



This is a repository copy of *Using biomarkers to predict healthcare costs: Evidence from a UK household panel*.

White Rose Research Online URL for this paper:
<https://eprints.whiterose.ac.uk/162718/>

Version: Published Version

Article:

Davillas, A. and Pudney, S. orcid.org/0000-0002-5697-0976 (2020) Using biomarkers to predict healthcare costs: Evidence from a UK household panel. *Journal of Health Economics*, 73. 102356. ISSN 0167-6296

<https://doi.org/10.1016/j.jhealeco.2020.102356>

Reuse

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:
<https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>



Using biomarkers to predict healthcare costs: Evidence from a UK household panel[☆]



Apostolos Davillas^a, Stephen Pudney^{b,*}

^a Health Economics Group, Norwich Medical School, University of East Anglia, Institute for Social and Economic Research, University of Essex, United Kingdom

^b School of Health and Related Research, University of Sheffield, United Kingdom

ARTICLE INFO

Article history:

Received 2 September 2019

Received in revised form 27 June 2020

Accepted 1 July 2020

Available online 2 July 2020

JEL classification:

C3

C8

I10

I18

Keywords:

Healthcare costs

Socioeconomic gradient

Biomarkers

Allostatic load

Understanding society

ABSTRACT

We investigate the extent to which healthcare service utilisation and costs can be predicted from biomarkers, using the UK Understanding Society panel. We use a sample of 2314 adults who reported no history of diagnosed long-lasting health conditions at baseline (2010/11), when biomarkers were collected. Five years later, their GP, outpatient (OP) and inpatient (IP) utilisation was observed. We develop an econometric technique for count data observed within ranges and a method of combining administrative reference cost data with the survey data without exact individual-level matching. Our composite biomarker index (allostatic load) is a powerful predictor of costs: for those with a baseline allostatic load of at least one standard deviation (1-s.d.) above mean, a 1-s.d. reduction reduces GP, OP and IP costs by around 18%.

© 2020 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Health care costs have risen faster than the rate of 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, about 10% of GDP is spent on

the National Health Service (NHS), which is broadly in line with other European countries. That proportion has doubled since the establishment of the NHS in 1948 (Charlesworth and Bloor, 2018). 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 preventive policy.

[☆] We are grateful to the editor and two anonymous reviewers for their valuable comments. 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). The funders, data creators and UK Data Service have no responsibility for the contents of this paper. 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), Understanding Society award no. ES/K005146/1) and the MiSoC research centre (award no. ES/L009153/1). We are grateful to members of the project teams for many helpful comments. Any remaining errors are our sole responsibility.

* Corresponding author at: SCHARR, University of Sheffield, 30 Regent Street, S1 4DA, United Kingdom.

E-mail address: steve.pudney@sheffield.ac.uk (S. Pudney).

Some recent studies have used administrative data from Spain, the UK and the Netherlands (Brilleman et al., 2014; Carreras et al., 2018; De Meijer et al., 2011; Howdon and Rice, 2018) to examine the “red herring” thesis (Zweifel et al., 1999). This suggests that extrapolation of simple age-health expenditure curves may not yield an accurate picture of current and future healthcare expenditures, which are primarily driven by proximity to death rather than age itself, with proximity to death in turn driven by health and disability status. In a similar context, Dalgaard and Strulik (2014), drawing on research in biology and medicine to develop an economic model of aging, argued that age of death is determined by optimal health investments and is relevant to biological aging; chronological aging and death are inevitable but individuals typically invest in their health, which slows down aging and prolongs life.

However, proximity to death is unobserved prior to death and therefore of no practical use for projecting health care expenditure for the needs of health policy. The development of forward-looking policy to control utilisation of healthcare services and the subsequent costs cannot be done solely on the basis of records accumulated by healthcare systems, since those records do not contain information about the full range of personal and socio-economic characteristics that may be required to account for confounding effects and do not have adequate coverage of individuals with latent health conditions that have not yet reached the stage of diagnosis. Analysis aiming to characterise individuals at risk of generating an increased burden on the healthcare system and identify potential cost savings requires data from the general population. Most relevant existing studies (Brilleman et al., 2014; Carreras et al., 2018; De Meijer et al., 2011; Howdon and Rice, 2018), although emphasising the importance of the current morbidity profile of the population for current healthcare spending, have not been able to identify individuals within the ‘apparently healthy’ population, who are nevertheless at risk of high future rates of healthcare utilisation and cost. The ability to do this would allow detailed targeting of health interventions, with the prospect of significant cost savings. A cumulative biological measure of wear and tear on the body may be particularly valuable as a proxy for future health risks.

The potential scope for prevention strategies to reduce health risks and contain healthcare costs has been central to health policy debates worldwide (Cohen et al., 2008) but the evidence for their effectiveness so far is mixed (Russell, 2009). Chernew and Newhouse (2011) argued that better targeting of preventative health services to high-risk groups is needed to ensure substantial cost savings for prevention strategies, otherwise potential cost savings tend to be offset by the high cost of unnecessary additional services offered to those with no need. This argues for more research characterising at-risk population groups – especially those not currently visible to the healthcare system – to target potential interventions effectively.

Using data from the UK Household Panel Study (UKHLS, also known as *Understanding Society*), the key feature of our

empirical modelling is the use of biomarkers to predict subsequent healthcare utilisation. UKHLS collects information on utilisation counts for the numbers of general practitioner (GP) and outpatient/day-patient (OP) consultations, and the number of days spent as a hospital inpatient (IP) within the preceding twelve months. Financial costs provide a natural metric for combining these different aspects of resource utilisation into an overall measure of the burden on the public healthcare system. We use data combination techniques to incorporate administrative data for England on GP, OP and IP utilisation and official reference costs to estimate the excess public costs generated by those with elevated biomarkers at baseline. By excluding individuals who reported any past or recently diagnosed long-lasting health condition, we focus on individuals who appeared, from a clinical point of view, to be healthy at baseline and who would not therefore already be prioritised by the healthcare system. To the best of our knowledge, our analysis is the first of its kind.

