Stress-related health depreciation

Using allostatic load to predict self-rated health

Highlights

- Chronic stress can lead to reduced physical health.
- Allostatic Load (AL) objectively measures stress-related "wear and tear" using biomarkers.
- We find that AL predicts future self-reported physical health and SF6D.
- Reductions in these measures are greatest for those reporting worse health at baseline.
- AL predicts health trajectories of which individual's may not be aware.

Abstract

Approximately one quarter of UK adults are currently diagnosed with two or more chronic conditions, often referred to as multimorbidity. Chronic stress has been implicated in the development of many diseases common to multimorbidity. Policymakers and clinicians have acknowledged the need for more preventative approaches to deal with the rise of multimorbidity and "early ageing". However divergence may occur between an individual's self-rated health and objectively measured health that may preclude preventative action. The use of biomarkers which look 'under the skin' provide crucial information on an individual's underlying health to facilitate lifestyle change or healthcare utilisation. The UK's *Understanding Society* dataset, was used to examine whether baseline variation in biomarkers measuring stress-related "wear and tear" – Allostatic Load (AL) – predict changes in future self-rated health (SRH) while adjusting for baseline SRH, socioeconomic and lifestyle factors, and healthcare inputs. An interaction between baseline AL and baseline SRH was included to test for differential rates of SRH change. We examined SRH using the SF6D instrument, measuring health-

related-quality of life (HRQoL), as well as it's physical and mental health components separately. We found that HRQoL and physical health decline faster for those with higher baseline AL (indicating greater "wear and tear") however the same pattern was not observed for mental health. These findings provide novel insights for clinicians and policymakers on the usefulness of AL in capturing health trajectories of which individual's may not be aware and its importance in targeting resilience enhancing measures earlier in the lifecourse to delay physical health decline.

Keywords: Stress, Ageing, Allostatic Load, Health Depreciation, Biomarkers, SF6D

Introduction

Approximately one quarter of UK adults are currently diagnosed with two or more chronic conditions, this often being referred to as multimorbidity (1, 2). Prevalence increases to approximately two thirds among individuals aged over 65 years, is concentrated in lower socioeconomic groups, and is projected to increase over the next 15 years (2, 3). As more people face the challenge of living with chronic disease, attempts to redefine "health" have made explicit the importance of fostering resilience to challenges that arise from one's environment (4, 5). Chronic stress is recognised as an important mechanism linking the socioeconomic environment to negative health outcomes (6).

Chronic stress has been implicated in the development of many chronic conditions, such as cancer, obesity, cardiovascular disease, diabetes, depression, and asthma (7, 8) - diseases which consistently account for large proportions of national health expenditure, disability and mortality (9, 10). Policymakers and clinicians have argued for more preventative approaches to deal with this rise in multimorbidity and "early ageing" (11, 12). Preventative approaches aimed at highlighting the long-term damage from stress, through teachable moments between doctors and patients (see (13, 14) for examples), or fostering resilience to stress, through the development of mutable non-cognitive skills like conscientiousness (15), may delay disease onset. Public Health England recently reported a positive return on investment for services to develop stress resilience although they highlighted a need to understand their long-term physical health consequences (16). Two recent Cochrane reviews although finding positive evidence of the effectiveness of psychological interventions to foster resilience in healthcare students and professionals on Self-Rated Health (SRH), noted a high level of uncertainty in this result and a paucity of longer-term data (17, 18).

The Food and Drug Administration and the National Institute for Health highlighted the use of susceptibility/risk biomarkers as valuable in guiding preventative strategies (19). An Allostatic Load

(AL) index combines a number of biomarkers to measure cumulative biological dysregulation or stress-related "wear and tear" (20, 21). Greater adversity in childhood and minority status have been linked to higher AL (22, 23), as have macroeconomic shocks (24). Those with higher AL experience earlier mortality (25, 26) and the development or exacerbation of many stress-related diseases (27). Objective health measures such as AL, where they represent credible surrogate endpoints (28), provide crucial information on an individual's underlying health or risk of health decline which may facilitate lifestyle change or healthcare utilisation.

This paper uses the *Understanding Society* dataset to examine whether baseline AL predicts future SRH over and above baseline SRH. A growing literature has demonstrated associations between SRH measures and AL however these studies are often cross-sectional and suffer from issues of reverse causality, small sample sizes and omitted variable bias (21, 27). Vie et al (2014) attempted to navigate the issue of reverse causality by examining the association of baseline SRH with future AL however they were unable to control for baseline AL (29). Engman (2019) argues that the potential for a rupture in the trajectory of one's life - biographical disruption - is dependent on the degree to which more subtle changes in health are embodied prior to the occurrence of a more severe change (30). Thus we employ Engman's description of biographical disruption with empirical analysis informed by Grossman's model for the demand for health from the health economics literature (31-33), to examine the converse scenario to that explored by Vie et al (2014) - AL predicting future SRH.

We explore whether this association varies according to differences in baseline SRH akin to the notion of increasing depreciation of health capital, i.e. how quickly health declines in relation to AL. The Grossman model of the demand for health and by extension healthcare (31, 32), posits that an individual maximises their utility subject to time and budget constraints through an iterative process that must be constantly updated according to, among other factors, depreciation of their health. The

depreciation rate was synonymous with the biological ageing rate (32, 34) however in recent years the model has been adapted to include stress as a source of depreciation (33) to help explain socioeconomic inequalities observed in health over the lifecourse.

