Is teaching bad for your health? New evidence from biomarker data

Sam Sims*
John Jerrim*
Hannah Taylor*
Rebecca Allen*

UCL Institute of Education

Teaching is a demanding job and research suggests that prolonged exposure to stress can affect physical health. While some studies have found that teachers do indeed report relatively poor physical health, the existing literature has important methodological limitations. In particular, no research exists comparing teachers to other occupations using objective biomarker data to measure health. We provide such evidence using two datasets: a representative, cross-sectional survey and a longitudinal convenience sample. We find no statistically significant overall association between teaching and physical health in any of our models or datasets. Teaching may therefore not be as bad for physical health as previously thought.

Key Words: Teachers; Health; Allostatic Load; Biobank; UK Household Longitudinal Study

1. Introduction

1.1 Stress and health among teachers

There is a long interdisciplinary tradition of investigating how individuals' occupation affects their physical and mental health, and in particular their levels of stress (Cooper & Marshall, 1976). Indeed, a recent review of the literature comparing teachers' mental health with other occupations (Van Droogenbroeck & Spruyt, 2015) found five studies that looked specifically at stress. Four of these concluded that that teaching was a relatively high stress occupation (Johnson et al., 2005; Schaufeli, Daamen & Van Mierlo, 1994; Smith et al., 2000; Health & Safety Executive, 2014), while the fifth did not find any difference (Pithers & Fogarty, 1995). Although all five studies have important methodological limitations, two large-scale surveys from the US (Gallup, 2014) and UK (Worth & Van den Brande, 2019) have also found that teaching is a relatively stressful occupation.

Occupational stress occurs when aspects of the work environment constrain or impose very high demands on an individual, threatening their ability to achieve their goals (Schuler, 1980). Research suggests that teachers experience high levels of demand in the form of pupil behaviour (Harmsen, Helms-Lorenz, Maulana, & van Veek, 2018; Cornell & Mayer, 2010), time pressure (Kovess-Masféty, Rios-Seidel, & Sevilla-Dedieu, 2007; Mujtaba & Reiss, 2013) and accountability reforms (Berryhill, Linney, & Fromewick, 2009; Hardy; Ronnerman & Beach, 2019). In addition, teachers report having less time control, lower participation in decision making, and less colleague support than those in other professions, which further contribute to the demands of the job (Heus & Diekstra, 1999).

The brain responds to the threats involved in stressful situations by sending messages along neuroendocrine pathways, resulting in metabolic and physiological changes within the body (Marmot & Wilkinson, 2005). These changes are normally adaptive, in that they help the individual respond to an acute threat. When an individual experiences prolonged exposure to stress, however, these changes can become maladaptive (McEwen, 1998), resulting in physical ill health (Beckie, 2012; Thoits, 2010). Prolonged stress can also prompt behavioural responses such as alcohol consumption, which themselves have implications for physical health (Head, Stansfeld, & Siegrist, 2004). If teachers do indeed experience higher levels of stress than other occupations, then it seems likely that this will translate into physical ill health.

1.2 Existing comparative research on teacher health

Some empirical research comparing teachers to non-teachers finds that the former do indeed experience worse general physical health, although these studies have important limitations. For example, Johnson et al. (2005) conducted a cross-sectional survey of individuals in 26 occupational groups and found that teachers had inferior self-reported general health than all but one of the other occupations. However, the paper does not employ representative data and provides only crude rankings. A similar survey of a representative group of secondary school teachers in Belgium also found that they reported worse health than a comparison group (Bogaert et al., 2014). However, this comparison group was drawn from a convenience sample, making the comparison hard to interpret. Studies comparing specific medical conditions also find evidence for higher incidence of certain health conditions among teachers. For example, a large cross-sectional survey in France found higher lifetime prevalence of infectious diseases such as rhinopharyngitis/laryngitis, conjunctivitis and bronchitis, which may reflect the large number of people that many teachers interact with in the workplace (Kovess-Masféty et al., 2006).

The literature is not entirely consensual, however. For example, large-scale survey research in Germany found that teachers report lower levels of cardiovascular disease than other occupational groups (Helmert, Shea, & Banmman, 1997). One plausible explanation for this is that teachers tend to stand up while delivering instruction and consequently experience greater levels of low-to-moderate physical activity at work than other occupations, which has in turn been linked to improved health (Stamatakis et al., 2013; Tudor-Locke, Ainsworth, Washington & Troiano, 2011). Teachers also display lower levels of tobacco smoking than other occupational groups (Gilbert et al., 2015), possibly because many schools ban onsite smoking altogether. Thus, certain characteristics of the job may compensate – partly or fully – for effects of stress on physical health in teachers. Kovess-Masféty et al. (2006) also find no difference between teachers and other occupations in terms of hypertension, which is an important indicator of overall health.

An important limitation of the existing literature is that it is dependent on self-report measures, either of perceived/subjective health or self-reports of diagnosed conditions. By contrast, few empirical studies have investigated the health of teachers using objective biomarker data. Such data has been used in related literature to show that workplace stressors predict higher concentrations of stress biomarkers, such as cortisol, among teachers (Bellingrath, Weigl and Kudielka., 2009; Masilamani et al., 2012; Qi et al., 2014; Wolfram, Bellingrath, Feuerhahn, & Kudielka, 2013). In addition, Bellingrath, Rohleder and Kudielka (2010) hypothesised that, because acute stress activates the immune system, chronic stress might also lead to persistent changes in the functioning of this system. Consistent with this, they found that teachers reporting certain sources of stress at work do indeed display a dampened immune response to acute stressors. However, no researchers have used objective biomarker data to compare health across different occupational groups. In addition, none of the research on teacher health has employed longitudinal data.

1.3 Aims

This study aims to address these limitations and provide new evidence as to whether teaching is associated with poor health outcomes using two datasets. The first is a cross-sectional household survey, which used nurse visits to collect biomarker data from a subsample of respondents broadly representative of the UK. The second dataset is a large convenience sample that incorporates a longitudinal component for a subsample of participants. We use these datasets to construct an index of teacher health similar to that in Bellingrath, Weigl and Kudielka (2009) for a large sample and compare this index across occupations (using the representative data) and over time (using the longitudinal data). Based on the existing literature, we tentatively hypothesised that teachers would have lower levels and faster deterioration in health relative to otherwise similar individuals in other occupations.

