Cureus
Part of SPRINGER NATURE

Open Access Original Article

Socioeconomic Gradient in Mortality and Temporal Variation in Health Behaviours Measured Over 28 Years: A Longitudinal Analysis of the Whitehall II Study

Review began 11/03/2026
Review ended 11/16/2026
Published 11/25/2026

© **Copyright** 2026
Sarwar et al. This is an open access article distributed under the terms of the Creative Commons Attribution License CC-BY 4.0., which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

DOI: 10.7759/cureus.120261

Danyal Sarwar [1], Rebecca E. Harounoff [1], Maria Perez Ortiz [2], Eric Brunner [3], Yuntao Chen [3], John Shawe-Taylor [2]

1. University College London (UCL) Medical School, University College London, London, GBR 2. Department of Computer Science, University College London, London, GBR 3. Department of Epidemiology and Public Health, University College London, London, GBR

Corresponding author: Danyal Sarwar, danyalsarwar@aol.com

Abstract

Background

All-cause mortality is persistently related to socioeconomic status in the form of an inverse gradient: lower status, higher mortality. Health behaviours are frequently proposed as explanatory factors but are typically assessed at a single timepoint, which does not reflect exposure across the life course. We aimed to extend a previous analysis of this cohort with longer mortality follow-up, and to further examine whether longitudinal assessment alters the estimated attenuation of this gradient by health behaviours.

Methodology

This was a retrospective analysis of data from the prospective Whitehall II cohort of British civil servants, recruited from 1985 to 1988 at age 35 to 55 years. Data were drawn from six phases of follow-up to 2013 (phases 1, 3, 5, 7, 9, and 11), with mortality follow-up to 2021. Socioeconomic status was defined by employment grade at baseline. Hazard ratios were calculated using Cox regression, adjusted for age and sex, for all-cause mortality in the lowest compared with the highest employment grade. Smoking, alcohol consumption, diet, and physical activity were added as covariates, both at baseline and as time-varying covariates, and the percentage attenuation of the log-hazard ratio was calculated.

Results

Of the 10,308 participants, 9,590 had complete baseline health behaviour data and mortality follow-up. Among these, 2,427 deaths occurred. The age- and sex-adjusted hazard ratio for the lowest employment grade compared with the highest was 1.55 (95% confidence interval (CI): 1.37 to 1.76). Adjusting for the core health behaviours at baseline attenuated this hazard ratio to 1.24 (95% CI: 1.08 to 1.41), a percentage attenuation of 51%. Longitudinal assessment, assessing the same health behaviours as time-varying covariates, attenuated the hazard ratio to 1.06 (95% CI: 0.92 to 1.21), a percentage attenuation of 87%. This CI implied the residual gradient in mortality was no longer statistically significant. An expanded diet score constructed by the authors increased this attenuation to 91%.

Conclusions

Repeated assessment of health behaviours during adult life, in contrast to a single baseline assessment, enables an improved and more naturalistic statistical model of the determinants of the socioeconomic gradient in all-cause mortality. In this cohort, smoking, alcohol consumption, diet, and physical activity attenuated the gradient by 51% when assessed at baseline and by 87% when assessed as time-varying covariates. While these estimates may move towards a causal decomposition, they remain statistical attenuations, and attenuation may be overestimated due to reverse causation: behaviour changes as a consequence of ill-health.

Categories: Public Health, Preventive Medicine, Epidemiology/Public Health
Keywords: exercise, health behaviours, health inequalities, lifestyle or behavioural risk factors, nutrition and diet, smoking & alcohol, social class

Introduction

With increasing life expectancy have come ageing populations [1]; it is expected that by 2050, one in six people will be over 65 years of age [2]. This has heightened the desire to understand the factors impacting longevity and quality of life in old age [3], which are complex and heterogeneous between individuals, encompassing lifestyle, psychosocial, and disease-related factors [4].

Socioeconomic status is associated with these factors, perhaps indicating why it shows a gradient in mortality: the gap in life expectancy in England between low- and high-income neighbourhoods, excluding the poorest and richest 5%, is six years [5]. As well as having shorter lives, those living in poorer areas spend longer in poor health [5]. Investing in prevention has been shown to reduce overall health costs and expenditure on welfare benefits [6].

Previous studies have aimed to investigate the contributions of health behaviours to social inequalities in mortality [7]. Health behaviours are associated with morbidity, mortality, cardiovascular disease, and cardiometabolic risk worldwide [8,9]. One study estimated that one in three premature deaths in England was attributable to socioeconomic inequalities [10]. Health behaviours are not the only suggested contributor to this gradient: material personal circumstances and psychosocial factors, such as housing and work conditions or psychological stress, are also thought to play a role [5,9].

However, important gaps remain. Typically, health behaviours are assessed at one point in time (baseline), which may only partially reflect behavioural exposure across the life course. A person's health behaviours are unlikely to remain constant throughout their life.

Stringhini et al. [7] investigated the role of health behaviours (smoking, alcohol consumption, diet, and physical activity) in the association between socioeconomic status and mortality. Adjustment for these behaviours statistically attenuated the gradient in mortality by 42% at baseline. When assessed longitudinally at several phases over 24 years, however, adjustment for these behaviours attenuated it by 72% [7].

Study objectives

This study had three aims. The primary aim was to estimate how much health behaviours attenuate the socioeconomic gradient in mortality when assessed repeatedly over follow-up, using a longitudinal model to account for behavioural changes, compared with a single assessment at baseline. The secondary aims were to compare the results reported by Stringhini et al. [7] using follow-up to 2009 with results from a comparable baseline model using follow-up to 2021, noting the differences in model specification described in the Materials and Methods section, and to expand the diet score to include a richer combination of dietary variables. The main outcome measure was all-cause mortality.