Our main contribution is to show the power of biomarker data at baseline for predicting future health service utilisation and concomitant future costs. Use of a long (five-year) prediction horizon is important, since public health initiatives and resource planning are not short-term processes. The biomarkers we use for prediction are more objective health measures than conventional self-reported or self-assessed health (Biomarkers Definitions Working Group, 2001), and several epidemiological studies have argued that biomarkers can predict future health and mortality risks even in the case of individuals without history of ill health (Goldman et al., 2006; Gleit et al., 2014; Zethelius et al., 2008; Wang et al., 2006). Biomarkers may provide direct information on pre-disease pathways, in particular by measuring physiological processes that are below the individual’s threshold of perception (McDade et al., 2007). However, these epidemiological studies have not considered cost implications explicitly and it is possible that the implications of health for future social costs and utilisation of healthcare services are quite different from their implications for any medical measure of health outcome. In this study, we use a measure of cumulative biological risk factors, often called allostatic load, which is an index combining biomarkers relevant to different biological systems, to give an overall assessment of physical health (Davillas and Pudney, 2017; Howard and Sparks, 2016; Seeman et al., 2004). The objective nature of biomarkers is also important for the design of targeted interventions; unlike self-assessed and self-reported health measures, biomarkers cannot normally be manipulated to achieve a desired outcome of the screening process and, thus, they prevent ‘gaming’ of the screening process.

We develop a data combination method to incorporate administrative data on caseload composition and unit costs in circumstances where administrative data matching at the level of the survey respondent is not feasible. Comprehensive individual-level administrative healthcare data are non-existent in many countries and, even where they do exist, there are legal and ethical data security difficulties that typically prevent matching to suitable longitudinal

surveys.¹ A further difficulty is that individual-level matching of survey data and administrative records normally requires informed consent, generating a possible source of bias (Jäckle et al., 2018; Riley, 2009). In our application, we attach average unit costs to GP and OP consultation counts, but use a more sophisticated data combination method for assigning expected costs to the predictions of the IP utilisation model, using individual-level survey information (on age, gender and duration of hospital stays) combined with administrative data aggregated within broad treatment/demographic patient groups. In applying the data combination method, we overcome some difficulties in the survey observation of healthcare utilisation, which yields count variables with highly skewed distributions featuring an excess mass of zero observations (Van Doorslaer et al., 2004) and counts observed within intervals rather than as exact figures. We use a zero-inflated interval-observed negative binomial specification, tailored to the form of the UKHLS data, together with appropriate prediction and simulation methods for conducting post-estimation analysis.²

It is beyond the scope of this paper to assess the cost-effectiveness of any particular screening or prevention strategies, but our results have potential implications for health policy in the UK and beyond. They can be used to indicate priority areas for interventions (such as screening programmes and health education initiatives) to control future treatment costs among individuals who have not yet reached the stage of diagnosis. Our analysis provides general evidence on the pattern of risk of future excess healthcare costs that is relevant to the many countries with healthcare systems similar to the UK (i.e., mostly publicly funded). These results contribute to the policy debate on the effectiveness of prevention strategies (WHO, 2018), and suggest that targeting interventions to reduce future costs may require more regular collection of biomarkers at a wider population level than is usual at present. For example, the NHS England Health Check offers only quinquennial check-ups, including blood tests (Schulein et al., 2017), to the 40–74 age group, and take-up (around 20%) is low. However, similar health checks in Japan are successfully offered annually to full-time employees and the 40–74 age group (OECD, 2019, chapter 3). Of course, implementing a high-frequency, universal check-up programme is costly but it may be cost-effective if substantial future healthcare costs come from the apparently healthy part of the population. This is related to the prevention paradox of

the epidemiology literature (Attwood et al., 2016), where the majority of cases of a disease may come from people at moderate risk of that disease, and only a minority from people at high risk of the disease.

Our results are also relevant to the design of capitation fee systems. Beyond the UK, capitation payment schemes are used in several European countries, New Zealand and Canada (Brilleman et al., 2014; Chambers, 1998; Sibley and Glazier, 2012; Shepherd, 2017), where they are used to pay providers prospectively for the treatment of patients to whom they are contracted to provide healthcare. 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).

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 6337 survey respondents with valid data on all the nurse-collected and blood-based biomarkers used in our analysis. From those, 4759 individuals had non-missing data on socioeconomic status 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. This allows us to follow a set of 2314 respondents in apparently good health at baseline, to a horizon up to five years later.

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 elevated risk of chronic conditions such as cardiovascular disease, impaired lung and liver functioning (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). It captures chronic physiological responses and dysregulations relevant to long-term chronic morbidity burdens, rather than acute infections. Thus the predictive power

¹ For example, in the USA, relevant administrative data are typically based on claims for a selection of insurance providers and do not usually represent total costs given co-payments (Riley, 2009) and, although the Medical Expenditure Panel Survey contains data on health care costs, it lacks objective health measures and baseline health status, including pre-symptomatic and pre-diagnosed conditions. The UK has no national-scale, individual-level administrative datasets on GP and community health service delivery; hospital episode statistics data have been linked to the Millennium Cohort Study but that is limited to a single birth cohort of pre-mature individuals.

² In these circumstances, some researchers have either used non-discrete interval regression models (Brown et al., 2015) or have sacrificed information by transforming outcomes into binary variables (Allin et al., 2011). There have been few applications of the more natural grouped count data models (but see Fu et al. (2018) for a Poisson-based example).

of our allostatic load measure relates more to healthcare demands arising from these chronic physical conditions, rather than those linked to transient infections, accidental injury, mental ill-health, etc.

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.³

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.⁴ 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, which rises as part of the immune response to infection and is associated with general chronic or systemic inflammation (Emerging Risk Factors Collaboration, 2010).⁵ 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 sulphate (DHEAS), a steroid hormone in the body, representing one of the primary mechanisms through which psychosocial stressors may affect health, with low levels associated with cardiovascular risk and all-cause mortality (Ohlsson et al., 2010). 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 of the nine biomarkers into a z-score and summing to produce the composite measure. The index is then scaled so that a 1-unit increase in allostatic load corresponds to an increase of one standard deviation.⁶

To illustrate the magnitudes involved, consider a healthy woman with waist 79 cm and average height 162 cm, normal systolic blood pressure of 105 mmHg, heart rate 70 bpm, an average HDL cholesterol level of 1.6 mmol/L, and average values for other biomarkers. If we compare her with a less healthy woman with waist 88, systolic blood pressure 140, heart rate 90 and low HDL cholesterol

1.0, then the difference in allostatic load between the two women is 1.5 standard deviations.⁷

2.2. Health care utilisation measures

Retrospective information was also collected at UKHLS wave 7, on the numbers of: GP consultations; attendances at a hospital or clinic as an out-patient or day patient (OP); and hospital in-patient (IP) days in the 12 months prior to interview. The GP and OP counts were recorded in five categories (0, 1-2, 3-5, 6-10, more than 10), whereas respondents were asked for the exact number of days spent in a hospital or clinic as an IP in the same period. We excluded women who reported any in-patient days for childbirth during this period (about 0.5% of our sample), so our cost analysis excludes services related to childbirth.

