Title: Exploring health in the UK Biobank: associations with sociodemographic characteristics, psychosocial factors, lifestyle and environmental exposures

Short title: Health in UK Biobank

Authors:

Julian Mutz[1*], Charlotte J. Roscoe[2,3], Cathryn M. Lewis[1,4]

Affiliations:

1. Social, Genetic & Developmental Psychiatry Centre, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, United Kingdom
2. Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London, London, United Kingdom
3. Harvard T.H. Chan School of Public Health, Boston, United States
4. Department of Medical and Molecular Genetics, Faculty of Life Sciences & Medicine, King's College London, London, United Kingdom

[*]**Corresponding author:**

Julian Mutz; Social, Genetic & Developmental Psychiatry Centre, Institute of Psychiatry, Psychology & Neuroscience, King's College London; Memory Lane, London SE5 8AF, United Kingdom. Email: julian.mutz@gmail.com. Twitter: @julianmutz.

Abstract

Objectives: Examine sociodemographic characteristics, psychosocial factors, lifestyle and environmental exposures associated with multiple indicators of health.

Design: Observational study of UK Biobank participants recruited between 2006–2010. Repeat assessments in subsets of the sample between 2012–2013 and 2014–2019.

Setting: United Kingdom.

Participants: 502,521 men and women, aged 37–73.

Main outcome measures: 81 cancer and 443 non-cancer illnesses to classify participants by health status (non-standard risk/standard risk). Long-standing illness (yes/no) and self-rated health (poor/fair/good/excellent) constituted secondary outcomes.

Results: 307,378 participants (mean age = 56.1 years, SD = 8.07; 51.91% female) selected for cross-sectional analyses. 16,058 and 32,617 participants selected for longitudinal analyses of self-rated health. Lower annual household income (OR = 0.74, 95% CI 0.726 to 0.764 [<£18,000 vs £31,000–51,999]), higher levels of neighbourhood deprivation (OR = 0.995, 95% CI 0.995 to 0.996), being male (OR = 0.88, 95% CI 0.870 to 0.899), loneliness (OR = 0.81, 95% CI 0.786 to 0.840) and social isolation (OR = 0.95, 95% CI 0.923 to 0.977) was associated with poor health. Walking more frequently (OR = 1.01, 95% CI 1.006 to 1.015) and engaging more frequently in vigorous physical activity (OR = 1.03, 95% CI 1.027 to 1.037) was associated with favourable health, whereas higher body mass index (OR = 0.968, 95% CI 0.967 to 0.970) and smoking (OR = 0.75, 95% CI 0.731-0.773 [current vs never]) was associated with poor health. Infrequent alcohol intake was associated with poor health (OR = 0.91, 95% CI 0.887 to 0.940 [1-3 times/month vs 1-2 times/week]), whereas a more frequent drinking pattern was associated with favourable health (OR = 1.06, 95% CI 1.035 to 1.084 [3-4 times/week vs 1-2 times/week]). There was some evidence that higher levels of airborne pollutants ($PM_{2.5}$, PM_{10} and NO_2) and noise (L_{den}) were associated with poor health, though findings were inconsistent after adjustment for other factors. Neighbourhood greenspace percentage was associated with positive self-rated health.

Conclusions: Public health medicine could put greater focus on non-medical factors, such as loneliness, further encourage healthy lifestyle behaviours and weight management, and examine efforts to improve health outcomes of individuals in the lowest income groups.

Introduction

Substantial improvements in human health and significant increases in average life expectancy are amongst the main achievements of civilization over the past two centuries[1]. Life expectancy at birth in England and Wales has doubled from about 40 years in 1850 to more than 80 years in 2013[2]. Future life expectancy is projected to increase across 35 industrialised nations, with a median increase in life expectancy at birth of approximately 3 years for women and 4 years for men in the United Kingdom (UK) between 2010–2030[3].

Many adults, however, now spend a substantial portion of their lives with late-life morbidities and decreased sensory, motor and cognitive functioning, which may lead to lower quality of life[4,5]. More than 80% of the population of England and Wales aged 85 and above report a disability and 50% require care or help with daily activities[6]. Nevertheless, some individuals experience very little functional decline in old age and rate their health as good or excellent[7]. A greater understanding of the factors associated with positive health may help increase longevity and healthspan, i.e. the length of time that a person is healthy[8].

The health effects of lifestyle factors such as smoking[9], diet[10], excessive alcohol intake[11] and physical activity[12] are well established. However, the number and type of potential confounders that have been adjusted for in previous analyses vary widely[13], and few studies have jointly examined sociodemographic characteristics, psychosocial factors, lifestyle factors and environmental exposures. In addition, research has primarily focussed on predicting mortality instead of healthy ageing, despite its potential to provide insights into the factors associated with a longer healthspan.

The UK Biobank study provides an unprecedented data resource to investigate determinants of health and ageing trajectories. One of its strengths is that data collection was not merely focussed on established predictors of health and disease but also included relevant non-medical factors such as social isolation and loneliness, as well as traffic-related noise and air pollution[14,15].

The aim of the present study was to explore potential predictors of health status. More specifically, we examined how health status was associated with (i) sociodemographic characteristics and psychosocial factors, (ii) lifestyle factors and (iii) environmental exposures, both independently and jointly. We also examined, as a secondary aim, whether there was evidence of similar associations between these factors and (i) long-standing illness and (ii) self-rated health.

Methods

Study description

The UK Biobank is a prospective study of 502,521[i] UK residents aged 37–73 at baseline, recruited between 2006–2010. Details of the study rationale and design have been reported elsewhere[16]. All variables used in the present study are listed in Supplement-e1.

Briefly, individuals registered with the UK National Health Service (NHS) and living within a 25-mile (~40 km) radius of one of 22 assessment centres were invited to participate (9,238,453 postal invitations sent). At the baseline assessment, participants completed questionnaires and were interviewed by nurses regarding sociodemographic characteristics, health behaviours and their medical history. They also underwent physical examination including height, weight and blood pressure measurement, and had blood and urine samples taken. A subset of participants completed repeat measurements: first revisit between 2012–2013; second revisit as part of the UK Biobank Imaging Study between 2014–2019. Participants consented to use of their de-identified data.

Sociodemographic characteristics

We considered individual-level sociodemographic characteristics including age at baseline assessment, sex, ethnicity (White, Asian, Black, Chinese, mixed-race or other), highest educational or professional qualification (four categories: (1) College/University degree; (2) A levels/AS levels or equivalent, NVQ/HND/HNC or equivalent, other professional qualifications; (3) O levels/GCSEs or equivalent, CSEs or equivalent; (4) none)[17] and gross annual household income (<£18,000, £18,000–30,999, £31,000–51,999, £52,000–100,000 or >£100 000). We also included the Index of Multiple Deprivation for England which is a small-area level measure derived from data on income deprivation, employment deprivation, health deprivation and disability, education skills and training

[i] Date: November 2019

deprivation, barriers to housing and services, living environment deprivation and crime[18]. Higher values on the index reflect greater deprivation.

Psychosocial factors

Loneliness was assessed using two questions: "Do you often feel lonely?" (no = 0 / yes = 1) and "How often are you able to confide in someone close to you?" (almost daily to about once a month = 0 / once every few months to never or almost never = 1). Individuals who received a sum score of 2 were classified as lonely[19].

Social isolation was assessed using three questions: "Including yourself, how many people are living together in your household?" (living alone = 1), "How often do you visit friends or family or have them visit you?" (less than once a month = 1) and "Which of the following [leisure/social activities] do you attend once a week or more often?" (none of the above = 1). Individuals who received a sum score of 2–3 were classified as socially isolated[19].

Lifestyle factors

Smoking status was assessed using two questions summarising current and past smoking behaviour. Individuals who responded "Yes, on most or all days" or "Only occasionally" to current tobacco smoking were coded as "current". Individuals who responded "Smoked on most or all days" or "Smoked occasionally" to past tobacco smoking were coded as "former". Individuals who responded "No" to current tobacco smoking and "Just tried once or twice" or "I have never smoked" to past tobacco smoking were coded as "never".

Alcohol intake frequency was assessed using one question: "About how often do you drink alcohol?". Response options included "Daily or almost daily", "Three or four times a week", "Once or twice a week", "One to three times a month", "Special occasions only" and "Never".

Sleep duration was assessed using one question: "About how many hours sleep do you get in every 24 hours? (please include naps)". Values below 1 hour or above 23 hours were rejected, and participants were asked to confirm values below 3 hours or above 12 hours.

Body mass index (BMI) was calculated as weight divided by height squared (kg/m^2). Weight measurements were obtained with a Tanita BC-418 MA body composition analyser. Standing height measurements were obtained using a Seca 202 height measure.

Physical activity was assessed using the International Physical Activity Questionnaire (IPAQ) short form[20]. Specifically, we included data on the number of days per week spent walking, engaging in moderate-intensity physical activity (e.g. "carrying light loads, cycling at normal pace") or engaging in vigorous-intensity physical activity (i.e. "activities that make you sweat or breathe hard such as fast cycling, aerobics, heavy lifting") for ≥10 minutes continuously.

Daily frequency of stair climbing at home was assessed using one question: "At home, during the last 4 weeks, about how many times a day do you climb a flight of stairs? (approx 10 steps)". These data were only collected from individuals who indicated that they were able to walk. Response options included "None", "1–5 times a day", "6–10 times a day", "11–15 times a day", "16–20 times a day" and "More than 20 times a day".

Environmental exposures

Exposure to traffic-related airborne pollutants was estimated using a Land Use Regression model developed as part of the European Study of Cohorts for Air Pollution Effects (ESCAPE) project[21,22]. We included annual average concentration of particulate matter with an aerodynamic diameter of <2.5 μm ($PM_{2.5}$) and <10 μm (PM_{10}), as well as nitrogen dioxide (NO_2) modelled at participants' residential addresses for the year 2010.

164 Residential road traffic noise was modelled for the year 2009 using the Common Noise Assessment
165 Methods (CNOSSOS-EU) algorithm[23,24]. We used L_{den} (day-evening-night noise level) which is an
166 annual average 24-hour sound pressure level in decibels with a 10-decibel penalty added between
167 11pm and 7am. This penalty has previously been added in epidemiological analyses to account for
168 annoyance/sleep disruption at night[25].
169
170 The percentage of the home location classed as greenspace, as a proportion of all land use types, was
171 modelled using 2005 data from the Generalized Land Use Database for England (GLUD)[26] for the
172 2001 Census Output Areas in England. Each residential address was allocated a circular distance
173 buffer of 1000m.
174
175 **Health status**
176 Data on 81 cancer and 443 non-cancer illnesses (past and current) were ascertained through
177 touchscreen self-report questionnaire and confirmed during a verbal interview by a trained nurse. In
178 order to provide a single health indicator ("health status") based on a previously defined algorithm,
179 we used a classification developed by the Reinsurance Group of America (RGA) from an insurance
180 underwriting perspective. Participants were classified as having "standard risk" or "non-standard risk"
181 health status based on their reported cancer and non-cancer illnesses (Supplement-e2).
182
183 Two secondary health outcomes were assessed. Firstly, whether patients had a long-standing illness,
184 disability or infirmity was assessed using the question "Do you have any long-standing illness,
185 disability or infirmity?" to which individuals could respond "Yes" or "No". Secondly, participants'
186 perceived health was assessed using the question "In general how would you rate your overall
187 health?". Response options included "Poor", "Fair", "Good" and "Excellent". These secondary health
188 outcomes will be termed "long-standing illness" and "self-rated health".
189
190 **Exclusion criteria**

Women who were pregnant at the time of assessment were excluded from the analysis based on the assumption that lifestyle patterns change during pregnancy[27]. Participants for whom their genetic sex, inferred from the relative intensity of biological markers on the Y and X chromosomes, and self-reported sex did not match were also excluded.