This article was previously posted to the medRxiv preprint server on October 2, 2024.

Materials And Methods

Study population

This was a retrospective analysis of data from Whitehall II, a prospective British longitudinal cohort study which investigated social determinants of health [11]. The target population was civil servants aged 35 to 55 years working in the London offices of 20 civil service departments. Civil servants were invited to participate in this study via letter; 73% consented in writing, giving a total of 10,308 civil servants. Whitehall II data from six phases at which clinical examinations were undertaken were used, giving six phases of health behaviour data (phase 1, 1985-1988; phase 3, 1991-1994; phase 5, 1997-1999; phase 7, 2002-2004; phase 9, 2007-2009; and phase 11, 2012-2013), with mortality data to 2021. Mortality follow-up was through the NHS Central Registry, which provides date and cause of death, with 99.9% of participants flagged [11].

Socioeconomic status was estimated from employment grade at baseline. Baseline grade was used throughout follow-up because participants were aged approximately 58 to 83 years by phase 11, when many had retired, and current grade (if any) would no longer reflect socioeconomic position, and because later grade could itself be affected by ill health. Participants were drawn from six Whitehall II study grades, mapped from civil service employment grades ranging from clerical and support staff through executive officer grades to unified administrator grades, and differing widely in salary [11]. Following the convention adopted in Whitehall II analyses, these were collapsed into three categories: high (administrative), intermediate (professional and executive), and low (clerical and support). Hazard ratios (HRs) compare the lowest employment grade with the highest, which was taken as the reference category. The study examined the socioeconomic gradient between the extremes of employment grade, and this contrast also allows comparison with Stringhini et al. [7]; HRs for the intermediate grade are therefore not reported.

Ethical approval and data access

Ethical approval for the Whitehall II study is held under REC reference 85/0938 (IRAS project ID 142374); recent amendments have been approved by the London - Harrow Research Ethics Committee, and research ethics approvals were renewed at each contact. All participants provided written informed consent. This analysis was conducted under the approval held by Eric Brunner's group at UCL School of Life and Medical Sciences, and falls within the study objectives set out in the approved Study Protocol, which include extending the evidence on social inequalities in ageing health outcomes and their explanations. Data were accessed through the Dementias Platform UK (DPUK) secure platform under DPUK study ID 0475 [12]. The

dataset contains a vast amount of confidential information; therefore, data cannot be imported or exported from the platform. As a consequence, the data underlying this analysis cannot be shared by the authors, and access must be sought directly from DPUK.

Health behaviours

Smoking status (never, former, current) was taken from participants' responses at each phase and entered as a covariate in the models described below.

Alcohol consumption in units was calculated as a combination of units of beer, wine, and spirits consumed in a week, and scoring allocated analogously to the scoring of Stringhini et al. [7]: never (0 units/week); moderate (1-21 units/week for men, 1-14 for women); heavy (>21 units/week for men, >14 for women).

For the purpose of comparison, diet was scored in accordance with the scoring devised by Stringhini et al. [7], where participants were classified as unhealthy if they ate white bread most frequently, consumed whole milk, and ate fruit and vegetables less than three times per month; healthy if they ate wholemeal, wheatmeal, or other brown bread most frequently, did not consume milk or only used skimmed or other types of milk, and ate fruit and vegetables daily or two or more times per day; or moderately healthy if their dietary pattern was in between these two descriptions.

Physical activity was assessed based on hours per week of moderate and vigorous physical activity, and scored as by Stringhini et al. [7]: active (>2.5 hours/week of moderate physical activity or >1 hour/week of vigorous physical activity); inactive (<1 hour/week of moderate physical activity and <1 hour/week of vigorous physical activity); or moderately active if their physical activity fell between these two descriptions.

The categories for smoking status, alcohol consumption, diet, and physical activity were entered into the models as categorical variables, so the order in which they are listed did not affect the estimates.

Removal of missing data

Of the 10,308 participants recruited to Whitehall II, 718 were excluded from analysis, 707 because of missing health behaviour data at baseline and 11 because of missing mortality follow-up, leaving 9,590 participants, the same analytic sample as Stringhini et al. [7], among whom 2,427 deaths were recorded by 2021. These rows were removed in Python using the Pandas module; 93% of participants had complete health behaviour data at baseline, motivating this decision. For comparison, Stringhini et al. [7] removed all rows corresponding to missing data in an analogous way. Across all six phases, approximately 25% of participants had missing data at one or more phases; the handling of these is described under the longitudinal model below.

Survival analysis and Cox regression

Where time-to-event is the main outcome under analysis (as in this study), it is referred to as survival time, which may differ between participants. Indeed, some, the majority in this study, may not undergo the event during follow-up and thus their true time-to-event is not observed. This is known as data censoring. This necessitates unique methods, the methods of survival analysis [13].

Cox regression is one of the most used methods in survival analysis because an HR can be calculated, and the estimated hazards are always non-negative [14]. The coefficients in the Cox regression relate to a hazard, where an HR above 1 indicates a higher estimated hazard relative to the reference group. Cox regression was therefore used to estimate the association between employment grade and mortality before and after adjustment for health behaviours.