There are clear demographic differences in utilisation. Fig. 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. GP and OP consultations are also more frequent among older people (Appendix Figures A1 and A2).

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 reporting errors and possibly biases in long-term recall (Bound et al., 2001). A fundamental difficulty in assessing the prevalence and nature of survey errors is the absence of an accurate external measure to validate the survey responses.⁸ Matched individual-level data on health service utilisation are not available to us, but 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, subject to variations in response rates across population groups.⁹

Overall, we find that the administrative GP consultation rates for England and Scotland are reasonably close to mean counts interpolated¹⁰ 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

³ Some authors include cortisol, in addition to the stress-related hormone dihydroepiandrosterone sulphate (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).

⁴ 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.

⁵ We follow common practice by excluding cases with CRP values over 10mg/L because such values normally reflect acute rather than systemic inflammatory processes (Pearson et al., 2003).

⁶ 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.

⁷ The equivalent comparison for men of average height 175cm would involve waist measurements of 93cm and 102.

⁸ See Meyer and Mittag (2019) for the use of validation data in the context of income.

⁹ 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.

¹⁰ 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.

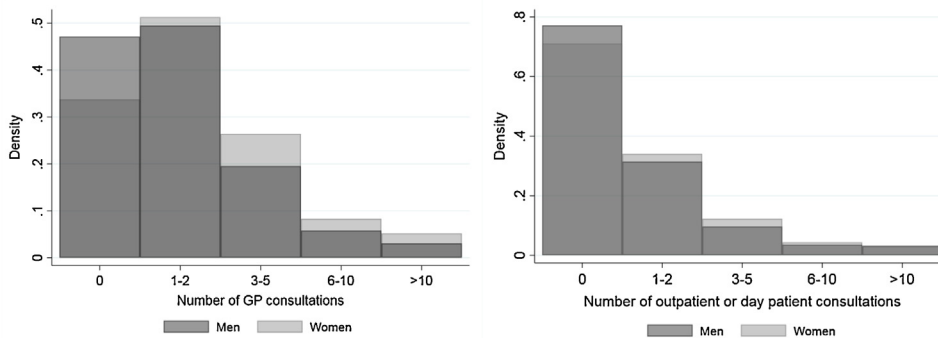


Fig. 1. Distribution of the numbers of GP and OP consultations in the preceding 12 months by gender.

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 utilisation, 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.¹¹ Instead, we pursue a data combination strategy, exploiting average cost data published in varying detail for the GP, OP and IP resource classes.¹² For simplicity, we use reference costs and caseload composi-

tion 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 (Department of Health, 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, maternity services and categories 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.¹³

For IP cases, reference costs relate to episodes of care, defined as “the time spent under the care of one consul-

¹¹ 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 bias.

¹² 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 (Department of Health, 2016, section 15).

¹³ Unlike the reference administrative IP costs data, which are classified by duration of hospital stay, the available unit cost data for GP and OP consultations is not classified by frequency. It is possible that patients who experience more complications more frequently may impose a higher GP and OP cost at each consultation than the average for those health services. Therefore, our analysis may under-predict GP and OP costs for such individuals and over-predict them for individuals with low-cost characteristics. If so, our estimates would give a lower bound for the gradient of future costs with respect to baseline health and other characteristics. If this bias exists at all, our focus on the population group who are apparently disease-free at baseline is likely to moderate its extent.

tant", 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). E and NEL episodes are further split to allow for excess bed days: health providers are required to provide *per diem* costs for the part of longer admissions that go beyond a 'trim point' set for each HRG. In our analysis, we treat E, NEL and NES episodes as separate categories within each HRG, 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 irrelevant and negligible categories and those with missing or invalid unit cost or mean stay data, we are left with 3827 IP categories, with a (caseload-weighted) mean stay length of 3.5 days, and a mean total of case-days of 9184 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. It should be borne in mind that these figures relate to all adult in-patients (thus excluding non-patients with zero costs), whereas our analysis of UKHLS survey data is concerned with average costs for the population of people with no diagnosed conditions at baseline, who are likely to have lower (and possibly zero) rates of healthcare utilisation than the average patient. Thus, our predicted costs are expected to be smaller than simple averages estimated from administrative data on patients as a whole. Descriptive statistics of the variables used in our analysis are presented in Table A3 of the Appendix.

3. Grouped count data models of healthcare utilisation

Let $Y_i \geq 0$ be the i th observation on a dependent variable (the GP, OP or IP utilisation count), which takes non-negative integer values, and \mathbf{X}_i a vector containing the explanatory covariates. We allow for the possibility of zero-inflation: a 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 probit¹⁴:

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

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

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

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

¹⁴ We also estimated logit specifications which gave almost identical estimates.

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 $L_i \leq Y_i \leq U_i$:

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

The model is completed by specifying a parameterised functional form for the non-degenerate component distribution $g(\cdot|\mathbf{X}_i)$. 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, 2019). It is derivable as a $\text{Poisson}(\lambda_i \nu)$ - $\text{gamma}(\alpha^{-1}, \alpha)$ mixture, where $\lambda_i = e^{\mathbf{X}_{i2}\boldsymbol{\beta}}$ and $\ln \alpha$ is treated as an unrestricted constant parameter. This gives a distribution for y with mean λ_i and variance $1 + \alpha\lambda_i$.¹⁵ The ML estimator is implemented using a Stata command `intcount` (Pudney, 2019).¹⁶

4. Parameter estimates

The explanatory covariates \mathbf{X} used in our healthcare utilisation model represent individual characteristics that have been shown to affect health outcomes directly or indirectly (Davillas and Pudney, 2020; Carrieri and Jones, 2017; Van 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 socio-economic status. Educational attainment is represented as a 3-category classification: degree-equivalent (reference category), 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 equalising 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. Age is measured in decades from an origin of 50 years, to reduce its scale and the correlation between the linear and quadratic term, to improve numerical conditioning of the likelihood function. The transformation of age has no effect on the predic-

¹⁵ In the terminology of Cameron and Trivedi (2013), this is the NB2 parameterisation of the regression model. Interval versions of the Poisson model (Dickman et al., 2004) and its zero-inflated version (Fu et al., 2018) have been used before, but we are not aware of any previous application of a zero-inflated negative binomial model under interval observation.