Statistical analyses

Participants with baseline data on all explanatory variables and health indicators were selected for cross-sectional analysis. Participants with missing data or who responded "do not know" or "prefer not to answer" were excluded.

Logistic regression analyses were used to examine associations between sociodemographic characteristics, psychosocial factors, lifestyle factors and environmental exposures with health status and long-standing illness. Ordinal logistic regression analyses were used to examined associations between these factors and self-rated health. For these outcomes, poor health was the reference group. In defining the reference groups for categorical explanatory variables, we focussed on interpretability of the results for UK citizens in this age group.

We fitted three incremental models using the baseline UK Biobank data: Model 1–individual explanatory variables; Model 2–individual explanatory variables plus age and sex; Model 3–all explanatory variables. All models were fitted in the analytical sample without missing data to ensure that any differences between the unadjusted and adjusted models were not due to inclusion of different participants. We calculated odds ratios and 95% confidence intervals. *P*-values were adjusted for multiple testing within each Model 3 fit using two methods: (1) Bonferroni and (2) Benjamini & Hochberg[28]. Multicollinearity was assessed using generalised variance inflation factors[29].

To compare the magnitudes of association between the explanatory variables and health indicators, we calculated standardised regression coefficients by rescaling all variables included in Model 3 to

219 have a mean of zero and by dividing the coefficients of numeric variables with more than two values
220 by twice their standard deviation[30].
221
222 To assess whether any associations between explanatory variables and health indicators were
223 modified by sex or age group, we stratified analyses by sex and by age at baseline assessment (<65
224 years and ≥65 years) and assessed potential age- and sex-interactions by adding cross-product terms
225 to Model 3. The choice of age strata was based on the current UK retirement age.
226
227 We used ordinal logistic regression analyses to model associations between the explanatory variables
228 at baseline and self-rated health at the first and second revisit (t1 and t2, respectively). We applied the
229 same modelling strategy, except that we additionally adjusted for the number of days between
230 baseline and follow-up assessment.
231
232 Additional analyses were performed in subsets of the data with further classification of risk factors.
233 For full details on methods, see Supplement-e3.
234
235 Statistical analyses were conducted using R (Version 3.5.0).

Results

Sample characteristics

Of the N = 502,521[ii] participants in the UK Biobank, 5.32% (n = 26,757) had missing data on health indicators and another 33.51% (n = 168,386) had missing data on explanatory variables or did not meet our inclusion criteria. We yielded an analytical sample of n = 307,378 adults (Figure-1). There were <2.5% missing data (excluding "do not know" and "prefer not to answer") for all variables except neighbourhood deprivation, particulate air pollution and greenspace. Descriptive statistics of the full and analytical sample are presented in Supplement-e4.

[insert Figure-1]

Most participants were White (95.51%, n = 293,565) and 51.91% (n = 159,574) were women. Their mean age was 56.1 years (SD = 8.07). 12.96% (n = 39,828) of participants reported no educational or professional qualification, while 35.96% (n = 110,518) had received a college/university degree. 26.79% (n = 82,338) of participants reported a household income between £31,000–51,999, while the other participants were relatively evenly spread across income groups, except that fewer individuals reported an income of >£100,000 (5.82%, n = 17,904). 8.60% (n = 26,447) of participants were classified as socially isolated, while 5.69% (n = 17,477) were classified as lonely.

For airborne pollutants, mean annual exposure estimates in $\mu g/m^3$ were 9.95 (SD = 1.04) for $PM_{2.5}$, 16.19 (SD = 1.88) for PM_{10} and 26.43 (SD = 7.56) for NO_2. The median duration of living at the current address was 16 years (IQR = 17).

8.18% (n = 25,142) of participants reported at least one cancer (range = 0–6) and 73.30% (n = 225,312) at least one non-cancer illness (range = 0–29) (Supplement-e5-A–B). 69.04% (n = 212,201) of participants were classified as standard risk health status, while 30.96% (n = 95,177) were classified

[ii] Date: November 2019

262 non-standard risk. A similar proportion of participants (30.5%, n = 93,757) reported that they had a
263 long-standing illness (72.9% agreement with health status). Finally, 3.60% (n = 11,066) rated their
264 health as poor, while 19.25% (n = 59,169), 59.44% (n = 182,699) and 17.71% (n = 54,444) rated their
265 health as fair, good and excellent, respectively.
266
267 Descriptive statistics of the analytical sample stratified by health indicators are presented in
268 Supplement-e6-A–C.
269
270 **Regression analyses**
271 Generalised variable inflation factor (VIF) values were within acceptable limits for 19/21 explanatory
272 variables (range = 1.02–2.62, except for NO_2 (VIF = 6.08) and $PM_{2.5}$ (VIF = 4.27)).
273
274 Results presented below correspond to multivariable-adjusted odds ratios for Model 3 including all
275 explanatory variables, unless indicated otherwise. A visual summary of our findings is presented in
276 Supplement-e7.
277
278 Increased age was associated with lower odds of being classified standard risk (OR = 0.953, 95% CI
279 0.952-0.954, $p_{\text{Bonf.}}$<0.001). For every 1-year increase in age, the log-odds of being classified standard
280 risk decreased by -0.048 (95%CI -0.050 to -0.047). Men had lower odds of being classified standard
281 risk (OR = 0.88, 95% CI 0.870-0.899, $p_{\text{Bonf.}}$<0.001). Individuals with higher levels of income had
282 higher odds of being classified standard risk (e.g. OR = 1.05, 95% CI 1.006-1.090, p_{BH}=0.028
283 [>£100,000 vs £31,000–51,999]), while those with lower levels of income had lower odds of being
284 classified standard risk (e.g. OR = 0.74, 95% CI 0.726-0.764, $p_{\text{Bonf.}}$<0.001 [<£18,000 vs £31,000–
285 51,999]). Increased neighbourhood deprivation was associated with lower odds of being classified
286 standard risk (OR = 0.995, 95% CI 0.994-0.996, $p_{\text{Bonf.}}$<0.001). Compared to Whites, individuals of
287 Black, Chinese, Mixed-race and "other" ethnic background had higher odds of being classified
288 standard risk in Model 1, but only Chinese ethnicity was associated with higher odds of being
289 classified standard risk across all models (Model 3: OR = 1.83, 95% CI 1.529-2.215, $p_{\text{Bonf.}}$<0.001).

290 Compared to individuals without educational or professional qualification, participants with any
291 qualification had higher odds of being classified standard risk in Model 1 and after adjustment for age
292 and sex (Model 2). However, findings were less consistent in Model 3 (Table-1-A).
293
294 [insert Table-1-A]
295
296 Findings regarding income, sex and neighbourhood deprivation were consistent across health
297 indicators, although with some variation in the magnitude of associations (Figure-2-A; Supplement-
298 e10-A). Compared to Whites, participants of Chinese ethnicity had lower odds of rating their health
299 favourable (OR = 0.68, 95% CI 0.591-0.778, $p_{\text{Bonf.}}<0.001$) but had higher odds of being classified
300 standard risk (OR = 1.83, 95% CI 1.529-2.215, $p_{\text{Bonf.}}<0.001$) and not having a long-standing illness
301 (OR = 1.62, 95% CI 1.362-1.939, $p_{\text{Bonf.}}<0.001$) (Supplement-e8-A). Individuals of non-white ethnic
302 backgrounds generally rated their health less favourable, but had higher odds of not having a long-
303 standing illness in Model 3 (Supplement-e9-A; e10-A). Individuals with any qualification had higher
304 odds of rating their health more favourable (e.g. OR = 1.50, 95% CI 1.465-1.542, $p_{\text{Bonf.}}<0.001$
305 [university/college degree vs no qualification]), while there was no consistent pattern for health status
306 and long-standing illness. Finally, older individuals tended to rate their health more favourable (OR =
307 1.011, 95% CI 1.010-1.012, $p_{\text{Bonf.}}<0.001$).
308
309 [insert Figure-2-B]
310
311 Loneliness was associated with lower odds of being classified standard risk (OR = 0.81, 95% CI
312 0.786-0.840, $p_{\text{Bonf.}}<0.001$). Socially isolated individuals had lower odds of being classified standard
313 risk as well, but there was greater attenuation in the strength in association in Model 3 (OR = 0.95,
314 95% CI 0.923-0.977, $p_{\text{Bonf.}}=0.013$) (Table-1-B).
315

316 These findings were consistent across health indicators, but the association was strongest for
317 loneliness and self-rated health (OR = 0.49, 95% CI 0.477-0.507, $p_{Bonf.}$<0.001) (Figure-2-B;
318 Supplement-e8-B; e9-B; e10-B).
319
320 [insert Table-1-B]
321
322 Longer sleep duration was associated with lower odds of being classified standard risk (OR = 0.97,
323 95% CI 0.968-0.982, $p_{Bonf.}$<0.001). For every 1-hour increase in sleep duration, the log-odds of being
324 classified standard risk decreased by -0.03 (95%CI -0.033 to -0.018). Walking more frequently and
325 engaging in more frequent vigorous physical activity was also associated with higher odds of being
326 classified standard risk (OR = 1.01, 95% CI 1.006-1.015, $p_{Bonf.}$<0.001 and OR = 1.03, 95% CI 1.027-
327 1.037, $p_{Bonf.}$<0.001, respectively). For every one-day increase in ≥10 minutes of vigorous physical
328 activity, the log-odds of standard risk increased by 0.03 (95%CI 0.027-0.036). While moderate
329 physical activity was associated with higher odds of being classified standard risk in Model 1–2, we
330 did not find evidence of an association in Model 3 (OR = 1.00, 95% CI 0.999-1.007, p=0.13). More
331 frequent daily stair climbing at home was associated with higher odds of being classified standard risk
332 (ranging from OR = 1.07, 95% CI 1.036-1.106, $p_{Bonf.}$=0.002 to OR = 1.19, 95% CI 1.140-1.239,
333 $p_{Bonf.}$<0.001 [1-5 times/day and >20 times/day vs none, respectively]). Reporting more frequent
334 alcohol intake was associated with higher odds of being classified standard risk (e.g. OR = 1.06, 95%
335 CI 1.035-1.084, $p_{Bonf.}$<0.001 [3-4 vs 1-2 times/week]), except daily alcohol drinking in Model 1 (OR
336 = 0.93, 95% CI 0.909-0.950), while less frequent alcohol intake was associated with lower odds of
337 being classified standard risk (ranging from OR = 0.91, 95% CI 0.887-0.939, $p_{Bonf.}$<0.001 to OR =
338 0.63, 95% CI 0.608-0.650, $p_{Bonf.}$<0.001 [1-3 times/month and never vs 1-2 times/week, respectively]).
339 Higher BMI was associated with lower odds of being classified standard risk (OR = 0.968, 95% CI
340 0.967-0.970, $p_{Bonf.}$<0.001). For every one-unit increase in kg/m^2, the log-odds of being classified
341 standard risk decreased by -0.03 (95%CI -0.034 to -0.030). Past and current tobacco smoking was
342 associated with lower odds of being classified standard risk (OR = 0.79, 95% CI 0.778-0.806,
343 $p_{Bonf.}$<0.001 and OR = 0.75, 95% CI 0.731-0.773, $p_{Bonf.}$<0.001, respectively) (Table-1-C).

