measurement error in survey data. In Heckman, J. and Leamer, E., editors, *Handbook of Econometrics. Vol. 5*, pages 3705–3843. Elsevier, Amsterdam.

Brilleman, S. L., Gravelle, H., Hollinghurst, S., Purdy, S., Salisbury, C., and Windmeijer, F. (2014). Keep it simple? predicting primary health care costs with clinical morbidity measures. *Journal of Health Economics*, 35:109–122.

Cameron, A. C. and Trivedi, P. K. (2013). *Regression Analysis of Count Data (2nd ed.)*. Cambridge University Press, Cambridge, UK.

Carreras, M., Ibern, P., and Inoriza, J. M. (2018). Ageing and healthcare expenditures: Exploring the role of individual health status. *Health economics*, 27(5):865–876.

Carrieri, V. and Jones, A. M. (2017). The income-health relationship beyond the mean: New evidence from biomarkers. *Health Economics*, 26(7):937–956.

Cawley, J. and Meyerhoefer, C. (2012). The medical care costs of obesity: an instrumental variables approach. *Journal of health economics*, 31(1):219–230.

Chang, K. C.-M., Lee, J. T., Vamos, E. P., Soljak, M., Johnston, D., Khunti, K., Majeed, A., and Millett, C. (2016). Impact of the National Health Service Health Check on cardiovascular disease risk: a difference-in-differences matching analysis. *Canadian Medical Association Journal*, pages cmaj–151201.

Charlesworth, A. and Bloor, K. (2018). 70 years of NHS funding: how do we know how much is enough? *BMJ*, 361:k2373.

Curtis, L. A. and Burns, A. (2017). Unit Costs of Health and Social Care 2017. https://doi.org/10.22024/UniKent/01.02/65559.

Davillas, A. and Pudney, S. (2017). Concordance of health states in couples: analysis of self-reported, nurse administered and blood-based biomarker data in the UK Understanding Society panel. *Journal of Health Economics*, 56:87–102.

Davillas, A. and Pudney, S. E. (2018). Biomarkers as precursors of disability. *Institute for Social and Economic Research*.

DOH (2016). *Reference costs guidance 2015-16*. Department of Health, Department of Health, London. https://www.gov.uk/government/publications/nhs-reference-costs-collection-guidance-for-2015-to-2016.

Doorslaer, E. v., Koolman, X., and Jones, A. M. (2004). Explaining income-related inequalities in doctor utilisation in europe. *Health Economics*, 13(7):629–647.

Geronimus, A. T., Hicken, M., Keene, D., and Bound, J. (2006). weathering and age patterns of allostatic load scores among blacks and whites in the United States. *American Journal of Public Health*, 96(5):826–833.

Goldman, N., Turra, C. M., Glei, D. A., Seplaki, C. L., Lin, Y.-H., and Weinstein, M. (2006). Predicting mortality from clinical and nonclinical biomarkers. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, 61(10):1070–1074.

Haider, A. W., Larson, M. G., Franklin, S. S., and Levy, D. (2003). Systolic blood pressure, diastolic blood pressure, and pulse pressure as predictors of risk for congestive heart failure in the Framingham Heart Study. *Annals of Internal Medicine*, 138(1):10–16.

Hinde, S., Bojke, L., Richardson, G., Retat, L., and Webber, L. (2017). The cost-effectiveness of population health checks: have the NHS Health Checks been unfairly maligned? *Journal of Public Health*, 25(4):425–431.

Hobbs, F. D. R., Bankhead, C., Mukhtar, T., Stevens, S., Perera-Salazar, R., Holt, T., and Salisbury, C. (2016). Clinical workload in UK primary care: a retrospective analysis of 100 million consultations in England, 2007-14. *The Lancet*, 387:2323–2330.

Howard, J. T. and Sparks, P. J. (2016). Does allostatic load calculation method matter? evaluation of different methods and individual biomarkers functioning by race/ethnicity and educational level. *American Journal of Human Biology*, 28(5):627–635.

ISD Scotland (2017). *Scottish health service costs, year ended 31 March 2017*. NHS National Services, Edinburgh. http://www.isdscotland.org/Health-Topics/Finance/Costs/Detailed-Tables.