Baseline Cox Model

The baseline Cox model uses data at a single point in time. In this case, we use health behaviour data from baseline. This method is used so we can compare how the socioeconomic gradient in mortality may change when data is assessed longitudinally. The central analysis was achieved using Stata statistical software, version 17 (StataCorp LLC, College Station, TX) to perform Cox regression with age as the time scale, adjusting for sex, and with employment grade entered as a categorical variable. Data were declared as survival data using the *stset* command, with age as the time scale, and Cox models were fitted using the *stcox* command, with employment grade and each health behaviour entered as categorical variables using the *i.* prefix. This specification differs from Stringhini et al. [7], who entered employment grade as a linear term and additionally adjusted for year of birth. The outputted HRs corresponded to the relative hazard of mortality of each socioeconomic group compared to the highest group, adjusted for age and sex. This would then be the HR unadjusted for health behaviours. Health behaviours were then added as covariates in turn to generate adjusted HRs.

Longitudinal Cox Model

The longitudinal Cox model incorporates data from different phases, with further years of follow-up data in comparison to Stringhini et al. [7]. It allows us to estimate the socioeconomic gradient in mortality after accounting for repeated measures of health behaviours. Where individuals had follow-up data available, the covariates for health behaviours may change with time and thus were time-dependent covariates [15], whereas socioeconomic status remained as a time-fixed exposure at baseline. Data were reshaped so that covariate values were represented at each phase, with participant identity serving as the multiple-record identifier in stset, so that each participant's follow-up was split by phase. Values from each phase were used until the next, and values from phase 11 were used until death or the end of follow-up in 2021. Where participants had missing data on health behaviours at one of the follow-up assessments, the value was carried forward, where available, from the preceding phase. Phases after a participant's death were removed. The intervals between phases were unequal, being either five or six years, and were specified accordingly rather than treated as uniform. The Cox regression was then run analogously to the above, estimating the socioeconomic gradient in mortality after accounting for changes in health behaviours across follow-up.

Percentage Attenuation

The percentage attenuation was calculated using the formula below [7] and was used for both baseline and longitudinal Cox regression models. This quantifies the extent to which the log HR is reduced when each health behaviour is added to the model, on a log scale.

$$\text{Percentage attenuation} = 100 \times \frac{\ln \text{HR}_{\text{unadjusted}} - \ln \text{HR}_{\text{adjusted}}}{\ln \text{HR}_{\text{unadjusted}}}$$

where $\text{HR}_{\text{unadjusted}}$ is the hazard ratio unadjusted for health behaviours and $\text{HR}_{\text{adjusted}}$ is the hazard ratio adjusted for health behaviours.

Results

Characteristics of the analytic sample

This analysis used the same 9,590 participants as Stringhini et al. [7]. Because baseline characteristics do not depend on the length of mortality follow-up, the baseline distributions they reported describe this sample (Table *1*). Current smoking, no alcohol consumption, unhealthy diet, and physical inactivity were each more common in the low than the high grade, whereas heavy alcohol consumption was more common in the high grade. For every behaviour, the intermediate grade lay between the two.

Characteristic	High	Intermediate	Low	Total
Participants, n (%)	2,867 (29.9)	4,663 (48.6)	2,060 (21.5)	9,590 (100)
Age, mean (SD), years	45.0 (5.8)	43.3 (6.0)	46.0 (6.0)	44.4 (6.0)
Men, n (%)	2,513 (87.7)	3,429 (73.5)	569 (27.6)	6,511 (67.9)
Current smoker, n (%)	290 (10.1)	845 (18.1)	612 (29.7)	1,747 (18.2)
No alcohol consumption, n (%)	227 (7.9)	745 (16.0)	747 (36.3)	1,719 (17.9)
Heavy alcohol consumption, n (%)	558 (19.5)	821 (17.6)	142 (6.9)	1,521 (15.9)
Unhealthy diet, n (%)	167 (5.8)	507 (10.9)	306 (14.9)	980 (10.2)
Physically inactive, n (%)	189 (6.6)	512 (11.0)	730 (35.4)	1,431 (14.9)

TABLE 1: Baseline characteristics of the analytic sample by employment grade.

Data from Stringhini et al. [7], whose analytic sample is identical to that of this study; baseline characteristics do not depend on the length of mortality follow-up. Total values were calculated from the grade-specific figures. Health behaviour categories are defined in the Materials and Methods section.

SD: standard deviation

These behaviours changed over follow-up, and differently by grade. Between baseline and the end of the follow-up reported by Stringhini et al., in which behaviours were last measured at phase 7 (2002 to 2004), current smoking fell from 10.1% to 4.8% in the high grade and from 29.7% to 16.5% in the low grade. Unhealthy diet fell from 5.8% to 1.0% and from 14.9% to 5.2%. Physical inactivity rose from 6.6% to 21.4% and from 35.4% to 41.6%. No alcohol consumption was almost unchanged in the high grade (7.9% to 7.7%)

but rose in the low grade (36.3% to 42.2%) [7]. Distributions at phases 9 and 11 are not reported here.

Comparison with the Stringhini et al. baseline results using extended follow-up

The baseline results reported by Stringhini et al. [7] using follow-up to 2009 were compared with results from the baseline model using follow-up to 2021. This meant more deaths: 654 deaths to 2009, increasing to 2,427 to 2021, from 9,590 participants without missing data values. The baseline HRs are compared in Table *2*.

Model	HR (95% CI) 2009	% Attenuation 2009	HR (95% CI) 2021	% Attenuation 2021
Model 1 (age and sex)	1.60 (1.26 to 2.04)	NA	1.55 (1.37 to 1.76)	NA
Plus smoking	1.36 (1.06 to 1.74)	32	1.33 (1.17 to 1.51)	35
Plus alcohol consumption	1.58 (1.24 to 2.03)	3	1.51 (1.34 to 1.72)	6
Plus diet	1.55 (1.21 to 1.98)	7	1.47 (1.26 to 1.69)	12
Plus physical activity	1.57 (1.23 to 2.00)	5	1.49 (1.31 to 1.69)	9
Fully adjusted	1.31 (1.02 to 1.69)	42	1.24 (1.08 to 1.41)	51

TABLE 2: Role of health behaviours in statistically accounting for the association between socioeconomic status and all-cause mortality: comparison at baseline between data from 2009 and from 2021.

