

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

3
4 **Short title:** Health in UK Biobank

5
6 **Authors:**

7 Julian Mutz^{1*}, Charlotte J. Roscoe^{2,3} and Cathryn M. Lewis^{1,4}

8
9 **Affiliations:**

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

17
18 **ORCID**

19 J Mutz: [0000-0001-5308-1957](https://orcid.org/0000-0001-5308-1957)

20 CM Lewis: [0000-0002-8249-8476](https://orcid.org/0000-0002-8249-8476)

21
22 ***Corresponding author:**

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

26 **Abstract**

27 A greater understanding of factors associated with good health may help increase longevity and
28 healthy life expectancy. Here we report associations between multiple health indicators and
29 sociodemographic (age, sex, ethnicity, education, income and deprivation), psychosocial (loneliness
30 and social isolation), lifestyle (smoking, alcohol intake, sleep, BMI, physical activity and stair
31 climbing) and environmental (air pollution, noise and greenspace) factors, using data from 307,378
32 UK Biobank participants. Low income, being male, neighbourhood deprivation, loneliness, social
33 isolation, short or long sleep duration, low or high BMI and smoking was associated with poor health.
34 Walking, vigorous-intensity physical activity and more frequent alcohol intake was associated with
35 good health. There was some evidence that airborne pollutants ($PM_{2.5}$, PM_{10} , and NO_2) and noise
36 (L_{den}) were associated with poor health, though findings were inconsistent in adjusted models. Our
37 findings highlight the multifactorial nature of health, the importance of non-medical factors, such as
38 loneliness, healthy lifestyle behaviours and weight management, and the need to examine efforts to
39 improve health outcomes of individuals with low income.

40 Introduction

41 Substantial improvements in human health and significant increases in average life expectancy are
42 amongst the main achievements of civilization over the past two centuries¹. Life expectancy at birth in
43 England and Wales has doubled from about 40 years in 1850 to more than 80 years in 2013². Future
44 life expectancy is projected to increase across 35 industrialised nations, with a median increase in life
45 expectancy at birth of approximately 3 years for women and 4 years for men in the United Kingdom
46 between 2010–2030³.

47
48 Many adults, however, now spend a substantial portion of their lives with late-life morbidities and
49 decreased sensory, motor and cognitive functioning, which may lead to lower quality of life^{4,5}. More
50 than 80% of the population of England and Wales aged 85 and above report a disability and 50%
51 require care or help with daily activities⁶. Nevertheless, some individuals experience very little
52 functional decline in old age and rate their health as good or excellent⁷. A greater understanding of the
53 factors associated with good health may help increase longevity and healthspan, i.e. the length of time
54 that a person is healthy⁸.

55
56 The health effects of lifestyle factors such as smoking⁹, diet¹⁰, excessive alcohol intake¹¹ and physical
57 activity¹² are well established. However, the number and type of potential confounders that have been
58 adjusted for in previous analyses vary widely¹³, and few studies have jointly examined
59 sociodemographic, psychosocial, lifestyle and environmental factors. In addition, research has
60 primarily focused on predicting mortality or disease incidence instead of overall health status, despite
61 its potential to provide insights into the factors associated with increased healthspan.

62
63 The UK Biobank study provides an unprecedented data resource to investigate determinants of health
64 and ageing trajectories. One of its strengths is that data collection was not merely focused on
65 established predictors of health and disease but also included relevant non-medical factors such as
66 social isolation and loneliness, as well as road traffic noise and ambient air pollution^{14,15}.

67
68 The aim of this study was to explore potential predictors of health status. More specifically, we
69 examined how health status was associated with (i) sociodemographic characteristics and
70 psychosocial factors, (ii) lifestyle factors, and (iii) environmental exposures, both independently and
71 jointly. We also examined, as a secondary aim, whether there was evidence of similar associations
72 between these factors and (i) long-standing illness and (ii) self-rated health.

73 **Methods**

74 **Study population**

75 The UK Biobank is a prospective study of >500,000 UK residents aged 37–73 at baseline, recruited
76 between 2006–2010. Details of the study rationale and design have been reported elsewhere¹⁶.

77 Briefly, individuals registered with the UK National Health Service (NHS) and living within a 25-
78 mile (~40 km) radius of one of 22 assessment centres were invited to participate (9,238,453 postal
79 invitations sent). At the baseline assessment, participants completed questionnaires and were
80 interviewed by nurses regarding sociodemographic characteristics, health behaviours and their
81 medical history. A small subset of participants completed repeat measurements: first revisit between
82 2012–2013 (20,344 participants); second revisit as part of the UK Biobank Imaging Study between
83 2014–2019 (43,190 participants). Participants consented to use of their de-identified data. See
84 Supplement-e1 for the UK Biobank data fields used.

85

86 **Exposures**

87 We considered individual-level sociodemographic characteristics including age at baseline
88 assessment, sex, ethnicity (White, Asian, Black, Chinese, mixed-race or other), highest educational or
89 professional qualification (four categories: (1) College/University degree; (2) A levels/AS levels or
90 equivalent, NVQ/HND/HNC or equivalent, other professional qualifications; (3) O levels/GCSEs or
91 equivalent, CSEs or equivalent; (4) none)¹⁷ and gross annual household income (<£18,000, £18,000–
92 30,999, £31,000–51,999, £52,000–100,000 or >£100 000). We also included the Index of Multiple
93 Deprivation for England which is a small-area level measure derived from data on income
94 deprivation, employment deprivation, health deprivation and disability, education skills and training
95 deprivation, barriers to housing and services, living environment deprivation and crime¹⁸. Higher
96 values on the index reflect greater deprivation.

97

98 Loneliness was assessed using two questions: "Do you often feel lonely?" (no=0/yes=1) and "How
99 often are you able to confide in someone close to you?" (almost daily to about once a month=0/once
100 every few months to never or almost never=1). Individuals who received a sum score of 2 were
101 classified as lonely¹⁹. Social isolation was assessed using three questions: "Including yourself, how
102 many people are living together in your household?" (living alone=1), "How often do you visit friends
103 or family or have them visit you?" (less than once a month=1) and "Which of the following
104 [leisure/social activities] do you attend once a week or more often?" (none of the above=1).
105 Individuals who received a sum score of 2–3 were classified as socially isolated¹⁹.

106

107 Smoking status was assessed using two questions summarising current and past smoking behaviour.

108 Individuals who responded "Yes, on most or all days" or "Only occasionally" to current tobacco

109 smoking were coded as "current". Individuals who responded "Smoked on most or all days" or
110 "Smoked occasionally" to past tobacco smoking were coded as "former". Individuals who responded
111 "No" to current tobacco smoking and "Just tried once or twice" or "I have never smoked" to past
112 tobacco smoking were coded as "never". Alcohol intake frequency was assessed using one question:
113 "About how often do you drink alcohol?". Response options included "Daily or almost daily", "Three
114 or four times a week", "Once or twice a week", "One to three times a month", "Special occasions
115 only" and "Never". Sleep duration was assessed using one question: "About how many hours sleep do
116 you get in every 24 hours? (please include naps)". Values below 1 hour or above 23 hours were
117 rejected, and participants were asked to confirm values below 3 hours or above 12 hours. Body mass
118 index (BMI) was calculated as weight divided by height squared (kg/m^2). Weight measurements were
119 obtained with a Tanita BC-418 MA body composition analyser. Standing height measurements were
120 obtained using a Seca 202 height measure. Physical activity was assessed using the International
121 Physical Activity Questionnaire (IPAQ) short form²⁰. Specifically, we included data on the number of
122 days per week spent walking, engaging in moderate-intensity physical activity (e.g. "carrying light
123 loads, cycling at normal pace") or engaging in vigorous-intensity physical activity (i.e. "activities that
124 make you sweat or breathe hard such as fast cycling, aerobics, heavy lifting") for ≥ 10 minutes
125 continuously. Daily frequency of stair climbing at home was assessed using one question: "At home,
126 during the last 4 weeks, about how many times a day do you climb a flight of stairs? (approx 10
127 steps)". These data were only collected from individuals who indicated that they were able to walk.
128 Response options included "None", "1–5 times a day", "6–10 times a day", "11–15 times a day", "16–
129 20 times a day" and "More than 20 times a day".

130
131 Exposure to airborne pollutants was estimated using a Land Use Regression model developed as part
132 of the European Study of Cohorts for Air Pollution Effects (ESCAPE) project^{21,22}. We included
133 annual average concentration of particulate matter with an aerodynamic diameter of $< 2.5 \mu\text{m}$ ($\text{PM}_{2.5}$)
134 and $< 10 \mu\text{m}$ (PM_{10}), as well as nitrogen dioxide (NO_2) modelled at participants' residential addresses
135 for the year 2010. Residential road traffic noise was modelled for the year 2009 using the Common
136 Noise Assessment Methods (CNOSSOS-EU) algorithm^{23,24}. We used L_{den} (day-evening-night noise
137 level) which is an annual average 24-hour sound pressure level in decibels with a 10-decibel penalty
138 added between 11pm and 7am. This penalty has previously been added in epidemiological analyses to
139 account for annoyance/sleep disruption at night²⁵. The percentage of the home location classed as
140 greenspace, as a proportion of all land use types, was modelled using 2005 data from the Generalized
141 Land Use Database for England (GLUD)²⁶ for the 2001 Census Output Areas in England. Each
142 residential address was allocated a circular distance buffer of 1000m, representing wider-area
143 greenspace.

144 **Health status**

145 Data on 81 cancer and 443 non-cancer illnesses (past and current) were ascertained through
146 touchscreen self-report questionnaire and confirmed during a verbal interview by a trained nurse. In
147 order to provide a single health indicator (“health status”) based on a previously defined algorithm,
148 we used a classification developed by the Reinsurance Group of America (RGA) in which an
149 experienced underwriter classified each illness according to whether it was “likely acceptable for
150 standard life insurance”²⁷. Participants were thus classified as healthy or unhealthy based on their
151 reported cancer and non-cancer illnesses (Supplement-e1; Supplementary file).

152

153 Two secondary health outcomes were assessed. Firstly, whether patients had a long-standing illness,
154 disability or infirmity was assessed using the question "Do you have any long-standing illness,
155 disability or infirmity?" to which individuals could respond "Yes" or "No". Secondly, participants’
156 perceived health was assessed using the question "In general how would you rate your overall
157 health?". Response options included "Poor", "Fair", "Good" and "Excellent". These health outcomes
158 will be termed "long-standing illness" and "self-rated health".

159

160 **Exclusion criteria**

161 Women who were pregnant at the time of assessment were excluded from the analysis based on the
162 assumption that lifestyle patterns change during pregnancy²⁸. Participants for whom their genetic sex,
163 inferred from the relative intensity of biological markers on the Y and X chromosomes, and self-
164 reported sex did not match were also excluded.