Jäckle, A., Beninger, K., Burton, J., and Couper, M. P. (2018). Understanding data linkage consent in longitudinal surveys. University of Essex: *Understanding Society* Working Paper 2018-07.

Labeit, A. M. and Peinemann, F. (2017). Determinants of a GP visit and cervical cancer screening examination in Great Britain. *PloS one*, 12(4):e0174363.

Manski, C. F. (2003). *Partial Identification of Probability Distributions*. Springer-Verlag, New York.

Morciano, M., Hancock, R. M., and Pudney, S. E. (2015). Birth-cohort trends in older-age functional disability and their relationship with socio-economic status: evidence from a pooling of repeated cross-sectional population-based studies for the UK. *Social Science and Medicine*, 136:1–9.

NHS Digital (2017). *Hospital Admitted Patient Care Activity 2016-17*. London. https://digital.nhs.uk/data-and-information/publications/statistical/hospital-admitted-patient-care-activity/2016-17.

NHS Improvement (2017). *Reference costs 2016/17: highlights, analysis and introduction to the data*. NHS Improvement, London. https://improvement.nhs.uk/resources/reference-costs/.

OECD (2015). *Fiscal sustainability of health systems: bridging health and finance perspectives*. Organisation for Economic Co-operation and Development.

Pearson, T. A., Mensah, G. A., Alexander, R. W., Anderson, J. L., Cannon III, R. O., Criqui, M., Fadl, Y. Y., Fortmann, S. P., Hong, Y., Myers, G. L., et al. (2003). Markers of inflammation and cardiovascular disease: application to clinical and public health practice: a statement for healthcare professionals from the Centers for Disease Control and Prevention and the American Heart Association. *Circulation*, 107(3):499–511.

Powdthavee, N. (2010). Does education reduce the risk of hypertension? estimating the biomarker effect of compulsory schooling in England. *Journal of Human Capital*, 4:173–202.

Pudney, S. E. (2018). IntCount: a Stata command for estimating count data models from interval data. University of Essex: *Understanding Society* Working Paper 2018-08.

Robson, J., Dostal, I., Sheikh, A., Eldridge, S., Madurasinghe, V., Griffiths, C., Coupland, C., and Hippisley-Cox, J. (2016). The NHS Health Check in England: an evaluation of the first 4 years. *BMJ open*, 6(1):e008840.

Sari, N. (2009). Physical inactivity and its impact on healthcare utilization. *Health Economics*, 18(8):885–901.

Seeman, T. E., Crimmins, E., Huang, M.-H., Singer, B., Bucur, A., Gruenewald, T., Berkman, L. F., and Reuben, D. B. (2004). Cumulative biological risk and socio-economic differences in mortality: Macarthur studies of successful aging. *Social Science and Medicine*, 58(10):1985–1997.

Shepherd, D. (2017). Capitation based funding of general practice is not fit for purpose. *BMJ*, 358:j4075.

Sturm, R. (2002). The effects of obesity, smoking, and drinking on medical problems and costs. *Health Affairs*, 21(2):245–253.

Terraneo, M. (2015). Inequities in health care utilization by people aged 50+: evidence from 12 European countries. *Social Science and Medicine*, 126:154–163.

Traczynski, J. and Udalova, V. (2018). Nurse practitioner independence, health care utilization, and health outcomes. *Journal of Health Economics*, 58:90–109.

Turner, R. J., Thomas, C. S., and Brown, T. H. (2016). Childhood adversity and adult health: Evaluating intervening mechanisms. *Social Science and Medicine*, 156:114–124.

Vie, T. L., Hufthammer, K. O., Holmen, T. L., Meland, E., and Breidablik, H. J. (2014). Is self-rated health a stable and predictive factor for allostatic load in early adulthood? Findings from the Nord Trøndelag Health Study (HUNT). *Social Science and Medicine*, 117:1–9.

Wang, Y., Hunt, K., Nazareth, I., Freemantle, N., and Petersen, I. (2013). Do men consult less than women? an analysis of routinely collected UK general practice data. *BMJ open*, 3(8):e003320.

WHO (2018). Strategy on the health and well-being of men in the WHO European Region.