2. Health

2.1 Operational definition

The World Health Organisation define health as "a state of complete physical, mental and social well-being and not merely the absence of disease or infirmity." This definition has, however, been criticised on the grounds that the word 'complete' means it can only be operationalised in a binary way and therefore fails to capture important variation (Larson, 1999). Drawing on the work of Georges Canguilhem (1978), it has been proposed that health instead be defined as "the ability of an organism to maintain a balance with its environment, with relative freedom from pain, disability, or limitations, including social abilities" (Larson, 1999, p. 131). By focusing on the *ability to adapt*, this definition does a better job of capturing the full range of variation in health, from pre-clinical declines in health, such as a weakened immune system, right through to fatal disease. This is the definition we adopt in the present research.

There are a number of options for operationalising this definition of health. As discussed, the approach most common in the social science and education literature is to use subjective self-reported data from questionnaires. However, this is arguably best suited to studying teacher stress and mental health (for a review, see: Droogenbroeck & Spruyt, 2015) and it is unclear whether such self-reported measures capture the *ability to adapt* aspect of our definition of health. In any case, self-reported health is problematic as a measure in longitudinal research due to ceiling effects and a lack of measurement invariance across age groups (Gunasekara, Carter, & Blakely, 2012; Zajacova & Woo, 2016). An alternative approach would be to use behavioural measures such as prescriptions taken from medical records. However, these measures are plausibly confounded by occupation, since certain groups may find it easier to leave work in order to visit a doctor, depending on the nature of their jobs. Moreover, they capture only clinical levels of ill-health, missing out on lower-level pre-clinical variation. Hence, we take a different approach, drawing instead on the concept of allostasis to measure health using an index of allostatic load.

2.2 Allostasis and allostatic load

Homeostasis refers to a stable equilibrium between interdependent elements of a system, such as the cardiovascular/respiratory, metabolic, immune and neuroendocrine systems in the human body. The term allostasis was coined to describe the way in which the body has to adapt in response to certain types of changes in the environment in order to maintain this homeostasis (Sterling & Eyer, 1988). Allostasis can therefore be thought of as *change in order to preserve* and is closely aligned with our definition of health. While allostasis is normally beneficial, McEwan and Stellar (1993) theorised that prolonged allostatic responses lead to overactivation or dysregulation of certain bodily systems. Allostatic load is therefore defined as "the wear and tear on the body and brain resulting from chronic overactivity or inactivity of physiological systems that are normally involved in adaptation to environmental challenge" (McEwen, 1998).

The sequence of events by which allostatic load accumulates is often referred to as a cascade. This begins with the perception of a threat, followed by a response from the brain in which chemical messengers (primary mediators) are released to the rest of the body, which produce cellular changes (primary effects), which in turn produce an integrated response to the threat (secondary outcome) and, in cases where the stress is chronic, can lead to disease (tertiary outcomes) (McEwan & Seaman, 1999; for a diagrammatic illustration of this cascade, see:

Beckie, 2012). For example, when the brain perceives an aggressor in the environment, it releases glucocorticoid hormones (such as cortisol) which help to mobilise and regulate the immune response to an injury (McEwen, 2003). Prolonged exposure to these hormones can lead cells to become insensitive to glucocorticoids, disinhibiting the release of inflammatory proteins from immune cells, which can bring about a chronic low-grade inflammatory state and, subsequently, autoimmune conditions (Sheilds & Slavich, 2017). For an overview of other such sequences mediating this process, see McEwan (2003).

2.3 Allostatic load and health

Allostatic load was first measured via an Allostatic Load Index (ALI) in the MacAthur Studies of Successful Aging by collecting data on ten biomarkers located at the primary mediator, primary effect and secondary outcome stages of the biological response cascade (Seeman et al., 2017). The ALI was found to correlate with increased risk of 7-year and 12-year mortality, as well as cognitive and physical decline in a sample aged 70-79 (Gruenewald et al., 2006; Seeman et al., 2001). Since then, higher ALI scores have been shown to be associated with increased risk of mortality in ageing studies in Taiwan, Sweden and the UK. A similar finding has also emerged from a separate general population survey in the USA (Beckie, 2012). Measures of ALI have also been shown to correlate with reduced self-rated health and a range of specific health conditions (for a review, see: Juster, McEwan, & Lupien, 2010). We construct a similar index to measure health in the present study.

3. Methods

3.1 Data

The first of the two datasets employed in this study is the UK Household Longitudinal Study; also known as Understanding Society (USoc) (University of Essex. Institute for Social and Economic Research, NatCen Social Research and Kantar Public, 2009-2015). USoc is a household panel survey, designed to be representative of the UK, which collects data through face-to-face interviews in participants' homes. The survey includes approximately 40,000 households and, at the time of writing, there were eight waves of data available, collected between 2009 and 2018. The data includes variables recording participants occupation at each wave (recorded as four-digit SOC codes), as well as extensive socio-demographic information on e.g. age, ethnicity and income. Between 2010 and 2012 trained nurses also visited a subsample of households in order to collect additional interview, anthropometric and blood sample data relating to participants' health. A total of 15,591 eligible adults (59% response rate) participated in the nurse health assessment with 10,175 (38%) providing a blood sample (McFall et al., 2014). We restricted the data to those of working age (21-60 years old) at the time of the nurse visit, leaving 7,286 observations.

The second dataset employed in this study is UK Biobank (UKB). This is a convenience sample survey comprising interview, cognitive, blood and urine sample measures collected from assessment centres located around the UK. Initial data collection took place between 2006 and 2010 and in total around half a million volunteers (all between the ages of 40 and 69) participated in the study. The dataset also includes important socio-demographic information including age and ethnicity. As with USoc, we restrict the data to those that are clearly of working age (below 60). This leaves us with 230,455 observations with a known occupation at the initial assessment centre. In 2016, 117,500 participants completed a follow-up 'occupational career' questionnaire in which they recorded the start and end dates for all of their previous spells of employment, allowing us to identify the years (if any) in which participants were working as teachers. In 2012 and 2013, all participants living within a 35km radius of the main assessment centre in Stockport (England) were asked to attend a follow-up assessment centre visit. Approximately 20,000 participants (21% response rate) attended and contributed a second wave of survey data and biomarker data, adding a longitudinal component to the UKB data.

Both USoc and UKB record occupation using standard occupation classification (SOC) codes. We defined teachers as anyone with one of the following four SOC codes: 2312 - further education teaching professionals; 2314 - secondary education teaching professionals; 2315 - primary and nursery education teaching professionals; 2316 - special needs education teaching professionals.

Table 1 shows counts and characteristics of both teachers and non-teachers in the USoc and UKB datasets, as well as in the UKB longitudinal subsample. Due to the age restricted sample design, UKB respondents have a higher average age than USoc respondents. In line with findings in the existing literature, teachers are more likely to be female, hold a degree and be born within the UK, relative to non-teachers (Author, 2018). Reassuringly, given that UKB is a pure convenience sample, the gender and ethnic makeup is very similar to USoc. The variable on which the two datasets clearly do diverge is the proportion with a degree. UKB respondents are substantially more likely to be graduates, whether or not they are teachers. Within the UKB, however, there are few differences between the cross-sectional and longitudinal subsamples.