165

166 **Statistical analyses**

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

170

171 Logistic regression analyses were used to estimate associations between sociodemographic
172 characteristics, psychosocial factors, lifestyle factors and environmental exposures with health status
173 and long-standing illness. Ordinal logistic regression analyses were used to estimate associations
174 between these factors and self-rated health. Across these outcomes, poor health was the reference
175 group. In defining the reference groups for categorical explanatory variables, we focused on
176 interpretability of the results for middle aged and older UK residents.

177

178 We fitted three incremental models using the baseline UK Biobank data: Model 1–individual
179 explanatory variables; Model 2–individual explanatory variables plus age and sex; Model 3–all
180 explanatory variables. All models were fitted in the analytical sample without missing data to ensure

181 that any differences between models were not due to inclusion of different participants. We calculated
182 odds ratios and Bonferroni-adjusted (~99.9%) confidence intervals. Adjusted *p*-values were calculated
183 to account for multiple testing within each model (i.e., typically for 39 tests). Two methods were
184 used: (1) Bonferroni and (2) Benjamini & Hochberg²⁹, all two-tailed with $\alpha = .05$. We report in-text
185 the most conservative correction at which statistical significance was reached. If the unadjusted *p*-
186 value was greater than .05, in which case neither Bonferroni nor Benjamini & Hochberg-adjusted *p*-
187 values would reach criteria for statistical significance, we report the unadjusted *p*-value.

188 Multicollinearity was assessed using generalised variance inflation factors³⁰.

189

190 To compare the magnitudes of association between the explanatory variables and health indicators,
191 we calculated standardised regression coefficients by rescaling all variables included in Model 3 to
192 have a mean of zero and by dividing the coefficients of numeric variables with more than two values
193 by twice their standard deviation³¹.

194

195 To assess whether any associations between explanatory variables and health indicators were
196 modified by sex or age, we stratified analyses by sex and by age at baseline assessment (<65 years
197 and ≥ 65 years) and assessed potential age- and sex-interactions by adding cross-product terms to
198 Model 3. The choice of age strata was based on the current UK retirement age.

199

200 We used ordinal logistic regression analyses to model associations between the explanatory variables
201 at baseline and self-rated health at the first and second revisit (t1 and t2, respectively). We applied the
202 same modelling strategy, except that we additionally adjusted for the number of days between
203 baseline and follow-up assessment.

204

205 As variable selection and classification might impact results, we conducted multiple additional
206 analyses to examine related variables, categorised continuous variables according to health guidelines
207 or previously determined cut-offs and tested the robustness of our findings to additional exclusion
208 criteria. We repeated the main analysis (i) with sleep duration as categorical variable (<7, 7-8
209 [reference] and ≥ 9 hours of sleep/day)³²; (ii) with BMI as categorical variable (<18.5, 18.5-24.9
210 [reference] and >24.9 kg/m²); (iii) with body fat percentage (estimated by electrical bio-impedance
211 measurement) instead of BMI as measure of body composition; (iv) excluding individuals who
212 reported that they stopped drinking alcohol because of "Illness or ill health" or "Doctor's advice"; (v)
213 sub-dividing those individuals who reported that they never drink alcohol into current and lifetime
214 abstainers; (vi) with current tobacco smoking only (three categories: "Yes, on most or all days", "Only
215 occasionally" and "No"); (vii) with Metabolic Equivalent Task (MET) minutes per week for walking,
216 moderate physical activity and vigorous physical activity³³ instead of the number of days per week
217 spent engaging in these activities for ≥ 10 minutes; (viii) reducing the greenspace percentage circular

218 distance buffer from 1000m to 300m to examine nearby greenspace; (ix) with air pollution and noise
219 exposure estimates dichotomised according to WHO recommendation thresholds^{34,35}: $PM_{2.5} \leq 10$
220 $\mu\text{g}/\text{m}^3$, $PM_{10} \leq 20 \mu\text{g}/\text{m}^3$, $NO_2 \leq 40 \mu\text{g}/\text{m}^3$ and $L_{\text{den}} \leq 53$ decibels; (x) restricting analyses to participants
221 assessed after 31 December 2008 (regarding noise estimates) and restricting analyses to participants
222 assessed after 31 December 2009 (regarding air pollution estimates); (xi) restricting analyses to those
223 individuals who had lived at their current address for at least 10 years; (xii) truncating continuous
224 explanatory variables at the 1st and 99th %ile of the distribution.
225
226 Statistical analyses were conducted using R (version 3.5.0).

227 **Results**

228 **Study population**

229 Of the $N = 502,521$ UK Biobank participants, 5.32% ($n = 26,757$) had missing health data and
230 33.51% ($n = 168,386$) had missing data on explanatory variables or did not meet our inclusion
231 criteria. Hence our analytical sample included $n = 307,378$ adults (Supplement-e2).

232
233 Descriptive statistics are presented in Supplement-e3–e6. There were no substantial differences
234 between the full and analytical sample. Of the participants in the analytical sample, 8.18% ($n =$
235 25,142) reported at least one cancer and 73.30% ($n = 225,312$) at least one non-cancer illness
236 (Supplement-e4). Approximately two thirds of participants (69.04%, $n = 212,201$) were classified as
237 healthy, while 30.96% ($n = 95,177$) were unhealthy. A similar percentage of participants (30.5%, $n =$
238 93,757) reported having a long-standing illness (72.9% agreement with health status). Finally, 3.60%
239 ($n = 11,066$) rated their health as poor, while 19.25% ($n = 59,169$), 59.44% ($n = 182,699$) and 17.71%
240 ($n = 54,444$) rated their health as fair, good and excellent, respectively.

241

242 **Regression analyses**

243 We found weak correlations between most continuous explanatory variables and moderate to strong
244 correlations between the environmental exposures (Supplement-e5). Generalised variance inflation
245 factor (VIF) values for Model 3 were within the range of 1.02–2.62 for 19/21 explanatory variables.
246 The VIFs for NO_2 and $\text{PM}_{2.5}$ were 6.08 and 4.27, respectively. Excluding NO_2 from the model
247 reduced the highest VIF to 2.42 (for $\text{PM}_{2.5}$). Fitting Model 3 without NO_2 or with each air pollutant
248 separately had little impact on their associations with health status. As such, we report the results for
249 Model 3 that included all explanatory variables. Unless indicated otherwise, results presented below
250 correspond to multivariable-adjusted odds ratios and Bonferroni-adjusted (~99.9%) confidence
251 intervals from Model 3. A simplified overview of our findings is presented in Supplement-e7.

252

253 Increased age was associated with lower odds of favourable health status (OR = 0.953, 99.9% CI
254 0.951-0.955, $p_{\text{Bonf.}} < 0.001$). For every 1-year increase in age, the log-odds of being healthy decreased
255 by -0.048 (99.9% CI -0.050 to -0.046). Men had lower odds of being healthy (OR = 0.88, 99.9% CI
256 0.861-0.909, $p_{\text{Bonf.}} < 0.001$). Individuals with high income had higher odds of being healthy (OR =
257 1.05, 99.9% CI 1.010-1.094, $p_{\text{Bonf.}} = 0.003$ [£52,000–100,000 vs £31,000–51,999]), while those with
258 lower levels of income had lower odds of being healthy (e.g. OR = 0.74, 99.9% CI 0.715-0.776,
259 $p_{\text{Bonf.}} < 0.001$ [$<£18,000$ vs £31,000–51,999]). Increased neighbourhood deprivation was associated
260 with lower odds of being healthy (OR = 0.995, 99.9% CI 0.994-0.996, $p_{\text{Bonf.}} < 0.001$). Compared to
261 Whites, individuals of Black, Chinese, Mixed-race, and “other” ethnic background had higher odds of
262 being healthy in Model 1, but only Chinese ethnicity was associated with higher odds of being healthy

263 across all models (Model 3: OR = 1.83, 99.9% CI 1.365-2.511, $p_{\text{Bonf.}} < 0.001$). Compared to individuals
264 without educational or professional qualification, participants with any qualification had higher odds
265 of being healthy in Model 1 and after adjustment for age and sex (Model 2). However, there was only
266 limited evidence of associations between highest qualification and health in Model 3 (Table-1).

267
268 [insert Table-1]

269
270 Loneliness was associated with lower odds of favourable health status (OR = 0.81, 99.9% CI 0.770-
271 0.859, $p_{\text{Bonf.}} < 0.001$). Socially isolated individuals also had lower odds of being healthy, but there was
272 greater attenuation in the strength of association in Model 3 (OR = 0.95, 99.9% CI 0.907-0.995, $p_{\text{Bonf.}}$
273 = 0.013) (Table-1).

274
275 Longer sleep duration was associated with lower odds of favourable health status (OR = 0.97, 99.9%
276 CI 0.963-0.987, $p_{\text{Bonf.}} < 0.001$). For every 1-hour increase in sleep duration, the log-odds of being
277 healthy decreased by -0.03 (99.9% CI -0.038 to -0.013). Walking frequently and engaging in frequent
278 vigorous physical activity was also associated with higher odds of being healthy (OR = 1.01, 99.9%
279 CI 1.003-1.018, $p_{\text{Bonf.}} < 0.001$ and OR = 1.03, 99.9% CI 1.024-1.040, $p_{\text{Bonf.}} < 0.001$, respectively). For
280 every one-day increase in ≥ 10 minutes of vigorous physical activity, the log-odds of favourable health
281 status increased by 0.03 (99.9% CI 0.024-0.039). Although moderate physical activity was associated
282 with higher odds of being healthy in Model 1–2, we did not find evidence of an association in Model
283 3 (OR = 1.00, 99.9% CI 0.996-1.010, $p_{\text{unadj.}} = 0.13$). Frequent daily stair climbing at home was
284 associated with higher odds of being healthy (ranging from OR = 1.07, 99.9% CI 1.015-1.129, $p_{\text{Bonf.}} =$
285 0.002 to OR = 1.19, 99.9% CI 1.110-1.272, $p_{\text{Bonf.}} < 0.001$ [1-5 times/day and >20 times/day vs none,
286 respectively]). We found some evidence that frequent alcohol intake was associated with higher odds
287 of being healthy (e.g. OR = 1.06, 99.9% CI 1.020-1.010, $p_{\text{Bonf.}} < 0.001$ [3-4 vs 1-2 times/week]),
288 although for daily/almost daily alcohol drinking only in Model 2 (OR = 1.06, 99.9% CI 1.023-1.104,
289 $p_{\text{Bonf.}} < 0.001$), while infrequent alcohol intake was associated with lower odds of favourable health
290 status (ranging from OR = 0.91, 99.9% CI 0.871-0.957, $p_{\text{Bonf.}} < 0.001$ to OR = 0.63, 99.9% CI 0.595-
291 0.664, $p_{\text{Bonf.}} < 0.001$ [1-3 times/month and never vs 1-2 times/week, respectively]). Higher BMI was
292 associated with lower odds of being healthy (OR = 0.968, 99.9% CI 0.966-0.971, $p_{\text{Bonf.}} < 0.001$). For
293 every one-unit increase in kg/m^2 , the log-odds of being healthy decreased by -0.03 (99.9% CI -0.035
294 to -0.029). Past and current tobacco smoking was associated with lower odds of being healthy (OR =
295 0.79, 99.9% CI 0.770-0.815, $p_{\text{Bonf.}} < 0.001$ and OR = 0.75, 99.9% CI 0.718-0.787, $p_{\text{Bonf.}} < 0.001$,
296 respectively) (Table-2).