344

345 [insert Table-1-C]

346

347 These findings were consistent across health indicators, except that more frequent moderate physical
348 activity was associated with better self-rated health (OR = 1.01, 95% CI 1.009-1.017, $p_{Bonf.}<0.001$).
349 Longer sleep duration was also associated with better self-rated health (OR = 1.07, 95% CI 1.066-
350 1.080, $p_{Bonf.}<0.001$) but there was no evidence of an association with long-standing illness (Figure-2-
351 B; Supplement-e8-C; e9-C; e10-C).

352

353 [insert Figure-2-B]

354

355 In our analyses of environmental exposures, higher $PM_{2.5}$ concentration was associated with lower
356 odds of being classified standard risk (OR = 0.97, 95% CI 0.951-0.981, $p_{Bonf.}<0.001$). PM_{10} was also
357 associated with lower odds of being classified standard risk in Model 1–2, but there was no evidence
358 of an association between PM_{10} and health status in Model 3 (OR = 1.002, 95% CI 0.997-1.007,
359 p=0.47). NO_2 was associated with lower odds of being classified standard risk in Model 1–2.
360 However, in Model 3, NO_2 was associated with higher odds of being classified standard risk (OR =
361 1.004, 95% CI 1.001-1.006, p_{BH}=0.009). Ambient sound level was associated with lower odds of
362 being classified standard risk in Model 1–2, but there was no evidence of an association with health
363 status in Model 3 (OR = 0.998, 95% CI 0.996-1.000, p=0.08). Finally, we did not find consistent
364 evidence of an association between percentage greenspace in a 1000m circular distance buffer and
365 health status (Table-1-D).

366

367 [insert Table-1-D]

368

369 Higher levels of $PM_{2.5}$ were also associated with lower odds of not having a long-standing illness (OR
370 = 0.97, 95% CI 0.953-0.983, $p_{Bonf.}<0.001$). While we found similar associations with self-rated health
371 in Model 1–2, $PM_{2.5}$ was associated with more favourable self-rated health in Model 3 (OR = 1.03,

95% CI 1.014-1.043, $p_{Bonf.}$=0.004) (Supplement-e9-D). There was no evidence of an association between PM_{10} and any health indicator in Model 3 (Figure-2-C; Supplement-e10-D), although PM_{10} was associated with poor health across outcomes in Model 1–2 (Table-1-D; Supplement-e8-D; e9-D). Higher NO_2 concentration was associated with less favourable self-rated health (OR = 0.99, 95% CI 0.993-0.997, $p_{Bonf.}$<0.001), while associations with health status and long-standing illness were inconsistent. Higher L_{den} was associated with poor health indicators in Model 1–2, but was associated with more favourable self-rated health in Model 3 (OR = 1.003, 95% CI 1.001-1.005, p_{BH}=0.013). There was no evidence of an association with long-standing illness in Model 3. Percentage greenspace was associated with better self-rated health (OR = 1.0007, 95% CI 1.0001-1.0012, p_{BH}=0.015) (Supplement-e9-D).

[insert Figure-2-C]

Magnitude of associations

Standardised regression coefficients for health status are presented in Figure-3. Most notably, the magnitude of association between BMI (β = -0.30, SE = 0.01) or very low income (β = -0.29, SE = 0.01) and health status was comparable to that of current smoking (β = -0.29, SE = 0.01). There was also a substantial difference in magnitude of association with health status between the very low and low income groups (β = -0.29, SE = 0.01 and β = -0.08, SE = 0.01, respectively). The magnitude of association between loneliness and health status was substantially larger than that of social isolation (β = -0.21, SE = 0.02 and β = -0.05, SE = 0.01, respectively). Finally, associations between environmental exposures and health status were relatively small compared to those of sociodemographic, psychosocial and lifestyle factors (Supplement-e11-A).

Standardised regression coefficients for long-standing illness and self-rated health are presented in Supplement-e11-B–C; e12-A–B.

[insert Figure-3]

Stratified analyses

Descriptive statistics of the analytical sample stratified by sex and age group are presented in Supplement-e13–e14.

The odds of being classified standard risk were lower for men in the lowest income group ($p_{\text{Bonf.(interaction)}}<0.001$). The association between increased age and lower odds of being classified standard risk was stronger in men ($p_{\text{Bonf.(interaction)}}<0.001$). Asian women had higher odds of being classified standard risk ($p_{\text{Bonf.(interaction)}}=0.017$). There was some evidence that the association between neighbourhood deprivation and lower odds of being classified standard risk was stronger in men ($p_{\text{BH(interaction)}}=0.012$). Men who were classified as lonely had lower odds of being classified standard risk ($p_{\text{BH(interaction)}}=0.005$). Longer sleep duration was associated with lower odds of being classified standard risk in men only ($p_{\text{Bonf.(interaction)}}=0.002$). Engaging more frequently in moderate physical activity was associated with higher odds of being classified standard risk in men ($p_{\text{BH(interaction)}}=0.015$). Reporting daily alcohol intake was associated with higher odds of being classified standard risk in men only ($p_{\text{BH(interaction)}}=0.026$). The association between higher BMI and lower odds of being classified standard risk was stronger in men ($p_{\text{Bonf.(interaction)}}<0.001$). The association between former smoking and health status was stronger in men ($p_{\text{Bonf.(interaction)}}<0.001$), while the association with current smoking and health status was stronger in women ($p_{\text{BH(interaction)}}=0.008$). In women, exposure to higher levels of noise was associated with lower odds of being classified standard risk ($p_{\text{BH(interaction)}}=0.029$), while in men there was evidence of an association between higher percentage of greenspace and lower odds of being classified standard risk ($p_{\text{BH(interaction)}}=0.013$) (Supplement-e15-A–D; e16-A–C).

The association between very low income and lower odds of being classified standard risk was stronger in participants below the age of 65 ($p_{\text{Bonf.(interaction)}}<0.001$). There was also some evidence that higher income was associated with higher odds of being classified standard risk only in participants younger than 65 ($p_{\text{BH(interaction)}}=0.046$ [£52,000–100,000]; although for >£100,000 $p_{\text{BH(interaction)}}=0.521$).

428 Men aged 65 and above had substantially lower odds of being classified standard risk
429 ($p_{Bonf.(interaction)}$<0.001). The association between loneliness and lower odds of being classified standard
430 risk was stronger in individuals younger than 65 ($p_{BH(interaction)}$=0.020). There was evidence that the
431 association between walking frequency and higher odds of being classified standard risk was stronger
432 in individuals aged 65 and above ($p_{BH(interaction)}$=0.019). We did not find evidence of an association
433 between climbing stairs 1-5 times/day and health status in individuals aged 65 and above
434 ($p_{Bonf.(interaction)}$=0.018). There was some evidence that the associations between climbing stairs 6-10
435 times/day or never drinking alcohol and health status were stronger in individuals younger than 65
436 ($p_{BH(interaction)}$=0.037 and $p_{BH(interaction)}$=0.037, respectively). The association between higher BMI and
437 higher odds of being classified standard risk was stronger in individuals aged 65 and above
438 ($p_{BH(interaction)}$=0.007). Finally, the association between current smoking and health status was stronger
439 in individuals younger than 65 ($p_{BH(interaction)}$=0.005), while the association with former smoking was
440 stronger in individuals aged 65 and above ($p_{BH(interaction)}$=0.027) (Supplement-e17-A–D; e18-A–C).
441
442 Results for long-standing illness and self-rated health stratified by sex and age group are presented in
443 Supplement-e19-A–D; e20-A–C; e21-A–D; e22-A–C; e23-A–D; e24-A–C; e25-A–D; e26-A–C.
444
445 **Prospective analyses**
446 We repeated our analyses of self-rated health in participants with follow-up data collected between
447 2012–2013 (*n* = 16,058; mean follow-up = 4.26 years, SD=0.87) and between 2014–2019 (*n* =
448 32,617; mean follow-up = 8.57 years, SD = 1.64). Descriptive statistics and results are presented in
449 Supplement-e27; e28-A–D; e29-A–D.
450
451 **Additional analyses**
452 Descriptive statistics are presented in Supplement-e30. Neuroticism was associated with lower odds
453 of being classified standard risk (*n* = 259,448; OR = 0.955, 95% CI 0.953-0.958). For sleep duration
454 and BMI, we found evidence of non-linearity in the association with health status (Supplement-e31-
455 D). As such, we also examined these variables as categorical explanatory variables. Compared to

456 individuals with optimal sleep duration (7-8 hours/day), those who slept <7 or ≥9 hours/day had lower
457 odds of being classified standard risk (*n* = 307,378; OR = 0.88, 95% CI 0.868-0.901 and OR = 0.69,
458 95% CI 0.672-0.714, respectively). We also found that both low BMI (<18.5 kg/m^2) and high BMI
459 (>24.9 kg/m^2) was associated with lower odds of being classified standard risk than a BMI of 18.5-
460 24.9 kg/m^2 (OR = 0.71, 95% CI 0.633-0.796 and OR = 0.86, 95% CI 0.842-0.873, respectively). For
461 all other numerical variables with more than 2 values, there was no evidence of substantial departure
462 from a linear association with health status (Supplement-e31). Repeating the main analysis with body
463 fat percentage as measure of body composition led to similar conclusions as for BMI: higher fat
464 percentage was associated with lower odds of being classified standard risk (*n* = 307,202; OR =
465 0.978, 95% CI 0.976-0.979). Excluding individuals who reported that they had stopped drinking
466 alcohol due to illness or their doctor's advice (*n* = 2,790) attenuated the association between never
467 drinking alcohol and health status (*n* = 304,588; OR = 0.74, 95% CI 0.718-0.772). When sub-dividing
468 individuals who reported that they never drink alcohol into current and lifetime abstainers, the
469 association with health status was stronger in current abstainers (*n* = 307,346; OR = 0.53, 95% CI
470 0.505-0.552 and OR = 0.75, 95% CI 0.718-0.787, respectively). Restricting analyses to current
471 smoking, we found that both regular smoking and smoking only occasionally was associated with
472 lower odds of being classified standard risk, although the magnitude of association was stronger for
473 regular smoking (*n* = 307,372; OR = 0.81, 95% CI 0.781-0.831 and OR = 0.90, 95% CI 0.861-0.949,
474 respectively). Examining MET minutes instead of the number of days per week spent walking,
475 engaging in moderate or vigorous-intensity physical activity led to similar results (*n* = 268,674; data
476 not shown). When we examined air pollution and noise exposure estimates dichotomised according to
477 WHO recommendation thresholds, we found that being exposed to ≤10 μg/m^3 $PM_{2.5}$ on average
478 annually was associated with higher odds of being classified standard risk (OR = 1.03, 95% CI 1.009-
479 1.049). There was no evidence of an association with health status for exposures to ≤20 μg/m^3 PM_{10},
480 ≤40 μg/m^3 NO_2 or ≤53 decibels L_{den} in Model 3. Restricting analyses to participants assessed after
481 2008 (regarding noise estimates; *n* = 182,342) or 2009 (regarding air pollution estimates; *n* = 60,521)
482 or to those who had lived at their current address for at least 10 years (*n* = 209,239) did not lead to
483 different conclusions regarding the associations between environmental exposures and health status.