AL may drive future SRH in a number of ways. Firstly, it may reduce the stock of health, as reflected in SRH in the baseline period, which would lower health in the future period since the individual begins from a lower level. If AL operated only through this channel, we would not expect baseline AL to predict future SRH after adjusting for baseline SRH and interventions targeting AL may have little added value over interventions targeting SRH. Secondly, baseline AL may directly influence SRH in future periods by influencing the health depreciation rate. If AL operated exclusively through this channel, we anticipate that AL would significantly predict future SRH even after controlling for baseline SRH. Here interventions targeting AL may be helpful in offsetting future SRH decline. Thirdly, it may be that the depreciation rate varies with the level of baseline health for instance those in poorer health may deteriorate more quickly than those in good health. This would imply an interaction between baseline SRH and AL. Here interventions targeting AL may also be helpful in offsetting future SRH decline and earlier intervention, for example providing resilience enhancing measures to younger individuals, may provide the greatest value over the lifecourse. Thus, exploring the relationship between AL and SRH provides insights into the pathways through which stress may influence health over the lifecourse.

We examine a SRH outcome commonly used in the generation of Quality-Adjusted Life Years (QALYs) (35, 36); the six-dimensional health state short form (SF6D, (37)) and it's physical and mental health component scores (PCS & MCS) separately. While we cannot claim that our estimates are causal we have endeavoured to control for a wide range of potential confounders while also examining the sensitivity of our results to unobserved confounding, missing observations, model misspecification,

and AL index configuration. This paper proceeds with: an outline of the data, empirical methodology, and sensitivity analysis in section two; followed by presentation of the results in section three; with a discussion and conclusion in sections four and five respectively.

Methods

2.1 Data

This study uses data from waves 1-7 of the UK Household Longitudinal Survey (UKHLS), also known as *Understanding Society.* This is a large national UK panel dataset beginning in 2009-10 with wave one and collected annually across 40,000 households covering approximately 100,000 individuals (38, 39). The British Household Panel Survey (BHPS) was absorbed into the UKHLS at wave two. A subset of the non-BHPS individuals received a nurse health visit on average five months after the mainstage survey in wave two while a subset of BHPS individuals were visited by the nurse after a similar period in wave three (39). The nurse measured physiological and biomarkers, e.g. height, blood glucose, which were used to form an AL index (see the 'Allostatic Load Index' section for details). These two subsamples were pooled and the baseline period (*t*) refers to wave two for non-BHPS and wave three for BHPS respondents. This presented a potential sample of 35,937 individuals, of which 20,700 underwent a nurse health assessment to collect physiological data (39) with 10,175 and 3,342 individuals from waves 2 and 3 respectively having blood sample data available.

The mainstage survey collected information on each individuals' socioeconomic status and SRH across multiple waves while the nurse health assessment was conducted only once for distinct subsets of waves two and three (non-BHPS vs BHPS). Therefore we observe a single set of biomarkers, and hence can measure the AL index, for each respondent in the sample albeit recorded at different timepoints. The final sample with complete nurse health assessment and mainstage survey data across individuals

was 7,712 (see Figure S1). This is slightly reduced when analysing SRH outcomes further from baseline (Figure 1) however we examine whether results vary when applying weights for longitudinal data and outlier adjustment, missing nurse health assessment data, and household composition and country-level differences (38, 39).

[INSERT FIGURE 1]

2.2 Allostatic Load Index

Twelve physiological markers and biomarkers – creatinine clearance rate (CCR), insulin-like growth factor one (IGF-1), DHEA-s, clauss fibrinogen, glycated haemoglobin (HbA1c), pulse, systolic and diastolic blood pressures (SBP and DBP), C-reactive protein (CRP), ratio of total to HDL cholesterol, triglycerides, and waist-to-height ratio (W2H) – were used in the AL index (table 1). For most biomarkers elevated levels reflect higher risk but for three of these (DHEA-s, IGF-1, and CCR) lower levels represent higher risk and so these were rescaled before combining with the others to create the AL index. Kolmogorov-Smirnov (KS) tests were used to test for differences in the distribution of each biomarker between males and females for all available observations (table 1) and significant differences ($p < 0.001$) supported the separate standardisation of each biomarker according to gender before being combined (see Figure S2 for further details). The nurse, during his/her visit, also collected information on medication use at the time of survey. Adjustments were made to biomarkers following Chandola et al (2017) while more detailed medication data was used in sensitivity analyses (see 'Covariates' section).

A 'z-score AL index' was used which sums each individual's z-score for each biomarker (21) to obtain a combined index that was then itself standardised. This makes use of the full distribution of each biomarker and allows the weight of each biomarker to vary according to its deviation from the mean.

However issues have been noted in the fidelity of AL indices; the choice of biomarkers and their configuration in an AL index vary across studies (21, 40). We thus examine model fit according to other AL index configurations ('Norm AL index' (21)) and subgroups of biomarkers capturing individual biological systems.

[INSERT TABLE 1]

2.3 Self-Reported Health

A number of SRH variables were examined in this analysis. For simplicity we will refer to all of these measures jointly as 'SRH' unless referring to one specifically. At each wave respondents answered the Short Form-12 (SF-12) which is a generic health measure (41). The components of SF-12 can be partitioned to create a single measure of physical health ranging from 0-100, the physical component score (PCS), with zero representing the worst physical health and 100 the best and a corresponding measure of mental health, the mental component score (MCS), also ranging from 0-100. Analysis was carried out using each measure to reflect two different aspects of SRH as well as using the combined health utility/HRQoL score (SF6D). SF6D is a combination of PCS and MCS weighted according to the UK public's preferences which estimates a HRQoL index with a value of one reflecting full health and zero being dead (37). However in practice a floor effect is commonly observed for SF6D values which in our data was 0.35. For the regression analyses, all SRH variables were standardised to allow for comparison of associations (expressed in terms of standard deviations of the outcome) across models but are presented in their raw form in the descriptive statistics in Table 2.