297
298 [insert Table-2]

299

300 In our analyses of environmental exposures, higher PM_{2.5} concentration was associated with lower
301 odds of favourable health status (OR = 0.97, 99.9% CI 0.941-0.991, $p_{\text{Bonf.}} < 0.001$). PM₁₀ was also
302 associated with lower odds of being healthy in Model 1–2, but there was no evidence of an
303 association between PM₁₀ and health status in Model 3 (OR = 1.002, 99.9% CI 0.993-1.010, $p_{\text{unadj.}} =$
304 0.47). NO₂ was associated with lower odds of being healthy in Model 1–2. However, in Model 3,
305 there was limited evidence that NO₂ was associated with higher odds of being healthy (OR = 1.004,
306 99.9% CI 0.999-1.008, $p_{\text{BH}} = 0.009$). Ambient sound level was associated with lower odds of being
307 healthy in Model 1–2, but there was no evidence of an association with health status in Model 3 (OR
308 = 0.998, 99.9% CI 0.995-1.002, $p_{\text{unadj.}} = 0.09$). Finally, we did not find consistent evidence of an
309 association between percentage greenspace within a 1000m circular distance buffer and health status
310 (Table-3).

311
312 [insert Table-3]

313
314 Results for long-standing illness and self-rated health are described in the Supplementary text and
315 presented in Supplement-e8–e10. Model 3 results across health indicators are presented in Figure-1–3.

316 317 **Magnitude of associations**

318 Standardised regression coefficients for health status are presented in Figure-4. Most notably, the
319 magnitude of association between BMI ($\beta = -0.30$, 99.9% CI -0.33 to -0.27) or very low income ($\beta = -$
320 0.29, 99.9% CI -0.34 to -0.25) and health status was comparable to that of current smoking ($\beta = -0.29$,
321 99.9% CI -0.33 to -0.24). There was also a substantial difference in magnitude of association with
322 health status between the very low and low income groups ($\beta = -0.29$, 99.9% CI -0.34 to -0.25, and β
323 = -0.08, 99.9% CI -0.12 to -0.05, respectively). The magnitude of association between loneliness and
324 health status was substantially larger than that of social isolation ($\beta = -0.21$, 99.9% CI -0.26 to -0.15,
325 and $\beta = -0.05$, 99.9% CI -0.10 to -0.01, respectively). Finally, associations between environmental
326 exposures and health status were relatively small compared to those observed for sociodemographic,
327 psychosocial and lifestyle factors (Supplement-e11).

328
329 [insert Figure-4]

330
331 Associations with long-standing illness and self-rated health compared to health status were stronger
332 for household income, sex, neighbourhood deprivation, loneliness, social isolation, walking
333 frequency, vigorous physical activity, BMI and stair climbing frequency (Supplement-e11–e12).

334 **Stratified and interaction analyses**

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

337

338 The odds of being healthy were lower for men than for women in the lowest income group
339 ($p_{\text{Bonf.}(interaction)} < 0.001$). The association between increased age and lower odds of being healthy was
340 stronger in men ($p_{\text{Bonf.}(interaction)} < 0.001$). Asian women had higher odds of being healthy ($p_{\text{BH}(interaction)} =$
341 0.007). The association between neighbourhood deprivation and lower odds of being healthy was
342 stronger in men ($p_{\text{Bonf.}(interaction)} < 0.001$). Men who were lonely had lower odds of being healthy
343 ($p_{\text{Bonf.}(interaction)} = 0.008$) and social isolation was associated with lower odds of being healthy only in
344 men ($p_{\text{BH}(interaction)} = 0.006$). Longer sleep duration was associated with lower odds of being healthy
345 only in men ($p_{\text{Bonf.}(interaction)} < 0.001$). Reporting frequent alcohol intake was associated with higher odds
346 of being healthy only in men ($p_{\text{BH}(interaction)} = 0.023$ for 3-4 times/week and $p_{\text{BH}(interaction)} = 0.039$ for
347 daily/almost daily). The association between higher BMI and lower odds of being healthy was
348 stronger in men ($p_{\text{Bonf.}(interaction)} < 0.001$). The association between former smoking and health status was
349 stronger in men ($p_{\text{Bonf.}(interaction)} < 0.001$), while the association between current smoking and health
350 status was stronger in women ($p_{\text{BH}(interaction)} = 0.005$). $\text{PM}_{2.5}$ was associated with lower odds of being
351 healthy only in men ($p_{\text{BH}(interaction)} = 0.011$). Finally, we found that NO_2 was associated with higher
352 odds of favourable health status only in women ($p_{\text{BH}(interaction)} = 0.024$) (Supplement-e15–e16).

353

354 The association between very low income and lower odds of being healthy was stronger in
355 participants below the age of 65 than in participants aged 65 and above ($p_{\text{Bonf.}(interaction)} < 0.001$).
356 Reporting an income of £52,000–100,000 was associated with higher odds of being healthy only in
357 participants younger than 65 ($p_{\text{BH}(interaction)} = 0.016$). Men aged 65 and above had lower odds of being
358 healthy ($p_{\text{Bonf.}(interaction)} < 0.001$). Loneliness and social isolation were associated with lower odds of
359 being healthy only in individuals younger than 65 ($p_{\text{BH}(interaction)} = 0.024$ and $p_{\text{Bonf.}(interaction)} = 0.011$,
360 respectively). The association between walking frequency and higher odds of being healthy was
361 stronger in individuals aged 65 and above ($p_{\text{Bonf.}(interaction)} < 0.001$). Climbing stairs 1-5 times/day was
362 associated with favourable health status only in individuals younger than 65 ($p_{\text{Bonf.}(interaction)} < 0.001$).
363 There was some evidence that the association between never drinking alcohol and poor health status
364 was stronger in individuals younger than 65 ($p_{\text{Bonf.}(interaction)} = 0.002$). The association between higher
365 BMI and lower odds of being healthy was stronger in individuals aged 65 and above
366 ($p_{\text{BH}(interaction)} < 0.001$). Finally, the association between current smoking and poor health status was
367 stronger in individuals younger than 65 ($p_{\text{BH}(interaction)} = 0.038$) (Supplement-e17–e18).

368

369 Although there were some differences, stratified analyses of long-standing illness and self-rated
370 health indicated a large degree of consistency with the results observed for health status (Supplement-
371 e19–e26).

372

373 **Longitudinal analyses**

374 We repeated our analyses of self-rated health in participants with follow-up data collected between
375 2012–2013 ($n = 16,058$; mean follow-up = 4.26 years, SD = 0.87) and between 2014–2019 ($n =$
376 $32,617$; mean follow-up = 8.57 years, SD = 1.64). Descriptive statistics and full results are presented
377 in Supplement-e27–e29. The findings were fully consistent with cross-sectional analyses for very low
378 and high to very high levels of income, sex, loneliness, sleep duration, walking frequency, vigorous
379 physical activity, infrequent alcohol intake, BMI and smoking status. Although many results were
380 consistent across timepoints, we found some differences between cross-sectional and prospective
381 analyses for low income, age, neighbourhood deprivation, ethnicity, highest qualification, social
382 isolation, moderate physical activity, stair climbing frequency, regular alcohol intake and
383 environmental exposures.

384

385 **Additional analyses**

386 Descriptive statistics are presented in Supplement-e30.

387

388 For sleep duration and BMI, we found evidence of non-linearity in the association with health status
389 (Supplement-e31). Hence, we also examined these explanatory variables as categorical variables.
390 Compared to individuals with optimal sleep duration (7-8 hours/day), those who slept <7 or ≥ 9
391 hours/day had lower odds of being healthy (OR = 0.90, 99.9% CI 0.868-0.923, $p_{\text{Bonf.}} < 0.001$ and OR =
392 0.70, 99.9% CI 0.667-0.736, $p_{\text{Bonf.}} < 0.001$, respectively). We also found that low ($< 18.5 \text{ kg/m}^2$) and
393 high ($> 24.9 \text{ kg/m}^2$) BMI was associated with lower odds of being healthy, compared to a BMI of
394 $18.5\text{-}24.9 \text{ kg/m}^2$ (OR = 0.70, 99.9% CI 0.584-0.852, $p_{\text{Bonf.}} < 0.001$ and OR = 0.85, 99.9% CI 0.827-
395 0.877, $p_{\text{Bonf.}} < 0.001$, respectively). There was no evidence of substantial departure from a linear
396 association with health status for all other numerical variables with more than two values
397 (Supplement-e31). Repeating the main analysis with body fat percentage as measure of body
398 composition led to similar conclusions as for BMI: higher fat percentage was associated with lower
399 odds of being healthy ($n = 307,202$; OR = 0.978, 99.9% CI 0.975-0.980, $p_{\text{Bonf.}} < 0.001$). Excluding
400 individuals who reported that they had stopped drinking alcohol due to illness or their doctor's advice
401 ($n = 2,790$) attenuated the association between never drinking alcohol and health status ($n = 304,588$;
402 OR = 0.74, 99.9% CI 0.701-0.790, $p_{\text{Bonf.}} < 0.001$). When sub-dividing individuals who reported that
403 they never drink alcohol into current and lifetime abstainers, the association with health status was
404 stronger in current abstainers ($n = 307,346$; OR = 0.53, 99.9% CI 0.491-0.568, $p_{\text{Bonf.}} < 0.001$ and OR =
405 0.75, 99.9% CI 0.698-0.810, $p_{\text{Bonf.}} < 0.001$, respectively). Restricting analyses to current smoking, we

406 found that regular smoking and smoking only occasionally was associated with lower odds of being
407 healthy, although the magnitude of association was stronger for regular smoking ($n = 307,372$; OR =
408 0.81, 99.9% CI 0.766-0.848, $p_{\text{Bonf.}} < 0.001$ and OR = 0.90, 99.9% CI 0.835-0.980, $p_{\text{Bonf.}} = 0.002$,
409 respectively). Examining MET minutes instead of the number of days per week spent walking,
410 engaging in moderate or vigorous physical activity led to similar results ($n = 268,674$; data not
411 shown).

