

Title: Cross-classification between self-rated health and health status: longitudinal analyses of all-cause mortality and leading causes of death in the UK

Short title: Health cross-classification and mortality

Authors: Julian Mutz^{1*} and Cathryn M. Lewis^{1,2}

Affiliations:

1. Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology & Neuroscience, King's College London, London, United Kingdom
2. Department of Medical and Molecular Genetics, Faculty of Life Sciences & Medicine, King's College London, London, United Kingdom

***Corresponding author:**

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

1 Abstract

2

3 **Background:** Risk stratification is an important public health priority that is central to clinical
4 decision making and resource allocation. The aim of the present study was to examine how different
5 combinations of self-rated and objective health status predict (i) all-cause mortality and (ii) cause-
6 specific mortality from leading causes of death in the UK.

7 **Methods:** The UK Biobank study recruited >500,000 participants, aged 37-73, between 2006–2010.
8 The health cross-classification examined incorporated self-rated health (poor, fair, good or excellent)
9 and health status derived from medical history and current disease status, including 81 cancer and 443
10 non-cancer illnesses. We examined all-cause mortality and six specific causes of death: ischaemic
11 heart disease, cerebrovascular disease, influenza and pneumonia, dementia and Alzheimer’s disease,
12 chronic lower respiratory disease and malignant neoplasm.

13 **Results:** Analyses included >370,000 middle-aged and older adults with a median follow-up of 11.75
14 (IQR = 1.4) years, yielding 4,320,270 person years of follow-up. Compared to excellent self-rated
15 health and favourable health status, all other levels of the health cross-classification were associated
16 with a greater risk of mortality, with hazard ratios ranging from 1.22 (95% CI 1.15-1.29, $p_{\text{Bonf.}} <$
17 0.001) for good self-rated health and favourable health status to 7.14 (95% CI 6.70-7.60, $p_{\text{Bonf.}} <$
18 0.001) for poor self-rated health and unfavourable health status.

19 **Conclusions:** Our findings highlight that self-rated health captures additional health-related
20 information and should be more widely assessed across settings. The cross-classification between
21 health status and self-rated health represents a straightforward metric for risk stratification, with
22 applications to population health, clinical decision making and resource allocation.

23

24 **Keywords:** Self-rated health; Health status; Mortality; UK Biobank; Leading causes of death

25 Introduction

26

27 Self-rated health is used extensively in epidemiological and public health research, and a plethora of
28 studies have found it to be predictive of morbidity and mortality¹. It constitutes a single item measure
29 of subjective health status that likely encompasses biological, psychological and social dimensions.

30

31 Despite its simplicity, self-rated health is correlated with objective assessments of health². For
32 example, research in more than 16,000 Chinese residents aged 18-80 found a higher prevalence of
33 diseases and abnormalities in laboratory tests in individuals with poor self-rated health³. A study of
34 1,322 community-dwelling elderly aged 60 or older who participated in the Bambui Cohort Study of
35 Aging in Brazil examined how well self-rated health predicted 10-year mortality, compared to a
36 comprehensive health score derived from objective clinical measures⁴. Individuals with poor self-
37 rated health had a two-fold increased risk of death during the follow-up period, compared to
38 individuals who rated their health as good or excellent. Self-rated health was comparable to the
39 comprehensive health score in predicting mortality and remained predictive of mortality after
40 adjustment for the health score, suggesting that it captures additional health-related information.

41

42 Electronic health records and other patient data that are available to health care providers could be
43 used to derive a measure of overall health status. Self-rated health can be assessed using a single-item
44 question and is for instance included in census questionnaires⁵ but not, to our knowledge, routinely
45 collected during primary care registrations or visits. The cross-classification between self-rated health
46 and objective health status based on medical history or clinical measures could represent a readily
47 available metric for risk stratification. The identification of at-risk populations is an important public
48 health priority that is central to clinical decision making and resource allocation.