<Table 1>

3.2 Allostatic load index

The original MacArther study (Seeman et al., 2001) employed ten biomarkers in an index of allostatic load: four primary mediators and six secondary outcomes. For each biomarker, an individual was given a score of 1 if they were located in the highest-risk quartile of the distribution and these scores where then summed to give an overall score between 0 and 10. In subsequent studies, a very wide range of biomarkers have been used. Indeed, a recent systematic review of workplace-based studies, found that across sixteen articles a total of 39 unique variables were used in different ALI, with between six and seventeen used in any one index (Mauss et al., 2015). Primary mediators were less likely to be measured than secondary outcomes, meaning that the indices tended to measure later stages of the biological stress-response cascade.

To our knowledge, there have been two other papers that have so far constructed an ALI from the USoc data: Chandola & Zhang (2018) and Chandola et al. (2019). We construct our index using the same set of 12 biomarkers (and adjustments for medication use) employed in Chandola & Zhang (2018), incorporating two primary mediators and ten secondary outcomes. Unfortunately, not all of these same biomarkers are available in the UKB data and we are not aware of any existing studies that use this UKB data to construct an ALI. We therefore selected our set of biomarkers by using all of those available in UKB which are also listed in the review of biomarkers by Juster et al., (2010). The one exception is glucose, which we exclude on the grounds that Hba1c provides a more reliable measure of long-run glucose dysregulation. This leaves us with a partially overlapping set of eleven UKB biomarkers, all of which constitute secondary outcomes. In both the USoc and UKB datasets, we follow the well-established convention of sum scoring a binary indicator of being in the highest-risk quartile for each biomarker (Beckie, 2012; Juster et al., 2010; Mauss et al., 2015).

Table 2 summarises the differences between our two ALIs and briefly elaborates on the biological significance of each biomarker. It is clear from this table that neither index measures the primary mediators particularly well. Indeed, the UKB index includes no primary mediators. Both of the ALI indices should therefore be interpreted primarily as indexing the secondary outcome stage of the stress-response cascade. Figure 1 shows the distribution of the ALI for the representative USoc data, for both teachers and non-teachers. This unadjusted comparison suggests that teachers have lower ALI (i.e. better health) than other working age adults. In

Figure 3 (Supplementary Online Material), we show the mean ALI for a range of occupations, which confirms that there is clear variation by occupational group.

<Table 2>

<Figure 1>

Since social scientists may be less familiar with the use of ALI as a measure of health, Figure 2 shows evidence of convergent validity for the index. The left-hand panel shows the relationship between ALI and age. As theory would predict, and as has been observed in other data, ALI is positively correlated with age in our USoc sample (Beckie, 2012). More precisely, a 10-year increase in age is associated with a 0.26 increase in the ALI (p<0.01). The right-hand panel shows the relationship between ALI and a self-reported measure of health in which USoc participants responded to the statement “In general, would you say your health is...” on a four-point categorical scale covering “excellent”, “very good”, “fair” or “poor”. Again, there is a clear relationship in the expected direction. This provides reassurance that our ALI captures variation in health in the intended way.

<Figure 2>

3.3 Statistical analysis

As can be seen from Figure 1, our outcome measures are count variables. More specifically, they are over-dispersed count variables (mean=2.69 variance=3.65 in UKB). We therefore employ negative binomial regression to estimate our models. Coefficients in the regression tables are reported as incidence rate ratios. Since ALI has been found to be correlated with age, sex, ethnicity and education (Beckie, 2012) - and these variables are all involved in the process of health and ageing - we also utilise these as controls. Our main model is therefore specified as:

$$ln(\widehat{ALI_i}) = \beta_0 + \beta_1 Teach_i + \beta_2 Age_i + \beta_3 Age_i^2 + \beta_3 Male_i + \beta_4 Ethnicity_i + \beta_5 Degree_i \qquad (1)$$

Where:

- ALI_i is the allostatic load index for individual i
- $Teach$ is either a binary or cumulative measure of exposure to teaching
- Age is a continuous variable
- $Male$ and $Degree$ are binary variables
- $Ethnicity$ is a categorical variable (see Table 2)

In our longitudinal model this is adapted to include a baseline ALI measure:

$$ln\left(\widehat{ALI_i^{t2}}\right) = \beta_0 + \beta_1 Teach_i^{t2} + \beta_2 X_i^{t2} + \beta_5 ALI_i^{t1} \quad (2)$$

Where superscript ^{t1} *and* ^{t2} indicates measurement at time 1 and time 2, with t2>t1. X_i^{t2} represents a vector of all the same control variables listed in the first model, for compactness.

The β_1 coefficient in models (1) and (2) compares health among teachers and *all* non-teachers with similar demographic characteristics and education levels. The results are therefore informative about the association between teaching and health, relative to a broad range of alternative occupations that recruit graduates. In order to provide more focused insights, we also run a slightly different model in which we replace the teaching dummy variable with a vector of dummy variables indicating whether an individual belongs to a set of specific occupations. The analysis here is conducted with the cross-sectional UKB subsample in order to maximise sample size within the occupational subgroups. We also cluster occupations within SOC Minor Groups for the same reason. In order to keep the regression table interpretable, we select six such occupational groups for comparison: accountants/consultants; health professionals; planners/surveyors; protective officers; research professionals and welfare professionals. While we recognise that the choice of comparators is somewhat arbitrary, we have tried to select a broad range of occupations that represent plausible and realistic alternatives for the average teacher.[1] Our comparators are also similar to those in existing research (e.g. Worth & Van den Brande, 2019).

$$ln\left(\widehat{ALI_i}\right) = \beta_0 + \beta_1 Occupation_i + \beta_2 Age_i + \beta_3 Age_i^2 + \beta_3 Male_i + \beta_4 Ethnicity_i \quad (3)$$

Throughout the analysis we apply the cross-sectional nurse visit weights provided with the USoc data in order to account for the complex survey design and observable patterns of nonresponse. UKB represents a convenience sample and weights are therefore not available.