2.4 Control variables

The log of equivalised household income and highest educational achievement were controlled for along with other individual socioeconomic characteristics of the individual outlined in the introduction as potentially influencing AL and SRH: age-group (quintiles), gender, marital status, employment status, self-reported ethnicity as white (reflecting 97% of the sample) and self-reported urban vs rural location. These factors were controlled for in our baseline models (Models 1 & 2) however a number of other potential confounders, described below and outlined in Figure 2, were also examined.

[INSERT FIGURE 2]

Case and Deaton (2005) note the important conceptual difference in the Grossman model between health depreciation (referred to by them as the rate of deterioration) and the actual change in the stock of health between periods (42). They suggest that medical care or other forms of health investment (e.g. exercise) may offset health deterioration. Therefore, we adjust for a broad range of other potential confounders (medication, lifestyle factors and non-cognitive skills). There does not appear to be a "gold standard" for how to account for medication in analyses involving AL. An individual may be selected into treatment because of their underlying health state while the treatment may then obscure this underlying health state (43). Consistent with the approach of Davillas and Pudney (2017), the main model was run with and without the inclusion of binary variables for whether an individual was in receipt of cardiovascular, gastrointestinal, respiratory, central nervous system, infection, endocrine, gynaecological/urinary, cytoxic, nutrition/blood, musculoskeletal, eye/ear, skin or other medications.

Stress may alter an individual's immune function directly through the disturbed regulation of hormones or indirectly by inducing unhealthy behaviours such as smoking or poor diet (8), which are

also risk factors for health decline. We adjust for smoking behaviour (ever smoked [Y/N]) and alcohol consumption (in the last 12 months), exercise (frequency and pace of walking in the last 4 weeks) and diet (quantity of fruit and vegetables consumed in last week) to account for correlation between lifestyle behaviours and the experience of stress or SES. For both the BHPS and non-BHPS samples information on diet, exercise, and alcohol consumption was available for all respondents in wave two. This reflects an earlier period for BHPS respondents rather than the baseline. However we expect this would reflect a reasonable approximation of lifestyle behaviours at baseline for all respondents.

Certain non-cognitive skills appear important for health capital formation and resilience to health shocks (15, 44). These can both shape and be shaped by an individual's AL (45) as well as affecting how they rate their own health (46). As such adjustment were made for the five factor (Big 5) model of personality types (47), modelled as a continuous variable with higher levels indicating greater belonging to a personality type. These have been used to capture non-cognitive skills in similar research (15).

2.5 Data Analysis

Our initial Ordinary Least Squares (OLS) model (Model 1), where *t* represents the baseline, examines the association of AL for individual *i* at baseline (AL_{it}) with SRH two periods post-baseline (SRH_{it+2}). The nurse visit occurred an average of 5 months after the mainstage survey so we focus on SRH at least two periods post-baseline (SRH_{it+2}) so as to ensure that no measurements at baseline could overlap with subsequent SRH. We estimate this model for three different SRH outcomes (PCS, MCS and SF6D) using robust standard errors (48) and also consider SRH outcomes 3 and 4 periods post-baseline (SRH_{it+3}, SRH_{it+4}). By focusing on SRH in periods after the collection of biomarkers (the baseline), we mitigate the issue of reverse causality, i.e. that poor SRH induces higher AL. To further mitigate this issue and to account for confounding factors we adjust for: SRH at baseline (SRH_{it}); SRH in the period

preceding baseline (SRH_{it-1}) to guard against period-specific fluctuations in the rating of health; a vector of socioeconomic variables at baseline (X_{it}); a binary indicator accounting for the wave at which the nurse health assessment was carried out to capture any period specific effects for BHPS and non-BHPS respondents; along with an idiosyncratic error term (ε_{it+q}).

$$SRH_{it+q} = \beta_0 + \beta_1 SRH_{it-1} + \beta_2 SRH_{it} + \beta_3 AL_{it} + \beta_4 X_{it} + e_{it+q}, \forall\, q = 2, 3, 4 \qquad (1)$$

This initial specification (Model 1) allows us to test whether higher AL at baseline is associated with an additive decline in SRH each period. However, we posit that this decline may differ for individuals in good health in the baseline period, i.e. higher depreciation for those with lower baseline SRH (33, 49), given epidemiological evidence suggesting augmented risk of morbidity and earlier mortality for those with higher AL and the representation of stress as depreciation in the Grossman model (33). To account for this, we adopt a second model specification (Model 2), that includes an interaction between baseline SRH and AL, where the coefficient on the interaction (α) can be thought of as varying the depreciation of baseline health across individuals with different AL:

$$SRH_{it+q} = \beta_0 + \beta_1 SRH_{it-1} + \beta_2 SRH_{it} + \beta_3 AL_{it} + \alpha AL_{it} * SRH_{it} + \beta_4 X_{it} + e_{it+q} \qquad (2)$$

The marginal effect of an increase in AL on current health in period *t+q*, thus depends on baseline health ($\frac{\partial SRH_{it+q}}{\partial AL_{it}} = \beta_3 + \alpha SRH_{it}$), while similarly the marginal effect of baseline health on current health depends on baseline AL ($\frac{\partial SRH_{it+q}}{\partial SRH_{it}} = \beta_2 + \alpha AL_{it}$). We hypothesise that $\beta_3 < 0$, $\beta_2 > 0$ and $\alpha > 0$, implying future SRH is decreasing with respect to AL but that this is moderated by higher baseline SRH, while increases in baseline health are associated with greater subsequent health, but

that this would be more beneficial for those with high baseline AL, who are likely to be more vulnerable to ill-health.