HRs compare the lowest socioeconomic status with the highest, adjusted for age and sex. 2009 figures are those reported by Stringhini et al. [7], whose model entered employment grade as a linear term and was adjusted for sex and year of birth. In the Whitehall II study, there were 654 deaths to 2009 and 2,427 deaths to 2021, from 9,590 participants. Fully adjusted includes all listed health behaviours.

CI: confidence interval; HR: hazard ratio; NA: not applicable

The HR for the unadjusted model represents the difference in hazard rate between individuals of the lowest and highest employment grade when adjusted for age and sex only. The listed health behaviours were then added individually as covariates, and adjusted HRs between the lowest and highest employment grades were calculated, and then all behaviours were added simultaneously. The associated percentage attenuations were then calculated (Table *2*).

Comparing follow-up to 2009 and to 2021, in the first model adjusting only for age and sex at baseline, the HR decreased slightly from 1.60 (95% confidence interval (CI): 1.26 to 2.04) to 1.55 (95% CI: 1.37 to 1.76). The attenuation obtained from the health behaviours at baseline was higher with longer follow-up for every listed behaviour, although differences in model specification mean this cannot be attributed to longer follow-up alone.

In addition, the CIs have narrowed, possibly due to several more individuals reaching the time-to-event (mortality) between 2009 and 2021. Overall, adjustment for health behaviours at baseline attenuated the gradient in mortality by 51%.

Longitudinal model accounting for behavioural change over follow-up

Following on from this, it was investigated how HRs considering changes in health behaviours across the period of follow-up compared with those solely assessing health behaviours at baseline (above). These longitudinal analyses, where the health behaviours were entered as time-varying covariates, are shown in Table *3*, with the percentage attenuation again alongside the HRs.

Model	HR (95% CI) baseline	% Att. baseline	HR (95% CI) longitudinal	% Att. longitudinal	% point difference
Model 1 (age and sex)	1.55 (1.37 to 1.76)	NA	1.55 (1.37 to 1.76)	NA	NA
Plus smoking	1.33 (1.17 to 1.51)	35	1.29 (1.13 to 1.46)	42	7
Plus alcohol consumption	1.51 (1.34 to 1.72)	6	1.37 (1.22 to 1.54)	28	22
Plus diet	1.47 (1.26 to 1.69)	12	1.35 (1.19 to 1.53)	32	20
Plus physical activity	1.49 (1.31 to 1.69)	9	1.33 (1.20 to 1.49)	35	26
Fully adjusted	1.24 (1.08 to 1.41)	51	1.06 (0.92 to 1.21)	87	36

TABLE 3: Role of health behaviours in statistically accounting for the association between socioeconomic status and all-cause mortality: comparison of baseline and longitudinal assessment up to and including 2021.

Longitudinal models include phases 1, 3, 5, 7, 9, and 11 and use person-phase records for the 9,590 participants, among whom there were 2,427 deaths. The final column gives the difference in percentage attenuation between the baseline model and the model with repeated assessment of health behaviours. The age- and sex-adjusted baseline Model 1 HR of 1.55 was reused as the reference for calculating longitudinal percentage attenuation and was not refitted to the reshaped person-phase data.

CI: confidence interval; HR: hazard ratio; NA: not applicable; % Att.: percentage attenuation

These percentage attenuations were then plotted (Figure *1*) against those obtained above from assessment solely at baseline to show how the attenuation differs.

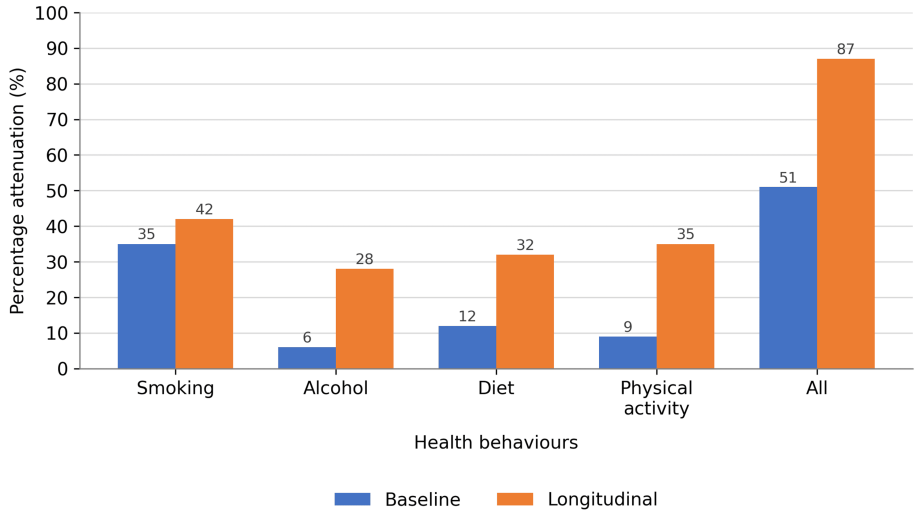


FIGURE 1: Comparison of percentage attenuation by health behaviour, baseline versus longitudinal assessment.