49

50 The aim of the present study was to examine the relative contributions of self-rated health and health
51 status based on medical history and current disease status to predict mortality in middle-aged and
52 older adults in the UK Biobank. We created a matrix of the cross-classification between self-rated
53 health and health status and examined how different combinations of self-rated and objective health
54 status predict (i) all-cause mortality and (ii) cause-specific mortality from leading causes of death in
55 the UK⁶ during a follow-up period of approximately 12 years.

56

57 We hypothesised that individuals with concordant self-rated health and health status would
58 demonstrate the longest (favourable health status and good or excellent self-rated health) and shortest
59 (unfavourable health status and poor or fair self-rated health) survival times, respectively. Individuals
60 with discordant self-rated health and health status were predicted to demonstrate intermediate survival
61 times, although no predictions were made about the precise order of survival times ([Supplement 1](#)).

62 The aim of comparing these survival times was to explore whether the discordant categories would
63 provide additional information about the relative importance of self-rated health and health status with
64 respect to future health outcomes.

65 **Methods**

66

67 **Study population**

68 The UK Biobank is a prospective study of > 500,000 UK residents aged 37–73 at baseline, recruited
69 between 2006–2010. Details of the study rationale and design have been reported elsewhere⁷. Briefly,
70 individuals registered with the UK National Health Service (NHS) and living within a 25-mile (~40
71 km) radius of one of 22 assessment centres were invited to participate (9,238,453 postal invitations
72 sent). At the baseline assessment, participants completed electronic questionnaires and nurse-led
73 interviews to provide data on sociodemographic characteristics, health behaviours and their medical
74 history. Linked hospital inpatient records are available for most participants and these data have been
75 linked to death registries.

76

77 **Exposures**

78 Data on 81 cancer and 443 non-cancer illnesses (past and current) were ascertained through
79 touchscreen self-report questionnaire and confirmed during a verbal interview by a trained nurse. In
80 order to provide a single health indicator (“health status”) based on a previously defined algorithm,
81 we used a classification developed by the Reinsurance Group of America (RGA) in which an
82 experienced underwriter classified each illness according to whether it was “likely acceptable for
83 standard life insurance”⁸. Participants were thus classified as having favourable or unfavourable
84 health status based on their reported cancer and non-cancer illnesses. Details of this classification
85 have been reported previously^{8,9}.

86

87 Participants’ self-rated health was assessed using the question “In general how would you rate your
88 overall health?”. Response options included “Poor”, “Fair”, “Good” and “Excellent”.

89

90 We derived a measure of the cross-classification between health status and self-rated health.

91 Individuals with missing data or who responded “prefer not to answer” or “do not know” to the self-
92 rated health question were excluded.

93

94 **Ascertainment of mortality**

95 Our primary outcome was all-cause mortality, i.e., death from any cause. All-cause mortality
96 represents a standard index used in clinical decision making that is easily assessed and concrete¹⁰. The
97 date of death was obtained through linkage with national death registries from NHS Digital for
98 participants in England and Wales and from the NHS Central Register for participants in Scotland.
99 The censoring date for mortality was 30 November 2020. The most recent death was recorded for 18
100 December 2020, although data were not complete for December 2020.

101

102 We also examined cause-specific mortality for leading causes of death in the UK⁶. The primary cause
103 of death was recorded based on the International Classification of Diseases (10th revision). The
104 following outcomes were examined: ischaemic heart diseases (I20-I25), cerebrovascular diseases
105 (I60-I69), influenza and pneumonia (J09-J18), dementia and Alzheimer’s disease (F01, F03 and G30),
106 chronic lower respiratory diseases (J40-J47) and malignant neoplasms (C00-C97). For each cause-
107 specific death, individuals who died of other causes were censored at their age at death.

108

109 **Covariates**

110 Potential confounders of the association between self-rated health or health status and all-cause or
111 cause-specific mortality were identified from the baseline assessment data: sociodemographic factors

112 (sex, ethnicity [6 levels], highest educational/professional qualification [4 levels]¹¹ and annual
113 household income [5 levels]), lifestyle (smoking [3 levels], alcohol intake frequency [6 levels],
114 physical activity [number of days per week spent walking, engaging in moderate-intensity physical
115 activity or engaging in vigorous-intensity physical activity for ≥ 10 minutes continuously]). All
116 covariates were assumed to be fixed.