Percentage greenspace within a 300m circular distance buffer was associated with lower odds of being classified standard risk (*n* = 307,378; OR = 0.999, 95% CI 0.9988-0.9998), although not of relevant magnitude. Truncating continuous explanatory variables did not materially change our results (*n* = 278,065; data not shown).

Discussion

Principal findings

Sociodemographic characteristics. We found that increased age was strongly associated with unfavourable health status and long-standing illness. However, older participants generally self-rated their health positively, which is consistent with several smaller studies[31,32], though not all[33]. Ageing might be the single most important factor underlying disease[34], with the expectation of declining health as people age being almost universally accepted. As attainable health states shift with age[35], older participants might evaluate their health more favourable, despite higher rates of illness and disability.

Being male was associated with unfavourable health indicators. Although women tend to report higher numbers of illnesses, disabilities and limitations in daily life[36], one of the most robust findings in human biology is that women live longer[37]. Our findings are consistent with previous research in which men rated their health less favourable[36], which could be due to differences in illness severity between the sexes.

Higher income and lower levels of neighbourhood deprivation were associated with favourable health indicators, broadly consistent with previous studies[38-40]. Notably, we found only a small difference in the strength of association with health status and long-standing illness between the high-income groups. The difference between the low-income groups, however, was substantial, supporting findings from a previous study in the United States which suggested a non-linear association between family income and mortality[41]. For self-rated health, we found evidence of substantial differences between the high-income groups as well.

Our study provides only limited evidence that education and professional qualification was associated with favourable health status. Once other factors had been adjusted for, no consistent pattern emerged, except that higher levels of qualification were generally associated with better self-rated health,

515 consistent with previous findings[33]. A recent analysis of data from the UK Biobank found that
516 remaining longer in school causally reduced participants' risk of diabetes and mortality[42]. A potential
517 explanation for why we did not find a consistent pattern in the present study is that differences in
518 health status result from educated individuals engaging in healthier lifestyle behaviours[43], or they
519 could be due to socioeconomic or genomic factors.
520
521 *Psychosocial factors*. Social isolation and loneliness were both associated with unfavourable health.
522 The strength of association was greater for loneliness, particularly in men and individuals below the
523 age of 65. In previous investigations, social isolation and loneliness were not always correlated[44,45],
524 likely because they measure somewhat different aspects of social relations. Social isolation reflects
525 scarcity of contact with others, while loneliness refers to perceptions of detachment or discrepancies
526 between the need for, and the fulfilment of, social interaction. In a meta-analysis of 70 studies, social
527 isolation, loneliness and living alone were associated with a 26-32% increased mortality risk[14]. There
528 was no evidence of differences in mortality between these measures, although the strength of
529 association was greater in individuals below the age of 65. A recent analysis of data from the UK
530 Biobank found that socially isolated and lonely individuals had an increased risk of death, but only
531 social isolation was a significant predictor of all-cause mortality when both variables were included in
532 the same model[19]. The discrepancy with our finding (the association between loneliness and health
533 was stronger than the association between social isolation and health) might reflect differences in
534 outcome measures (general health in the present study vs mortality in previous investigations).
535
536 *Lifestyle factors*. Long sleep duration, high BMI as well as past and current smoking was associated
537 with unfavourable health. Sleeping less than 7 hours/day was also associated with unfavourable
538 health, which is consistent with a meta-analysis that provided evidence of a U-shaped association
539 between sleep duration and all-cause mortality[46]. We also found that a BMI outside the optimal range
540 of 18.5-24.9 kg/m^2 was associated with unfavourable health.
541

542 Physical activity is a key lifestyle factor recommended for primary and secondary prevention of
543 chronic health conditions[47], and is associated with lower mortality risk[48]. In the present study, walking
544 frequency, especially in individuals aged 65 and above, stair climbing, and engaging in vigorous-
545 intensity physical activity was associated with favourable health. Moderate-intensity physical activity
546 was associated with better self-rated health, and there was some evidence of favourable health status
547 and not having a long-standing illness in men. A previous study in middle-aged British men found
548 evidence of an association between vigorous, but not moderate, physical activity and reduced
549 mortality[49]. Reviews of the literature include mixed findings on the relative contributions of moderate
550 and vigorous-intensity physical activity[50,51], with some evidence suggesting stronger associations for
551 vigorous activity[52].
552
553 A more regular drinking pattern was associated with favourable health. Alcohol drinking often occurs
554 in a social context and might therefore constitute a proxy for social wellness, supported by the finding
555 that non-drinkers tend to be characterised by poor psychosocial health and low socioeconomic status.
556 Moderate drinkers also performed more sports than lifelong abstainers in previous research[53].
557 Drinking less frequent than 1-2 times per week was associated with unfavourable health in the present
558 study and could not be fully accounted for by excluding individuals who had discontinued alcohol
559 intake for health reasons or because of their doctor's advice. However, we found that the association
560 with health was stronger for current abstainers, suggesting that many non-drinkers in later life stopped
561 due to illness[54]. Current abstainers who quit drinking for health reasons could exaggerate poor health
562 outcomes associated with not drinking alcohol[55]. Those who never drink might also differ from
563 current drinkers in other characteristics[56].
564
565 *Environmental exposures*. Traffic-related air pollution and noise are major environmental risk factors
566 for health. While we found some evidence that higher levels of all airborne pollutants were associated
567 with unfavourable health, only $PM_{2.5}$ was associated with unfavourable health status and long-
568 standing illness after adjustment for other factors. Higher levels of NO_2 were associated with less
569 favourable self-rated health. A joint analysis of UK Biobank, HUNT and EPIC-Oxford data did not

find evidence of an association between NO_2 and any disease outcome, although both $PM_{2.5}$ and PM_{10} were associated with higher incident cardiovascular disease[57]. Additional adjustment for physical activity and neighbourhood deprivation in the present study could have contributed to differences in findings between studies. Few individuals in our study were exposed to high levels of PM_{10} and NO_2, which might explain why we did not find strong associations between these pollutants and health indicators. The only exception was $PM_{2.5}$, for which 45% of our sample were exposed to levels above the WHO recommendation threshold of ≤ 10 μg/m^3. Positive associations between $PM_{2.5}$ and self-rated health and between NO_2 and favourable health status might represent chance findings; we are not aware of a plausible mechanism linking higher levels of pollution to favourable health.

While higher levels of residential noise were associated with unfavourable health indicators in Model 1-2, there was no evidence of an association with health status or long-standing illness after adjustment for other factors, except that higher levels of noise were associated with unfavourable health status in women. Our findings are consistent with a previous analysis that did not find evidence of an independent association between L_{den} and cardiovascular disease, ischemic heart disease or cerebrovascular disease[57].

Greenspace was associated with better self-rated health, consistent with a recent meta-analysis[58]. However, there were some inconsistencies in the prospective analyses and regarding health status and long-standing illness. Previous studies examining associations between greenspace and health often did not adjust for individual-level confounding or other environmental exposures[59], which might explain why we did not find consistent associations after adjustment in the present study.

Strengths and weaknesses of the study

The present study has several strengths. The large analytical sample size enabled high precision in the estimation of associations, and it allowed us to explore a wide range of explanatory variables, sex and age-specific associations, interactions, and multiple sensitivity analyses with further classification of

risk factors. In addition to cross-sectional analyses, we modelled self-rated health at follow-up in subsets of participants with repeat measurements.

We examined three health indicators that had slightly different ascertainment. While our findings were fairly consistent across these outcomes, we identified some differences between objective and subjective health. Interpretation of these outcomes might seem challenging in the current departmentalised framework of public health medicine. However, overall health is arguably what matters most to people, and exploring factors associated with health indicators that transcend traditional disease-boundaries could represent an effective strategy to increase longevity and improve healthy life expectancy in the general population.

Previous studies have often used composite scores of healthy lifestyles[13] in which individual behaviours were weighted equally, even though the strength of their respective association with health might differ. In the present study, we estimated associations separately for each lifestyle factor. We also incorporated a wide range of other factors in our models, including sociodemographic characteristics, loneliness and a multi-item assessment of social isolation with good predictive validity for mortality[14]. Finally, we included several environmental exposures in joint analyses, and repeated our main analysis using policy-relevant thresholds for traffic-related air pollution and noise.

Nevertheless, our study has limitations. Cross-sectional analyses prevent examination of temporal sequences. Reverse causality such as poor health and disability affecting behaviour or socioeconomic outcomes needs to be considered. However, this does not apply to fixed characteristics such as sex or ethnicity, and there was substantial consistency in prospective analyses of self-rated health.

As many explanatory variables were self-reported, we cannot exclude the possibility that social desirability might have influenced responses. For example, previous research has shown that up to half of participants who claim to never have had an alcoholic drink reported alcohol intake during previous surveys[60]. For most participants, explanatory variables were measured at a single time point,

625 which could increase the potential for measurement error and did not allow us to model changes
626 across the life course. However, there is some evidence that lifestyle is fairly stable in this age group,
627 with few people newly adopting a healthy lifestyle in middle age[61].
628
629 The finding that more frequent alcohol intake was associated with favourable health should be treated
630 with caution as we did not take into account the amount and type of alcohol consumed. We
631 specifically included alcohol intake frequency because drinking patterns are important to study in
632 addition to quantity[62]. More regular drinking is often associated with consuming less alcohol per
633 occasion, less binge drinking as well as consuming different types of alcohol (i.e. more wine, less beer
634 and spirits)[62], potentially reflecting safer drinking habits, wellness and social advantage[63]. Positive
635 associations between frequent alcohol intake and health might also be due to different risk-factor
636 profiles between regular drinkers and infrequent or non-drinkers, although associations were robust to
637 adjustment for sociodemographic and psychosocial factors.
638
639 There might be some misclassification in the reporting of medical illnesses, for instance that
640 participants engaging in a healthy lifestyle might underreport illnesses. However, participants were
641 asked to report illnesses that had been diagnosed by a doctor and these were confirmed during a
642 nurse-led interview.
643
644 While we incorporated a multifactorial index of area-level deprivation, perceptions of the
645 neighbourhood environment, which might affect behaviour and well-being, were not assessed in the
646 present study.
647
648 Outdoor pollutants were modelled at residential addresses and do not take into account the time spent
649 at home or in traffic, the degree of pollutant infiltration into buildings or exposure at the workplace.
650 These estimates also do not take into account pollution from indoor sources, even though we spend
651 about 90% of our time indoors[64]. Data on exposure to ultrafine particles ($PM_{0.1}$) were also not
652 available and noise from other sources including railway, aircraft or the workplace were not included