4. Results

Table 3 presents the results from our regression models. Column 1 shows the results of regressing an indicator of whether an individual was employed as a teacher at the time of the nurse visit in the USoc data on their ALI score. Column 2 uses a very similar cross-sectional regression but using the much larger Biobank sample. Because ALI is a cumulative measure, we might be concerned that a binary measure of teaching at the time of the nurse visit (in USoc) or the initial assessment centre (in Biobank) is ignoring variation in lifetime exposure to teaching. Fortunately, the career histories in Biobank allow us to construct a measure of total years of teaching and Column 3 shows the results from regressing this on ALI. The Biobank sample is also older, which should make it easier to detect the cumulative wear-and-tear involved in allostatic load[2]. Column 4 assesses cumulative effects in a different way by looking at whether those who worked as teachers both at the initial Biobank assessment centre (2006-2010) and at the follow-up assessment centre (2012-2013) showed worse health than those who did not work as a teacher across this period, conditional on their health at the time of the initial assessment centre (see Model 2 in section 3.3)[3].

[1] We have not included elite occupations such as barristers and architects, for example.
[2] The number of observations in Column 3 drops because only a subsample completed the separate occupational history questionnaire.
[3] The number of observations in Column 4 drops again because only the subsample who both completed the occupational history questionnaire and attended the follow-up assessment centre are included.

<Table 3>

Across the columns, the covariates enter the models with the expected signs. For example, graduates and those born in the UK have a lower ALI, conditional on the other variables (Beckie, 2012). In Column 1, working as a teacher is associated with a small decrease in the ALI. The incidence rate ratio of 0.938 indicates that teaching is associated with 6% reduction in the ALI, however this is not statistically significant at conventional levels. The analogous regression using the UKB data (column 2) suggests that the difference between teachers and other working age adults is effectively zero (incidence rate ratio = 1.004). Taken together, these findings suggest that the raw difference in ALI depicted in Figure 1 (where teachers have slightly lower ALI scores than other working age adults) is accounted for by the other covariates.

If teaching damages health through cumulative exposure to stress, then we might expect longer periods of teaching to be associated with worse health. In order to investigate this, in Column 3, the ALI is regressed upon number of years spent teaching for the subsample of individuals who completed the occupational history questionnaire, including those who have never taught. Again, there is effectively no association (incidence rate ratio = 1.0001 for each additional year of teaching). One explanation for a lack of any association between teaching and health is that more or less healthy people select into the profession, thus offsetting any causal effect of teaching. As a check on this, in Column 4, we regress the ALI at the follow-up assessment centre visit on an indicator of whether an individual was teaching at both the initial and follow-up assessment centre, controlling for the ALI at the initial assessment centre. The sample for this analysis is restricted to those who attended more than one assessment centre. Again, there is very little association between teaching and change in the ALI (incidence rate ratio = 0.954).

Table 4 focuses the comparison on specific occupational groups, with the sample size dropping to 9,356 as a result.[4] Teaching serves as the reference category and the six comparator occupational groups can be seen down the left-hand side of the table (a detailed breakdown of each can be found in the notes to the table). Four of the six groups – accountants/consultants, planners/surveyors, protective officers and research professionals – show coefficients very close to one (0.98-1.03) and are not statistically significantly different to teachers. The other two groups show a statistically significant difference with teachers. However, these go in opposite directions: health professionals are slightly healthier than teachers and welfare professionals are slightly less healthy. Overall, teachers are no more or less healthy than these comparators.

<Table 4>

4.1 Sensitivity checks

In order to check the sensitivity of our main results, we run several alternative specifications of our models. Column 5 in Table 5 (Supplementary Online Material) includes a number of additional controls for family history of specific medical conditions available exclusively in the UKB data. This is intended as a check on whether those who enter teaching might be more or less prone to certain types of disease. Column 6 and 7 uses the first principal component rather than the sum score to calculate the ALI and column 8 and 9 uses the sum scored ALI but with OLS rather than negative binomial regression. These specifications are intended as a check on whether our results are driven by our choice of measurement model or the specification of our regression models. Table 6 (Supplementary Online Material) replaces the ALI with each of the individual allostatic load indicators in turn. This is intended as a check on whether teaching harms some aspects of health, but this is masked by providing benefits in

[4] Thanks to anonymous reviewer for suggesting this additional analysis.

other areas. Across all of these alternative specifications, we find no association between teaching and health.

5. Discussion

Teaching is a demanding job and prolonged exposure to stress tends to result in physiological dysregulation and ill-health (Marmot & Wilkinson, 2005; McEwen, 1998; Beckie, 2012; Thoits, 2010). With this in mind, we set out to conduct the first study comparing the health of teachers and non-teachers using objective biomarker measures. Contrary to our hypothesis, we found no statistically significant overall association between teaching and health. Furthermore, our large datasets mean that our coefficient estimates of zero are precisely estimated, allowing us to rule out even very small associations. This overall finding held across two datasets, amongst a representative sample of teachers and among a group of older teachers, for both binary and cumulative measures of exposure to teaching, with the addition of parental medical history variables, across different measurement and regression models, and across the individual components of the allostatic load index.

The lack of association could in principle be accounted for by teaching being an overwhelmingly graduate profession in England, given that education is known to have an independent positive effect on health (Eide & Shawalter, 2011). However, our models controlled for graduate status and our comparisons with other graduate occupations did not reveal inferior health amongst teachers. Likewise, the lack of association in our cross-sectional models could be explained by selection of individuals into the profession cancelling out any underlying causal effect. However, our longitudinal analysis also found no difference in the change in ALI across time among those who always taught compared to those who never taught during the period, which helps reduce concerns about selection effects.

Another possibility is that we find no association because we are mis-measuring health, and this is attenuating the underlying relationship. Our indices of allostatic load do have limitations. For one, they do not utilise as many markers of primary mediators as the index used in the original MacArthur studies. The simple method of aggregating the biomarkers by sum scoring indicators of being in the highest risk quartile is also somewhat crude and is unlikely to represent the optimal measurement model. Having said that, it seems unlikely that measurement error could entirely explain our findings since sum-scored ALI have managed to detect differences in allostatic load across occupational groups in other research (Hasson, Von Thiele Schwarz, & Lindfords, 2009).

Another important limitation of our ALI is that it captures only very late stages of the stress-response cascade. Indeed, a recent systematic review has pointed out that the primary mediators are an important part of the allostatic load concept, suggesting that indices which omit biomarkers at the primary mediator stage may be missing the theoretical point (Johnson, Cavallaro, & Leon, 2017). However, we again think this is unlikely to explain our results since (self-rated) health and ALI are correlated across the full range of our ALI measure (see Figure 1, Panel B) and our null finding holds even among the older sample in the UKB, who will have cumulatively experienced more biological wear and tear of the sort captured by our ALI (see Figure 1, Panel A).