¹⁶ This is a quite standard application of maximum likelihood, but see Fu et al. (2018) for a detailed discussion of the asymptotic properties in the more special Poisson case.

tions nor the coefficient and standard error of the quadratic coefficient. Finally, we also allow for differences between the three nations of Great Britain (England as the reference category and Scotland and Wales), since NHS policy is determined on a national basis.¹⁷ The demographic and socioeconomic covariates are included in the healthcare utilisation models to allow us to assess the predictive power of allostatic load net of the potential confounding effect of individuals' demographic characteristics and socioeconomic status. They may proxy a wide range of socioeconomic factors and, thus, do not necessarily have a simple direct causal interpretation.

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 (Van 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 \mathbf{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.¹⁸

Table 1 shows a significant 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 one 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, much higher, at 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.

Specifying higher order polynomials of allostatic load (second or third order) revealed no significant nonlinearity in any of our model specifications (P -values between 0.205 and 0.985). To determine whether the predictive role of allostatic load varies by age and gender, we tested the relevant interaction terms in all model specifications for GP, OP and IP utilisation counts. There were no significant differences: P -values for the interactions of allostatic load with gender range between 0.105 and 0.830 across the different models; those for age interactions range between 0.155 and 0.699.¹⁹

We now consider the sub-sample of individuals with elevated levels of allostatic load at baseline (defined as at least one standard deviation above the mean), and examine the expected savings in future healthcare resource utilisation and cost generated by that group, which would result under a counterfactual where each had their allostatic load lowered by one standard deviation. Since we are interested in the effect of variations in allostatic load on the aggregate burden of treatment within this part of the population, it is appropriate to examine the mean of the counterfactual effect over all members of the group, rather than the effect calculated for an 'average' individual to represent the group.²⁰ To do this, we use the following sequential Monte Carlo simulation, conditional on the observed baseline covariates, where $r = 1 \dots R$ indexes pseudo-random replication sequences.²¹

(i) For each member i of the group with elevated allostatic load at baseline, construct counterfactual covariate vectors \mathbf{X}_i^* identical to the observed \mathbf{X}_i , but with the allostatic load variable reduced by 1 standard deviation unit for each individual.

¹⁷ 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. We dropped covariates only where their coefficients were statistically insignificant at the 10% level in all models, resulting in the smaller set of covariates listed in Table 1. To curtail pre-test bias, we retained the remaining covariates in all models, irrespective of statistical significance in particular models. 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 utilisation beyond what can be achieved using biomarkers.

¹⁸ 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 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. Demographic and socioeconomic covariates were not significantly significant in the zero-inflation probit component of the conditional models for OP and IP.

¹⁹ As a further sensitivity check, we also re-estimated our preferred utilisation models using age-gender dummies instead of age polynomials; we still failed to reject the null hypothesis of linearity in allostatic load (P -values ranging from 0.142 to 0.893).

²⁰ Of course, this is a quantitative summary of the importance of allostatic load as an indicator of the risk of future excess costs, not a policy simulation. Any feasible public health intervention would have a much more complex outcome than an immediate one standard deviation reduction in allostatic load. Note also that, unlike a marginal effect for an 'average individual', these average effects are not generally symmetric across groups – for example, the effect of transforming those observed to be male into counterfactual females is not the negative of the effect of transforming sampled females into counterfactual males.

²¹ It is possible to construct confidence intervals for simulated resource utilisation and cost, using bootstrap resampling. This requires repetition of the model estimation and utilisation/cost simulation and is very intensive in computer time. For this reason, we report confidence intervals only for the key estimates of the effects of allostatic load in Tables 2 and 4.

Table 1

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

Parameter	OP model			IP model	
	GP (1)	Marginal (2)	Conditional (3)	Marginal (4)	Conditional (5)
Allostatic load	0.210*** (0.036)	0.129** (0.053)	-0.030 (0.055)	0.555*** (0.162)	0.219 (0.175)
Age [§]	-0.053* (0.030)	0.033 (0.046)	0.068 (0.045)	0.241 (0.154)	0.177 (0.137)
Age ² [§]	0.017 (0.016)	0.018 (0.026)	0.002 (0.026)	-0.024 (0.081)	-0.051 (0.072)
Male	-0.198*** (0.076)	-0.327*** (0.121)	-0.319*** (0.110)	-0.285 (0.397)	-0.493 (0.370)
Age × male [§]	0.146*** (0.042)	0.187*** (0.069)	0.119* (0.067)	-0.367* (0.209)	-0.165 (0.201)
Age ² × male [§]	-0.025 (0.025)	0.024 (0.038)	0.061* (0.037)	0.182 (0.117)	0.247** (0.111)
No qualifications [†]	0.092 (0.092)	0.033 (0.142)	-0.016 (0.137)	0.744 (0.566)	1.481** (0.718)
Intermediate qualifications [†]	0.016 (0.065)	0.104 (0.103)	0.112 (0.096)	0.691** (0.321)	0.745** (0.290)
ln(income)	-0.140*** (0.045)	-0.106* (0.062)	-0.002 (0.059)	-0.209 (0.237)	0.101 (0.206)
Wales ^a	0.555*** (0.192)	0.140 (0.332)	-0.365 (0.249)	-0.061 (0.694)	-1.062* (0.590)
Scotland ^a	-0.018 (0.135)	-0.246 (0.231)	-0.311 (0.198)	0.278 (0.776)	-0.405 (0.680)
Intercept	1.793*** (0.351)	1.001** (0.483)	0.165 (0.481)	-0.270 (1.883)	-1.573 (1.670)
<i>Impact of conditioning on GP consultation count (reference level = 0)</i>					
1-2			-0.237 (0.206)		-1.384* (0.794)
3-5			0.534*** (0.206)		-1.248 (0.817)
6-10			0.952*** (0.214)		-0.203 (0.860)
more than 10			1.977*** (0.270)		1.640** (0.832)
ln(α)	0.012 (0.056)	1.099*** (0.058)	0.489*** (0.075)	3.895*** (0.121)	3.307*** (0.150)
<i>Zero-inflation parameters</i>					
zero GP consultation count			16.68*** (0.322)		14.97*** (0.722)
intercept			-16.17*** (0.233)		-12.86*** (0.558)
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 $N = 2314$.