Appendix 1: Additional Figures and Tables

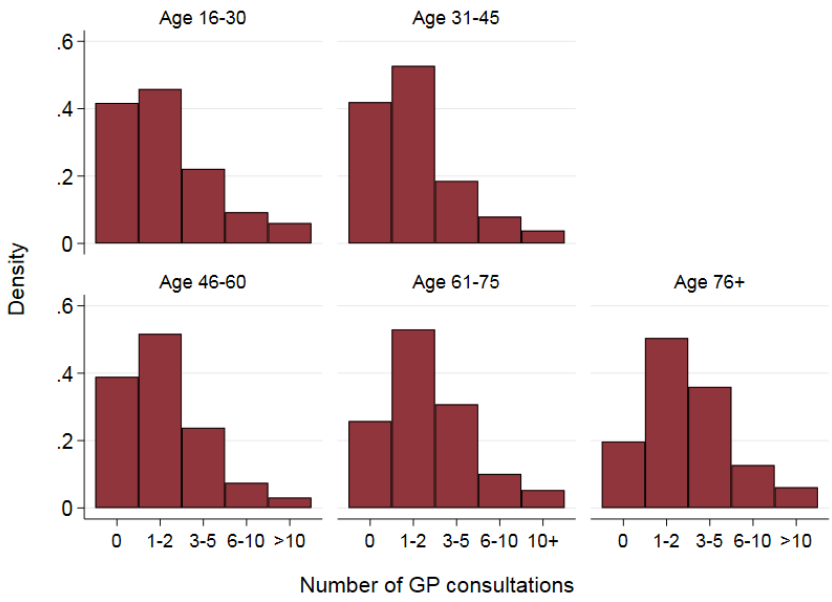

Figure A1: Distribution of the number of GP consultations in the preceding 12 months by age

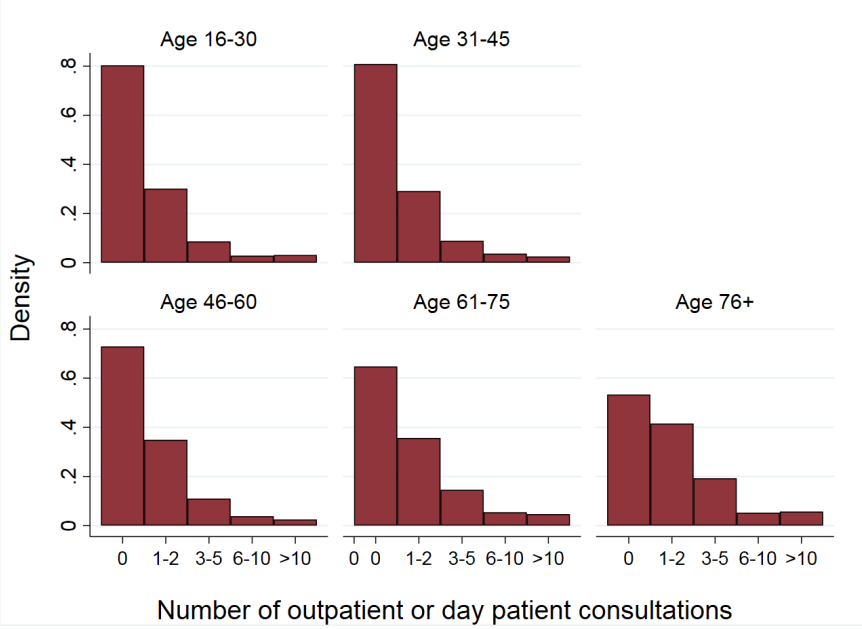

Figure A2: Distribution of the number of OP consultations in the preceding 12 months by age