117

118 **Exclusion criteria**

119 Participants for whom their genetic sex, inferred from the genotype information on the Y and X
120 chromosomes, and self-reported sex did not match were excluded. Individuals with missing data or
121 who responded “do not know” or “prefer not to answer” to any of the assessed covariates were also
122 excluded from analyses.

123

124 **Statistical analyses**

125 All analyses were pre-specified prior to inspection of the data (preregistration: osf.io/qvm39) and
126 algorithms were tested on simulated data. Statistical analyses were conducted using R (version 3.6.0).

127

128 Characteristics of the full and analytical sample were summarised using means and standard
129 deviations or counts and percentages. The total number of self-reported diseases and the frequencies
130 of illnesses by disease group were summarised for each level of the cross-classification between
131 health status and self-rated health.

132

133 We present the number of individuals who died during follow-up of any cause (all-cause mortality)
134 and of specific causes (cause-specific mortality). The minimum number of observed deaths was set a
135 priori at 20 for each level of our primary exposure. This criterion was based on a previous UK
136 Biobank study of 5-year mortality¹². Finally, we calculated person-years of follow-up and the median
137 duration of follow-up of censored individuals.

138

139 Unadjusted survival probabilities by health status, self-rated health and health cross-classification
140 were estimated non-parametrically from observed survival times using the Kaplan-Meier (KM)
141 method¹³. We present KM survival curves and *p*-values from log-rank tests. Hazard ratios (HRs) and
142 95% confidence intervals were estimated using Cox proportional hazards models¹⁴ to examine
143 associations between the health cross-classification and mortality adjusted for potential confounders.
144 Age (in years) was used as the underlying time axis, conditional on living to age 40. We fitted a
145 sequence of three models: Model 1 – unadjusted; Model 2 – adjusted for sociodemographic
146 characteristics; Model 3 – additionally adjusted for lifestyle factors. We present plots of the estimated
147 survival probability and cumulative hazards. Continuous covariates were fixed at their mean value
148 while categorical covariates were fixed at the largest group. When presenting results in tables and
149 figures, the levels of the health cross-classification are shown in ascending order by their HR for all-
150 cause mortality from the results of Model 1. Scaled Schoenfeld residuals and Log(-log(survival))
151 curves as a function of time were examined to assess the assumption of proportional hazards.
152 Martingale residuals were visually inspected to assess the assumption of log-linearity for continuous
153 covariates.

154

155 Adjusted *p*-values were calculated using the `p.adjust()` command in R to account for multiple testing
156 across models, separately for each outcome. *P*-values were corrected for 21 tests (three models \times
157 seven estimated parameters). Two methods were used: (1) Bonferroni and (2) Benjamini &
158 Hochberg¹⁵, all two-tailed, with $\alpha = .05$ and false discovery rate of 5%, respectively.

159

160 **Sensitivity analysis**

161 In a sensitivity analysis we examined a simplified health cross-classification in which fair and poor
162 self-rated health and good and excellent were merged prior to analysis ([Supplement 1](#)). Self-rated
163 health levels have been combined previously, although with 5 initial response options¹⁶.