Only the baseline values of the covariates were controlled for, since controlling for values in subsequent years may induce post-treatment bias (50). For the associations we observe to reflect causal effects, we would require that conditional on past health and baseline covariates, there are no unobserved confounders influencing both baseline AL and current health. Thus we assume that SRH in future periods is conditionally independent from AL after adjusting for baseline and pre-baseline SRH, and our other control variables:

$$E\left(SRH_{it+q} \perp AL_{it} \,\middle|\, SRH_{it}, SRH_{it-1}, X_{it}\right) \quad (3)$$

This assumption would be violated if for instance unobserved variables such as adverse childhood experiences influenced AL in previous periods but also had a persistent effect that influenced health over and above its effect on health in previous periods (see Figure 2). Ideally one would use an instrumental variable (IV) approach to address this concern, however this would require an IV that would be correlated with AL but which does not directly influence health in the current period. Genetic characteristics may provide one such IV, however this was not available here and, given an AL index is a collection of multiple biomarkers, it would be difficult to rule out horizontal pleiotropy (i.e. unobserved confounding) in a Mendelian Randomisation analysis using genetic characteristics as instruments (51). We explore the robustness of this assumption in sensitivity analysis.

2.5 Sensitivity Analysis

We adjust Model 2 for additional vectors of medication (M_{it}), lifestyle (L_{it}), and non-cognitive variables (P_{it}) at baseline (Models 3-6). To account for bias due to sample selection we apply sample weights to Model 2 (38, 39). While we control for a rich set of baseline covariates, we cannot be confident that there are no material omitted variables or that past values do not have dynamic effects. As such, the inclusion of parents' highest educational achievement was also tested where data were available (N = 5,979). Nonetheless there remains the possibility for observed and unobserved confounding and bias through model misspecification.

In order to account for potential observed confounding and model misspecification, we applied entropy balancing for continuous treatments (52, 53) such that all covariates (socioeconomic, lifestyle, medication and non-cognitive skills) were balanced to remove correlation with the treatment variable (AL) before regression adjustment using the same covariates with extreme weights trimmed (Model 7). Imbens (2004) suggests truncating excessive weights beyond 4% as these may increase the variance of estimates (54); we conservatively trimmed extreme weights at 2% although results are not sensitive to greater trimming. It was not possible to balance "other medications" due to the low number of individuals in receipt of these (< 0.03% of the sample), however this variable is controlled for in the regression adjustment stage.

To examine the extent to which unobserved confounding may explain the association between AL and future SRH, we conduct a bounds analysis on the fully restricted model (Model 6). This involved estimation of the degree of proportional selection required (delta) to imply that our coefficient of interest (AL_{it}) is driven entirely by selection on observables (55, 56).

We examine model fit according to different configurations of the AL index. In particular whether model fit is improved with the inclusion of a stress-related biomarker - in this case dehydroepiandrosterone-sulphate (DHEA-s) - reflecting the original conceptual framework of AL as capturing heightened neural or neuroendocrine activity from repeated or chronic environmental challenge (40, 57). We use Model 6 to test whether our 'z-score AL index' which captures primary mediators and effects, and secondary and tertiary outcomes as part of the cascade of activity in the stress-response (25, 26, 58) demonstrates better fit (based on AIC and BIC) than indices reflecting individual biological systems or whether the 'norm AL index' (21) provides better fit than the z-scored approach.

Results

A total of 7,712 adults (aged over 16 years) were available for analysis, though this was reduced when including information from periods further from the baseline or when including additional variables as part of the sensitivity analysis. In Table 2 we see that socioeconomic variables vary according to AL, with those who are younger, female, higher earners, non-white, never married, more educated, and employed, self-employed, a student, in family care, on maternity leave, or live in urban areas are more likely to have an AL below the median, where lower AL suggests lower dysregulation (i.e. is better). Those with better physical health (PCS) and HRQoL (SF6D) tend to have below median AL, while those with worse mental health (MCS) tend to have above median AL.

[INSERT TABLE 2]

As shown in Figure 3, there is a clear negative correlation between AL at baseline and PCS two periods post-baseline, and slightly less so for SF6D, while there appears to be much less variation in AL across

levels of MCS. The relationship tends to be most pronounced at high values of AL, although sample sizes are somewhat smaller there, as reflected in the widening confidence intervals.

[INSERT FIGURE 3]

Figure 4 presents average SRH (with 95% CI) measured two periods post-baseline across age deciles for those above and below median baseline AL. Those with a lower AL tend to report better health relative to those with higher AL at baseline. While SF6D tends to be stable for both groups across age deciles, PCS appears to decline gradually across age deciles, while MCS increases. For MCS, there is some evidence of convergence between the high and low AL groups. While for PCS, there is evidence of a divergence between the high and low AL groups in older age with signs of narrowing again in the oldest age-groups.

[INSERT FIGURE 4]

Table 3 displays the regression results across SRH measures according to the model laid out in equation one above (Model 1); and with an AL and SRH interaction – reflecting equation two (Model 2). For PCS and SF6D it appears that higher AL at baseline predicts lower SRH ($p < 0.001$). According to Model 1, a one standard deviation (SD) increase in AL is associated with a 0.07 SD reduction in PCS, a 0.07 SD reduction in SF6D, and a 0.04 SD reduction in MCS. In the case of PCS and SF6D a significant interaction is observed between baseline AL and SRH ($p < 0.001$) with the higher R^2 for this specification (Model 2) suggesting their inclusion is warranted, while this is not the case for MCS.