Table A1: Comparison of GP utilisation counts in UKHLS with administrative data

	Men			Women		
	Administrative data	UKHLS data		Administrative data	UKHLS data	
	mean	mean	95% CI	mean	mean	95% CI
			England			
Age15-24	1.362	1.339	1.244; 1.435	2.965	2.783	2.632; 2.935
Age 25-44	1.772	1.688	1.608; 1.768	3.535	3.269	3.156; 3.383
Age 45-64	2.840	2.639	2.534; 2.742	3.976	3.406	3.294; 3.518
Age 65-74	4.529	3.523	3.334; 3.711	5.034	3.528	3.353; 3.703
Age 75+	6.245	4.423	4.146; 4.700	6.521	4.336	4.091; 4.581
Total (age 15+)	2.717	2.451	2.383; 2.518	4.045	3.373	3.309; 3.436
			Scotland			
15-24	1.428	1.445	1.094; 1.797	3.228	2.814	2.292; 3.335
25-34	1.612	1.901	1.252; 2.550	3.373	3.259	2.510; 4.008
35-44	1.942	1.540	1.126; 1.952	3.418	3.493	2.863; 4.123
45-54	2.207	2.253	1.809; 2.698	3.467	3.135	2.605; 3.664
55-64	2.871	3.102	2.572; 3.633	3.731	3.194	2.747; 3.643
65-74	3.719	3.307	2.703; 3.913	4.228	3.600	3.014; 4.179
75+	5.329	4.352	3.361; 5.344	5.502	3.613	2.836; 4.390
Total (age 15+)	2.465	2.585	2.363; 2.806	3.768	3.293	3.067; 3.518

The aggregate administrative data by age groups and gender for England, based on the most recent data from the Clinical Practice Research Datalink (2013-2014 period), are extracted from Hobbs et al. (2016). These rates are increased by 1.3 per cent (an annualised increase based on a linear trend derived by GP utilisation data for the 2007/2008 to 2013/14 period) to project the corresponding GP consultations rates for the 2016/2017 period (to be comparable to the UKHLS wave 7 time period). The GP consultation rates for Scotland are administrative data for 2017 (ISD Scotland, 2017). UKHLS means are based on imputations from grouped NB models estimated separately for England and Scotland using the UKHLS wave 7 data.

Table A2: Comparison of OP and IP utilisation counts in UKHLS with administrative data

	Administrative data	UKHLS mean	95% CI
Outpatient and day-patient cases			
England (age 15+)	1.667	1.600	1.561; 1.638
Scotland (age 15+)	1.726	1.520	1.404; 1.636
Number of inpatient days			
England (age 15+)	0.644	0.706	0.638; 0.775

OP rates are based on administrative data, extracted from national schedules of reference costs NHS Improvement (2017) and Scottish Health Service Costs (ISD Scotland, 2017) for England and Scotland, respectively. Administrative data on IP days for England are extracted from NHS Digital (2017). UKHLS means for OP consultations are based on imputations from grouped NB models estimated separately for England and Scotland using the UKHLS wave 7 data.

Table A3: Estimated parameters for grouped negative binomial models of GP, OP and IP utilisation: English sub-sample

		OP model		IP model	
Parameter	GP	marginal	conditional	marginal	conditional
Allostatic load	0.213***	0.125**	-0.014	0.544***	0.261
	(0.038)	(0.055)	(0.071)	(0.167)	(0.181)
age50	-0.046	0.050	0.059	0.269*	0.158
	(0.031)	(0.047)	(0.048)	(0.159)	(0.143)
agesq	0.019	0.025	0.007	-0.019	-0.050
	(0.017)	(0.026)	(0.028)	(0.083)	(0.076)
agemale	0.133***	0.178**	0.143*	-0.444**	-0.237
	(0.045)	(0.071)	(0.075)	(0.220)	(0.203)
agesqmale	-0.028	0.021	0.061	0.239*	0.258**
	(0.026)	(0.039)	(0.042)	(0.129)	(0.115)
male	-0.205***	-0.348***	-0.341***	-0.560	-0.712*
	(0.079)	(0.125)	(0.122)	(0.407)	(0.386)
intermed	-0.004	0.070	0.107	0.502	0.718**
	(0.068)	(0.106)	(0.106)	(0.326)	(0.304)
noqual	0.052	-0.051	-0.054	0.703	1.442**
	(0.094)	(0.149)	(0.154)	(0.576)	(0.718)
lnincome	-0.147***	-0.117*	0.006	-0.182	0.104
	(0.047)	(0.064)	(0.062)	(0.232)	(0.204)
GPcat2			0.811***		-1.745**
			(0.151)		(0.743)
GPcat3			1.571***		-1.439*
			(0.156)		(0.764)
GPcat4			1.994***		-0.471
			(0.171)		(0.822)
GPcat5			2.974***		1.097
			(0.243)		(0.787)
cons	1.859***	1.108**	-0.927*	-0.336	-1.206
	(0.363)	(0.495)	(0.476)	(1.836)	(1.671)
$ln(\alpha)$	0.017	1.076***	0.669***	3.858***	3.278***
	(0.058)	(0.060)	(0.084)	(0.125)	(0.156)
GP0					14.095***
					(0.749)
cons					-11.783***
					(0.558)
N	2144.000	2144.000	2144.000	2144.000	2144.000
AIC	5610.102	4574.433	4247.375	1410.677	1347.004
BIC	5672.476	4636.808	4332.432	1473.052	1443.402