412
413 When we examined air pollution and noise exposure estimates dichotomised according to WHO
414 recommendation thresholds, we found some evidence that being exposed to $\leq 10 \mu\text{g}/\text{m}^3$ PM_{2.5} on
415 average annually was associated with higher odds of being healthy (OR = 1.03, 99.9% CI 0.997-
416 1.062, $p_{\text{BH}} = 0.005$). There was no evidence of an association with health status for exposures to ≤ 20
417 $\mu\text{g}/\text{m}^3$ PM₁₀, $\leq 40 \mu\text{g}/\text{m}^3$ NO₂ or ≤ 53 decibels L_{den} in Model 3. We also did not find evidence of a
418 threshold in the association between PM₁₀ or NO₂ and health status when including quintiles of these
419 exposures in Model 3 (data not shown). Restricting analyses to participants assessed after 2008
420 (regarding noise estimates; $n = 182,342$) or 2009 (regarding air pollution estimates; $n = 60,521$) or to
421 those who had lived at their current address for at least 10 years ($n = 209,239$) did not lead to different
422 conclusions regarding the associations between environmental exposures and health status. There was
423 some evidence that percentage greenspace within a 300m circular distance buffer was associated with
424 health status ($n = 307,378$; OR = 0.999, 99.9% CI 0.9985-1.0001, $p_{\text{BH}} = 0.009$). Truncating
425 continuous explanatory variables at the 1st and 99th %ile did not materially change our results ($n =$
426 278,065; data not shown).

427 **Discussion**

428 Increased age was strongly associated with unfavourable health status and reporting a long-standing
429 illness. However, older participants generally rated their health positive, which is consistent with
430 several smaller studies^{36,37}, though not all^{38,39}. Ageing might be the single most important factor
431 underlying disease⁴⁰, with an almost universally-accepted expectation of declining health as people
432 get older. As attainable health states shift with age⁴¹, older participants might evaluate their health
433 more favourable, despite higher rates of illness and disability.

434

435 Although women, on average, report more illnesses, disabilities, and limitations in daily life⁴²⁻⁴⁴, one
436 of the most robust findings in human biology is that they live longer⁴⁵. Our findings are consistent
437 with results from the Newcastle 85+ cohort study in which women rated their health more favourably
438 than men⁴². However, other studies reported that women rated their health less favourably than men⁴⁶⁻
439 ⁴⁸, or did not find evidence of sex differences in self-rated health^{49,50}. Sex differences in health status
440 could result from differences in the frequency of specific illnesses, sex-specific reporting patterns, or
441 biological and social factors⁴⁸. Discrepancies in findings between studies might reflect differences in
442 age group or socio-cultural factors.

443

444 High income and low levels of neighbourhood deprivation were associated with better health, broadly
445 consistent with previous studies⁵¹⁻⁵³. Notably, we found only a small difference in the strength of
446 association with health status and long-standing illness between the high-income groups. The
447 difference between the low-income groups, however, was substantial, supporting previous findings
448 which suggested a non-linear association between family income and mortality⁵⁴. For self-rated
449 health, we also found evidence of substantial differences between the high-income groups.

450

451 Our study provides limited evidence that education was associated with favourable health status once
452 other factors had been accounted for, although higher levels of qualification were associated with
453 better self-rated health, consistent with previous research³⁸. A recent UK Biobank analysis showed
454 that remaining longer in school causally reduced participants' risk of diabetes and mortality⁵⁵. A
455 potential explanation for why we did not find a consistent pattern in the full model is that differences
456 in health status result from educated individuals engaging in healthier lifestyle behaviours⁵⁶ that we
457 had accounted for, or they could be due to socioeconomic or genomic factors.

458

459 Social isolation and loneliness were associated with poor health. The strength of association was
460 greater for loneliness, particularly in men and individuals below the age of 65. Social isolation and
461 loneliness are not always correlated^{57,58} and represent different aspects of social relations (scarcity of
462 contact with others and discrepancies between the need for, and the fulfilment of, social interaction,

463 respectively). In a meta-analysis of 70 studies, social isolation, loneliness, and living alone were
464 associated with a 26-32% increased mortality risk¹⁴. There was no evidence of differences in mortality
465 between these measures, although the strength of association was greater in individuals below the age
466 of 65. A recent UK Biobank analysis found that socially isolated and lonely individuals had an
467 increased risk of death, but only social isolation predicted all-cause mortality in a joint model¹⁹. The
468 discrepancy with our finding (loneliness was more strongly associated with poor health) might reflect
469 differences in outcome measures (general health in the present study vs mortality in previous
470 investigations).

471
472 Long sleep duration, higher BMI as well as past and current smoking was associated with poor health.
473 Sleeping less than 7 hours/day was also associated with poor health, consistent with a meta-analysis
474 that provided evidence of a U-shaped association between sleep duration and all-cause mortality³². A
475 BMI outside the optimal range of 18.5-24.9 kg/m² was also associated with poor health.

476
477 Physical activity is a key lifestyle factor recommended for primary and secondary prevention of
478 chronic health conditions⁵⁹, and is associated with lower mortality risk⁶⁰. In this study, walking
479 frequency, especially in individuals aged 65 and above, stair climbing, and engaging in vigorous
480 physical activity was associated with good health. Moderate physical activity was associated with
481 better self-rated health, especially in men. A study in middle-aged British men found evidence of an
482 association between vigorous, but not moderate, physical activity and reduced mortality⁶¹. Reviews of
483 the literature report mixed findings on the relative contributions of moderate and vigorous physical
484 activity^{62,63}, with some evidence suggesting stronger associations for vigorous activity⁶⁴.

485
486 A more frequent drinking pattern was associated with good health. Alcohol drinking often occurs in a
487 social context and might therefore constitute a proxy for social wellness, supported by the finding that
488 non-drinkers tend to be characterised by poor psychosocial health and low socioeconomic status.
489 Moderate drinkers also perform more sports than lifelong abstainers⁶⁵. Drinking less frequent than 1-2
490 times/week was associated with poor health and could not be fully accounted for by excluding
491 individuals who had discontinued alcohol intake for health reasons or because of their doctor's
492 advice. However, the association with health status was stronger for current abstainers, suggesting
493 that some individuals are non-drinkers in later life due to illness⁶⁶. Current abstainers who quit
494 drinking for health reasons could exaggerate poor health outcomes associated with not drinking
495 alcohol⁶⁷. Those who never drink might also differ from current drinkers in other characteristics⁶⁸.

496
497 Road traffic noise and ambient air pollution are environmental risk factors for health. While we found
498 some evidence that higher levels of all airborne pollutants were associated with poor health, only
499 PM_{2.5} was associated with unfavourable health status and long-standing illness after adjustment for

500 other factors. Higher levels of NO₂ were associated with less favourable self-rated health. A joint
501 analysis of UK Biobank, HUNT, and EPIC-Oxford data did not find evidence of an association
502 between NO₂ and any disease outcome, although both PM_{2.5} and PM₁₀ were associated with higher
503 incident cardiovascular disease⁶⁹. Adjustment for physical activity and neighbourhood deprivation in
504 the present study could have contributed to differences in findings between studies. Few individuals
505 in our study were exposed to high levels of PM₁₀ and NO₂, which might explain why we did not find
506 strong associations between these pollutants and health. The only exception was PM_{2.5}, for which
507 45% of our sample were exposed to levels above the WHO recommendation threshold of $\leq 10 \mu\text{g}/\text{m}^3$.
508 Positive associations between PM_{2.5} and self-rated health and between NO₂ and favourable health
509 status might represent spurious associations or statistical artefacts; we are not aware of any
510 mechanisms linking higher levels of pollution to good health. While higher levels of residential noise
511 were associated with poor health in Model 1-2, there was no evidence of an association after
512 adjustment for other factors. Our findings are consistent with a previous analysis that did not find
513 evidence of an independent association between L_{den} and cardiovascular disease, ischemic heart
514 disease or cerebrovascular disease⁶⁹.

515

516 We found some evidence that greenspace was associated with better self-rated health, consistent with
517 a recent meta-analysis⁷⁰, although not after adjustment for other factors. There were also some
518 inconsistencies in the prospective analyses and regarding health status and long-standing illness.
519 Several previous studies examining associations between greenspace and health did not adjust for
520 physical activity and most did not include other environmental exposures such as air pollution or
521 noise⁷⁰. These differences could explain why we did not find consistent associations in this study.

522

523 **Strengths and limitations**

524 The large sample size enabled high precision in the estimation of associations and it allowed us to
525 explore a range of explanatory variables, sex and age-specific associations, interactions and additional
526 analyses with further classification of explanatory variables. Many of the findings from previous
527 studies that we replicated in our analyses were conducted in smaller populations and with one
528 exposure at a time, often providing limited insight into the multifactorial nature of health. One of the
529 contributions of this study is that it allows for systematic comparisons of a broad range of factors
530 associated with health. We examined three health indicators with slightly different ascertainment.
531 While our findings were fairly consistent across these outcomes, we also identified differences
532 between objective and subjective health. Overall health is arguably what matters most to people, and
533 exploring factors associated with health indicators that transcend traditional disease-boundaries could
534 represent an effective strategy to increase longevity and improve healthy life expectancy. Previous
535 studies have often used composite scores of healthy lifestyles¹³ in which individual behaviours were

536 weighted equally, even though the strength of their respective associations with health might differ. In
537 the present study, we estimated associations separately for each lifestyle factor.

538

539 Nevertheless, our study has limitations. Cross-sectional analyses prevent examination of temporal
540 sequences. Reverse causality such as poor health and disability leading to unhealthy lifestyle, less
541 favourable socioeconomic outcomes, fewer social interactions or living in more deprived and polluted
542 areas needs to be considered, except for fixed characteristics such as sex and ethnicity. Associations
543 reported here cannot be interpreted as causal effects, given that several variables included in Model 3
544 could lie on the causal pathway of other variables. Model 3 might also be over-adjusted. Hence, we
545 have focussed on associations that are consistent across Model 1-3 in the main body of the text and
546 highlighted inconsistencies. As many explanatory variables were self-reported, we cannot exclude the
547 possibility that social desirability might have influenced responses. For example, previous research
548 has shown that up to half of participants who claim to never have had an alcoholic drink reported
549 alcohol intake during previous surveys⁷¹. For most participants, explanatory variables were measured
550 at a single time point, which could increase the potential for measurement error and did not allow us
551 to model changes across the life course. However, there is evidence that lifestyle is fairly stable in this
552 age group, with few people newly adopting a healthy lifestyle in middle age⁷². There might be some
553 misclassification in the reporting of medical illnesses. However, participants were asked to report
554 illnesses that had been diagnosed by a doctor and these were confirmed during a nurse-led interview.