[INSERT TABLE 3]

For SF6D and PCS , this is better illustrated in figure 5 showing the average partial effect of AL on SRH with baseline SRH held at its mean and +/- 1 SD from the mean using multiple SRH outcomes post-baseline (*t+q,* $\forall$ *q = 2, 3 and 4*). For individuals with high baseline SRH (+ 1 SD), a one SD increase in AL has a small and often insignificant association with SRH across all outcomes post-baseline; for SF6D at *t+2* this is -0.02 (95% CI: -0.05 to 0.01) while for PCS this -0.04 (95% CI: -0.07 to -0.02). While for individuals with lower baseline SRH (- 1 SD) a one SD increase in AL has a much larger and significant association with SRH across post-baseline outcomes; for SF6D at *t+2* this is -0.11 (95% CI: -0.14 to -0.08) while for PCS this -0.13 (95% CI: -0.17 to -0.10). Put simply, AL predicts larger and larger reductions in an individuals' SRH (PCS and SF6D) as their baseline SRH deteriorates but not so for MCS.

[INSERT FIGURE 5]

Sensitivity Analyses

Table 4 demonstrates how the inclusion of lifestyle, medication and non-cognitive skills variables (Models 3-6) impacts upon the association between AL and SRH. For SF6D and PCS, a significant association and interaction is still observed. For MCS there remains little evidence of an association between AL and SRH. Although not presented here, there was no meaningful difference in these results when adjusting for sample weights accounting for sample selection or when adjusting for highest parental education.

Model 7, which uses entropy balancing in addition to regression adjustment is broadly consistent with previous models. While the interaction between AL and PCS remains significant, this is not the case for SF6D. Table S2 shows the improvement in covariate balance according to AL while figure S3

illustrates the general consistency across model coefficients by plotting the marginal effects of AL on each outcome for models 2-7.

[INSERT TABLE 4]

Table S4 provides the relative weight of unobservables to observables (delta) required to drive the main coefficient (AL_{it}) to zero. Oster (2019) suggests a maximum R^2 of 1.3 times the R^2 from the fully controlled regression (Model 6). At this level, the degree of selection on unobservables would need to be 64% (SF6D), 26% (PCS) and 85% (MCS) of observables for the coefficient on AL to be equal to zero. Table S5 presents the AIC and BIC for Model 6 while substituting AL for subgroups of biomarkers or the 'Norm AL index'. For SF6D and PCS the full z-score AL index, which accounts for multiple biological systems activated as part of the stress-response, provides the best fit for predicting Sf6D and PCS, though this is not the case for MCS.

Discussion

We examined whether baseline AL predicts future SRH over and above baseline SRH. Others have examined the use of SRH in predicting future AL (29) however an inability to control for baseline AL makes it more difficult to rule out reverse causality. We consider Engman's description of biographical disruption from the medical sociology literature (30), with empirical analysis informed by Grossmans model for the demand for health from the health economics literature (31-33), to examine the converse scenario - AL predicting future SRH. We found that higher AL predicts lower SRH consistent with expectations; an increase in AL by one SD at baseline is associated with a 0.07 ($p < 0.001$) SD decrease in physical health (PCS) and in HRQoL (SF6D) two periods post-baseline. Furthermore the

significant interaction, *ceteris paribus*, between baseline AL and SRH (PCS and SF6D) suggests that SRH decline according to AL is greatest for those already reporting lower physical health or HRQoL, supporting our hypothesis.

This is in line with the original approach by Seeman et al (2001), who formulated the AL index (20), and showed that baseline AL predicted lower physical functioning 7.5 years later over and above baseline measures (25). These authors estimated a physical decline of approximately 0.12 SD (0.06/0.49) for a unit increase in AL. Caution is required in comparing results given this was for a smaller sample of 720 US adults aged 70-79, physical health measures differed, different biomarkers were used, follow-up was longer and information on a number of important confounders was not available. However, that this estimate for a group that would have lower SRH compared to the national average is closer to our estimate for those with baseline PCS 1 SD below the mean (estimate=0.13) is reassuring and further supports our hypothesised interaction between baseline PCS and AL. This is also consistent with evidence demonstrating that higher AL is associated with greater morbidity risk and earlier mortality and supports the notion of AL as reflecting stress-related health depreciation (33, 49). That AL is most significant in predicting physical health depreciation is also consistent with research suggesting that AL biomarkers predict long-term healthcare utilisation and cost via the onset of disability (59, 60). We do not observe a significant association between AL and mental health (MCS) perhaps suggesting that AL does not translate well into an objective measure of mental and social health dimensions.

This observed depreciation is attenuated when adjusting for lifestyle factors, such as smoking, diet and exercise (which has face validity), and when balancing on these factors, baseline SRH and sociodemographic characteristics alongside regression adjustment consistent with intuition. The interaction with SF6D is no longer significant which is unsurprising given this is a weighted

combination of MCS and PCS and that estimates are noisier (larger SE) for entropy balancing. A significant association is still observed in relation to PCS, which is robust to the inclusion of prescribed medications, non-cognitive skills and lifestyle factors.

Our findings have implications for health promotion and education whereby measurement of an individual's AL index predicts longer-term health outcomes and, as shown elsewhere, healthcare utilisation (59, 60). While there are many social, economic, political and cultural factors which underlie over- or under-use of healthcare, behavioural factors such as present bias, symptom salience, and false beliefs are important in this context (61). Providing individuals with information on their underlying health, through indices like AL, may allow them to adjust their health expectations or may be a catalyst for lifestyle change to reduce their AL and potentially avoid biographical disruption. Encouragingly, where reductions in AL have been achieved, individuals have been shown to lower their mortality risk (62). Empowering physicians with tools to measure stress-related "wear and tear" offers information for potential use in patient/doctor interactions – a teachable moment - on the need for lifestyle change (13, 14). This may be especially so for younger individuals for whom symptoms might be least apparent but for whom the lifetime benefit is greatest. While the provision of information on the mortality risks of health-harming behaviours such as smoking have been demonstrated (63), the opportunity to promote broader lifestyle changes supported with personalised evidence likely extends far beyond this example.