§ Age measured in decades from an origin of 50. Standard errors in parentheses. Sample size $N = 2314$.
Statistical significance: * = 10%, ** = 5%, *** = 1%.

Table A4: Estimated parameters for grouped negative binomial models of GP, OP and IP utilisation: No medications at baseline

		OP model		IP model	
Parameter	GP	marginal	conditional	marginal	conditional
Allostatic load	0.214***	0.149***	-0.002	0.472***	-0.062
	(0.039)	(0.057)	(0.073)	(0.159)	(0.197)
age50	-0.081**	0.024	0.065	0.376**	0.428***
	(0.034)	(0.051)	(0.052)	(0.157)	(0.159)
agesq	0.010	0.025	0.018	0.026	0.035
	(0.018)	(0.028)	(0.030)	(0.084)	(0.085)
male	-0.218***	-0.346***	-0.333***	-0.272	-0.239
	(0.081)	(0.129)	(0.125)	(0.356)	(0.378)
agemale	0.166***	0.212***	0.151*	-0.408*	-0.311
	(0.047)	(0.078)	(0.080)	(0.220)	(0.236)
agesqmale	-0.020	0.037	0.066	0.194	0.192
	(0.027)	(0.043)	(0.045)	(0.124)	(0.121)
noqual	0.052	-0.056	-0.088	-0.572	-0.200
	(0.102)	(0.154)	(0.159)	(0.471)	(0.451)
intermed	0.014	0.086	0.118	0.723**	1.030***
	(0.069)	(0.110)	(0.110)	(0.348)	(0.330)
lnincome	-0.142***	-0.105	0.031	-0.183	0.393**
	(0.048)	(0.067)	(0.066)	(0.243)	(0.185)
wales	0.577***	-0.047	-0.619**	-0.238	-2.253***
	(0.216)	(0.424)	(0.315)	(1.012)	(0.759)
scot	0.048	-0.207	-0.350*	0.278	-1.501**
	(0.143)	(0.252)	(0.211)	(0.890)	(0.637)
1.loGP			0.793***		-0.701
			(0.148)		(0.868)
3.loGP			1.533***		-0.253
			(0.155)		(0.896)
6.loGP			1.840***		0.752
			(0.173)		(0.965)
11.loGP			3.188***		3.514***
			(0.253)		(0.925)
cons	1.775***	0.951*	-1.129**	-0.610	-4.954***
	(0.371)	(0.519)	(0.509)	(1.963)	(1.780)
$ln(\alpha)$	0.040	1.123***	0.717***	3.949***	3.132***
	(0.060)	(0.062)	(0.085)	(0.130)	(0.162)
inflate					
GP0					12.896***
					(2.017)
cons					-11.049***
					(1.914)
N	2109.000	2109.000	2109.000	2109.000	2109.000
AIC	5449.543	4361.401	4047.284	1231.593	1150.856
BIC	5523.045	4434.903	4143.402	1305.094	1258.281

§ Age measured in decades from an origin of 50. Standard errors in parentheses. Sample size $N = 2314$. Statistical significance: * = 10%, ** = 5%, *** = 1%.