In sum, we believe that the most appropriate interpretation of our findings is that teaching is not bad for ones health. How can we explain this finding given the theory around stress response cascades set out in section 1 and 2 of this paper? One potential explanation is that teaching is not, after all, a particularly stressful occupation. As previously noted, the literature on teacher stress has important limitations, particularly with respect to the use of non-representative data. Alternatively, certain aspects of teaching could compensate for the generally stressful nature of the profession. In particular, teaching is known to be less sedentary than many other office-based, graduate occupations (Tudor-Locke et al., 2011). Epidemiological research generally finds a relationship between prolonged sedentary behaviour and long-run health outcomes (Owen et al., 2020) and experimental evidence suggests that that the underlying relationship is causal in nature (Benatti & Ried-Larsen, 2015). The reduced incidence of smoking amongst teachers (Gilbert et al., 2015), which is in part a result of official guidance that all schools in England should be smoke free, is also likely to be an important part of any countervailing effect of teaching on health. Either way, it appears that teaching is not an unhealthy career choice.

References

Barron, E., Lara, J., White, M., & Mathers, J. C. (2015). Blood-borne biomarkers of mortality risk: systematic review of cohort studies. *PloS one*, *10*(6), e0127550.

Beckie, T. M. (2012). A systematic review of allostatic load, health, and health disparities. *Biological Research for Nursing*, *14*(4), 311-346.

Bellingrath, S., Weigl, T., & Kudielka, B. M. (2009). Chronic work stress and exhaustion is associated with higher allostastic load in female school teachers: Original research report. *Stress*, *12*(1), 37-48.

Bellingrath, S., Rohleder, N., & Kudielka, B. M. (2010). Healthy working school teachers with high effort–reward-imbalance and overcommitment show increased pro-inflammatory immune activity and a dampened innate immune defence. *Brain, Behavior, and Immunity*, *24*(8), 1332-1339.

Benatti, F. B., & Ried-Larsen, M. (2015). The effects of breaking up prolonged sitting time: a review of experimental studies. *Medicine & Science in Sports & Exercise*, *47*(10), 2053-61.

Berryhill, J., Linney, J., & Fromewick, J. (2009). The Effects of Education Accountability on Teachers: Are Policies Too-Stress Provoking for Their Own Good?. *International Journal of Education Policy and Leadership*, *4*(5), 1-14.

Bogaert, I., De Martelaer, K., Deforche, B., Clarys, P., & Zinzen, E. (2014). Associations between different types of physical activity and teachers' perceived mental, physical, and work-related health. *BMC Public Health*, *14*(1), 534.

Canguilhem, G. (1978) *The normal and the pathological*. D. Reidel Publishing Company, Dordrecht, Holland.

Chandola, T., & Zhang, N. (2017). Re-employment, job quality, health and allostatic load biomarkers: prospective evidence from the UK Household Longitudinal Study. *International Journal of Epidemiology*, *47*(1), 47-57.

Chandola, T., Booker, C. L., Kumari, M., & Benzeval, M. (2019). Are Flexible Work Arrangements Associated with Lower Levels of Chronic Stress-Related Biomarkers? A Study of 6025 Employees in the UK Household Longitudinal Study. *Sociology*, *53(*4), 779-799.

Clayton, P. E., Banerjee, I., Murray, P. G., & Renehan, A. G. (2011). Growth hormone, the insulin-like growth factor axis, insulin and cancer risk. *Nature Reviews Endocrinology*, *7*(1), 11-24.

Cooper, C. L., & Marshall, J. (1976). Occupational sources of stress: A review of the literature relating to coronary heart disease and mental ill health. *Journal of Occupational Psychology*, *49*(1), 11-28.

Cornell, D. G., & Mayer, M. J. (2010). Why do school order and safety matter?. *Educational Researcher*, *39*(1), 7-15.

Eide, E. R., & Showalter, M. H. (2011). Estimating the relation between health and education: What do we know and what do we need to know?. *Economics of Education Review*, *30*(5), 778-791.

Health & Safety Executive (2014). Work related stress, anxiety and depression in Great Britain 2014. Retrieved from: https://consult-smp.com/archive/2016/06/hse-work-related-stress-anxiety-and-depression-statistics-in-great-britain-2015.html#:~:text=demands%20placed%20on%20them%20at,of%201380%20per%20100%2C000%20workers.

Helmert, U., Shea, S., & Bammann, K. (1997). The impact of occupation on self-reported cardiovascular morbidity in western Germany: gender differences. *Review of Environmental Health*, *12*, 25-42.

Heus, P. D., & Diekstra, R. F. W. (1999). Do teachers burn out more easily? A comparison of teachers with other social professions on work stress and burnout symptoms. In R. Vandenberghe, A. M. Huberman, R. Vandenberghe, & A. M. Huberman (Eds.), Understanding and preventing teacher burnout (pp. 269e284). Cambridge: Cambridge University Press.

Droogenbroeck, F., & Spruyt, B. (2015). Do teachers have worse mental health? Review of the existing comparative research and results from the Belgian Health Interview Survey. *Teaching and Teacher Education*, *51*, 88-100.

Ettehad, D., Emdin, C. A., Kiran, A., Anderson, S. G., Callender, T., Emberson, J., ... & Rahimi, K. (2016). Blood pressure lowering for prevention of cardiovascular disease and death: a systematic review and meta-analysis. *The Lancet*, *387*(10022), 957-967.

Johnson, S., Cooper, C., Cartwright, S., Donald, I., Taylor, P., & Millet, C. (2005). The experience of work-related stress across occupations. *Journal of Managerial Psychology*, *20*(2), 178-187.

Gilbert, F., Richard, J. B., Lapie-Legouis, P., Beck, F., & Vercambre, M. N. (2015). Health behaviors: Is there any distinction for teachers? A cross-sectional nationwide study. *PloS one*, *10*(3), e0120040.

Gruenewald, T. L., Seeman, T. E., Ryff, C. D., Karlamangla, A. S., & Singer, B. H. (2006). Combinations of biomarkers predictive of later life mortality. *Proceedings of the National Academy of Sciences of the United States of America*, *103*, 14158–14163.

Hardy, I., Rönnerman, K., & Beach, D. (2019). Teachers' work in complex times: the 'fast policy' of Swedish school reform. *Oxford Review of Education*, *45*(3), 350-366.

Harmsen, R., Helms-Lorenz, M., Maulana, R., & van Veen, K. (2018). The relationship between beginning teachers' stress causes, stress responses, teaching behaviour and attrition. *Teachers and Teaching*, *24*(6), 626-643.

Hasson, D., Von Thiele Schwarz, U., & Lindfors, P. (2009). Self-rated health and allostatic load in women working in two occupational sectors. *Journal of Health Psychology*, *14*(4), 568-577.