[†] Reference categories: degree-level education.

^a England.

* Statistical significance: 10%.

** Statistical significance: 5%.

*** Statistical significance: 1%.

(ii) At each replication r , draw a pair of pseudo-random counts Y_{ir}^{GP} , Y_{ir}^{GP*} from the conditional distributions $Pr(Y_{ir}^{GP} | \mathbf{X}_i)$ and $Pr(Y_{ir}^{GP*} | \mathbf{X}_i^*)$, derived from the fitted model for GP consultations (column (1), Table 1). The same underlying set of pseudo-random numbers is used to generate both Y_{ir}^{GP} and Y_{ir}^{GP*} .

(iii) At each replication r , use Y_{ir}^{GP} to construct a vector \mathbf{Z}_{ir} containing the five additional dummy variables representing GP utilisation which appear in the conditional OP and IP models (columns (3) and (5) of Table 1), and construct an extended covariate vector $\mathbf{X}_{ir}^\dagger = (\mathbf{X}_i, \mathbf{Z}_{ir})$. Then evaluate $Pr(Y_{ir}^{OP} | \mathbf{X}_{ir}^\dagger)$ and $Pr(Y_{ir}^{IP} | \mathbf{X}_{ir}^\dagger)$ for each individual, using the fitted conditional models for OP and IP consultation

counts, and draw pseudo-random Y_{ir}^{OP} and Y_{ir}^{IP} from those distributions.

(iv) Repeat step (iii), substituting the counterfactual \mathbf{X}_i^* , Y_{ir}^{GP*} for \mathbf{X}_i , Y_{ir}^{GP} but re-using the same underlying pseudo-random numbers, to generate counterfactual outcomes Y_{ir}^{OP*} , Y_{ir}^{IP*} .

(v) Compute any means and probabilities of positive counts for Y^{GP} , Y^{OP} , Y^{IP} over all individuals i in the elevated-allostatic load group and all R replications. For each utilisation count Y , average marginal effects are expressed as % differences between base and counterfactual outcomes. For mean counts, they are computed as $100[\sum_i \sum_r (Y_{ir}^* - Y_{ir})] / \sum_i \sum_r Y_{ir}$, while the % difference at

the extensive margin is $100 [\sum_i \sum_r (\mathbb{1}(Y_{ir}^* > 0) - \mathbb{1}(Y_{ir} > 0))] / \sum_i \sum_r \mathbb{1}(Y_{ir} > 0)$ and $\mathbb{1}(\cdot)$ is the indicator function.

The results are shown in Table 2. Confidence intervals are wider for the OP and IP impacts, because OP and IP consultations are relatively uncommon and because the estimates relate to averages over a small subsample of individuals with allostatic load at least one standard deviation above mean at baseline.²² Allostatic load proves to be a very strong predictor of future healthcare demand, suggesting potential for a substantial reduction in resource usage if effective interventions could be targeted on those with high allostatic load. Most of the effect is at the intensive rather than extensive margin, with the proportion of individuals calling on GP, OP and IP resources reduced much less proportionately than is the mean consultation count. A 1σ reduction among the group with high allostatic load is predicted to reduce GP and OP consultations by 19% and 12% respectively. Results for the more costly IP resource are stronger still, with a 1σ reduction in allostatic load reducing expected future resource usage by over 40%. Allostatic load is a particularly effective predictor for relatively serious conditions requiring hospital stays.

5. 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 $\bar{c}^{GP} = \text{£}66.20$ 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:

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

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 $E(C_i^{OP} | \mathbf{X}_{ir}^{OP}) = \bar{c}^{OP} E(Y_i^{OP} | \mathbf{X}_{ir}^{OP})$, where $\bar{c}^{OP} = \sum_j \pi^{OP}(j) \bar{c}^{OP}(j)$ and \mathbf{X}_{ir}^{OP} is the

²² Our use of 90% confidence intervals is motivated by a belief that the relevant alternative hypothesis is 1-sided rather than 2-sided – it is implausible to suggest that worsening health reduces the expected healthcare burden. Thus the lower limit of the 90% confidence interval can be interpreted either as part of a 2-sided 90% interval, or more appropriately as a 95% 1-sided confidence limit on the magnitude of the effect. The bootstrap calculation is burdensome because it involves simulation within simulation. We attempted 300 bootstrap replications; 23 were lost due to non-convergence of one or more of the re-estimated models. We have not attempted to modify specifications or explore alternative starting values for the ML optimisation in those 23 cases.

covariate vector in the OP consultation model, constructed at replication r using Y_{ir}^{GP} .

Table 3 summarises the cost simulations for each of nine hypothetical variations: a 1σ reduction in allostatic load for all those who are at least 1σ above the mean; an increase in age of 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 the unqualified and the intermediate group; and a universal 10% increase in equivalised income. This analysis allows us to interpret the results by comparing the predicted impact of a reduced allostatic load with demographic and socioeconomic differences in healthcare costs.

The expected GP costs in Table 3 are particularly high for the group with elevated allostatic load compared to all the comparator demographic and socioeconomic groups. For example, the mean base cost for those with high allostatic load is about £ 207, exceeding even the GP costs incurred for the over-75 age group. A 1σ reduction in baseline allostatic load for those with elevated values would reduce GP costs by almost one-fifth, holding other characteristics constant. This almost four times larger in magnitude than the proportional increase in the GP costs from an additional 10 years of age among the oldest group, and is only exceeded by the proportional difference in mean GP costs between men and women.

The mean OP costs in Table 3 are uniformly higher than GP costs across the set of illustrative baseline population groups. Predicted mean OP costs for the group with elevated allostatic load (£ 295) exceed those of every other group except for the over-75s (£ 420). For the former group, a 1σ reduction in allostatic load reduces predicted mean OP costs by over 12%, which is larger in magnitude than the proportionate impacts of differences in education and income, but smaller than the effects of gender or of ageing by ten years among the over-45 population.