653 in the modelling. We also could not take into account address history, which might result in some
654 exposure misclassification, although there was a considerable degree of residential stability, with
655 participants living on average 17 years at their current address. Findings were robust to sensitivity
656 analyses in which we restricted the sample to participants who had lived at their current address for at
657 least 10 years. It is worth noting that while absolute levels of environmental pollutants in the UK have
658 decreased over time, relative differences between different geographical areas have remained stable[65].
659
660 Data on percentage greenspace were available only for participants with home location in England,
661 limiting generalisability to other parts of the UK. Our measure of greenspace was derived from a land
662 use database which might not capture small scale variability in vegetation, private gardens and trees.
663
664 **Generalisability**
665 The overall response rate in the present study was low (5.5%) and compared with non-responders,
666 participants were older, more likely to be female and lived in less deprived neighbourhoods.
667 Participants were also less likely to be obese, to smoke, to never drink or to drink alcohol daily, and
668 they reported fewer health conditions compared with data from a nationally representative survey[66].
669 While the UK Biobank states that "valid assessment of exposure-disease relationships are nonetheless
670 widely generalizable and do not require participants to be representative of the population at large"[67],
671 concordant with the finding that there is little evidence of considerable bias due to non-participation
672 in epidemiological research[68], the magnitude of exposure-disease associations may depend on
673 prevalence of effect modifiers[69]. A recent empirical investigation comparing the UK Biobank with
674 data from 18 prospective cohort studies with conventional response rates showed that the direction of
675 risk factor associations were similar, although with differences in magnitude[70].
676
677 **Implications**
678 The findings of our research help shed light on some of the behavioural, psychosocial and
679 environmental pathways to health and are relevant to public health policies aimed at promoting health
680 in later life. For example, public health medicine could put greater focus on non-medical factors such

681 as loneliness, further encourage healthy lifestyle behaviours and weight management, and examine
682 efforts to improve the health outcomes of individuals with low income. Our findings support the view
683 that promoting individual responsibility for one's health (e.g. engaging in healthy lifestyle behaviours
684 and reducing emissions) and policy-level commitments (e.g. reducing long-term exposure to
685 environmental pollutants[64] or improving neighbourhoods) both need to be considered in population
686 health. While associations between lifestyle and health might reflect a life-long commitment to
687 healthy behaviours, research has shown that it is not too late to newly adopt healthy behaviours later
688 in life[61]. Finally, prevention, in addition to drug discovery and disease treatment, should be the top
689 priority for public health policy.

690 Authorship contributions

691 CML acquired the studentship funding. JM and CML conceived the idea of the study; JM acquired
692 the data; JM carried out the statistical analysis; JM and CML interpreted the findings; JM wrote the
693 manuscript. CJR and CML critically reviewed the manuscript and JM revised the manuscript for final
694 submission. All authors read and approved the final manuscript. JM had full access to all the data in
695 the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

696

697 Funding and disclosure

698 JM gratefully acknowledges current funding from the Biotechnology and Biological Sciences
699 Research Council (BBSRC) (ref: 2050702) and Eli Lilly and Company Limited in support of this
700 work. CML is part-funded by the National Institute for Health Research (NIHR) Biomedical Research
701 Centre at South London and Maudsley NHS Foundation Trust and King's College London. The views
702 expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department
703 of Health and Social Care. CML is a member of the Scientific Advisory Board of Myriad
704 Neuroscience.

705

706 Acknowledgments

707 This project made use of time on Rosalind HPC, funded by Guy's & St Thomas' Hospital NHS Trust
708 Biomedical Research Centre (GSTT-BRC), South London & Maudsley NHS Trust Biomedical
709 Research Centre (SLAM-BRC), and Faculty of Natural Mathematics & Science (NMS) at King's
710 College London.

711

712 Ethics

713 Ethical approval for the UK Biobank study has been granted by the National Information Governance
714 Board for Health and Social Care and the NHS North West Multicentre Research Ethics Committee
715 (11/NW/0382). No project-specific ethical approval is needed. Data access permission has been
716 granted under UK Biobank application 45514.

717

Data sharing statement

The data used in the present study are available to all bona fide researchers for health-related research that is in the public interest, subject to an application process and approval criteria. Study materials are publicly available online at http://www.ukbiobank.ac.uk.

Supplementary material

Supplementary information is available online.

References

1. Partridge L, Deelen J, Slagboom PE. Facing up to the global challenges of ageing. *Nature.* 2018;561(7721):45.
2. Roser M. Life expectancy. Our World in Data Web site. Published 2013. Accessed August 22, 2019.
3. Kontis V, Bennett JE, Mathers CD, Li G, Foreman K, Ezzati M. Future life expectancy in 35 industrialised countries: projections with a Bayesian model ensemble. *The Lancet.* 2017;389(10076):1323-1335.
4. Netuveli G, Wiggins RD, Hildon Z, Montgomery SM, Blane D. Quality of life at older ages: evidence from the English longitudinal study of aging (wave 1). *Journal of Epidemiology & Community Health.* 2006;60(4):357-363.
5. Van Hecke O, Torrance N, Smith B. Chronic pain epidemiology and its clinical relevance. *British journal of anaesthesia.* 2013;111(1):13-18.
6. Brown GC. Living too long. *EMBO reports.* 2015;16(2):137-141.
7. Nybo H, Gaist D, Jeune B, McGue M, Vaupel JW, Christensen K. Functional status and self-rated health in 2,262 nonagenarians: the Danish 1905 Cohort Survey. *Journal of the American Geriatrics Society.* 2001;49(5):601-609.
8. Olshansky SJ. From Lifespan to Healthspan. *JAMA.* 2018.
9. Carter BD, Abnet CC, Feskanich D, et al. Smoking and mortality—beyond established causes. *New England journal of medicine.* 2015;372(7):631-640.
10. Lai HT, de Oliveira Otto MC, Lemaitre RN, et al. Serial circulating omega 3 polyunsaturated fatty acids and healthy ageing among older adults in the Cardiovascular Health Study: prospective cohort study. *BMJ.* 2018;363:k4067.
11. Room R, Babor T, Rehm J. Alcohol and public health. *The lancet.* 2005;365(9458):519-530.
12. Cooper R, Kuh D, Hardy R. Objectively measured physical capability levels and mortality: systematic review and meta-analysis. *Bmj.* 2010;341:c4467.
13. Loef M, Walach H. The combined effects of healthy lifestyle behaviors on all cause mortality: a systematic review and meta-analysis. *Preventive medicine.* 2012;55(3):163-170.
14. Holt-Lunstad J, Smith TB, Baker M, Harris T, Stephenson D. Loneliness and social isolation as risk factors for mortality: a meta-analytic review. *Perspectives on Psychological Science.* 2015;10(2):227-237.
15. Smith RB, Fecht D, Gulliver J, et al. Impact of London's road traffic air and noise pollution on birth weight: retrospective population based cohort study. *bmj.* 2017;359:j5299.
16. Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature.* 2018;562(7726):203.

17. Guggenheim JA, Williams C. Childhood febrile illness and the risk of myopia in UK Biobank participants. *Eye.* 2016;30(4):608.
18. *The English indices of deprivation 2010.* Department for Communities and Local Government2011.
19. Elovainio M, Hakulinen C, Pulkki-Råback L, et al. Contribution of risk factors to excess mortality in isolated and lonely individuals: an analysis of data from the UK Biobank cohort study. *The Lancet Public Health.* 2017;2(6):e260-e266.
20. IPAQ-Group. IPAQ scoring protocol - International Physical Activity Questionnaire. https://sites.google.com/site/theipaq/scoring-protocol. Accessed 26 Mar, 2018.
21. Beelen R, Hoek G, Vienneau D, et al. Development of NO2 and NOx land use regression models for estimating air pollution exposure in 36 study areas in Europe–The ESCAPE project. *Atmospheric Environment.* 2013;72:10-23.
22. Eeftens M, Beelen R, de Hoogh K, et al. Development of land use regression models for PM2. 5, PM2. 5 absorbance, PM10 and PMcoarse in 20 European study areas; results of the ESCAPE project. *Environmental science & technology.* 2012;46(20):11195-11205.
23. Kephalopoulos S, Paviotti M, Anfosso-Lédée F, Van Maercke D, Shilton S, Jones N. Advances in the development of common noise assessment methods in Europe: The CNOSSOS-EU framework for strategic environmental noise mapping. *Science of the Total Environment.* 2014;482:400-410.
24. Morley D, De Hoogh K, Fecht D, et al. International scale implementation of the CNOSSOS-EU road traffic noise prediction model for epidemiological studies. *Environmental pollution.* 2015;206:332-341.
25. Vienneau D, Schindler C, Perez L, Probst-Hensch N, Röösli M. The relationship between transportation noise exposure and ischemic heart disease: a meta-analysis. *Environmental research.* 2015;138:372-380.
26. Department-for-Communities-and-Local-Government. Generalised Land Use Database Statistics for England 2005 (Enhanced Basemap). In. https://data.gov.uk/dataset/land_use_statistics_generalised_land_use_database2007.
27. Fell DB, Joseph K, Armson BA, Dodds L. The impact of pregnancy on physical activity level. *Maternal and child health journal.* 2009;13(5):597.
28. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological).* 1995;57(1):289-300.
29. Fox J, Monette G. Generalized collinearity diagnostics. *Journal of the American Statistical Association.* 1992;87(417):178-183.
30. Gelman A. Scaling regression inputs by dividing by two standard deviations. *Statistics in medicine.* 2008;27(15):2865-2873.

31. Cockerham WC, Sharp K, Wilcox JA. Aging and perceived health status. *Journal of Gerontology.* 1983;38(3):349-355.
32. Ferraro KF. Self-ratings of health among the old and the old-old. *Journal of Health and Social Behavior.* 1980:377-383.
33. Wu S, Wang R, Zhao Y, et al. The relationship between self-rated health and objective health status: a population-based study. *BMC public health.* 2013;13(1):320.
34. Sinclair DA, LaPlante MD. *Lifespan: Why We Age—and Why We Don't Have To.* Atria Books; 2019.
35. Brouwer WB, van Exel NJA, Stolk EA. Acceptability of less than perfect health states. *Social Science & Medicine.* 2005;60(2):237-246.
36. Collerton J, Davies K, Jagger C, et al. Health and disease in 85 year olds: baseline findings from the Newcastle 85+ cohort study. *Bmj.* 2009;339:b4904.
37. Austad SN, Bartke A. Sex differences in longevity and in responses to anti-aging interventions: a mini-review. *Gerontology.* 2016;62(1):40-46.
38. Shibuya K, Hashimoto H, Yano E. Individual income, income distribution, and self rated health in Japan: cross sectional analysis of nationally representative sample. *Bmj.* 2002;324(7328):16.
39. Der G, Macintyre S, Ford G, Hunt K, West P. The relationship of household income to a range of health measures in three age cohorts from the West of Scotland. *The European Journal of Public Health.* 1999;9(4):271-277.
40. Godhwani S, Jivraj S, Marshall A, Bécares L. Comparing subjective and objective neighbourhood deprivation and their association with health over time among older adults in England. *Health & place.* 2019;55:51-58.
41. Rehkopf DH, Berkman LF, Coull B, Krieger N. The non-linear risk of mortality by income level in a healthy population: US National Health and Nutrition Examination Survey mortality follow-up cohort, 1988–2001. *BMC Public Health.* 2008;8(1):383.
42. Davies NM, Dickson M, Smith GD, Van Den Berg GJ, Windmeijer F. The causal effects of education on health outcomes in the UK Biobank. *Nature human behaviour.* 2018;2(2):117.
43. Ding D, Rogers K, van der Ploeg H, Stamatakis E, Bauman AE. Traditional and emerging lifestyle risk behaviors and all-cause mortality in middle-aged and older adults: evidence from a large population-based Australian cohort. *PLoS medicine.* 2015;12(12).
44. Coyle CE, Dugan E. Social isolation, loneliness and health among older adults. *Journal of aging and health.* 2012;24(8):1346-1363.
45. Perissinotto CM, Cenzer IS, Covinsky KE. Loneliness in older persons: a predictor of functional decline and death. *Archives of internal medicine.* 2012;172(14):1078-1084.
46. Cappuccio FP, D'Elia L, Strazzullo P, Miller MA. Sleep duration and all-cause mortality: a systematic review and meta-analysis of prospective studies. *Sleep.* 2010;33(5):585-592.