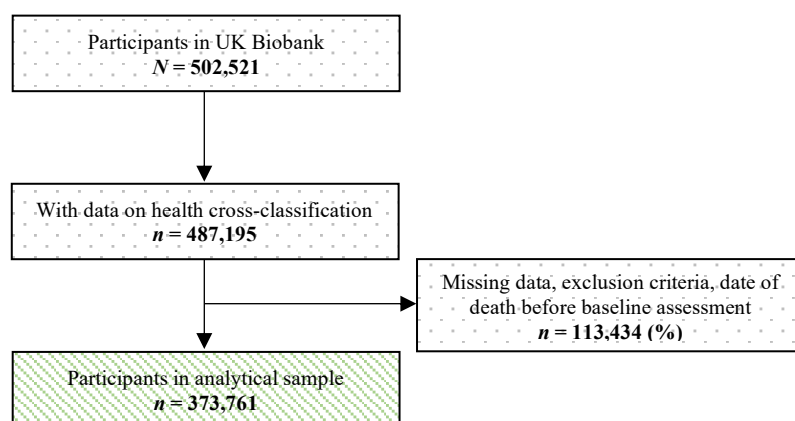
164 Results

165

166 Sample characteristics

167 Of the 502,521 UK Biobank participants, 487,195 (96.95%) had complete data on health cross-
 168 classification. We retained an analytical sample of 373,761 participants after removing individuals
 169 with missing data on covariates ($n = 118,190$), who did not meet our inclusion criteria ($n = 372$) or
 170 whose recorded date of death was before or on the same day as the baseline assessment ($n = 495$)
 171 (Figure 1).

172



173

174

175 **Figure 1.** Flowchart of study sample.

176

177 Descriptive statistics of the full and analytical samples are presented in [Supplement Table 1](#). The
 178 mean age at baseline assessment of participants included in our analytical sample was 56.02 (SD =
 179 8.08) years and 51.8% of these participants were female. We classified 117,212 (31.36%) participants
 180 as having an unfavourable health status and 256,549 (68.64%) participants as having a favourable
 181 health status. Most participants ($n = 219,628$, 58.76%) had good self-rated health and 14,185
 182 (3.80%), 73,138 (19.57%) and 66,810 (17.88%) participants had poor, fair and excellent self-rated
 183 health, respectively ([Table 1](#)).

184

Table 1. Cross-classification of health status and self-rated health

Health status	Self-rated health			
	Poor $N = 14185$	Fair $n = 73138$	Good $n = 219628$	Excellent $n = 66810$
Favourable $n = 256549$	4004 (1.07%)	37863 (10.13%)	159173 (42.59%)	55509 (14.85%)
Unfavourable $n = 117212$	10181 (2.72%)	35275 (9.44%)	60455 (16.17%)	11301 (3.02%)

Note: Percentages in cells are based on complete analytical sample of $n = 373,761$.