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 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 give 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 j th 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

Table 2

Estimated effects of a one standard deviation reduction in allostatic load on expected future resource utilisation counts among individuals with elevated allostatic load at baseline.

Service type	Mean base count [†]	Marginal effect [90% confidence interval] [‡]	Base proportion positive [†]	Marginal effect at extensive margin [‡] [90% confidence interval] [§]
GP consultations	3.12	-18.9 [-27.8%, -15.6%]	0.75	-5.4% [-7.9%, -4.4%]
OP consultations	1.80	-12.4 [-24.4%, -2.4%]	0.46	-4.9% [-9.3%, -1.1%]
IP days	0.94	-43.5 [-63.0%, 10.2%]	0.07	-15.4% [-44.2%, 4.5%]

[†] Predicted mean count and probability of positive utilisation averaged over the 359 individuals with allostatic load at least 1σ above mean at baseline (15.5% sample proportion).

[‡] Percentage difference in the proportion with positive utilisation within the high-allostatic load group.

[§] Nonparametric bootstrap confidence intervals, with both model estimation and utilisation simulation repeated at each of 277 bootstrap replications.

Table 3

Personal characteristics and annual GP and OP costs five years later.

Base sample and hypothetical variation	GP consultations		OP consultations	
	Mean base cost	% change in cost	Mean base cost	% Change in cost
High allostatic load - 1σ	£ 207	-18.9 [†]	£ 295	-12.4 [†]
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

Sample sizes: high allostatic load: 359; age 16–29: 239; age 30–44: 740; age 45–59: 823; age 60–74: 444; age over 75: 68; males: 1018; females: 1296; no qualifications: 345; intermediate qualifications: 945; all incomes: 2314.

[†] Predicted costs are proportional to predicted utilisation for GP and OP services, so the % change in cost is identical to the corresponding figure in Table 2.

treatment type, conditional on the simulated treatment duration and observed characteristics of each individual. Those probabilities are 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. Table 4 gives two alternative sets of cost estimates. The first (columns 1 and 2) uses only the simulated IP duration to tailor treatment type probabilities to individuals; the second (columns 3 and 4) uses duration, age group and gender to tailor the treatment probabilities. Perhaps surprisingly, the additional 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 implication would be a reduction of almost a quarter in their future IP costs: roughly comparable to ten years' ageing in middle age and early old age.²³

²³ The estimated % change in cost for the 45–59, 60–74 and over 75 age groups (16.3%, 28.3% and 52.4% respectively), weighted by group sample numbers, average to 22%.

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 major influence of allostatic load. Those with elevated biomarkers at baseline are predicted to generate a mean total cost of £ 823 five years later. If allostatic load were hypothetically reduced by one standard deviation for each member of this group, combined future GP, OP and IP costs are predicted to be reduced by 18%, so there is clear scope for an effectively targeted intervention to curtail future healthcare costs significantly. This projected % reduction in total cost is broadly comparable with the predicted cost differences stemming from ageing (19% weighted average effect of an additional 10 years for the 45–59, 60–74 and 75+ groups), gender (approximately 22%) and one category of education (approximately 17%).

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, 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

Table 4

Personal characteristics and expected IP costs and total costs five years later.

Base group and hypothetical variation	Statistical allocation on...			
	Duration only		Duration + age + gender	
	Mean base cost (1)	% change in cost (2)	Mean base cost (3)	% change in cost (4)
	<i>Conditional mean IP costs</i>			
High allostatic load - 1σ [90% confidence interval] [§]	£ 321	-23.2	£ 323	-23.8 [-49.2, 7.5]
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
	<i>Conditional mean total (GP+OP+IP) costs</i>			
High allostatic load - 1σ [90% confidence interval] [§]	£ 823	-18.2	£ 825	-18.5 [-28.5, -8.7]
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

[§] Nonparametric bootstrap confidence intervals, with both model estimation and utilisation simulation repeated at each of 277 bootstrap replications.

attach costs to healthcare utilisation involves accounting and recording errors inherent in the available reference cost data, which are averages across groups of cases rather than true individual-specific costs and exclude some cost elements (such as most community health services).

Our methodology of statistical cost allocation rests on assumptions that may seem strong at first sight, although we would argue that they are more innocent than they appear. In estimating IP costs, we assume an individual's reported number of days in the hospital (that is, each admission) stem from a single episode. It should be noted here that the payment by results reimbursement scheme, adopted by the NHS since 2004, provides no financial incentive for healthcare providers to increase the number of episodes per admission, aiming to promote both efficiency and the accuracy of the initial diagnosis. According to NHS statistics, the vast majority of the patient spells (a continuous period of time spent as a patient within a trust, which is equivalent to an admission) have only one episode, while there is only a tiny proportion of admissions with three or more episodes (Department of Health, 2012). Note also that 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, only indirectly through our use of duration in the calculation of individual-specific probabilities of alternative disease/treatment types.

The outcomes that we study are necessarily 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. Since we do not have biomarker measurement repeated over time, we do not observe change in biomarkers and can therefore say nothing definite about the optimal frequency of measurement within a real-world screening programme. 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.

We carried out sensitivity checks to determine whether our results are sensitive to our use of English NHS reference cost data to the whole sample, which includes individual's resident in Wales and Scotland. Restricting the analysis to English residents (93% of our full sample) gives results that are practically identical to those for the full GB sample both in terms of model estimates (Appendix Table A4) and cost implications (Table A5). We also assessed sensitivity to the role of medications by repeating our analysis excluding the few individuals (4% of the sample) who, at baseline, were taking medication for high blood pressure, cardiovascular

conditions, diabetes or respiratory conditions.²⁴ The model estimates (Table A6) and cost simulation results (Table A5) are very close to our primary results.

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 for 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 reflecting pre-diagnosed and pre-symptomatic pathways is a powerful predictor of the future burden of service utilisation for the primary and hospital healthcare systems. Combining the prediction models with healthcare cost data for England, we have quantified the excess healthcare costs generated by those individuals with elevated biomarkers at baseline. To achieve this, we developed a simulation-based method of assigning administrative cost estimates to the service utilisation levels predicted by count data models, incorporating both duration and demographic characteristics to personalise the assignment in the case of inpatient treatment episodes.