555

556 **Generalisability**

557 The overall response rate of the UK Biobank was low (5.5%) and compared with non-responders,
558 participants were older, more likely to be female and lived in less deprived neighbourhoods.
559 Participants were also less likely to be obese, to smoke, to never drink or to drink alcohol daily and
560 they reported fewer health conditions compared with data from a nationally representative survey⁷³.
561 While the UK Biobank states that "valid assessment of exposure-disease relationships are nonetheless
562 widely generalizable and do not require participants to be representative of the population at large"⁷⁴,
563 concordant with the finding that there is little evidence of considerable bias due to non-participation
564 in epidemiological research⁷⁵, the magnitude of exposure-disease associations may depend on
565 prevalence of effect modifiers⁷⁶. A recent empirical investigation comparing the UK Biobank with
566 data from 18 prospective cohort studies with conventional response rates showed that the direction of
567 risk factor associations were similar, although with differences in magnitude⁷⁷.

568

569 **Implications**

570 Our findings, while partly confirmatory in nature, shed light on some of the behavioural, psychosocial
571 and environmental pathways to health and are thus relevant to public health policies aimed at
572 promoting health in later life. Public health and medicine could put greater focus on non-medical

573 factors such as loneliness, further encourage healthy lifestyle behaviours and weight management, and
574 examine efforts to improve the health outcomes of individuals with low income. Our findings support
575 the view that promoting individual responsibility for one's health (e.g. engaging in healthy lifestyle
576 behaviours) and policy-level commitments (e.g. reducing long-term exposure to environmental
577 pollutants⁷⁸ or improving neighbourhoods) both need to be considered in population health. While
578 associations between lifestyle and health might reflect a life-long commitment to healthy behaviours,
579 it is not too late to newly adopt healthy behaviours later in life⁷². Prevention, in addition to drug
580 discovery and disease treatment, should be the top priority for health policy.

581 **Acknowledgments**

582 JM acknowledges current studentship funding from the Biotechnology and Biological Sciences
583 Research Council (BBSRC) (ref: 2050702) and Eli Lilly and Company Limited in support of this
584 work. CML is part-funded by the National Institute for Health Research (NIHR) Biomedical Research
585 Centre at South London and Maudsley NHS Foundation Trust and Kings College London. The views
586 expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department
587 of Health and Social Care. This project made use of time on Rosalind HPC, funded by Guy's & St
588 Thomas' Hospital NHS Trust Biomedical Research Centre (GSTT-BRC), South London & Maudsley
589 NHS Trust Biomedical Research Centre (SLAM-BRC), and Faculty of Natural Mathematics &
590 Science (NMS) at King's College London.

591

592 **Authorship contributions**

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

598

599 **Competing interests**

600 JM receives studentship funding from the Biotechnology and Biological Sciences Research Council
601 (BBSRC) and Eli Lilly and Company Limited. CML is a member of the Scientific Advisory Board of
602 Myriad Neuroscience. CJR does not declare any potential conflict of interest.

603

604 **Ethics**

605 Ethical approval for the UK Biobank study has been granted by the National Information Governance
606 Board for Health and Social Care and the NHS North West Multicentre Research Ethics Committee
607 (11/NW/0382). No project-specific ethical approval is needed. Data access permission has been
608 granted under UK Biobank application 45514.

609

610 **Data sharing statement**

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

614 References

- 615 1 Partridge, L., Deelen, J. & Slagboom, P. E. Facing up to the global challenges of ageing.
616 *Nature* **561**, 45 (2018).
- 617 2 Roser, M. *Life expectancy*, 2013).
- 618 3 Kontis, V. *et al.* Future life expectancy in 35 industrialised countries: projections with a
619 Bayesian model ensemble. *The Lancet* **389**, 1323-1335 (2017).
- 620 4 Netuveli, G., Wiggins, R. D., Hildon, Z., Montgomery, S. M. & Blane, D. Quality of life at
621 older ages: evidence from the English longitudinal study of aging (wave 1). *Journal of*
622 *Epidemiology & Community Health* **60**, 357-363 (2006).
- 623 5 Van Hecke, O., Torrance, N. & Smith, B. Chronic pain epidemiology and its clinical
624 relevance. *British journal of anaesthesia* **111**, 13-18 (2013).
- 625 6 Brown, G. C. Living too long. *EMBO reports* **16**, 137-141 (2015).
- 626 7 Nybo, H. *et al.* Functional status and self-rated health in 2,262 nonagenarians: the Danish
627 1905 Cohort Survey. *Journal of the American Geriatrics Society* **49**, 601-609 (2001).
- 628 8 Olshansky, S. J. From Lifespan to Healthspan. *JAMA* (2018).
- 629 9 Carter, B. D. *et al.* Smoking and mortality—beyond established causes. *New England journal*
630 *of medicine* **372**, 631-640 (2015).
- 631 10 Lai, H. T. *et al.* Serial circulating omega 3 polyunsaturated fatty acids and healthy ageing
632 among older adults in the Cardiovascular Health Study: prospective cohort study. *BMJ* **363**,
633 k4067 (2018).
- 634 11 Room, R., Babor, T. & Rehm, J. Alcohol and public health. *The lancet* **365**, 519-530 (2005).
- 635 12 Cooper, R., Kuh, D. & Hardy, R. Objectively measured physical capability levels and
636 mortality: systematic review and meta-analysis. *Bmj* **341**, c4467 (2010).
- 637 13 Loeff, M. & Walach, H. The combined effects of healthy lifestyle behaviors on all cause
638 mortality: a systematic review and meta-analysis. *Preventive medicine* **55**, 163-170 (2012).
- 639 14 Holt-Lunstad, J., Smith, T. B., Baker, M., Harris, T. & Stephenson, D. Loneliness and social
640 isolation as risk factors for mortality: a meta-analytic review. *Perspectives on Psychological*
641 *Science* **10**, 227-237 (2015).
- 642 15 Smith, R. B. *et al.* Impact of London's road traffic air and noise pollution on birth weight:
643 retrospective population based cohort study. *bmj* **359**, j5299 (2017).
- 644 16 Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature*
645 **562**, 203 (2018).
- 646 17 Guggenheim, J. A. & Williams, C. Childhood febrile illness and the risk of myopia in UK
647 Biobank participants. *Eye* **30**, 608 (2016).
- 648 18 *The English indices of deprivation 2010*. (2011).

- 649 19 Elovainio, M. *et al.* Contribution of risk factors to excess mortality in isolated and lonely
650 individuals: an analysis of data from the UK Biobank cohort study. *The Lancet Public Health*
651 **2**, e260-e266 (2017).
- 652 20 IPAQ-Group. *IPAQ scoring protocol - International Physical Activity Questionnaire*,
653 <<https://sites.google.com/site/theipaq/scoring-protocol>> (
654 21 Beelen, R. *et al.* Development of NO₂ and NO_x land use regression models for estimating air
655 pollution exposure in 36 study areas in Europe–The ESCAPE project. *Atmospheric*
656 *Environment* **72**, 10-23 (2013).
- 657 22 Eeftens, M. *et al.* Development of land use regression models for PM_{2.5}, PM_{2.5} absorbance,
658 PM₁₀ and PM_{coarse} in 20 European study areas; results of the ESCAPE project.
659 *Environmental science & technology* **46**, 11195-11205 (2012).
- 660 23 Kephelopoulos, S. *et al.* Advances in the development of common noise assessment methods
661 in Europe: The CNOSSOS-EU framework for strategic environmental noise mapping.
662 *Science of the Total Environment* **482**, 400-410 (2014).
- 663 24 Morley, D. *et al.* International scale implementation of the CNOSSOS-EU road traffic noise
664 prediction model for epidemiological studies. *Environmental pollution* **206**, 332-341 (2015).
- 665 25 Vienneau, D., Schindler, C., Perez, L., Probst-Hensch, N. & Rösli, M. The relationship
666 between transportation noise exposure and ischemic heart disease: a meta-analysis.
667 *Environmental research* **138**, 372-380 (2015).
- 668 26 Department-for-Communities-and-Local-Government.
669 (https://data.gov.uk/dataset/land_use_statistics_generalised_land_use_database, 2007).
- 670 27 Maxwell, J. M. *et al.* Multifactorial disorders and polygenic risk scores: predicting common
671 diseases and the possibility of adverse selection in life and protection insurance. *Annals of*
672 *Actuarial Science*, 1-16 (2020).
- 673 28 Fell, D. B., Joseph, K., Armson, B. A. & Dodds, L. The impact of pregnancy on physical
674 activity level. *Maternal and child health journal* **13**, 597 (2009).
- 675 29 Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful
676 approach to multiple testing. *Journal of the Royal statistical society: series B*
677 *(Methodological)* **57**, 289-300 (1995).
- 678 30 Fox, J. & Monette, G. Generalized collinearity diagnostics. *Journal of the American*
679 *Statistical Association* **87**, 178-183 (1992).
- 680 31 Gelman, A. Scaling regression inputs by dividing by two standard deviations. *Statistics in*
681 *medicine* **27**, 2865-2873 (2008).
- 682 32 Cappuccio, F. P., D'Elia, L., Strazzullo, P. & Miller, M. A. Sleep duration and all-cause
683 mortality: a systematic review and meta-analysis of prospective studies. *Sleep* **33**, 585-592
684 (2010).