Mutable psychosocial traits and teachable non-cognitive skills, for example, can be important in developing resilience to environmental stressors (15, 64, 65) thereby reducing the potential for stress-related "wear and tear" of which AL is a measure. While more research is needed into the reliability and consistency of measures of non-cognitive skills (66), that these may be altered is supported by the effectiveness of psychotherapeutic techniques (e.g. mindfulness-based cognitive therapy) in

randomised controlled trials aimed at enhancing resilience in individuals or improving physical health outcomes (67, 68).

Finally that some individuals unnecessarily face more stressors than others should also be a focus of policy. Increasing income inequality appears to play a causal role in poorer health outcomes, of which stress may be one mechanism along this pathway (6). Socioeconomic disparities in AL are largely governed by circumstances beyond an individual's control rather than their own efforts (69). Focusing on clinical endpoints (i.e. the diagnosis of disease) may fail to identify potentially effective policies aimed to reduce health disparities over relatively short-term time horizons that might otherwise be missed (70); an argument that could now be extended to SRH measures.

Our study has limitations which provide areas of future research. Firstly, we were limited to a single timepoint for measurement of AL here; a common limitation in analyses using AL biomarkers in relation to health outcomes (15, 43, 59, 71). Other data sources may collect data for the construction of an AL index at multiple time periods which would allow a richer modelling of the inter-relationship between AL and SRH. Additionally even with the strengths of this dataset we note that a large majority (97%) of respondent reported their ethnicity as white making inference relating to other ethnicities more challenging.

We cannot rule out unobserved confounding hence estimates do not necessarily represent causal effects. The extent to which unobserved confounding may nullify the coefficient on AL depends on the extent to which the variation in our SRH outcome can be explained fully (R^2=1) (Table S4). This is unrealistic as measurement error is likely to occur. Oster (2019) suggests an R^2-max of 1.3 times the R^2 from the controlled regression. In the case of PCS, R^2-max would be 0.75 and the corresponding delta for the AL coefficient is 0.26. Oster (2019) suggests a lower bound of delta=1 however other

studies argue for a lower delta (72, 73). Our ability to control for baseline and pre-baseline SRH alongside AL and the interaction (Degree of Freedom [DoF]=4) is critical and results in an R^2=0.55, while the inclusion of all other covariates (DoF=60) results in an R^2=0.58. Thus it is difficult to imagine unobserved characteristics with such great predictive power as to overturn our results. Future work could explore the use of instrumental variables (IV) to obtain estimates that are more credibly causal, albeit it may be difficult to ensure potential IVs do not directly impact health, even after conditioning on baseline health.

Finally, alternative measurements of AL and SRH are possible and could capture different aspects of the relationship of interest. While we focus on AL as a measure to examine health depreciation it is important to note that the measurement of AL is evolving (74) and research is ongoing as to which biomarkers should be combined and how to form an AL index (58). Other separate although perhaps inter-related measures also exist which attempt to measure health depreciation beyond ageing (75, 76). Thus we do not propose that the measurement of AL used here is the definitive measure of stress-related health depreciation though, from the biomarkers available in this dataset, we do find that the z-score AL index which captures the cascade of activity as part of the stress response is the best predictor of future physical health and HRQoL (Table S5).

Conclusion

Allostatic Load, as a measure of stress-related "wear and tear", predicts subsequent declines in self-rated physical health and to some extent HRQoL; a decline which appears to accelerate as health declines further. This is akin to the notion of depreciation of health. In this case, it is thought to be a result of individuals' experiences of stress over their lifetime which leads to dysregulation across multiple biological systems and, as we have shown, accelerated physical health decline. This provides potentially valuable insights that may be useful in clinical practice, or in monitoring certain policy

outcomes or interventions especially with shorter time horizons. While the causes of such stress may be multi-faceted and require further research, policies which target younger individuals in general, through school curricula for example, to develop non-cognitive skills may be effective in delaying an individual's entry into the disease span of their lifetime.

References

1. Barnett K, Mercer SW, Norbury M, Watt G, Wyke S, Guthrie B. Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. The Lancet. 2012;380(9836):37-43.
2. Cassell A, Edwards D, Harshfield A, Rhodes K, Brimicombe J, Payne R, et al. The epidemiology of multimorbidity in primary care: a retrospective cohort study. Br J Gen Pract. 2018;68(669):e245-e51.
3. Kingston A, Robinson L, Booth H, Knapp M, Jagger C. Projections of multi-morbidity in the older population in England to 2035: estimates from the Population Ageing and Care Simulation (PACSim) model. Age and Ageing. 2018;47(3):374-80.
4. Huber M, Knottnerus JA, Green L, van der Horst H, Jadad AR, Kromhout D, et al. How should we define health? BMJ: British Medical Journal (Online). 2011;343.
5. Jambroes M, Nederland T, Kaljouw M, van Vliet K, Essink-Bot M-L, Ruwaard D. A new concept of health—implications for public health policy and practice: a qualitative analysis. The Lancet. 2014;384:S39.
6. Pickett KE, Wilkinson RG. Income inequality and health: A causal review. Social Science & Medicine. 2015;128:316-26.
7. Kajantie E. Fetal Origins of Stress-Related Adult Disease. Annals of the New York Academy of Sciences. 2006;1083(1):11-27.
8. Cohen S, Janicki-Deverts D, Miller GE. Psychological Stress and Disease. JAMA. 2007;298(14):1685-7.
9. Steel N, Ford JA, Newton JN, Davis ACJ, Vos T, Naghavi M, et al. Changes in health in the countries of the UK and 150 English Local Authority areas 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. The Lancet. 2018;392(10158):1647-61.
10. Dieleman JL, Baral R, Birger M, Bui AL, Bulchis A, Chapin A, et al. US Spending on Personal Health Care and Public Health, 1996-2013Spending on US Health Care, 1996-2013Spending on US Health Care, 1996-2013. JAMA. 2016;316(24):2627-46.
11. Whitty CJM, MacEwen C, Goddard A, Alderson D, Marshall M, Calderwood C, et al. Rising to the challenge of multimorbidity. BMJ. 2020;368:l6964.
12. Centre for Ageing Better. A consensus on health ageing. Public Health England,; 2019.
13. Westmaas JL, Newton CC, Stevens VL, Flanders WD, Gapstur SM, Jacobs EJ. Does a Recent Cancer Diagnosis Predict Smoking Cessation? An Analysis From a Large Prospective US Cohort. Journal of Clinical Oncology. 2015;33(15):1647-52.