Overall, we found that those with elevated biomarkers are predicted to generate a mean annual total healthcare cost (GP, OP consultations and IP days) of £ 823 five years later. A standard deviation reduction in allostatic load for each member of this group resulted in an 18% reduction in predicted total costs, with the largest impact for hospital inpatient treatment. These results suggest that there is clear scope for interventions targeted effectively on those with elevated biomarkers to reduce future health care costs significantly – if such interventions can be designed. The magnitude of the effects of allostatic load on caseloads and costs is broadly comparable with the substantial estimated influence we found for age and education. Although the main focus of our paper is not the socioeconomic gradient in healthcare and costs, our results extend previous evidence of the impact of education on health (Conti et al., 2010) to healthcare utilisation and costs.

The predictive power of the biomarker-based health measures gives a possible basis for sophisticated tailoring of preventive interventions. Our findings suggest that effective targeting reflecting the biological pattern of risk using biomarker data could identify better the population groups with highest potential future healthcare needs and costs. A measure similar to our allostatic load proxy could be constructed from information gathered in screening check-up programmes, like the NHS England Health Checks and similar programs in other countries including Australia, Germany, the Netherlands and Japan (OECD, 2019; Schülein et al., 2017). Recent technical developments have expanded the options for obtaining biomarkers

at the population level. As an alternative to conventional venepuncture, dried blood spot (DBS) sampling (drops of whole blood collected on filter paper from a simple finger prick) offers a minimally invasive method for sampling blood, at a significantly lower collection cost (Martial et al., 2016; McDade et al., 2007). Recent evidence alleviates concerns over the validity of DBS-derived biomarkers (Samuelsson et al., 2015), suggesting that the DBS method has become an effective way to implement blood-based screening tests on a large scale in practice. We hope that technical improvements and evidence of the kind presented here may contribute to the development of these programmes to improve their cost-effectiveness.

There is a continuing debate on capitation-based payments that are currently used to allocate budgets to GPs in a number of countries (Brilleman et al., 2014; Chambers, 1998; Sibley and Glazier, 2012; Shepherd, 2017). A potential policy application of our findings is in refining the design of these systems by reorienting the capitation formula to match more closely patient level morbidity data and other individual characteristics (Shepherd, 2017). This offers the prospect of improved allocation of resources and better health outcomes by reducing incentives for health providers to “skim the cream” from the patient population by selecting patient sub-groups with lower expected future healthcare costs.

Appendix A. Supplementary Data

Supplementary data associated with this article can be found, in the online version, at <https://doi.org/10.1016/j.jhealeco.2020.102356>.

References

- Allin, S., Masseria, C., Mossialos, E., 2011. Equity in health care use among older people in the UK: an analysis of panel data. *Appl. Econ.* 43 (18), 2229–2239.
- Attwood, S., Morton, K., Sutton, S., 2016. Exploring equity in uptake of the NHS health check and a nested physical activity intervention trial. *J. Public Health* 38 (3), 560–568.
- Biomarkers Definitions Working Group, 2001. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clin. Pharmacol. Therapeut.* 69 (3), 89–95.
- Bound, J., Brown, C., Mathiowetz, N., 2001. Measurement error in survey data. In: Heckman, J., Leamer, E. (Eds.), *Handbook of Econometrics*. Vol. 5. Elsevier, Amsterdam, pp. 3705–3843.
- Brilleman, S.L., Gravelle, H., Hollinghurst, S., Purdy, S., Salisbury, C., Windmeijer, F., 2014. Keep it simple? Predicting primary health care costs with clinical morbidity measures. *J. Health Econ.* 35, 109–122.
- Brown, S., Duncan, A., Harris, M.N., Roberts, J., Taylor, K., 2015. A zero-inflated regression model for grouped data. *Oxford Bull. Econ. Stat.* 77 (6), 822–831.
- Cameron, A.C., Trivedi, P.K., 2013. *Regression Analysis of Count Data*, 2nd ed. Cambridge University Press, Cambridge, UK.
- Carreras, M., Ibern, P., Inoriza, J.M., 2018. Ageing and healthcare expenditures: Exploring the role of individual health status. *Health Econ.* 27 (5), 865–876.
- Carrieri, V., Jones, A.M., 2017. The income-health relationship beyond the mean: New evidence from biomarkers. *Health Econ.* 26 (7), 937–956.
- Chambers, G.R., 1998. *Health Care Systems in The EU a Comparative Study*. European Parliament.
- Charlesworth, A., Bloor, K., 2018. 70 years of NHS funding: how do we know how much is enough? *BMJ* 361, k2373.
- Chernew, M.E., Newhouse, J.P., 2011. Health care spending growth. In *Handbook of Health Economics*, volume 2, 1–43.
- Cohen, J.T., Neumann, P.J., Weinstein, M.C., 2008. Does preventive care save money? *Health economics and the presidential candidates*. *N. Engl. J. Med.* 358 (7), 661–663.

²⁴ These individuals reported taking medication despite having reported no such diagnosed conditions. It is not certain which of the two reports is erroneous in any given case.