- 685 33 Cassidy, S., Chau, J. Y., Catt, M., Bauman, A. & Trenell, M. I. Cross-sectional study of diet,
686 physical activity, television viewing and sleep duration in 233 110 adults from the UK
687 Biobank; the behavioural phenotype of cardiovascular disease and type 2 diabetes. *BMJ open*
688 **6**, e010038 (2016).
- 689 34 WorldHealthOrganization. *Air quality guidelines: global update 2005: particulate matter,*
690 *ozone, nitrogen dioxide, and sulfur dioxide.* (World Health Organization, 2006).
- 691 35 Mayor, S. (British Medical Journal Publishing Group, 2018).
- 692 36 Cockerham, W. C., Sharp, K. & Wilcox, J. A. Aging and perceived health status. *Journal of*
693 *Gerontology* **38**, 349-355 (1983).
- 694 37 Ferraro, K. F. Self-ratings of health among the old and the old-old. *Journal of Health and*
695 *Social Behavior*, 377-383 (1980).
- 696 38 Wu, S. *et al.* The relationship between self-rated health and objective health status: a
697 population-based study. *BMC public health* **13**, 320 (2013).
- 698 39 Kelleher, C., Friel, S., Gabhainn, S. N. & Tay, J. B. Socio-demographic predictors of self-
699 rated health in the Republic of Ireland: findings from the National Survey on Lifestyle,
700 Attitudes and Nutrition, SLAN. *Social science & medicine* **57**, 477-486 (2003).
- 701 40 Sinclair, D. A. & LaPlante, M. D. *Lifespan: Why We Age—and Why We Don't Have To.*
702 (Atria Books, 2019).
- 703 41 Brouwer, W. B., van Exel, N. J. A. & Stolk, E. A. Acceptability of less than perfect health
704 states. *Social Science & Medicine* **60**, 237-246 (2005).
- 705 42 Collerton, J. *et al.* Health and disease in 85 year olds: baseline findings from the Newcastle
706 85+ cohort study. *Bmj* **339**, b4904 (2009).
- 707 43 Gorman, B. K. & Read, J. n. G. Gender disparities in adult health: an examination of three
708 measures of morbidity. *Journal of health and social behavior* **47**, 95-110 (2006).
- 709 44 McDonough, P. & Walters, V. Gender and health: reassessing patterns and explanations.
710 *Social science & medicine* **52**, 547-559 (2001).
- 711 45 Austad, S. N. & Bartke, A. Sex differences in longevity and in responses to anti-aging
712 interventions: a mini-review. *Gerontology* **62**, 40-46 (2016).
- 713 46 Phaswana-Mafuya, N. *et al.* Self-rated health and associated factors among older South
714 Africans: evidence from the study on global ageing and adult health. *Global Health Action* **6**,
715 19880 (2013).
- 716 47 Benyamini, Y., Blumstein, T., Lusky, A. & Modan, B. Gender differences in the self-rated
717 health–mortality association: Is it poor self-rated health that predicts mortality or excellent
718 self-rated health that predicts survival? *The gerontologist* **43**, 396-405 (2003).
- 719 48 Singh, L., Arokiasamy, P., Singh, P. K. & Rai, R. K. Determinants of gender differences in
720 self-rated health among older population: evidence from India. *Sage Open* **3**,
721 2158244013487914 (2013).

- 722 49 Jylhä, M., Guralnik, J. M., Ferrucci, L., Jokela, J. & Heikkinen, E. Is self-rated health
723 comparable across cultures and genders? *The Journals of Gerontology Series B:*
724 *Psychological Sciences and Social Sciences* **53**, S144-S152 (1998).
- 725 50 Gold, C. H., Malmberg, B., McClearn, G. E., Pedersen, N. L. & Berg, S. Gender and health: a
726 study of older unlike-sex twins. *The Journals of Gerontology Series B: Psychological*
727 *Sciences and Social Sciences* **57**, S168-S176 (2002).
- 728 51 Shibuya, K., Hashimoto, H. & Yano, E. Individual income, income distribution, and self rated
729 health in Japan: cross sectional analysis of nationally representative sample. *Bmj* **324**, 16
730 (2002).
- 731 52 Der, G., Macintyre, S., Ford, G., Hunt, K. & West, P. The relationship of household income
732 to a range of health measures in three age cohorts from the West of Scotland. *The European*
733 *Journal of Public Health* **9**, 271-277 (1999).
- 734 53 Godhwani, S., Jivraj, S., Marshall, A. & Bécaries, L. Comparing subjective and objective
735 neighbourhood deprivation and their association with health over time among older adults in
736 England. *Health & place* **55**, 51-58 (2019).
- 737 54 Rehkopf, D. H., Berkman, L. F., Coull, B. & Krieger, N. The non-linear risk of mortality by
738 income level in a healthy population: US National Health and Nutrition Examination Survey
739 mortality follow-up cohort, 1988–2001. *BMC Public Health* **8**, 383 (2008).
- 740 55 Davies, N. M., Dickson, M., Smith, G. D., Van Den Berg, G. J. & Windmeijer, F. The causal
741 effects of education on health outcomes in the UK Biobank. *Nature human behaviour* **2**, 117
742 (2018).
- 743 56 Ding, D., Rogers, K., van der Ploeg, H., Stamatakis, E. & Bauman, A. E. Traditional and
744 emerging lifestyle risk behaviors and all-cause mortality in middle-aged and older adults:
745 evidence from a large population-based Australian cohort. *PLoS medicine* **12** (2015).
- 746 57 Coyle, C. E. & Dugan, E. Social isolation, loneliness and health among older adults. *Journal*
747 *of aging and health* **24**, 1346-1363 (2012).
- 748 58 Perissinotto, C. M., Cenzer, I. S. & Covinsky, K. E. Loneliness in older persons: a predictor
749 of functional decline and death. *Archives of internal medicine* **172**, 1078-1084 (2012).
- 750 59 Warburton, D. E., Nicol, C. W. & Bredin, S. S. Health benefits of physical activity: the
751 evidence. *Cmaj* **174**, 801-809 (2006).
- 752 60 Arem, H. *et al.* Leisure time physical activity and mortality: a detailed pooled analysis of the
753 dose-response relationship. *JAMA internal medicine* **175**, 959-967 (2015).
- 754 61 Yu, S., Yarnell, J., Sweetnam, P. & Murray, L. What level of physical activity protects
755 against premature cardiovascular death? The Caerphilly study. *Heart* **89**, 502-506 (2003).
- 756 62 Lee, I., Paffenbarger, R. & Hennekens, C. Physical activity, physical fitness and longevity.
757 *Aging Clinical and Experimental Research* **9**, 2-11 (1997).

- 758 63 Oguma, Y., Sesso, H., Paffenbarger, R. & Lee, I. Physical activity and all cause mortality in
759 women: a review of the evidence. *British journal of sports medicine* **36**, 162-172 (2002).
- 760 64 Lee, I.-M. & Paffenbarger Jr, R. S. Associations of light, moderate, and vigorous intensity
761 physical activity with longevity: the Harvard Alumni Health Study. *American journal of*
762 *epidemiology* **151**, 293-299 (2000).
- 763 65 Piazza-Gardner, A. K. & Barry, A. E. Examining physical activity levels and alcohol
764 consumption: are people who drink more active? *American Journal of Health Promotion* **26**,
765 e95-e104 (2012).
- 766 66 Shaper, A. G., Wannamethee, G. & Walker, M. Alcohol and mortality in British men:
767 explaining the U-shaped curve. *The Lancet* **332**, 1267-1273 (1988).
- 768 67 Stockwell, T. *et al.* Do “moderate” drinkers have reduced mortality risk? A systematic review
769 and meta-analysis of alcohol consumption and all-cause mortality. *Journal of studies on*
770 *alcohol and drugs* **77**, 185-198 (2016).
- 771 68 Fat, L. N., Cable, N., Marmot, M. & Shelton, N. Persistent long-standing illness and non-
772 drinking over time, implications for the use of lifetime abstainers as a control group. *J*
773 *Epidemiol Community Health* **68**, 71-77 (2014).
- 774 69 Cai, Y. *et al.* Road traffic noise, air pollution and incident cardiovascular disease: a joint
775 analysis of the HUNT, EPIC-Oxford and UK Biobank cohorts. *Environment international*
776 **114**, 191-201 (2018).
- 777 70 Twohig-Bennett, C. & Jones, A. The health benefits of the great outdoors: A systematic
778 review and meta-analysis of greenspace exposure and health outcomes. *Environmental*
779 *research* **166**, 628-637 (2018).
- 780 71 Rehm, J. *et al.* Are lifetime abstainers the best control group in alcohol epidemiology? On the
781 stability and validity of reported lifetime abstention. *American journal of epidemiology* **168**,
782 866-871 (2008).
- 783 72 King, D. E., Mainous III, A. G. & Geesey, M. E. Turning back the clock: adopting a healthy
784 lifestyle in middle age. *The American journal of medicine* **120**, 598-603 (2007).
- 785 73 Fry, A. *et al.* Comparison of sociodemographic and health-related characteristics of UK
786 Biobank participants with those of the general population. *American Journal of Epidemiology*
787 **186**, 1026-1034 (2017).
- 788 74 UK-Biobank. [http://www.ukbiobank.ac.uk/wp-content/uploads/2017/03/access-matters-](http://www.ukbiobank.ac.uk/wp-content/uploads/2017/03/access-matters-representativeness-1.pdf)
789 [representativeness-1.pdf](http://www.ukbiobank.ac.uk/wp-content/uploads/2017/03/access-matters-representativeness-1.pdf), <[http://www.ukbiobank.ac.uk/wp-content/uploads/2017/03/access-](http://www.ukbiobank.ac.uk/wp-content/uploads/2017/03/access-matters-representativeness-1.pdf)
790 [matters-representativeness-1.pdf](http://www.ukbiobank.ac.uk/wp-content/uploads/2017/03/access-matters-representativeness-1.pdf)> (
- 791 75 Galea, S. & Tracy, M. Participation rates in epidemiologic studies. *Annals of epidemiology*
792 **17**, 643-653 (2007).
- 793 76 Keyes, K. M. & Westreich, D. UK Biobank, big data, and the consequences of non-
794 representativeness. *The Lancet* **393**, 1297 (2019).

- 795 77 Battya, G. D., Galec, C. R., Kivimäkia, M., Dearyd, I. J. & Belle, S. Generalisability of
796 Results from UK Biobank: Comparison With a Pooling of 18 Cohort Studies.
797 78 Holgate, S. T. ‘Every breath we take: the lifelong impact of air pollution’—a call for action.
798 *Clinical Medicine* **17**, 8-12 (2017).