14. Xiang X. Chronic Disease Diagnosis as a Teachable Moment for Health Behavior Changes Among Middle-Aged and Older Adults. Journal of Aging and Health. 2016;28(6):995-1015.
15. Atkins R, Turner AJ, Chandola T, Sutton M. Going beyond the mean in examining relationships of adolescent non-cognitive skills with health-related quality of life and biomarkers in later-life. Economics & Human Biology. 2020;39:100923.
16. London School of Economics, Personal Social Services Research Unit. Commissioning cost-effective services for promotion of mental health and wellbeing and prevention of mental ill-health. London: Public Health England; 2017.
17. Kunzler AM, Helmreich I, König J, Chmitorz A, Wessa M, Binder H, et al. Psychological interventions to foster resilience in healthcare students. Cochrane Database of Systematic Reviews. 2020(7).
18. Kunzler AM, Helmreich I, Chmitorz A, König J, Binder H, Wessa M, et al. Psychological interventions to foster resilience in healthcare professionals. Cochrane Database of Systematic Reviews. 2020(7).
19. FDA-NIH BiomarkerWorking Group. BEST (Biomarkers, EndpointS, and other Tools) Resource, 2016 [Available from: https://preview.ncbi.nlm.nih.gov/books/NBK326791/.
20. McEwen BS, Seeman T. Protective and Damaging Effects of Mediators of Stress: Elaborating and Testing the Concepts of Allostasis and Allostatic Load. Annals of the New York Academy of Sciences. 1999;896(1):30-47.
21. Juster R-P, McEwen BS, Lupien SJ. Allostatic load biomarkers of chronic stress and impact on health and cognition. Neuroscience & Biobehavioral Reviews. 2010;35(1):2-16.
22. Geronimus AT, Hicken M, Keene D, Bound J. "Weathering" and Age Patterns of Allostatic Load Scores Among Blacks and Whites in the United States. American Journal of Public Health. 2006;96(5):826-33.
23. Danese A, McEwen BS. Adverse childhood experiences, allostasis, allostatic load, and age-related disease. Physiology & Behavior. 2012;106(1):29-39.
24. Lipowicz A, Szklarska A, Mitas AW. Biological costs of economic transition: Stress levels during the transition from communism to capitalism in Poland. Economics & Human Biology. 2016;21:90-9.
25. Seeman TE, McEwen BS, Rowe JW, Singer BH. Allostatic load as a marker of cumulative biological risk: MacArthur studies of successful aging. Proceedings of the National Academy of Sciences of the United States of America. 2001;98(8):4770-5.
26. Robertson T, Beveridge G, Bromley C. Allostatic load as a predictor of all-cause and cause-specific mortality in the general population: Evidence from the Scottish Health Survey. PLOS ONE. 2017;12(8):e0183297.
27. Guidi J, Lucente M, Sonino N, Fava GA. Allostatic Load and Its Impact on Health: A Systematic Review. Psychotherapy and Psychosomatics. 2021;90(1):11-27.
28. Robb MA, McInnes PM, Califf RM. Biomarkers and Surrogate Endpoints: Developing Common Terminology and Definitions. JAMA. 2016;315(11):1107-8.
29. Vie TL, Hufthammer KO, Holmen TL, Meland E, Breidablik HJ. 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 & Medicine. 2014;117:1-9.
30. Engman A. Embodiment and the foundation of biographical disruption. Social Science & Medicine. 2019;225:120-7.
31. Grossman M. The demand for health: a theoretical and empirical investigation. NBER Books. 1972.

32. Grossman M. The human capital model. Handbook of health economics. 1: Elsevier; 2000. p. 347-408.
33. Galama TJ, van Kippersluis H. A Theory of Socio-economic Disparities in Health over the Life Cycle. The Economic Journal. 2018;129(617):338-74.
34. Galama T, Kapteyn A. Grossman's missing health threshold. Journal of health economics. 2011;30(5):1044-56.
35. National Institute for Health and Care Excellence. NICE technology appraisal guidance 2019 [Available from: https://www.nice.org.uk/About/What-we-do/Our-Programmes/NICE-guidance/NICE-technology-appraisal-guidance.
36. Health Information and Quality Authority. Guidelines for the Economic Evaluation of Health Technologies in Ireland 2018. 2018.
37. Brazier JE, Roberts J. The Estimation of a Preference-Based Measure of Health from the SF-12. Medical Care. 2004;42(9):851-9.
38. McFall S, Petersen J, Kaminska O, Lynn P. Understanding Society—The UK Household Longitudinal Study: Waves 2 and 3 Nurse Health Assessment, 2010–2012 Guide to Nurse Health Assessment. Colchester, UK: Institute for Social and Economic Research, University of Essex; 2013.
39. Benzeval M, Davillas A, Kumari M, Lynn P. Understanding society: the UK household longitudinal study biomarker user guide and glossary. Institute for Social and Economic Research, University of Essex. 2014.
40. Johnson SC, Cavallaro FL, Leon DA. A systematic review of allostatic load in relation to socioeconomic position: Poor fidelity and major inconsistencies in biomarkers employed. Social Science and Medicine. 2017;192:66-73.
41. Ware JE, Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: Construction of Scales and Preliminary Tests of Reliability and Validity. Medical Care. 1996;34(3):220-33.
42. Case A, Deaton AS. Broken down by work and sex: How our health declines. Analyses in the Economics of Aging: University of Chicago Press; 2005. p. 185-212.
43. Davillas A, Pudney S. 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. 2017;56:87-102.
44. Heckman JJ. The developmental origins of health. Health economics. 2012;21(1):24-9.
45. Stephan Y, Sutin AR, Luchetti M, Terracciano A. Allostatic Load and Personality: A 4-Year Longitudinal Study. Psychosomatic medicine. 2016;78(3):302-10.
46. Stephan Y, Sutin AR, Luchetti M, Hognon L, Canada B, Terracciano A. Personality and self-rated health across eight cohort studies. Social Science & Medicine. 2020;263:113245.
47. McCrae RR, John OP. An introduction to the five-factor model and its applications. Journal of personality. 1992;60(2):175-215.
48. StataCorp. Stata Statistical Software: Release 14. College Station: TX: StataCorp LP; 2015.
49. Bhattacharya J, Hyde T, Tu P. Health economics: Macmillan International Higher Education; 2013.
50. Gelman A, Hill J. Data analysis using regression and multilevel/hierarchical models: Cambridge university press; 2006.
51. Didelez V, Sheehan N. Mendelian randomization as an instrumental variable approach to causal inference. Statistical Methods in Medical Research. 2007;16(4):309-30.