47. Warburton DE, Nicol CW, Bredin SS. Health benefits of physical activity: the evidence. *Cmaj.* 2006;174(6):801-809.
48. Arem H, Moore SC, Patel A, et al. Leisure time physical activity and mortality: a detailed pooled analysis of the dose-response relationship. *JAMA internal medicine.* 2015;175(6):959-967.
49. Yu S, Yarnell J, Sweetnam P, Murray L. What level of physical activity protects against premature cardiovascular death? The Caerphilly study. *Heart.* 2003;89(5):502-506.
50. Lee I, Paffenbarger R, Hennekens C. Physical activity, physical fitness and longevity. *Aging Clinical and Experimental Research.* 1997;9(1-2):2-11.
51. Oguma Y, Sesso H, Paffenbarger R, Lee I. Physical activity and all cause mortality in women: a review of the evidence. *British journal of sports medicine.* 2002;36(3):162-172.
52. Lee I-M, Paffenbarger Jr RS. Associations of light, moderate, and vigorous intensity physical activity with longevity: the Harvard Alumni Health Study. *American journal of epidemiology.* 2000;151(3):293-299.
53. Piazza-Gardner AK, Barry AE. Examining physical activity levels and alcohol consumption: are people who drink more active? *American Journal of Health Promotion.* 2012;26(3):e95-e104.
54. Shaper AG, Wannamethee G, Walker M. Alcohol and mortality in British men: explaining the U-shaped curve. *The Lancet.* 1988;332(8623):1267-1273.
55. Stockwell T, Zhao J, Panwar S, Roemer A, Naimi T, Chikritzhs T. Do "moderate" drinkers have reduced mortality risk? A systematic review and meta-analysis of alcohol consumption and all-cause mortality. *Journal of studies on alcohol and drugs.* 2016;77(2):185-198.
56. Fat LN, Cable N, Marmot M, Shelton N. Persistent long-standing illness and non-drinking over time, implications for the use of lifetime abstainers as a control group. *J Epidemiol Community Health.* 2014;68(1):71-77.
57. Cai Y, Hodgson S, Blangiardo M, et al. Road traffic noise, air pollution and incident cardiovascular disease: a joint analysis of the HUNT, EPIC-Oxford and UK Biobank cohorts. *Environment international.* 2018;114:191-201.
58. Twohig-Bennett C, Jones A. The health benefits of the great outdoors: A systematic review and meta-analysis of greenspace exposure and health outcomes. *Environmental research.* 2018;166:628-637.
59. Sarkar C, Webster C, Gallacher J. Residential greenness and prevalence of major depressive disorders: a cross-sectional, observational, associational study of 94 879 adult UK Biobank participants. *The Lancet Planetary Health.* 2018;2(4):e162-e173.
60. Rehm J, Irving H, Ye Y, Kerr WC, Bond J, Greenfield TK. Are lifetime abstainers the best control group in alcohol epidemiology? On the stability and validity of reported lifetime abstention. *American journal of epidemiology.* 2008;168(8):866-871.

871 61. King DE, Mainous III AG, Geesey ME. Turning back the clock: adopting a healthy lifestyle
872 in middle age. *The American journal of medicine.* 2007;120(7):598-603.
873 62. Holst C, Becker U, Jørgensen ME, Grønbæk M, Tolstrup JS. Alcohol drinking patterns and
874 risk of diabetes: a cohort study of 70,551 men and women from the general Danish
875 population. *Diabetologia.* 2017;60(10):1941-1950.
876 63. Naimi TS, Xuan Z, Brown DW, Saitz R. Confounding and studies of 'moderate'alcohol
877 consumption: the case of drinking frequency and implications for low-risk drinking
878 guidelines. *Addiction.* 2013;108(9):1534-1543.
879 64. Holgate ST. 'Every breath we take: the lifelong impact of air pollution'–a call for action.
880 *Clinical Medicine.* 2017;17(1):8-12.
881 65. Gulliver J, de Hoogh K, Hansell A, Vienneau D. Development and back-extrapolation of
882 NO2 land use regression models for historic exposure assessment in Great Britain.
883 *Environmental science & technology.* 2013;47(14):7804-7811.
884 66. Fry A, Littlejohns TJ, Sudlow C, et al. Comparison of sociodemographic and health-related
885 characteristics of UK Biobank participants with those of the general population. *American*
886 *Journal of Epidemiology.* 2017;186(9):1026-1034.
887 67. UK-Biobank. http://www.ukbiobank.ac.uk/wp-content/uploads/2017/03/access-matters-
888 representativeness-1.pdf. http://www.ukbiobank.ac.uk/wp-content/uploads/2017/03/access-
889 matters-representativeness-1.pdf. Accessed 12 September, 2019.
890 68. Galea S, Tracy M. Participation rates in epidemiologic studies. *Annals of epidemiology.*
891 2007;17(9):643-653.
892 69. Keyes KM, Westreich D. UK Biobank, big data, and the consequences of non-
893 representativeness. *The Lancet.* 2019;393(10178):1297.
894 70. Battya GD, Galec CR, Kivimäkia M, Dearyd IJ, Belle S. Generalisability of Results from UK
895 Biobank: Comparison With a Pooling of 18 Cohort Studies.
896

Tables

Table 1-A. **Sociodemographic characteristics associated with health status**

	Model 1			Model 2			Model 3		
Term	OR	95% CI		OR	95% CI		OR	95% CI	
Household income[1]									
Very low	0.5161	0.5049	0.5276	0.6298	0.6156	0.6443	0.7447	0.7263	0.7635
Low	0.7548	0.7389	0.7711	0.8728	0.8539	0.8921	0.9196	0.8992	0.9404
Middle	Ref	–	–	Ref	–	–	Ref	–	–
High	1.2084	1.1803	1.2372	1.1016	1.0755	1.1283	1.0510	1.0256	1.0770
Very high	1.2867	1.2386	1.3369	1.1483	1.1048	1.1937	1.0472	1.0064	1.0897
Sex									
Female	Ref	–	–	Ref	–	–	Ref	–	–
Male	0.8642	0.8511	0.8775	–	–	–	0.8847	0.8702	0.8994
Age	0.9462	0.9453	0.9472	–	–	–	0.9528	0.9517	0.9539
Multiple deprivation	0.9912	0.9906	0.9917	0.9885	0.9879	0.9891	0.9952	0.9945	0.9958
Ethnicity									
White	Ref	–	–	Ref	–	–	Ref	–	–
Mixed-race	1.2991	1.1685	1.4470	0.9743	0.8744	1.0876	1.1037	0.9888	1.2341
Asian	1.0196	0.9584	1.0852	0.8560	0.8034	0.9126	1.0188	0.9541	1.0884
Black	1.2994	1.2132	1.3929	0.9831	0.9165	1.0553	1.3125	1.2208	1.4120
Chinese	2.1528	1.8023	2.5898	1.7413	1.4541	2.0996	1.8339	1.5289	2.2148
Other	1.2077	1.1002	1.3273	0.9952	0.9050	1.0959	1.2570	1.1409	1.3865
Highest qualification									
None	Ref	–	–	Ref	–	–	Ref	–	–

O levels/GCSEs/CSEs	1.6519	1.6114	1.6934	1.2266	1.1954	1.2586	1.0315	1.0042	1.0595
A levels/NVQ/HND/HNC[2]	1.5125	1.4746	1.5514	1.2102	1.1791	1.2422	0.9879	0.9613	1.0152
Degree	1.7949	1.7525	1.8383	1.3373	1.3045	1.3709	0.9558	0.9300	0.9823

Note: OR = odds ratio; CI = confidence interval; GCSEs = general certificate of secondary education; CSE = certificate of secondary education; NVQ = national vocational qualification; HND = higher national diploma; HNC = higher national certificate. [1]Annual household income groups: very low (<£18 000), low (£18 000–30 999), middle (£31 000–51 999), high (£52 000–100 000) and very high (>£100 000). [2]also includes 'other professional qualifications'.

Model 1 – only individual explanatory variables.

Model 2 – adjusted for age and sex.

Model 3 – all explanatory variables.

Table 1-B. **Psychosocial factors associated with health status**

	Model 1			Model 2			Model 3		
Term	OR	95% CI		OR	95% CI		OR	95% CI	
Loneliness									
Not lonely	Ref	–	–	Ref	–	–	Ref	–	–
Lonely	0.7283	0.7057	0.7517	0.6947	0.6726	0.7175	0.8129	0.7863	0.8404
Social isolation									
Not isolated	Ref	–	–	Ref	–	–	Ref	–	–
Isolated	0.7544	0.7348	0.7745	0.7542	0.7343	0.7747	0.9496	0.9232	0.9768

Note: OR = odds ratio; CI = confidence interval.

Model 1 – only individual explanatory variables.

Model 2 – adjusted for age and sex.

Model 3 – all explanatory variables.