Head, J., Stansfeld, S., Siegrist, J. (2004). The psychosocial work environment and alcohol dependence: a prospective study. *Occupational and Environmental Medicine*, 61**,** 219-224.

Gallup (2014). State of America's schools: A path to winning again in education. Washington, DC: Author. Retrieved from http://www.gallup.com/services/178709/state-america-schools-report.aspx?aysn

Genest, J. (2010). C-reactive protein: risk factor, biomarker and/or therapeutic target?. *Canadian Journal of Cardiology*, *26*, 41A-44A.

Gunasekara, F. I., Carter, K., & Blakely, T. (2012). Comparing self-rated health and self-assessed change in health in a longitudinal survey: Which is more valid?. *Social Science & Medicine*, *74*(7), 1117-1124.

Johnson, S. C., Cavallaro, F. L., & Leon, D. A. (2017). A systematic review of allostatic load in relation to socioeconomic position: Poor fidelity and major inconsistencies in biomarkers employed. *Social Science & Medicine*, *192*, 66-73.

Juster, R. P., McEwen, B. S., & Lupien, S. J. (2010). Allostatic load biomarkers of chronic stress and impact on health and cognition. *Neuroscience & Biobehavioral Reviews*, *35*(1), 2-16.

Kovess-Masféty, V., Sevilla-Dedieu, C., Rios-Seidel, C., Nerrière, E., & Chee, C. C. (2006). Do teachers have more health problems? Results from a French cross-sectional survey. *BMC Public Health*, *6*(1), 101.

Kovess-Masféty, V., Rios-Seidel, C., & Sevilla-Dedieu, C. (2007). Teachers' mental health and teaching levels. *Teaching and Teacher Education*, *23*(7), 1177-1192.

Larson, J. S. (1999). The conceptualization of health. *Medical Care Research and Review*, *56*(2), 123-136.

Lewitt, M. S., Dent, M. S., & Hall, K. (2014). The insulin-like growth factor system in obesity, insulin resistance and type 2 diabetes mellitus. *Journal of Clinical Medicine*, *3*(4), 1561-1574.

Lyons, T. J., & Basu, A. (2012). Biomarkers in diabetes: hemoglobin A1c, vascular and tissue markers. *Translational Research*, *159*(4), 303-312.

Mannic, T., Viguie, J., & Rossier, M. F. (2015). In vivo and in vitro evidences of dehydroepiandrosterone protective role on the cardiovascular system. *International Journal of Endocrinology and Metabolism*, *13*(2), e24660.

Marmot, M., & Wilkinson, R. (Eds.). (2005). *Social determinants of health*. OUP Oxford.

Masilamani, R., Darus, A., Ting, A. S., Ali, R., Mahmud, A. B. A., & David, K. (2012). Salivary biomarkers of stress among teachers in an urban setting. *Asia Pacific Journal of Public Health*, *24*(2), 278-287.

Mauss, D., Li, J., Schmidt, B., Angerer, P., & Jarczok, M. N. (2015). Measuring allostatic load in the workforce—a systematic review. *Industrial Health*, *53*, 5–20.

McEwen, B. S. (1998). Stress, adaptation, and disease: Allostasis and allostatic load. *Annals of the New York Academy of Sciences*, *840*(1), 33-44.

McEwen, B. S. (2003). Interacting mediators of allostasis and allostatic load: towards an understanding of resilience in aging. *Metabolism*, *52*, 10-16.

McFall, S., Petersen, J., Kaminska, O., Lynn, P. (2014). *Understanding Society –UK Household Longitudinal Study: Waves 2 and 3 Nurse Health Assessment, 2010-2012, Guide to Nurse Health Assessment*. Colchester: University of Essex.

Mujtaba, T., & Reiss, M. (2013). Factors that lead to positive or negative stress in secondary school teachers of mathematics and science. *Oxford Review of Education*, *39*(5), 627-648.

Owen, N., Healy, G. N., Dempsey, P. C., Salmon, J., Timperio, A., Clark, B. K., ... & Lambert, G. (2020). Sedentary behavior and public health: integrating the evidence and identifying potential solutions. *Annual Review of Public Health*, *41*, 265-287.

Pithers, R. T., & Fogarty, G. J. (1995). Symposium on teacher stress. *British Journal of Educational Psychology*, *65*(1), 3-14.

Qi, X., Zhang, J., Liu, Y., Ji, S., Chen, Z., Sluiter, J. K., & Deng, H. (2014). Relationship between effort–reward imbalance and hair cortisol concentration in female kindergarten teachers. *Journal of Psychosomatic Research*, *76*(4), 329-332.

Rutkowski, K., Sowa, P., Rutkowska-Talipska, J., Kuryliszyn-Moskal, A., & Rutkowski, R. (2014). Dehydroepiandrosterone (DHEA): hypes and hopes. *Drugs*, *74*(11), 1195-1207.

Schaufeli, W. B., Daamen, J., & Van Mierlo, H. (1994). Burnout among Dutch teachers: an Mbi-validity study. *Educational and Psychological Measurement, 54*(3), 803-812.

Schuler, R. S. (1980). Definition and conceptualization of stress in organizations. *Organizational Behavior and Human Performance*, *25*(2), 184-215.

Seeman, T. E., McEwen, B. S., Rowe, J. W., & Singer, B. H. (2001). Allostatic load as a marker of cumulative biological risk: MacArthur studies of successful aging. *Proceedings of the National Academy of Sciences*, *98*(8), 4770-4775.

Shields, G. S., & Slavich, G. M. (2017). Lifetime stress exposure and health: A review of contemporary assessment methods and biological mechanisms. *Social and Personality Psychology Compass*, *11*(8), e12335.

Author. (2018). [Details removed for peer review]

Smith, A., Brice, C., Collins, A., Matthews, V., & McNamara, R. (2000). *The scale of occupational stress: A further analysis of the impact of demographic factors and type of job (Contract Research Report 311/2000)*. Sudbury: Health and Safety Executive. HSE Books.

Stamatakis, E., Chau, J. Y., Pedisic, Z., Bauman, A., Macniven, R., Coombs, N., & Hamer, M. (2013). Are sitting occupations associated with increased all-cause, cancer, and cardiovascular disease mortality risk? A pooled analysis of seven British population cohorts. *PloS one*, *8*(9), e73753.

Sterling, P., Eyer, J. (1988). *Handbook of Life Stress, Cognition and Health. Allostasis: A new paradigm to explain arousal pathology*. John Wiley & Sons, New York.