Percentage attenuation of the age- and sex-adjusted hazard ratio for all-cause mortality (lowest versus highest employment grade) after adjustment for each health behaviour, and for all four together, assessed at baseline only or as time-varying covariates across phases 1, 3, 5, 7, 9, and 11.

The attenuation increase for the time-varying exposure model compared to the baseline model was modest for smoking (35% to 42%) compared to diet (12% to 32%), physical activity (9% to 35%), and alcohol consumption (6% to 28%). This implies that the attenuation attributable to smoking was not substantially increased when entered as a time-dependent covariate, unlike for diet, physical activity, and alcohol consumption. More broadly, attenuation reached 87% when all four behaviours were entered together as time-varying covariates. At this point, the residual HR was 1.06 (95% CI: 0.92 to 1.21), and the association between employment grade and mortality was no longer statistically significant. Overall, smoking produced

the greatest attenuation, followed by physical activity, diet, and alcohol.

Incorporating more variables into the diet score

In the analyses above, the number of variables used in the scores for behaviours was low compared to the number available. Diet was classified using a scoring metric devised by Stringhini et al. [7] which incorporated only the type of bread and milk, and amount of fruit and vegetables consumed. To consider the combined effect of more variables on the HR, a new diet score was formulated (Figure *2*) where consumption of meat, cheese, and fat were added. This doubled the number of dietary factors considered, and included at least one variable corresponding to each of the five food groups in the NHS Eatwell Guide [16]. The new score retained the broad structure of the original, with the healthy and unhealthy categories describing clear patterns and the intermediate category covering everything lying in between these. Each added item was also in keeping with the Eatwell Guide, which advises choosing unsaturated over saturated fats, eating less red and processed meat, and choosing lower-fat dairy [16].

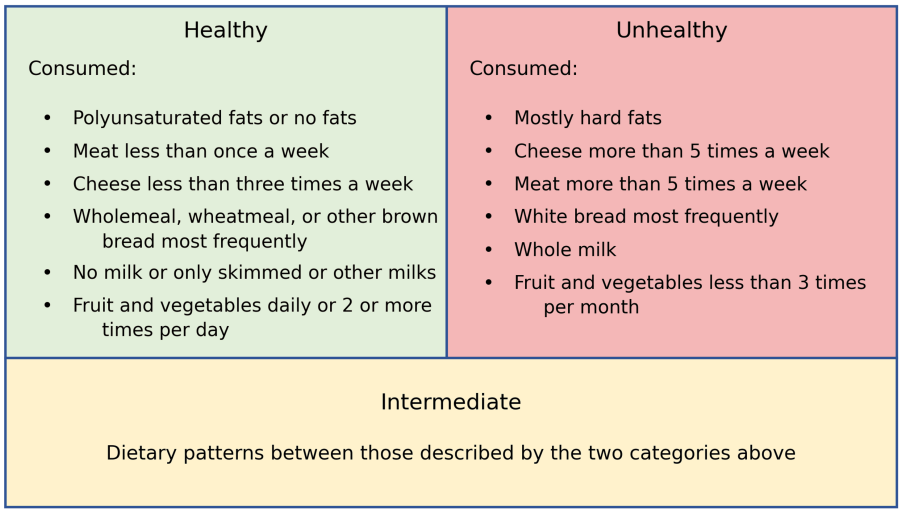


FIGURE 2: New diet scoring system, showing healthy, unhealthy, and intermediate categories.

The effects of this new diet score on the HRs for the longitudinal analyses, and associated percentage attenuations, are displayed in Table *4*.

Model	HR (95% CI)	% Attenuation
Model 1 (age and sex)	1.55 (1.37 to 1.76)	NA
Plus Stringhini et al. diet score	1.35 (1.19 to 1.53)	32
Plus new diet score	1.30 (1.15 to 1.48)	40
Full adjustment including Stringhini diet score	1.06 (0.92 to 1.21)	87
Full adjustment including new diet score	1.04 (0.91 to 1.18)	91

TABLE 4: Effect of incorporating more variables into the diet component of the health behaviours on the longitudinal analysis of their role in accounting for the association between socioeconomic status and all-cause mortality, deaths to 2021.

Models include phases 1, 3, 5, 7, 9, and 11. Of a total of 9,590 participants, there were 2,427 deaths to 2021. Full adjustment includes all health behaviours as in Table 2. The age- and sex-adjusted baseline Model 1 HR of 1.55 was reused as the reference for calculating longitudinal percentage attenuation and was not refitted to the reshaped person-phase data.

CI: confidence interval; HR: hazard ratio; NA: not applicable

When entered as a time-varying covariate, the new diet scoring attenuated the socioeconomic gradient in mortality more than the previous scoring: by 40% compared with 32%, a relative increase of 25%. Furthermore, when the new diet scores replace the old in the full adjustment along with the other health behaviours, attenuation increased to 91%. The construction of this score is described in Figure *2*; its thresholds were selected by the authors and have not been externally validated, and this is considered further in the Discussion section.

Discussion

The link between socioeconomic status and mortality is long established; among British civil servants, the first Whitehall study demonstrated a gradient in mortality by employment grade [17]. In this study, most of the association between socioeconomic status and mortality was attenuated by adjustment for health behaviours. This could be due to the increased prevalence of detrimental health behaviours in lower socioeconomic groups [18].

Considering changes in exposure to certain health behaviours across follow-up was found to increase the attenuation obtained. It has been suggested that those of a lower socioeconomic status are less willing or able to change health behaviours [19]. This would support our findings.