Appendix 2: Allocation of costs for inpatient days

The aim here is to construct an estimate of the individual's expected treatment cost C, conditional on his or her personal characteristics $\boldsymbol{X}$. There are two cases; the simpler setting is where we have data from administrative records on the full distribution of $\boldsymbol{X}$ conditional on the treatment category j. This is the case if we use only age variables or only gender variables as covariates in the IP count data model, since we the NHS administrative data give separate age and gender breakdowns of unit cost and caseload. A more problematic setting is where $\boldsymbol{X}$ is a vector and we only have univariate marginal distributions of a subset of the variables in X-variable. This is relevant if we use a full range of demographic and SES as covariates in $\boldsymbol{X}$.

Let Y be the episode duration and $c_j(Y)$ be the cost function (5). The probability function $f(Y, j, \boldsymbol{X})$ represents the distribution across IP episodes of treatment type and duration and personal characteristics of the patient and is defined as the probability of a randomly-selected episode being treatment type j for length of stay Y received by a person with characteristics $\boldsymbol{X}$. We also use $f(.)$ as generic notation for any conditional or marginal distribution derived from $f(Y, j, \boldsymbol{X})$, with the type of distribution defined by the argument list. The count data model for duration in days gives a distribution $h(Y|\boldsymbol{X})$ which is conditional on $\boldsymbol{X}$ but not on the treatment type which is unobserved in the UKHLS.

Write $\boldsymbol{X} = (A, G, S)$, where A represents age variables, G is a gender indicator and S is a set of SES descriptors. The difficulty here is that we observe separate treatment-specific marginal distributions $Y|j$, $A|j$ and $G|j$ but not their joint distribution, and we have no information at all on the distribution of S within treatment categories. The simplest way of using this information is to make the conditional independence assumption $Y \perp\!\!\!\perp A \perp\!\!\!\perp G \perp\!\!\!\perp |j$, which implies $f(Y, j, A, G, S) = f(Y|j)f(A|j)f(G|j)f(S|j)f(j)$. This is much less restrictive than full independence, since it allows the duration, demographic and SES composition of patient groups to be completely different in different treatment categories.

Under these assumptions, the conditional expectation of cost is:

$$E(C|Y=y,\boldsymbol{X}) = \frac{\sum_{j=1}^{J} c_j(y)f(y|j)f(A|j)f(G|j)f(S|j)f(j)}{\sum_{j=1}^{J} f(y|j)f(A|j)f(G|j)f(S|j)f(j)} \tag{A1}$$

This structure is not operational, since $f(S|j)$ is unobserved. To overcome this, make the further assumption that the weighted covariance over treatment categories between $c_j(y)$ and $f(S|j)$ is zero for any y and any S:

$$\sum_j w_j c_j(y) f(S|j) - \left(\sum_j w_j c_j(y)\right)\left(\sum_j w_j f(S|j)\right) = 0 \tag{A2}$$

where the weight w_j is:

$$w_j = \frac{f(y|j)f(A|j)f(G|j)}{\sum_j f(y|j)f(A|j)f(G|j)} \tag{A3}$$

Assumptions (A1) and (A2) imply $E(C|Y=y,\boldsymbol{X}) = \sum_j w_j c_j(y) f(S|j) / \sum_j w_j f(S|j) = \sum_j w_j c_j(y)$ and thus $E(C|Y=y,\boldsymbol{X})$ does not depend on S:

$$E(C|Y=y,\boldsymbol{X}) = \frac{\sum_{j=1}^{J} c_j(y)f(y|j)f(A|j)f(G|j)f(j)}{\sum_{j=1}^{J} f(y|j)f(A|j)f(G|j)f(j)} \tag{A4}$$

In other words, expected cost may depend on SES characteristics, but that dependence disappears if we also condition on treatment duration - people with different SES characteristics incur different costs only because they tend to have different IP durations.

Overall expected cost is:

$$E(C|\boldsymbol{X}) = \sum_y h(y|\boldsymbol{X}) \frac{\sum_{j=1}^{J} c_j(y)f(y|j)f(A|j)f(G|j)f(j)}{\sum_{j=1}^{J} f(y|j)f(A|j)f(G|j)f(j)} \tag{A5}$$

Exactly the same approach can be used without exploiting age and gender information, by simply removing the terms $f(A|j)f(G|j)$ from (A5). This allows us to assess the role of demographic characteristics in converting utilisation counts into expected costs.