Thoits, P. A. (2010). Stress and health: Major findings and policy implications. *Journal of Health and Social Behavior*, *51,* S41-S53.

Tudor-Locke, C., Ainsworth, B. E., Washington, T. L., & Troiano, R. (2011). Assigning metabolic equivalent values to the 2002 census occupational classification system. *Journal of Physical Activity and Health*, *8*(4), 581-586.

University of Essex. Institute for Social and Economic Research, NatCen Social Research and Kantar Public, [producers]: Understanding Society: Waves 1-6, 2009-2015 [computer file]. 8th Edition. Colchester, Essex: UK Data Service [distributor], November 2016. SN: 6614.

Van Droogenbroeck, F., & Spruyt, B. (2015). Do teachers have worse mental health? Review of the existing comparative research and results from the Belgian Health Interview Survey. *Teaching and Teacher Education*, *51*, 88-100.

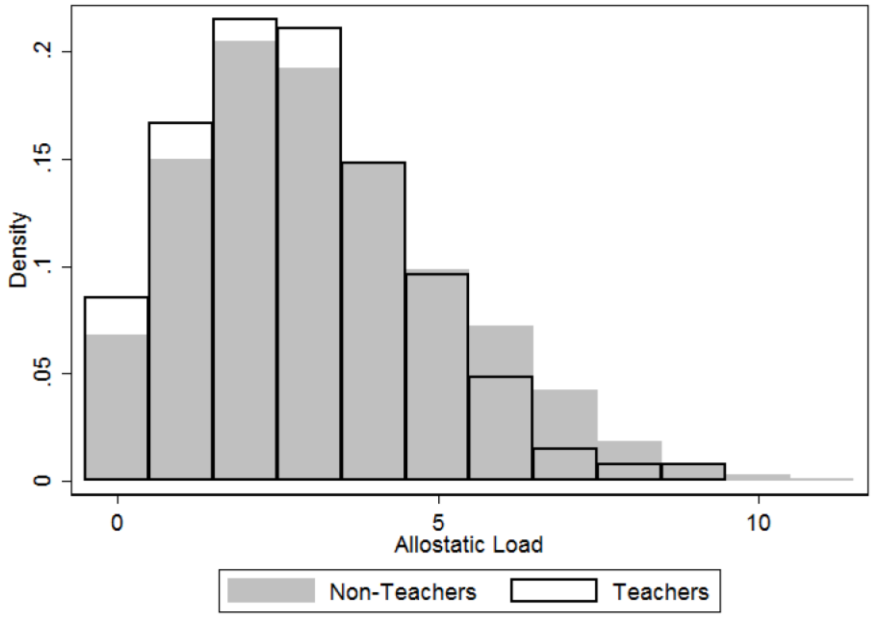
Wolfram, M., Bellingrath, S., Feuerhahn, N., & Kudielka, B. M. (2013). Cortisol responses to naturalistic and laboratory stress in student teachers: Comparison with a non-stress control day. *Stress and Health*, *29*(2), 143-149.

Worth, J., & Van den Brande, J. (2019). *Teacher Labour Market in England Annual Report 2019*. Slough: National Foundation for Educational Research.

Zajacova, A., & Woo, H. (2015). Examination of age variations in the predictive validity of self-rated health. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *71*(3), 551-557.

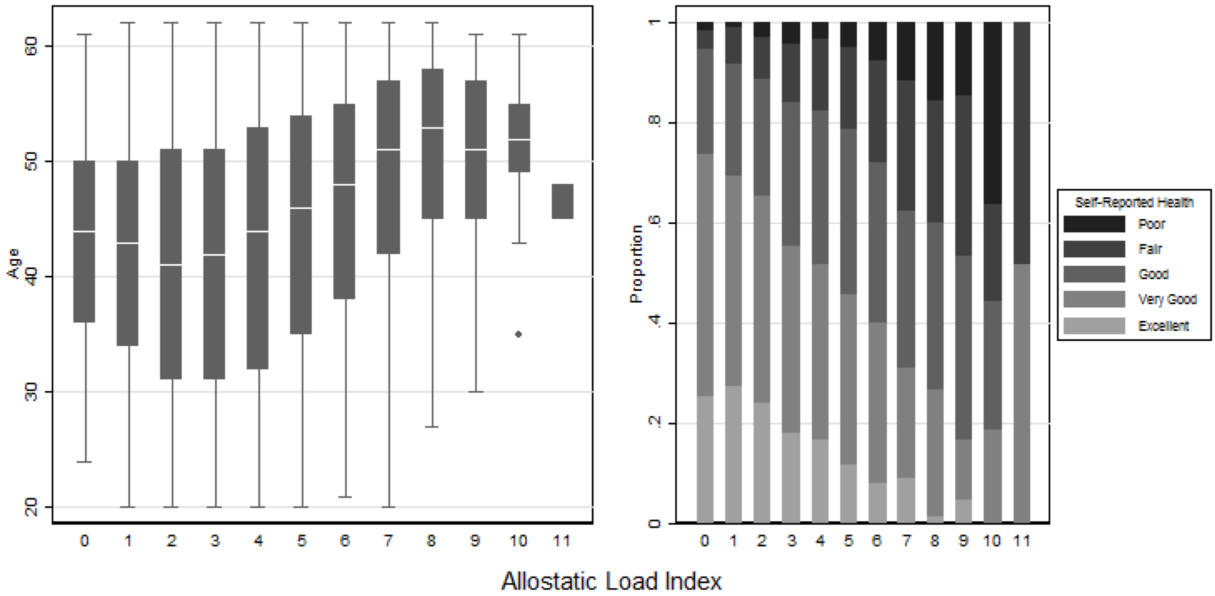
Figures

Figure 1. Histogram of the unadjusted allostatic load index for teachers and non-teachers



Notes: Uses only the Understanding Society data and allostatic load index. n=7,286.

Figure 2: Evidence of convergent validity for the allostatic load index



Notes: Uses only the Understanding Society data and allostatic load index. n=7,286.