The longitudinal attenuation we report, 87%, is larger than the 72% first reported by Stringhini et al. [7]. The inclusion of two additional phases allows the covariates to track exposure more closely, and thereby absorb more of the gradient in mortality associated with socioeconomic status. However, the two models also differ in specification, as described in the Materials and Methods section, so the difference cannot be attributed to the additional phases and longer follow-up alone. In addition, the lengthened follow-up may increase the number of older participants who developed ill health and changed their health behaviours as a consequence, which would represent reverse causation, as discussed below. The increase in the total number of deaths from 654 to 2,427 improved precision, narrowing the CIs on the HRs obtained.

The increase in attenuation owing to longitudinal consideration seen for smoking was less than that seen for other health behaviours. This may be because smoking exerts its adverse effects over a long period of adult life, so baseline smoking status already captures much of this exposure. Hence, time-varying data adds little new information for this.

Building a more detailed dietary analysis increased the attenuation seen. There is evidence that those from lower socioeconomic backgrounds have, in general, poorer dietary behaviours than those from higher socioeconomic backgrounds [20]. This analysis demonstrates an association between this pattern and the gradient in mortality that likewise exists between these socioeconomic backgrounds.

Interpretation of the percentage attenuation

The percentage attenuation does not represent a causal pathway between the health behaviours and mortality, nor the proportion mediated by the health behaviours. The non-collapsibility of the HRs means that when covariates are entered, attenuation can arise for non-substantive reasons [21]. This is because conditioning changes the estimand even when the covariate does not lie on a causal pathway.

When a time-varying covariate is simultaneously dependent on prior exposure and predictive of the outcome, conventional Cox regression with time-dependent covariates may give biased estimates, and marginal structural models using inverse probability weighting have been developed for this setting [22]. Health behaviours plausibly operate in this way, and as such the use of conventional Cox regression is a limitation. In a later analysis of the Whitehall II and French GAZEL cohorts, Stringhini et al. reported that health behaviours assessed over follow-up attenuated the association between socioeconomic status and mortality by 75% in Whitehall II [23].

Percentage attenuations were calculated against the age- and sex-adjusted model as fitted at baseline, following the approach of Stringhini et al. [7]. The model was not refitted after the phasic reshaping of the data. However, the model contains only variables that are fixed over time (employment grade and sex, with age as the time scale), and the reshaping only split each participant's follow-up by phase, without changing who was followed or for how long. Refitting it would therefore, in principle, return the same estimate, so the longitudinal attenuation is unlikely to be affected by the restructuring of the data.

Reverse causation

The effect of reverse causation should also be considered: poorer health at earlier phases could lead to physical inactivity at later phases, or to inability to work [24], and thus to afford what constitutes a healthy diet at later phases [25]. This may partly explain why longitudinal consideration attenuated the gradient more than baseline consideration, and we therefore interpret the increase in attenuation with caution. This could be examined further by lagging the time-varying covariates by one phase, or excluding deaths soon after exposure measurement. As this was not done, the longitudinal analysis remains limited.

Limitations

There are no CIs on the percentage attenuations achieved. This makes the results of the study more difficult to interpret because CIs provide information about whether differences observed are significant. For instance, Stringhini et al. [7] reported intervals of 21% to 94% around their baseline attenuation of 42%, and 42% to 154% around their longitudinal attenuation of 72%. Widths of this order imply that attenuations are imprecisely estimated. As such, differences between individual attenuation estimates should not be taken as statistically significant, although the increase in attenuation from baseline to longitudinal assessment was consistent across all four behaviours, and in the same direction as reported by Stringhini et al. [7]. This limitation could have been addressed by bootstrapping, which was not undertaken here.

The baseline model also differed from that of Stringhini et al. [7] in its specification, as described in the Materials and Methods section. As access to the dataset and hosting platform had ended, the model could not be refitted using the original specification, so the effect of longer follow-up could not be fully distinguished from that of the change in model specification.

Is Whitehall II representative of the wider population? It contains information pertaining to mostly male, mostly white, white-collar workers living in London and does not account for those unemployed or living in poverty. Whitehall II was established with a view to examining the social gradient in health and disease [11], following the first Whitehall study's finding that these gradients ran in the same direction as national social class mortality data, though larger. The magnitude of the associations may not generalise exactly to the wider population, but the cohort remains informative about relationships between socioeconomic status, health behaviours, and mortality.

Socioeconomic status was measured only at baseline. Although this was a deliberate choice, for the reasons discussed in the Materials and Methods section, later changes in employment, income, or wider circumstances were not captured, and may have altered some participants' socioeconomic position over follow-up.

Health behaviours were also grouped into broad categories, following Stringhini et al. [7], to allow for comparison. This may have led to a loss of detail, for example, on the amount of alcohol consumed within each category, and more precise measures may have attenuated the gradient further.

Responding to questionnaires about health behaviours could lead to self-report bias, such as recall bias and social desirability bias. Participants may not accurately remember behaviours such as diet and physical activity, and social pressure may lead to an unwillingness to admit to not having acted in a socially desirable manner. Self-administered questionnaires reduce susceptibility to information bias such as social desirability bias; however, they are more prone to missing data on sensitive information.

There is no accounting for health behaviours that existed prior to the study, but which may have changed prior to the commencement. For instance, ex-smokers who had ceased for health reasons, or people who had stopped drinking for health reasons rather than being lifelong abstainers. In the latter case, the reference category for alcohol does not comprise only lifelong abstainers, which may explain its comparatively low attenuation.

Additionally, this study does not enable us to draw conclusions on the relative importance of these selected health behaviours in comparison with other factors, including psychosocial factors, because this analysis was not included. To improve this study, we could include psychosocial and financial factors in the models and compare their effects.