Tables

Table 1. Sociodemographic characteristics and psychosocial factors associated with health status									
Term	Model 1			Model 2			Model 3		
	OR	Bonferroni-corrected CI		OR	Bonferroni-corrected CI		OR	Bonferroni-corrected CI	
<i>Sociodemographic characteristics</i>									
Household income¹									
Very low	0.5161	0.4978	0.5351	0.6298	0.6066	0.6538	0.7447	0.7148	0.7759
Low	0.7548	0.7288	0.7818	0.8728	0.8419	0.9048	0.9196	0.8864	0.9541
Middle	Ref	–	–	Ref	–	–	Ref	–	–
High	1.2084	1.1626	1.2561	1.1016	1.0591	1.1458	1.0510	1.0096	1.0941
Very high	1.2867	1.2088	1.3703	1.1483	1.0778	1.2239	1.0472	0.9812	1.1181
Sex									
Female	Ref	–	–	Ref	–	–	Ref	–	–
Male	0.8642	0.8428	0.8862	NA	NA	NA	0.8847	0.8610	0.9090
Age	0.9462	0.9446	0.9478	NA	NA	NA	0.9528	0.9510	0.9546
Multiple deprivation	0.9912	0.9903	0.9921	0.9885	0.9876	0.9894	0.9952	0.9941	0.9963
Ethnicity									
White	Ref	–	–	Ref	–	–	Ref	–	–
Mixed-race	1.2991	1.0925	1.5524	0.9743	0.8163	1.1684	1.1037	0.9221	1.3273
Asian	1.0196	0.9213	1.1298	0.8560	0.7714	0.9511	1.0188	0.9149	1.1359
Black	1.2994	1.1612	1.4571	0.9831	0.8764	1.1050	1.3125	1.1657	1.4805
Chinese	2.1528	1.6132	2.9286	1.7413	1.2994	2.3777	1.8339	1.3646	2.5106
Other	1.2077	1.0369	1.4115	0.9952	0.8519	1.1667	1.2570	1.0727	1.4779
Highest qualification									
None	Ref	–	–	Ref	–	–	Ref	–	–
O levels/GCSEs/CSEs	1.6519	1.5859	1.7207	1.2266	1.1757	1.2796	1.0315	0.9870	1.0779
A levels/NVQ/HND/HNC ²	1.5125	1.4508	1.5769	1.2102	1.1595	1.2631	0.9879	0.9445	1.0332
Degree	1.7949	1.7257	1.8667	1.3373	1.2838	1.3929	0.9558	0.9137	0.9997
<i>Psychosocial factors</i>									
Loneliness									
Not lonely	Ref	–	–	Ref	–	–	Ref	–	–
Lonely	0.7283	0.6915	0.7672	0.6947	0.6588	0.7326	0.8129	0.7697	0.8587
Social isolation									
Not isolated	Ref	–	–	Ref	–	–	Ref	–	–
Isolated	0.7544	0.7226	0.7877	0.7542	0.7218	0.7882	0.9496	0.9067	0.9947

Note: Bonferroni-adjusted (~99.9%) confidence intervals. 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; NA = not applicable. ¹Annual 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). ²also includes 'other professional qualifications'.

Model 1 – only individual explanatory variables.

Model 2 – adjusted for age and sex.

Model 3 – all explanatory variables.

Table 2. Lifestyle factors associated with health status									
Term	Model 1			Model 2			Model 3		
	OR	Bonferroni-corrected CI		OR	Bonferroni-corrected CI		OR	Bonferroni-corrected CI	
Sleep duration (hours/day)	0.9574	0.9460	0.9688	0.9817	0.9699	0.9937	0.9748	0.9631	0.9868
Physical activity (days/week) ¹									
Walking	1.0099	1.0034	1.0164	1.0258	1.0190	1.0326	1.0105	1.0032	1.0178
Moderate activity	1.0095	1.0040	1.0149	1.0253	1.0197	1.0310	1.0032	0.9964	1.0100
Vigorous activity	1.0571	1.0502	1.0642	1.0502	1.0431	1.0573	1.0319	1.0238	1.0400
Stair climbing frequency									
None	Ref	–	–	Ref	–	–	Ref	–	–
1-5/day	1.3009	1.2362	1.3688	1.0980	1.0422	1.1568	1.0704	1.0148	1.1289
6-10/day	1.5529	1.4813	1.6278	1.3354	1.2725	1.4013	1.1756	1.1188	1.2352
11-15/day	1.6271	1.5457	1.7128	1.3923	1.3211	1.4672	1.1714	1.1098	1.2363
16-20/day	1.6494	1.5530	1.7520	1.4123	1.3280	1.5020	1.1650	1.0937	1.2410
20+/day	1.7494	1.6390	1.8675	1.4402	1.3474	1.5396	1.1881	1.1097	1.2723
Alcohol intake frequency									
Never	0.5651	0.5364	0.5954	0.5849	0.5545	0.6171	0.6285	0.5946	0.6644
Special occasions	0.7015	0.6705	0.7341	0.7109	0.6785	0.7448	0.7792	0.7429	0.8174
1-3/month	0.9082	0.8678	0.9505	0.8730	0.8333	0.9146	0.9129	0.8708	0.9570
1-2/week	Ref	–	–	Ref	–	–	Ref	–	–
3-4/week	1.0652	1.0271	1.1047	1.1059	1.0656	1.1478	1.0590	1.0198	1.0998
Daily/almost daily	0.9293	0.8957	0.9643	1.0627	1.0232	1.1037	1.0334	0.9939	1.0745
BMI (kg/m ²)	0.9561	0.9536	0.9587	0.9580	0.9554	0.9606	0.9683	0.9656	0.9711
Smoking status									
Never	Ref	–	–	Ref	–	–	Ref	–	–
Former	0.6977	0.6792	0.7168	0.7879	0.7663	0.8100	0.7922	0.7698	0.8152
Current	0.7309	0.7003	0.7629	0.6850	0.6556	0.7158	0.7518	0.7183	0.7869

Note: Bonferroni-adjusted (~99.9%) confidence intervals. OR = odds ratio; CI = confidence interval; BMI = body mass index. ¹number 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 3. Environmental exposures associated with health status

Term	Model 1			Model 2			Model 3		
	OR	Bonferroni-corrected CI		OR	Bonferroni-corrected CI		OR	Bonferroni-corrected CI	
PM_{2.5}	0.9640	0.9525	0.9757	0.9357	0.9243	0.9473	0.9656	0.9411	0.9908
PM₁₀	0.9917	0.9851	0.9983	0.9842	0.9775	0.9909	1.0019	0.9935	1.0104
NO₂	0.9972	0.9955	0.9988	0.9929	0.9912	0.9945	1.0036	0.9993	1.0078
L_{den}	0.9967	0.9938	0.9996	0.9945	0.9916	0.9975	0.9981	0.9945	1.0017
Greenspace 1000m	1.0001	0.9995	1.0006	1.0013	1.0007	1.0019	0.9995	0.9985	1.0004

Note: Bonferroni-adjusted (~99.9%) confidence intervals. OR = odds ratio; CI = confidence interval; PM = particulate matter; NO₂ = 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.

Figures

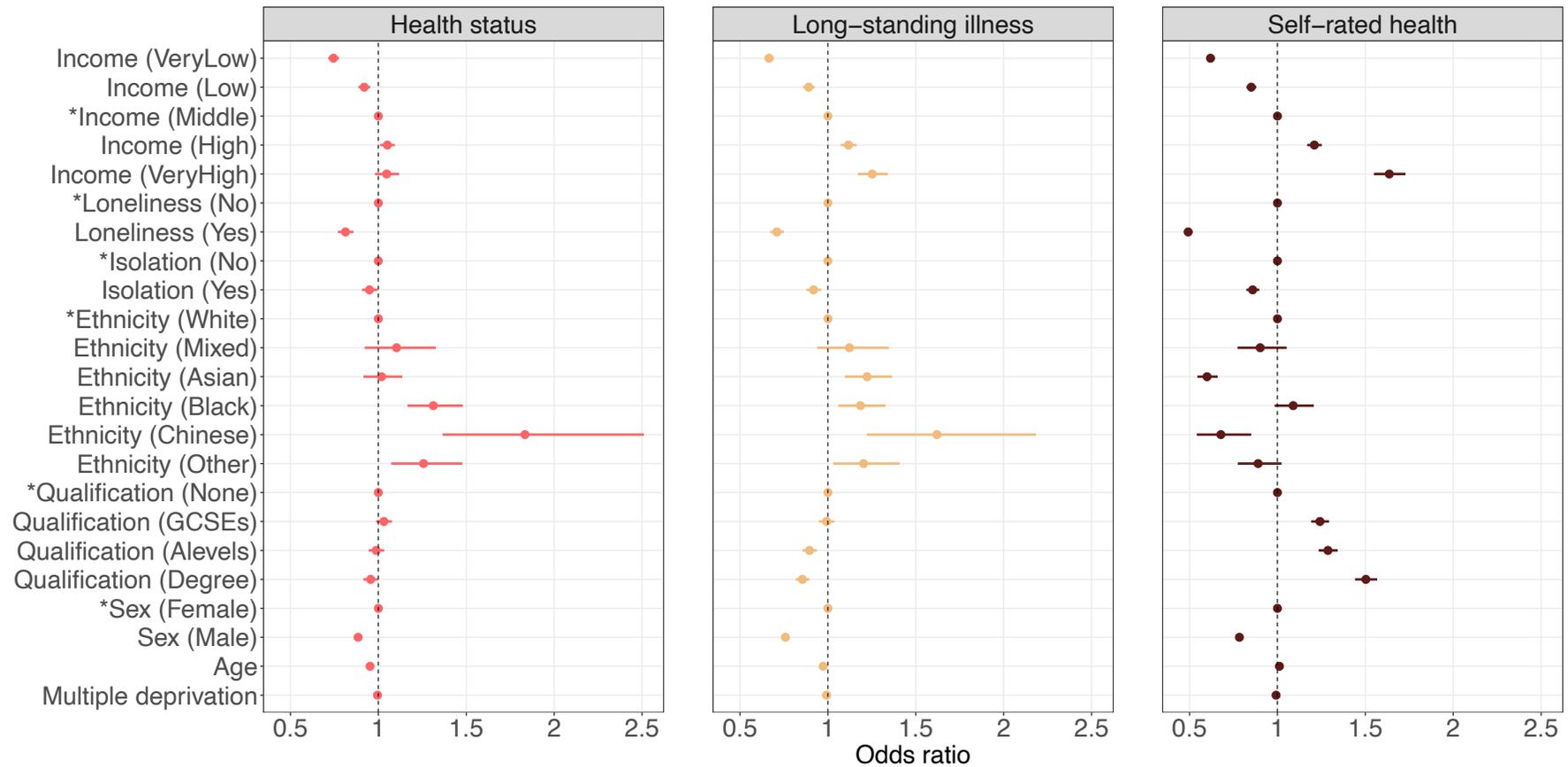


Figure 1. Sociodemographic characteristics and psychosocial factors associated with health indicators. Confidence interval plot (odds ratio \pm Bonferroni-adjusted (~99.9%) confidence intervals) for Model 3 (i.e. including all explanatory variables). GCSEs = general certificate of secondary education. *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'.