Tables

Table 1. Descriptive statistics for the different datasets and subsamples

	USoc		UKB Cross-Section		UKB Occup. History		UKB Follow-Up	
	Teachers	**Others**	**Teachers**	**Others**	**Ever Teachers**	**Others**	**Always Teachers**	**Others**
Age (mean)	45.3	42.6	52.3	50.7	52.0	51.5	52.5	51.9
Male	30.4%	45.5%	26.0%	47.6%	25.5%	44.3%	25.7%	47.4%
White	95.8%	91.4%	95.9%	93.3%	97.8%	96.9%	97.9%	98.0%
Degree	69.5%	19.6%	84.8%	35.5%	84.4%	44.4%	84.6%	48.0%
Count (N)	270	7,016	14,651	215,804	11,542	68,074	280	5,607
	7,286		230,455		79,616		5,887	

Notes: USoc = Understanding Society dataset. UKB = UK Biobank dataset. “Occup. History” = subsample of UKB who completed the occupational history questionnaire. “Follow-Up” = subsample of UKB who attended a second assessment centre. “Ever Teachers” = those who report working as a teacher at any stage in their occupational history. “Always Teachers” = those who were working as a teacher at both their first and second UKB Assessment Centre visit. Nurse visit weights applied to the Understanding Society data. UK Biobank is a convenience sample, so no weights have been applied. N is unweighted. Percentages have been rounded and may not sum to zero. Ethnicity is not shown in greater detail to guard against disclosure. Samples restricted to those of working age (<60).

Table 2. Biomarkers included in the Allostatic Load Index in the two datasets

Stage	Understanding Society	UK Biobank	Notes
Primary Mediators	Insulin-like growth factor		Hormones that regulate blood glucose levels. Biomarker for diabetes and cancer (Clayton et al., 2011; Lewitt, Dent, & Hall, 2013.)
	DHEA-S		Adrenal hormone and functional HPA-axis antagonist. Biomarker for cardiovascular disease (Mannic, Viguie, & Rossier, 2015; Rutkowski et al., 2014).
Secondary Outcomes	Resting pulse rate		Heart rate. Indicator of cardiovascular fitness.
	Waist to height/hip ratio		Indicator of location of adipose tissue deposits.
	HbA1c	HbA1c	Average glucose level over previous 12 weeks. Biomarker for poorly managed diabetes (Lyons & Basu, 2012).
	Systolic BP	Systolic BP	Indicator of intravascular pressure at end of left ventricular contraction. Biomarker for hypertension and cardiovascular disease (Ettehad et al., 2016).
	Diastolic BP	Diastolic BP	Indicator of intravascular pressure at end of left ventricular relaxation. Biomarker for hypertension and cardiovascular disease (Ettehad et al., 2016).
	Cholesterol to HDL	Cholesterol to HDL	Cholesterol is a basic element of steroid hormones. HDL is a cardioprotective form of cholesterol. Biomarker for heart disease (Barron, 2015; Upadhyay, 2015).
	Triglycerides	Triglycerides	Cardio-damaging form of fat. Biomarker for heart disease (Upadhyay, 2015).
	Creatinine clearance rate	Creatinine clearance rate	Volume of blood plasma that is cleared of creatinine per unit of time. Biomarker for kidney disease (Tesch, 2010).
	C-reactive Protein	C-reactive Protein	Acute phase inflammatory protein. Biomarker for inflammation due to injury or infection and cardiovascular disease (Barron, 2015; Genest, 2010).
	Fibrinogen	Fibrinogen	Protein and factor of blood coagulation. Biomarker for inflammation due to injury or infection and cardiovascular disease (Barron, 2015).
		Albumin	Protein made by the liver. Biomarker for sub-clinical renal damage and liver dysfunction (Tesch, 2010).
		BMI	Indicator of obesity.

Notes: Based in part on Mouss et al. (2015)

Table 3. Regressing teaching on allostatic load

	(1) USoc	(2) UKB	(3) UKB	(4) UKB
Teaching at baseline assessment centre	0.938 (0.0390)	1.004 (0.00640)	-	-
Total years teaching at baseline assessment centre	-	-	1.001** (0.00028)	-
Teaching at baseline & follow-up assessment centre	-	-	-	0.954 (0.0387)
Age	0.966** (0.00493)	1.055** (0.00494)	1.047** (0.00882)	1.018 (0.0280)
Age Squared	1.001** (5.93e-05)	1.000** (4.61e-05)	1.000** (8.21e-05)	1.000 (0.000269)
Male	0.940** (0.0140)	1.617** (0.00463)	1.624** (0.00818)	1.251** (0.0214)
Ethnicity: (ref=white)				
Mixed	1.031 (0.0766)	1.080** (0.0176)	1.058 (0.0345)	0.971 (0.129)
Asian	1.125** (0.0452)	1.272** (0.0102)	1.223** (0.0270)	1.110 (0.0857)
Black	1.227** (0.0775)	1.322** (0.0121)	1.302** (0.0360)	1.099 (0.149)
Arab	1.344 (0.244)	-	-	-
Other	1.177 (0.186)	1.192** (0.0160)	1.158** (0.0379)	0.944 (0.117)
Graduate	0.849** (0.0172)	0.842** (0.00257)	0.855** (0.00443)	0.978 (0.0159)
Baseline ALI				1.166 (0.005)
N	7,173	229,503	79,384	5,876

Notes: USoc = Understanding Society. UKB = UK Biobank. ALI = Allostatic Load Index. Baseline in Understanding Society refers to data collected between 2010 and 2012. Baseline in UK Biobank refers to data collected between 2006 and 2010. Follow-up relates only to UK Biobank data and refers to data collected between 2012 and 2013. Each column is a separate regression. All columns are negative binomial regressions and coefficients are incidence rate ratios. Column 4 includes a control for ALI measured at a timepoint prior to which the outcome variable was measured. is an ordinary least squares regression. Parentheses contain standard errors. * = p<0.05, **<0.01. Nurse visit weights applied to the Understanding Society data. UK Biobank is a convenience sample, so no weights have been applied.

Table 4. Regression of occupation on allostatic load (UK Biobank)

	(5)	
	Coeff.	**S.E.**
Occupation: (ref=teacher)		
Accountants/Consultants	0.979	0.028
Health Professionals	0.921*	0.032
Planners/Surveyors	0.998	0.039
Protective Officers	1.031	0.039
Research Professionals	0.986	0.032
Welfare Professionals	1.078**	0.027
Age	1.111**	0.030
Age Squared	0.999**	0.030
Male	1.621**	0.027
Ethnicity: (ref=white)		
Mixed	1.181	0.115
Asian	1.246**	0.089
Black	1.264**	0.101
Other	1.149	0.129
Graduate	0.876**	0.017
N	9,356	

Notes: Results from a single negative binomial regression. Coeff. = coefficients reported as incidence rate ratios. S.S. = standard error. * = p<0.05, **<0.01. Accountants/Consultants = chartered and certified accountants; management consultant and business analysts. Health Professional = psychologists; pharmacists. Planners and Surveyors = town planning officers; quantity surveyors; chartered surveyors. Protective Officers = police officers, fire officers, prison officers. Research Professionals = chemical scientists; biological scientists and biochemists; physical scientists; social and humanities scientists. Welfare Professionals: social workers; probation officers.