Table 1-C. **Lifestyle factors associated with health status**

	Model 1			Model 2			Model 3		
Term	OR	95% CI		OR	95% CI		OR	95% CI	
Sleep duration (hours/day)	0.9574	0.9505	0.9643	0.9817	0.9745	0.9890	0.9748	0.9677	0.9821
Physical activity (days/week)[1]									
Walking	1.0099	1.0060	1.0139	1.0258	1.0217	1.0299	1.0105	1.0060	1.0149
Moderate activity	1.0095	1.0061	1.0128	1.0253	1.0219	1.0288	1.0032	0.9990	1.0073
Vigorous activity	1.0571	1.0529	1.0614	1.0502	1.0459	1.0545	1.0319	1.0270	1.0368
Stair climbing frequency									
None	Ref	–	–	Ref	–	–	Ref	–	–
1-5/day	1.3009	1.2611	1.3418	1.0980	1.0637	1.1334	1.0704	1.0362	1.1057
6-10/day	1.5529	1.5089	1.5981	1.3354	1.2968	1.3752	1.1756	1.1407	1.2116
11-15/day	1.6271	1.5771	1.6788	1.3923	1.3485	1.4374	1.1714	1.1335	1.2105
16-20/day	1.6494	1.5900	1.7111	1.4123	1.3603	1.4662	1.1650	1.1210	1.2106
20+/day	1.7494	1.6813	1.8203	1.4402	1.3829	1.4999	1.1881	1.1397	1.2387
Alcohol intake frequency									
Never	0.5651	0.5475	0.5834	0.5849	0.5662	0.6043	0.6285	0.6076	0.6501
Special occasions	0.7015	0.6825	0.7212	0.7109	0.6910	0.7313	0.7792	0.7569	0.8022
1-3/month	0.9082	0.8834	0.9337	0.8730	0.8486	0.8981	0.9129	0.8870	0.9395
1-2/week	Ref	–	–	Ref	–	–	Ref	–	–
3-4/week	1.0652	1.0418	1.0891	1.1059	1.0812	1.1312	1.0590	1.0350	1.0837
Daily/almost daily	0.9293	0.9087	0.9504	1.0627	1.0385	1.0875	1.0334	1.0091	1.0582
BMI (kg/m^2)	0.9561	0.9546	0.9577	0.9580	0.9564	0.9596	0.9683	0.9667	0.9700
Smoking status									
Never	Ref	–	–	Ref	–	–	Ref	–	–

Former	0.6977	0.6864	0.7092	0.7879	0.7747	0.8013	0.7922	0.7785	0.8061
Current	0.7309	0.7121	0.7502	0.6850	0.6670	0.7036	0.7518	0.7312	0.7730

Note: OR = odds ratio; CI = confidence interval; BMI = body mass index. [1]number of days per week engaging in these activities for 10+ minutes continuously.

Model 1 – only individual explanatory variables.

Model 2 – adjusted for age and sex.

Model 3 – all explanatory variables.

Table 1-D. **Environmental exposures associated with health status**

	Model 1			Model 2			Model 3		
Term	OR	95% CI		OR	95% CI		OR	95% CI	
$PM_{2.5}$	0.9640	0.9570	0.9711	0.9357	0.9287	0.9428	0.9656	0.9506	0.9809
PM_{10}	0.9917	0.9877	0.9957	0.9842	0.9801	0.9883	1.0019	0.9968	1.0070
NO_2	0.9972	0.9962	0.9982	0.9929	0.9918	0.9939	1.0036	1.0010	1.0061
L_{den}	0.9967	0.9949	0.9985	0.9945	0.9927	0.9964	0.9981	0.9959	1.0003
Greenspace 1000m	1.0001	0.9997	1.0004	1.0013	1.0010	1.0017	0.9995	0.9989	1.0001

Note: OR = odds ratio; CI = confidence interval; PM = particulate matter; NO_2 = nitrogen dioxide; L_{den} = day-evening-night noise level.

Model 1 – only individual explanatory variables.

Model 2 – adjusted for age and sex.

Model 3 – all explanatory variables.

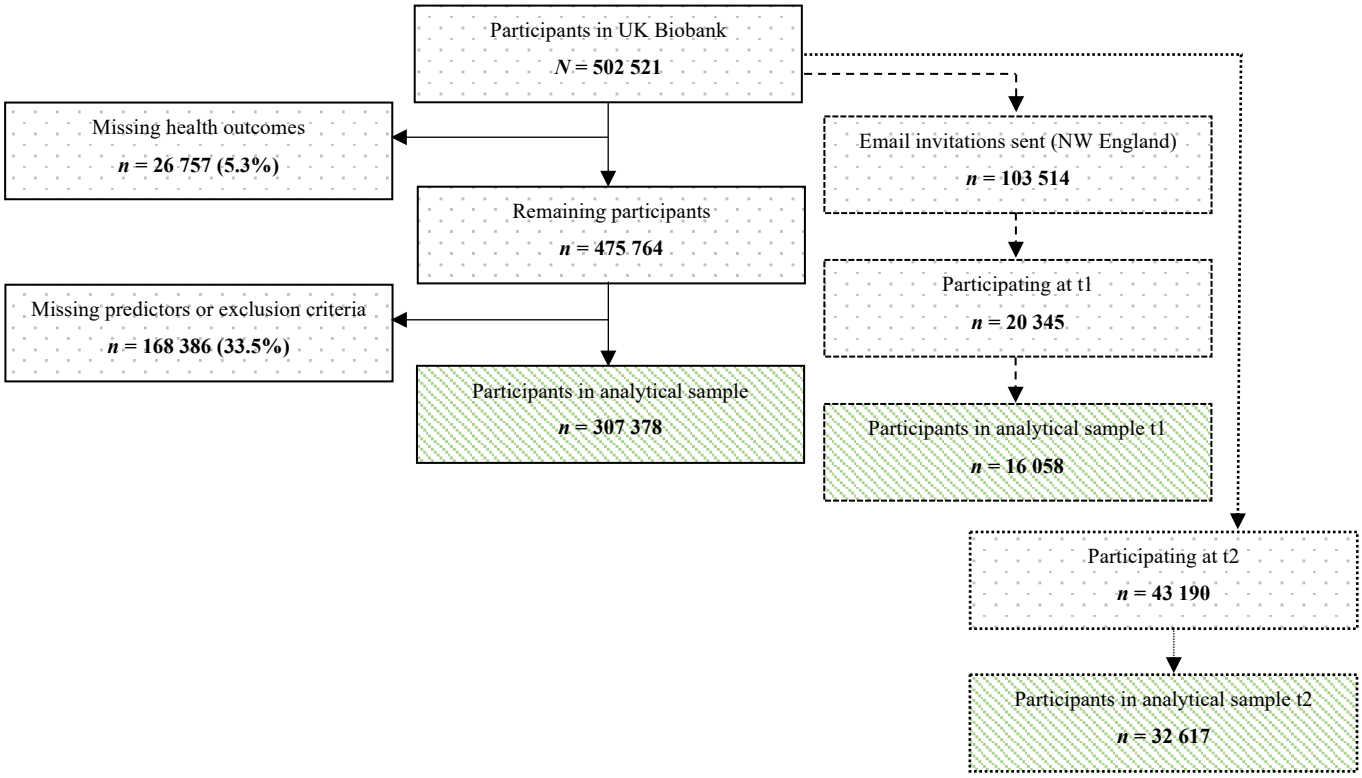

Figure 1. Flowchart of study population. t1 = follow-up during first revisit between 2012–2013; t2 = follow-up during UK Biobank Imaging Study since 2014.

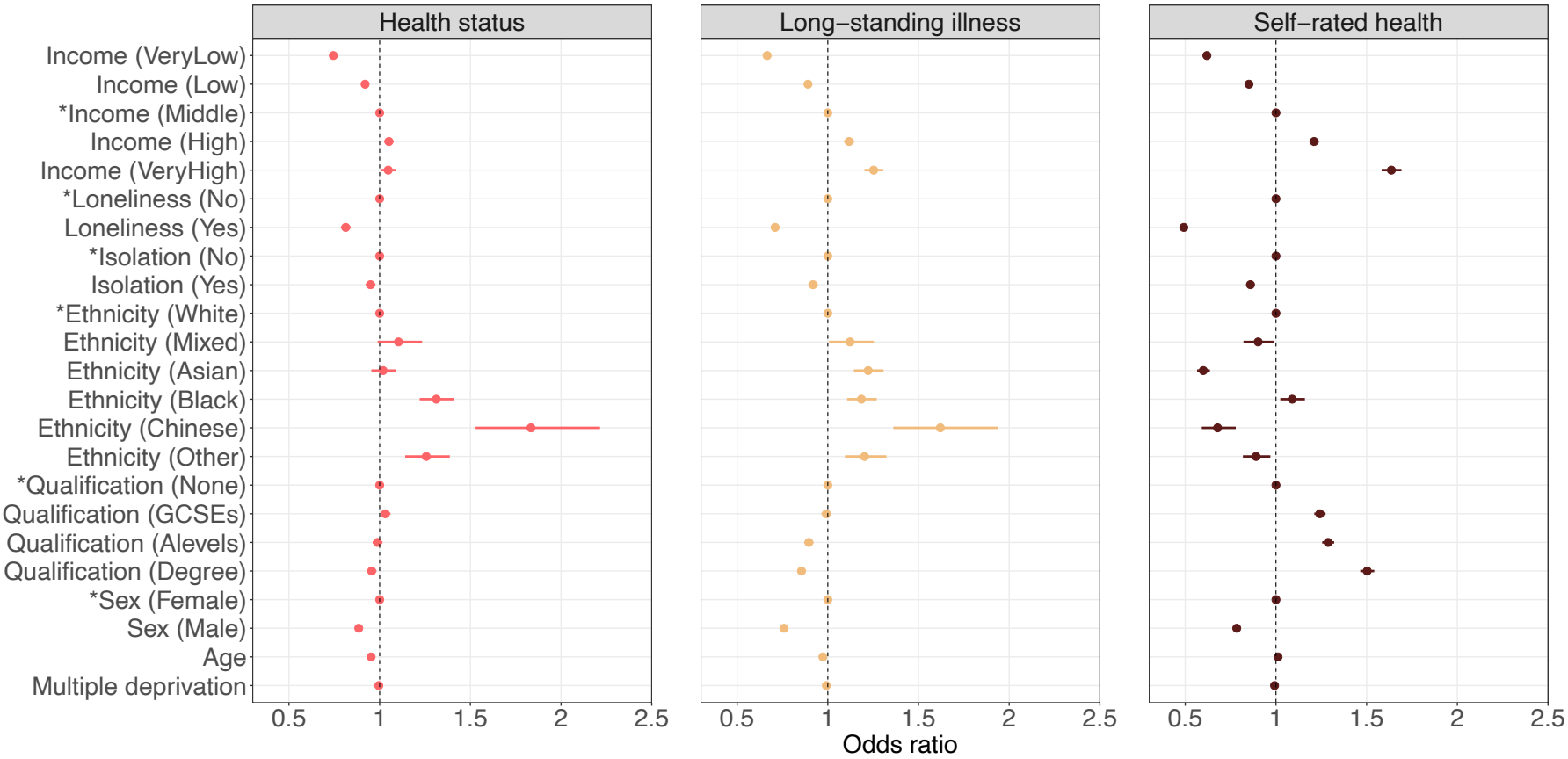

Figure 2-A. Sociodemographic characteristics and psychosocial factors associated with health indicators. Confidence interval plot (odds ratio ± 95% confidence interval) for Model 3 (i.e. including all explanatory variables). GCSEs = general certificate of secondary education. *Indicates reference group for categorical explanatory variables. For categorical explanatory variables the odds ratios for self-rated health indicate the changes in odds of reporting better self-rated health associated with the explanatory variable group relative to the reference group. Odds ratios for continuous explanatory variables for self-rated

health indicate proportional odds ratios for a 1-unit increase in the explanatory variable on level of self-rated health. Annual household income groups: very low (<£18 000), low (£18 000–30 999), middle (£31 000–51 999), high (£52 000–100 000) and very high (>£100 000). 'GCSEs' also includes O levels and certificate of secondary education (CSE). 'A levels' also includes national vocational qualification (NVQ), higher national diploma (HND), higher national certificate (HNC) and 'other professional qualifications'.