- Conti, G., Heckman, J., Urzua, S., 2010. The education-health gradient. *Am. Econ. Rev.* 100 (2), 234–238.
- Curtis, L.A., Burns, A., 2017. Unit Costs of Health and Social Care 2017. <https://doi.org/10.22024/UniKent/01.02/65559>.
- Dalgaard, C.-J., Strulik, H., 2014. Optimal aging and death: understanding the preston curve. *J. Eur. Econ. Assoc.* 12 (3), 672–701.
- Davillas, A., 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. *J. Health Econ.* 56, 87–102.
- Davillas, A., Pudney, S.E., 2020. Biomarkers as precursors of disability. *Economics and Human Biology*, pp. 36.
- De Meijer, C., Koopmanschap, M., d'Uva, T.B., Van Doorslaer, E., 2011. Determinants of long-term care spending: age, time to death or disability? *J. Health Econ.* 30 (2), 425–438.
- Department of Health, 2012. A Simple Guide to Payment by Results. Department of Health, Leeds, UK.
- Department of Health 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>.
- Dickman, P.W., Sloggett, A., Hills, M., Hakulinen, T., 2004. Regression models for relative survival. *Stat. Med.* 23, 51–64.
- Emerging Risk Factors Collaboration 2010. C-reactive protein concentration and risk of coronary heart disease, stroke, and mortality: an individual participant meta-analysis. *The Lancet* 375 (9709), 132–140.
- Fu, Q., Guo, X., Land, K.C., 2018. A Poisson-multinomial mixture approach to grouped and right-censored counts. *Commun. Stat. Theory Methods* 47, 427–447.
- Geronimus, A.T., Hicken, M., Keene, D., Bound, J., 2006. Weathering" and age patterns of allostatic load scores among blacks and whites in the United States. *Am. J. Public Health* 96 (5), 826–833.
- Glei, D.A., Goldman, N., Rodríguez, G., Weinstein, M., 2014. Beyond self-reports: changes in biomarkers as predictors of mortality. *Popul. Develop. Rev.* 40 (2), 331–360.
- Goldman, N., Turra, C.M., Gleib, D.A., Seplaki, C.L., Lin, Y.-H., Weinstein, M., 2006. Predicting mortality from clinical and nonclinical biomarkers. *J. Gerontol. Ser. A: Biol. Sci. Med. Sci.* 61 (10), 1070–1074.
- Haider, A.W., Larson, M.G., Franklin, S.S., 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. *Ann. Inter. Med.* 138 (1), 10–16.
- Hobbs, F.D.R., Bankhead, C., Mukhtar, T., Stevens, S., Perera-Salazar, R., Holt, T., Salisbury, C., 2016. Clinical workload in UK primary care: a retrospective analysis of 100 million consultations in England, 2007–14. *Lancet* 387, 2323–2330.
- Howard, J.T., Sparks, P.J., 2016. Does allostatic load calculation method matter? Evaluation of different methods and individual biomarkers functioning by race/ethnicity and educational level. *Am. J. Human Biol.* 28 (5), 627–635.
- Howdon, D., Rice, N., 2018. Health care expenditures, age, proximity to death and morbidity: implications for an ageing population. *J. Health Econ.* 57, 60–74.
- 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., Couper, M.P., 2018. Understanding data linkage consent in longitudinal surveys. University of Essex: *UnderstandingSociety Working Paper* 2018-07.
- Martial, L.C., Aarnoutse, R.E., Schreuder, M.F., Henriët, S.S., Brüggemann, R.J., Joore, M.A., 2016. Cost evaluation of dried blood spot home sampling as compared to conventional sampling for therapeutic drug monitoring in children. *PLOS ONE* 11 (12), e0167433.
- McDade, T.W., Williams, S., Snodgrass, J.J., 2007. What a drop can do: dried blood spots as a minimally invasive method for integrating biomarkers into population-based research. *Demography* 44 (4), 899–925.
- Meyer, B.D., Mittag, N., 2019. Using linked survey and administrative data to better measure income: Implications for poverty, program effectiveness, and holes in the safety net. *Am. Econ. J.: Appl. Econ.* 11 (2), 176–204.
- Morciano, M., Hancock, R.M., 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. *Soc. Sci. Med.* 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.
- OECD 2019. OECD Reviews of Public Health: Japan: A Healthier Tomorrow. Organisation for Economic Co-operation and Development.
- Ohlsson, C., Labrie, F., Barrett-Connor, E., Karlsson, M.K., Ljunggren, O., Vandenput, L., Mellstrom, D., Tivesten, A., 2010. Low serum levels of dehydroepiandrosterone sulfate predict all-cause and cardiovascular mortality in elderly Swedish men. *J. Clin. Endocrinol. Metabol.* 95 (9), 4406–4414.
- 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.
- Pudney, S.E., 2019. intcount: a Stata command for estimating count data models from interval data. *Stata J.* 19, 645–666.
- Riley, G.F., 2009. Administrative and claims records as sources of health care cost data. *Med. Care* 47 (Supplement 1), S51–S55.
- Russell, L.B., 2009. Preventing chronic disease: and important investment, but don't count on cost savings. *Health Affairs* 28, 42–45.
- Samuelsson, L.B., Hall, M.H., McLean, S., Porter, J.H., Berkman, L., Marino, M., Sembajwe, G., McDade, T.W., Buxton, O.M., 2015. Validation of biomarkers of CVD risk from dried blood spots in community-based research: Methodologies and study-specific serum equivalencies. *Biodemogr. Soc. Biol.* 61 (3), 285–297.
- Schüle, S., Taylor, K.J., Schriefer, D., Blettner, M., Klug, S.J., 2017. Participation in preventive health check-ups among 19,351 women in Germany. *Prevent. Med. Reports* 6, 23–26.
- Seeman, T.E., Crimmins, E., Huang, M.-H., Singer, B., Bucur, A., Gruenewald, T., Berkman, L.F., Reuben, D.B., 2004. Cumulative biological risk and socio-economic differences in mortality: MacArthur studies of successful aging. *Soc. Sci. Med.* 58 (10), 1985–1997.
- Shepherd, D., 2017. Capitation based funding of general practice is not fit for purpose. *BMJ* 358, j4075.
- Sibley, L.M., Glazier, R.H., 2012. Evaluation of the equity of age-sex adjusted primary care capitation payments in Ontario, Canada. *Health Policy* 104 (2), 186–192.
- Turner, R.J., Thomas, C.S., Brown, T.H., 2016. Childhood adversity and adult health: Evaluating intervening mechanisms. *Soc. Sci. Med.* 156, 114–124.
- Van Doorslaer, E., Koolman, X., Jones, A.M., 2004. Explaining income-related inequalities in doctor utilisation in Europe. *Health Econ.* 13 (7), 629–647.
- Vie, T.L., Hufthammer, K.O., Holmen, T.L., Meland, E., 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). *Soc. Sci. Med.* 117, 1–9.
- Wang, T.J., Gona, P., Larson, M.G., Tofler, G.H., Levy, D., Newton-Cheh, C., Jacques, P.F., Rifai, N., Selhub, J., Robins, S.J., et al., 2006. Multiple biomarkers for the prediction of first major cardiovascular events and death. *N. Engl. J. Med.* 355 (25), 2631–2639.
- WHO, 2018. Strategy on the health and well-being of men in the WHO European Region. World Health Organization, European Region.
- Zethelius, B., Berglund, L., Sundström, J., Ingelsson, E., Basu, S., Larsson, A., Venge, P., Ärnlöv, J., 2008. Use of multiple biomarkers to improve the prediction of death from cardiovascular causes. *N. Engl. J. Med.* 358 (20), 2107–2116.
- Zweifel, P., Felder, S., Meiers, M., 1999. Ageing of population and health care expenditure: a red herring? *Health Econ.* 8 (6), 485–496.