The proportional hazards assumption was made. This implies that the HRs associated with the variables remain constant over follow-up. If this were not the case, then inferences of the analysis might be false, which would be detrimental to the validity of the study. This was not formally assessed, and with follow-up of this duration, it is plausible that the effect of a variable such as smoking status might vary. To assess for violation of the assumption, Schoenfeld residuals could be calculated.

Imputation with data from the previous phase for longitudinal analysis assumes that behaviours did not change between phases, which may have led to misclassification of some participants and thus to underrepresentation of changes in health behaviours in this study. However, around 25% of the participants had missing data at one or more phases, and therefore removing these participants entirely as at baseline would mean loss of a substantial section of data; those with missing data are also likely to differ from those without. Understating behavioural change would be expected to reduce, rather than inflate, the longitudinal attenuation, provided missing data were similar across grades. Multiple imputation would nonetheless have been a preferable approach.

The thresholds used in the expanded diet score were selected by the authors and have not been externally validated, so the additional attenuation it produced should be regarded as exploratory.

Future implications

Our findings may have important public health implications. There is certainly a plausible and strong association between the groups of factors studied (smoking, diet, exercise, alcohol consumption) and morbidity or mortality [26-30]. This suggests that public health interventions addressing health behaviours may attenuate health inequalities and improve population health, though this analysis does not estimate the effect of any such interventions.

Conclusions

Assessing health behaviours at a single timepoint understates their attenuation of the socioeconomic gradient in mortality. In this cohort, adjusting for smoking, alcohol consumption, diet, and physical activity assessed repeatedly across adult life, with longer mortality follow-up, attenuated substantially more of the gradient than assessment at baseline alone, to the extent that the residual gradient was no longer statistically significant. A more detailed diet score increased this attenuation further. Information on exposure at one point in the life course does not represent the whole life experience of the population, although part of the additional attenuation from repeated assessment may reflect reverse causation. Different socioeconomic groups are subject to an array of social, cultural, and economic influences over their life course, which shape their health behaviours, and how these change over time is also influenced by this array of factors. Together, these findings suggest that targeting actionable factors through public health interventions may contribute to lessening some of the social inequalities in mortality.

Additional Information

Author Contributions

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work.

Concept and design: Danyal Sarwar, Rebecca E. Harounoff, Eric Brunner, John Shawe-Taylor, Maria Perez Ortiz

Acquisition, analysis, or interpretation of data: Danyal Sarwar, Rebecca E. Harounoff, Eric Brunner, Yuntao Chen

Drafting of the manuscript: Danyal Sarwar, Rebecca E. Harounoff

Critical review of the manuscript for important intellectual content: Danyal Sarwar, Rebecca E. Harounoff, Eric Brunner, John Shawe-Taylor, Maria Perez Ortiz, Yuntao Chen

Supervision: John Shawe-Taylor, Maria Perez Ortiz

Disclosures

Human subjects: Informed consent for treatment and open access publication was obtained or waived by all participants in this study. London - Harrow Research Ethics Committee issued approval 85/0938. The Whitehall II study was approved by the London - Harrow Research Ethics Committee (REC reference 85/0938; IRAS project ID 142374), with ethics approvals renewed at each phase of data collection. All participants provided written informed consent. This secondary analysis was conducted within the objectives of the approved study protocol, using data accessed through Dementias Platform UK (DPUK study ID 0475). **Animal subjects:** All authors have confirmed that this study did not involve animal subjects or tissue. **Conflicts of interest:** In compliance with the ICMJE uniform disclosure form, all authors declare the following: **Payment/services info:** The Whitehall II study has been supported by grants from the Wellcome Trust (221854/Z/20/Z); the UK Medical Research Council (Y014154, K013351, R024227, S011676); the British Heart Foundation (PG/11/63/29011, RG/13/2/30098, RG/16/11/32334); the British Health and Safety Executive; the British Department of Health; the National Heart, Lung, and Blood Institute (R01HL036310); the National Institute on Aging, National Institutes of Health (R01AG013196, R01AG034454); and the Economic and Social Research Council (ES/J023299). Data access was provided through Dementias Platform UK, which is funded by the Medical Research Council (MR/T033371/1). Yuntao Chen is supported by the Wellcome Trust through a Wellcome Early-Career Award (227639/Z/23/Z). Eric Brunner was supported by the Economic and Social Research Council (ES/T014377). This secondary analysis received no specific funding. **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work. **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

Acknowledgements

Danyal Sarwar and Rebecca Harounoff contributed equally to the work and should be considered co-first authors. The data underlying this study were accessed through the Dementias Platform UK (DPUK) secure

platform under DPUK study ID 0475 and cannot be shared by the authors; access to Whitehall II data can be requested directly from DPUK. DPUK is funded by the Medical Research Council (MR/T033371/1). We thank all participating civil service departments and their welfare, personnel, and establishment officers; the British Occupational Health and Safety Agency; the British Council of Civil Service Unions; all participating civil servants in the Whitehall II study; and all members of the Whitehall II study team.