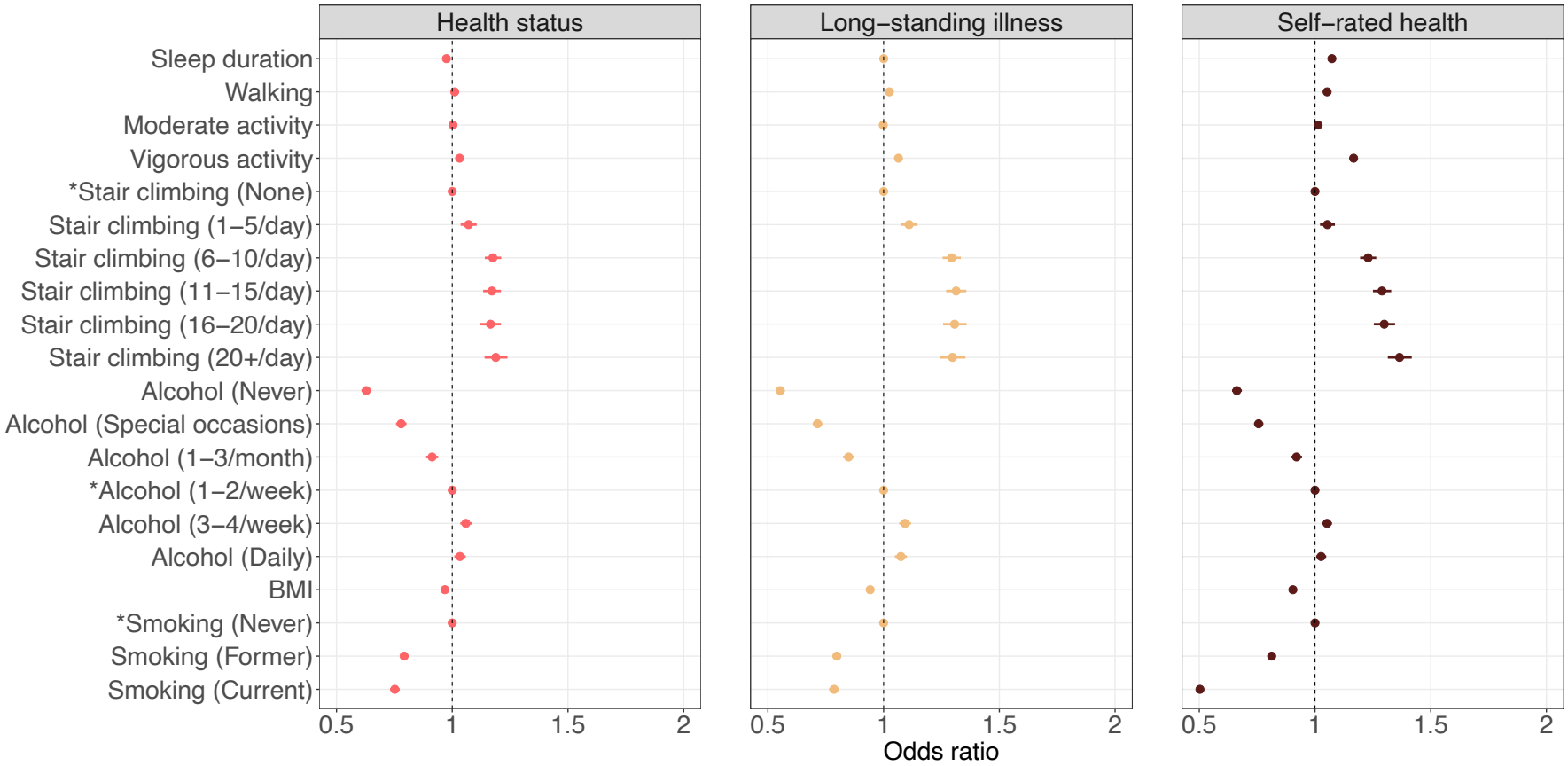

Figure 2-B. Lifestyle factors associated with health indicators. Confidence interval plot (odds ratio ± 95% confidence interval) for Model 3 (i.e. including all explanatory variables). BMI = body mass index. *Indicates reference group for categorical explanatory variables. For categorical explanatory variables the odds ratios for self-rated health indicate the changes in odds of reporting better self-rated health associated with the explanatory variable group relative to the

reference group. Odds ratios for continuous explanatory variables for self-rated health indicate proportional odds ratios for a 1-unit increase in the explanatory variable on level of self-rated health.

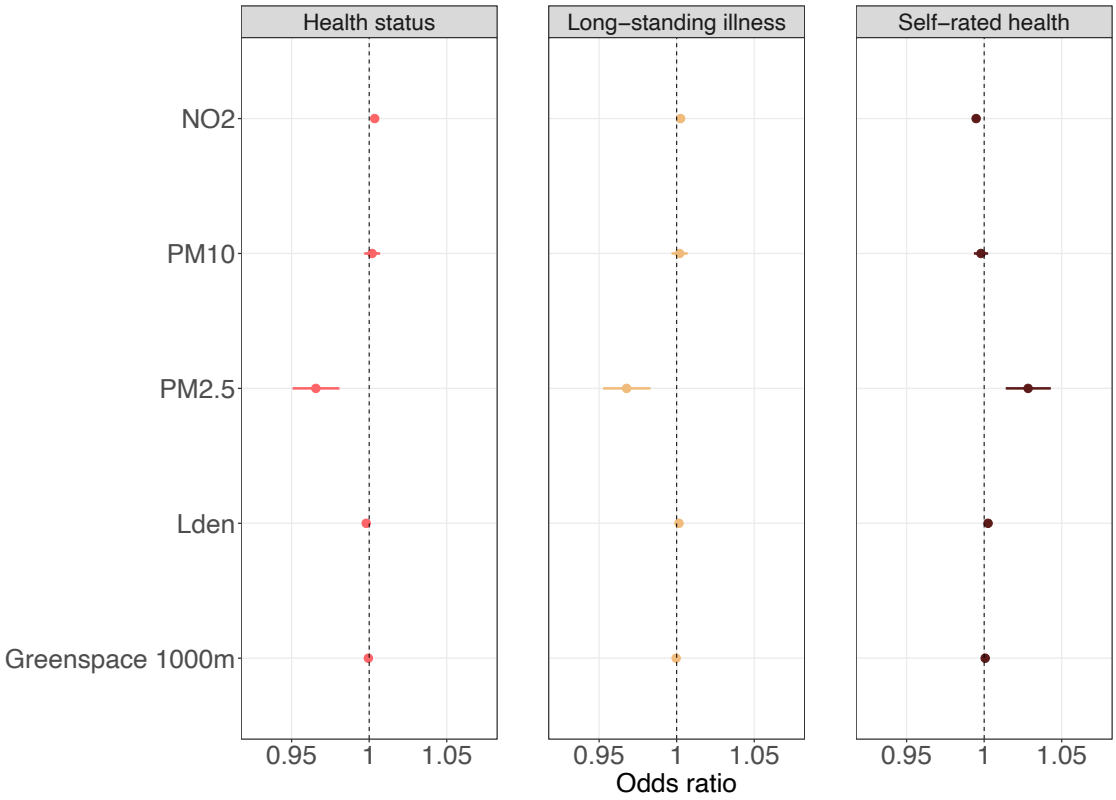

Figure 2-C. Environmental exposures associated with health indicators. Confidence interval plot (odds ratio ± 95% confidence interval) for Model 3 (i.e. including all explanatory variables). PM = particulate matter; NO_2 = nitrogen dioxide; L_{den} = day-evening-night noise level. Odds ratios for self-rated health indicate proportional odds ratios for a 1-unit increase in the explanatory variable on level of self-rated health.

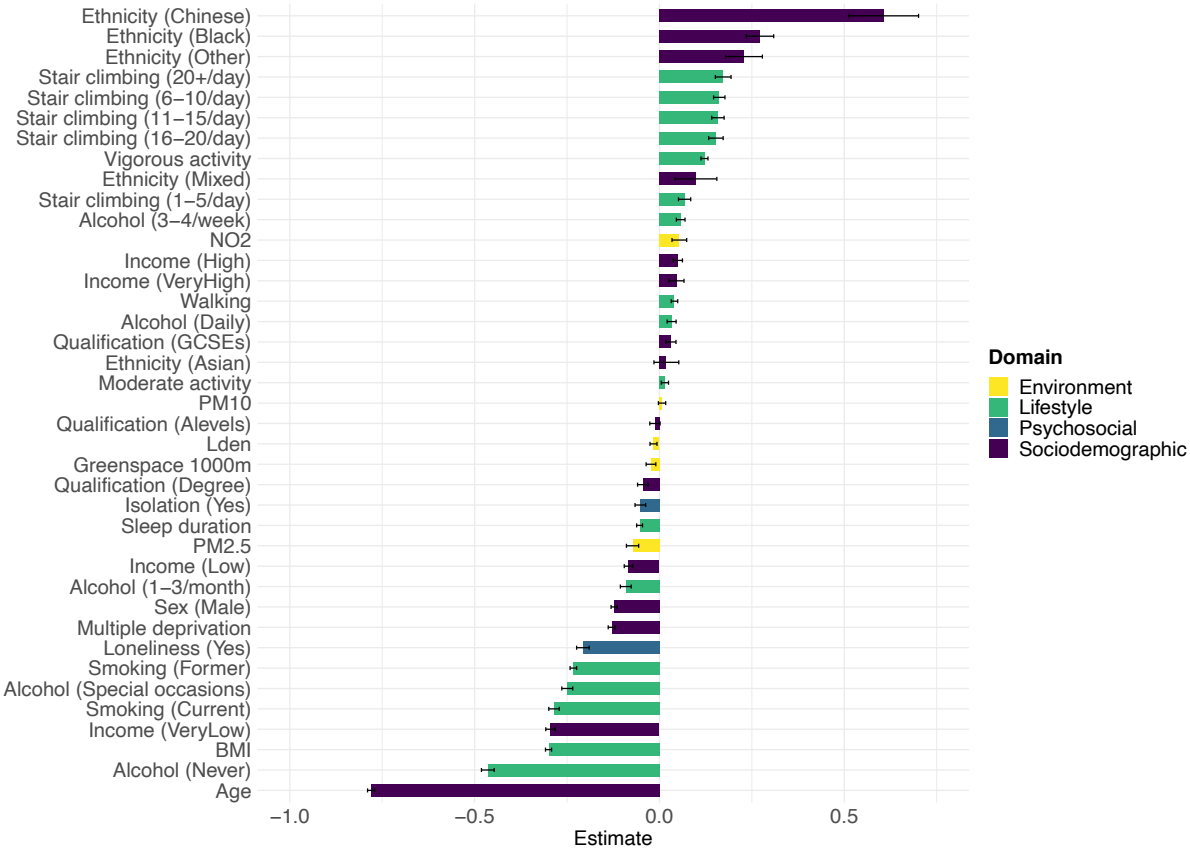

Figure 3. Horizontal bar plot for health status ($\beta \pm$ standard error). Model 3 regression coefficients rescaled to have a mean equal to zero and, for numeric variables with more than two values, divided by two standard deviations.