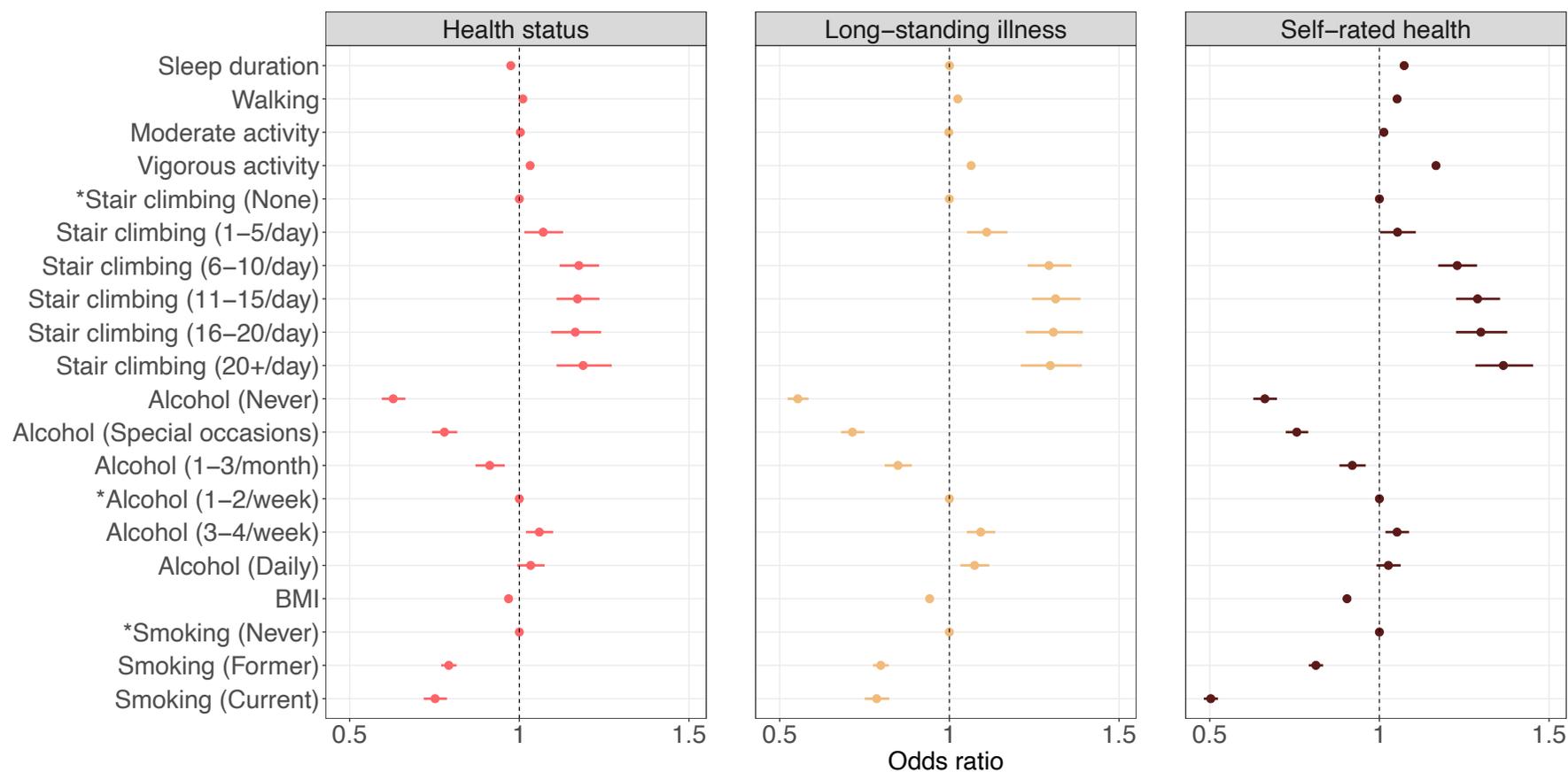


Figure 2. Lifestyle factors associated with health indicators. Confidence interval plot (odds ratio \pm Bonferroni-adjusted (~99.9%) confidence intervals) for Model 3 (i.e. including all explanatory variables). BMI = body mass index. *Indicates reference group for categorical explanatory variables. For 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.

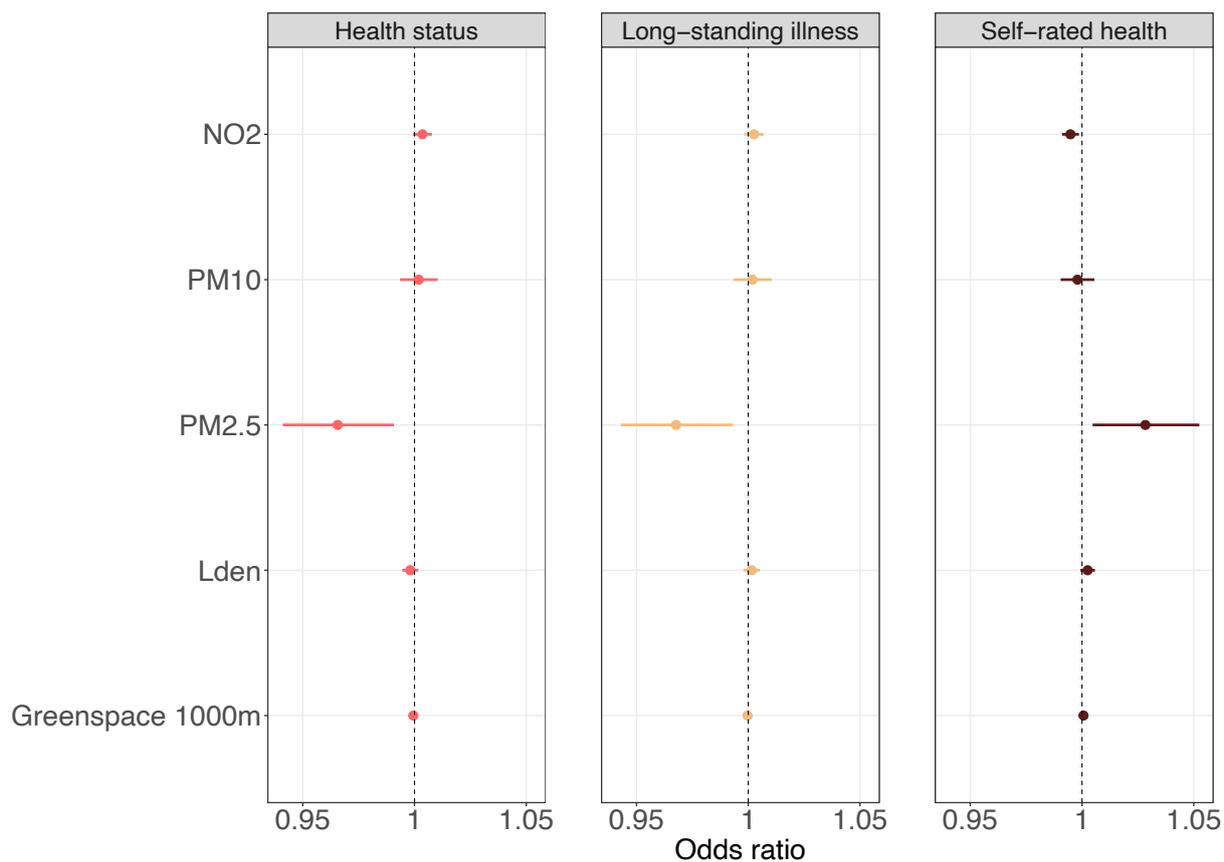


Figure 3. Environmental exposures associated with health indicators. Confidence interval plot (odds ratio \pm Bonferroni-adjusted (\sim 99.9%) confidence intervals) for Model 3 (i.e. including all explanatory variables). PM = particulate matter; NO₂ = nitrogen dioxide; L_{den} = day-evening-night noise 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.

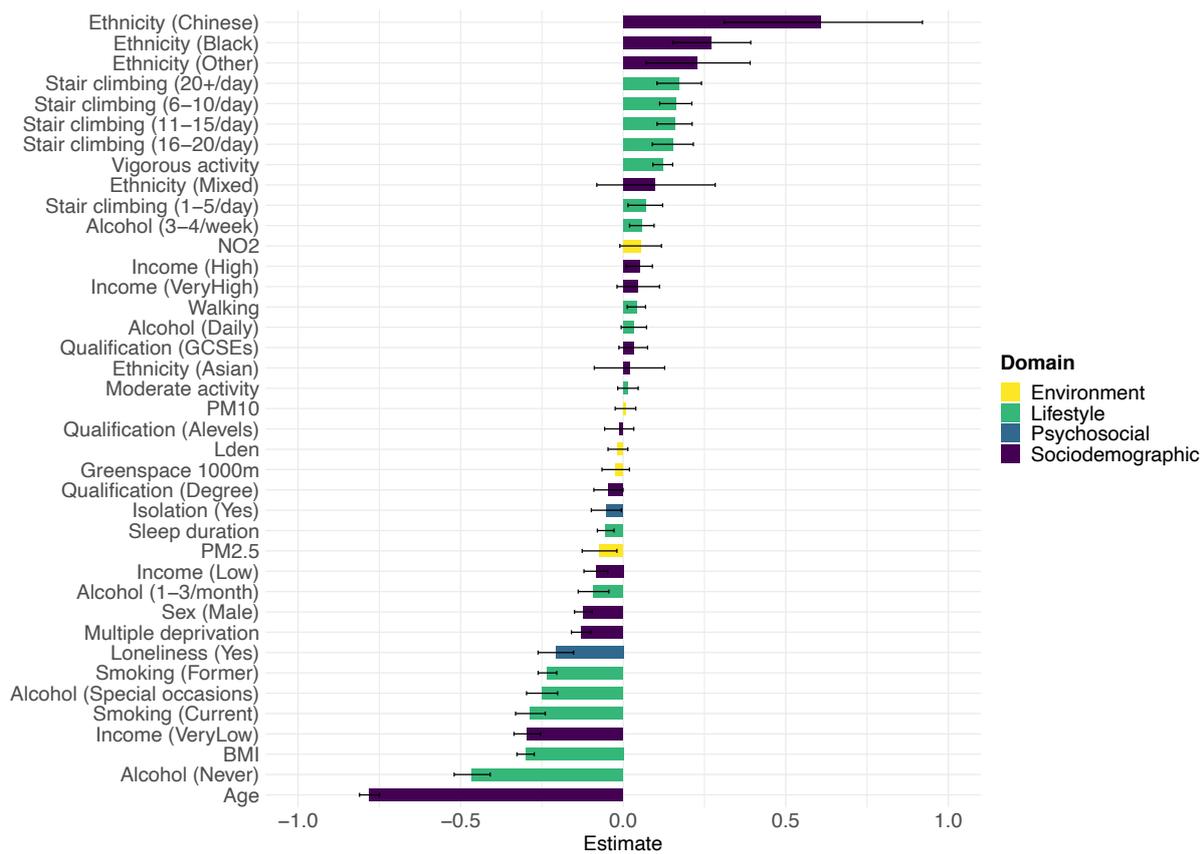


Figure 4. Horizontal bar plot for health status. β estimates and Bonferroni-adjusted ($\sim 99.9\%$) confidence intervals. Model 3 regression coefficients rescaled to have a mean equal to zero and, for numeric variables with more than two values, divided by two standard deviations.