References

1. Living longer: how our population is changing and why it matters . (2018). Accessed: November 1, 2026: https://www.ons.gov.uk/releases/livinglongerhowourpopulationischangingandwhyitmatters.
2. World Population Prospects 2019: Highlights. (2019). Accessed: November 1, 2026: https://population.un.org/wpp/Publications/Files/WPP2019_Highlights.pdf.
3. Ahmadi-Abhari S, Guzman-Castillo M, Bandosz P, et al.: Temporal trend in dementia incidence since 2002 and projections for prevalence in England and Wales to 2040: modelling study. BMJ. 2017, 358:j2856. 10.1136/bmj.j2856
4. Pandey S: Factors contributing of ageing: factors in ageing . Handbook of Research on Geriatric Health, Treatment, and Care. Vijaya Prasad B, Akbar S (ed): IGI Global, Hershey, PA; 2018. 393-408. 10.4018/978-1-5225-3480-8.ch022
5. Marmot M: Fair Society, Healthy Lives: The Marmot Review. Strategic Review of Health Inequalities in England Post-2010. The Marmot Review, London; 2010.
6. Nurse J, Dorey S, Yao L, et al.: The Case for Investing in Public Health: A Public Health Summary Report for EPHO 8. WHO Regional Office for Europe, Copenhagen; 2014.
7. Stringhini S, Sabia S, Shipley M, Brunner E, Nabi H, Kivimaki M, Singh-Manoux A: Association of socioeconomic position with health behaviors and mortality. JAMA. 2010, 303:1159-66. 10.1001/jama.2010.297
8. Mokdad AH, Marks JS, Stroup DF, Gerberding JL: Actual causes of death in the United States, 2000 . JAMA. 2004, 291:1238-45. 10.1001/jama.291.10.1238
9. Brunner EJ: Social factors and cardiovascular morbidity. Neurosci Biobehav Rev. 2017, 74:260-8. 10.1016/j.neubiorev.2016.05.004
10. Lewer D, Jayatunga W, Aldridge RW, Edge C, Marmot M, Story A, Hayward A: Premature mortality attributable to socioeconomic inequality in England between 2003 and 2018: an observational study. Lancet Public Health. 2020, 5:e33-41. 10.1016/S2468-2667(19)30219-1
11. Marmot M, Brunner E: Cohort profile: the Whitehall II study. Int J Epidemiol. 2005, 34:251-6. 10.1093/ije/dyh372
12. Kivimaki M: Whitehall II (Data Set). Dementias Platform UK, Oxford; 2017. 10.48532/048000
13. Clark TG, Bradburn MJ, Love SB, Altman DG: Survival analysis part I: basic concepts and first analyses . Br J Cancer. 2003, 89:232-8. 10.1038/sj.bjc.6601118
14. Singh R, Mukhopadhyay K: Survival analysis in clinical trials: basics and must know areas . Perspect Clin Res. 2011, 2:145-8. 10.4103/2229-3485.86872
15. Stel VS, Dekker FW, Tripepi G, Zoccali C, Jager KJ: Survival analysis II: Cox regression. Nephron Clin Pract. 2011, 119:c255-60. 10.1159/000328916
16. Public Health England: The Eatwell Guide. Public Health England, London; 2016. https://www.gov.uk/government/publications/the-eatwell-guide.
17. Marmot MG, Shipley MJ, Rose G: Inequalities in death--specific explanations of a general pattern? . Lancet. 1984, 1:1003-6. 10.1016/s0140-6736(84)92337-7
18. Lynch JW, Kaplan GA, Salonen JT: Why do poor people behave poorly? Variation in adult health behaviours and psychosocial characteristics by stages of the socioeconomic lifecourse. Soc Sci Med. 1997, 44:809-19. 10.1016/s0277-9536(96)00191-8
19. Winkleby MA, Flora JA, Kraemer HC: A community-based heart disease intervention: predictors of change . Am J Public Health. 1994, 84:767-72. 10.2105/ajph.84.5.767
20. Darmon N, Drewnowski A: Does social class predict diet quality?. Am J Clin Nutr. 2008, 87:1107-17. 10.1093/ajcn/87.5.1107
21. Martinussen T, Vansteelandt S: On collapsibility and confounding bias in Cox and Aalen regression models . Lifetime Data Anal. 2013, 19:279-96. 10.1007/s10985-013-9242-z
22. Robins JM, Hernán MA, Brumback B: Marginal structural models and causal inference in epidemiology . Epidemiology. 2000, 11:550-60. 10.1097/00001648-200009000-00011
23. Stringhini S, Dugravot A, Shipley M, et al.: Health behaviours, socioeconomic status, and mortality: further analyses of the British Whitehall II and the French GAZEL prospective cohorts. PLoS Med. 2011, 8:e1000419. 10.1371/journal.pmed.1000419
24. García-Gómez P, van Kippersluis H, O'Donnell O, van Doorslaer E: Long term and spillover effects of health shocks on employment and income. J Hum Resour. 2013, 48:873-909. 10.1353/jhr.2013.0031
25. Darmon N, Drewnowski A: Contribution of food prices and diet cost to socioeconomic disparities in diet quality and health: a systematic review and analysis. Nutr Rev. 2015, 73:643-60. 10.1093/nutrit/nuv027
26. Doll R, Peto R, Boreham J, Sutherland I: Mortality in relation to smoking: 50 years' observations on male British doctors. BMJ. 2004, 328:1519. 10.1136/bmj.38142.554479.AE
27. Rehm J, Gmel G, Sempos CT, Trevisan M: Alcohol-related morbidity and mortality. Alcohol Res Health. 2003, 27:39-51.
28. Kenfield SA, Stampfer MJ, Rosner BA, Colditz GA: Smoking and smoking cessation in relation to mortality in women. JAMA. 2008, 299:2037-47. 10.1001/jama.299.17.2037
29. Biddle SJ, Bennie JA, Bauman AE, et al.: Too much sitting and all-cause mortality: is there a causal link? . BMC Public Health. 2016, 16:635. 10.1186/s12889-016-3307-3
30. Health effects of dietary risks in 195 countries, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet. 2019, 393:1958-72. 10.1016/S0140-6736(19)30041-8