52. Tübbicke S. Entropy Balancing for Continuous Treatments. arXiv preprint arXiv:200106281. 2020.
53. Crump RK, Hotz VJ, Imbens GW, Mitnik OA. Dealing with limited overlap in estimation of average treatment effects. Biometrika. 2009;96(1):187-99.
54. Imbens GW. Nonparametric estimation of average treatment effects under exogeneity: A review. Review of Economics and statistics. 2004;86(1):4-29.
55. Oster E. Unobservable Selection and Coefficient Stability: Theory and Evidence. Journal of Business & Economic Statistics. 2019;37(2):187-204.
56. Oster E. PSACALC: Stata module to calculate treatment effects and relative degree of selection under proportional selection of observables and unobservables. 2016.
57. McEwen BS, Stellar E. Stress and the individual: Mechanisms leading to disease. Archives of Internal Medicine. 1993;153(18):2093-101.
58. Wiley JF, Gruenewald TL, Karlamangla AS, Seeman TE. Modeling Multisystem Physiological Dysregulation. Psychosomatic medicine. 2016;78(3):290-301.
59. Davillas A, Pudney S. Biomarkers, disability and health care demand. Economics & Human Biology. 2020:100929.
60. Davillas A, Pudney S. Using biomarkers to predict healthcare costs: Evidence from a UK household panel. Journal of Health Economics. 2020;73:102356.
61. Baicker K, Mullainathan S, Schwartzstein J. Behavioral hazard in health insurance. The Quarterly Journal of Economics. 2015;130(4):1623-67.
62. Karlamangla AS, Singer BH, Seeman TE. Reduction in Allostatic Load in Older Adults Is Associated With Lower All-Cause Mortality Risk: MacArthur Studies of Successful Aging. Psychosomatic Medicine. 2006;68(3):500-7.
63. Carbone JC, Kverndokk S, Røgeberg OJ. Smoking, health, risk, and perception. Journal of Health Economics. 2005;24(4):631-53.
64. Chen E, Miller GE, Lachman ME, Gruenewald TL, Seeman TE. Protective factors for adults from low-childhood socioeconomic circumstances: the benefits of shift-and-persist for allostatic load. Psychosomatic medicine. 2012;74(2):178-86.
65. Kautz T, Heckman JJ, Diris R, Ter Weel B, Borghans L. Fostering and measuring skills: Improving cognitive and non-cognitive skills to promote lifetime success. National Bureau of Economic Research; 2014. Report No.: 0898-2937.
66. Morris TT, Smith GD, van Den Berg G, Davies NM. Investigating the longitudinal consistency and genetic architecture of non-cognitive skills, and their relation to educational attainment. bioRxiv. 2019:470682.
67. Creswell JD, Lindsay EK, Villalba DK, Chin B. Mindfulness Training and Physical Health: Mechanisms and Outcomes. Psychosomatic Medicine. 2019;81(3):224-32.
68. Fava GA, Guidi J. The pursuit of euthymia. World Psychiatry. 2020;19(1):40-50.
69. Carrieri V, Davillas A, Jones AM. A latent class approach to inequity in health using biomarker data. Health Economics. 2020;29(7):808-26.
70. Seeman T, Epel E, Gruenewald T, Karlamangla A, McEwen BS. Socio-economic differentials in peripheral biology: Cumulative allostatic load. Annals of the New York Academy of Sciences. 2010;1186(1):223-39.
71. Davillas A, Pudney S. Biomarkers as precursors of disability. Economics & Human Biology. 2019:100814.
72. Evans BJ. How college students use advanced placement credit. American Educational Research Journal. 2019;56(3):925-54.

73. Dackehag M, Ellegård LM, Gerdtham U-G, Nilsson T. Social assistance and mental health: evidence from longitudinal administrative data on pharmaceutical consumption. Applied Economics. 2020;52(20):2165-77.
74. Howard JT, Sparks PJ. 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. 2016;28(5):627-35.
75. Cawthon RM, Smith KR, O'Brien E, Sivatchenko A, Kerber RA. Association between telomere length in blood and mortality in people aged 60 years or older. The Lancet. 2003;361(9355):393-5.
76. Horvath S. DNA methylation age of human tissues and cell types. Genome biology. 2013;14(10):R115-R.