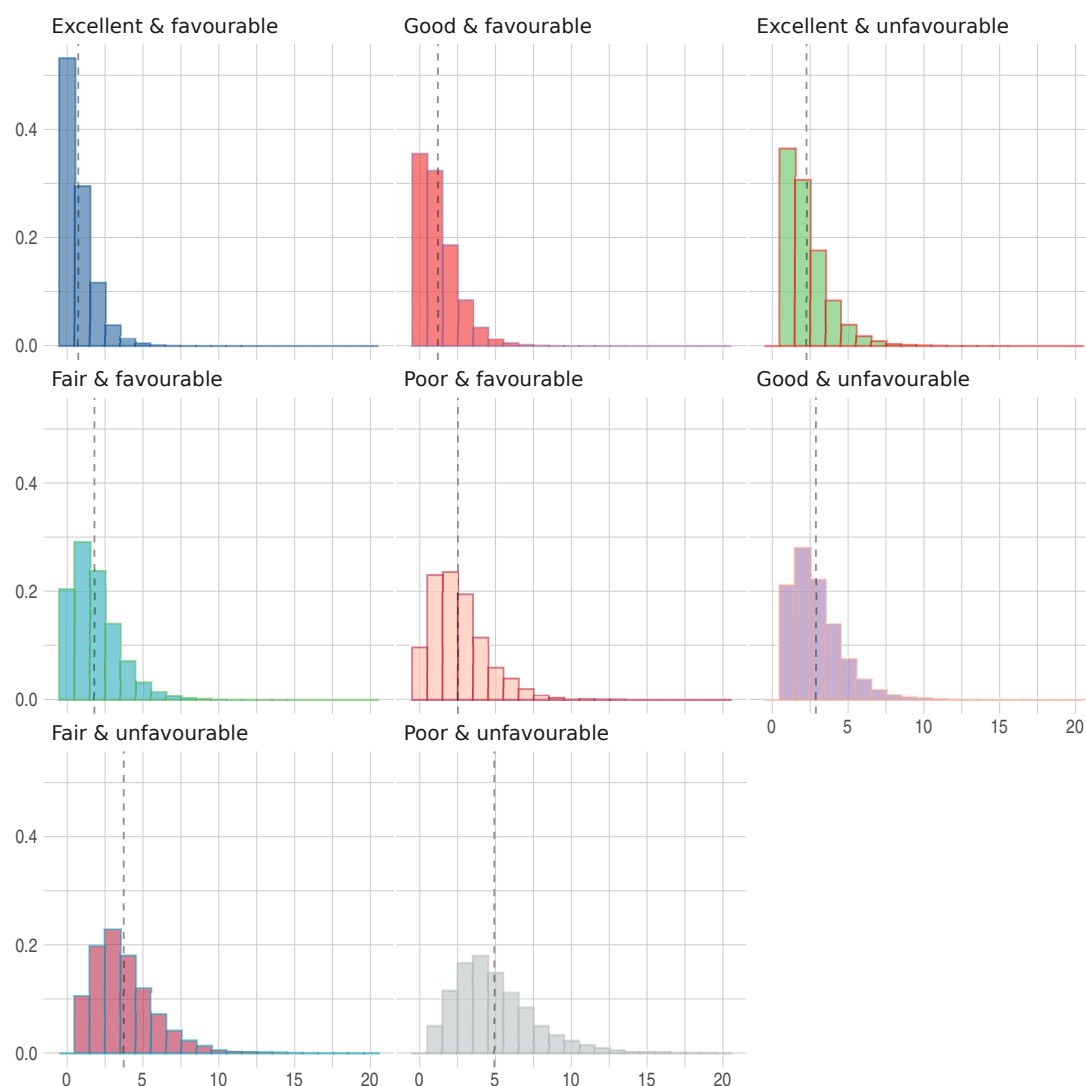
185

186 The average number of illnesses by health cross-classification ranged from 0.73 (SD = 0.99) in
 187 individuals with excellent self-rated health and favourable health status to 4.93 (SD = 2.70) in
 188 individuals with poor self-rated health and unfavourable health status ([Table 2](#)). The distributions of
 189 the number of illnesses by health cross-classification are presented in [Figure 2](#) and show that there
 190 were substantial differences between groups. The proportions of individuals with at least one cancer
 191 or non-cancer illness within several broader illness groups are presented by health cross classification
 192 in [Supplementary Figure 1](#) and [Supplementary Figure 2](#).

Table 2. Number of illnesses by health cross-classification

Self-rated health	Health status	Mean	SD
Excellent	Favourable	0.73	0.99
Good	Favourable	1.19	1.27
Excellent	Unfavourable	2.26	1.43
Fair	Favourable	1.80	1.54
Good	Unfavourable	2.88	1.70
Poor	Favourable	2.52	1.81
Fair	Unfavourable	3.73	2.09
Poor	Unfavourable	4.93	2.70

Note: SD = standard deviation.



193
 194 **Figure 2.** Number of illnesses by health cross-classification. 9 observations were above graph
 195 maximum of 20 illnesses (5 poor & unfavourable, 3 fair & unfavourable and 1 excellent &
 196 unfavourable).

197
 198 The median follow-up was 11.75 (IQR = 1.4) years, yielding 4,320,270 person years of follow-up.
 199 The median follow-up of censored individuals was 11.81 (IQR = 1.35) years and the potential
 200 median survival time, calculated using the reverse Kaplan-Meier method, was 11.819 days (95% CI
 201 = 11.814-11.822). We observed 21,980 (5.88%) deaths from any cause. Deaths observed for specific
 202 causes were 2,432 (0.65%) for ischaemic heart disease, 908 (0.24%) for cerebrovascular disease, 409

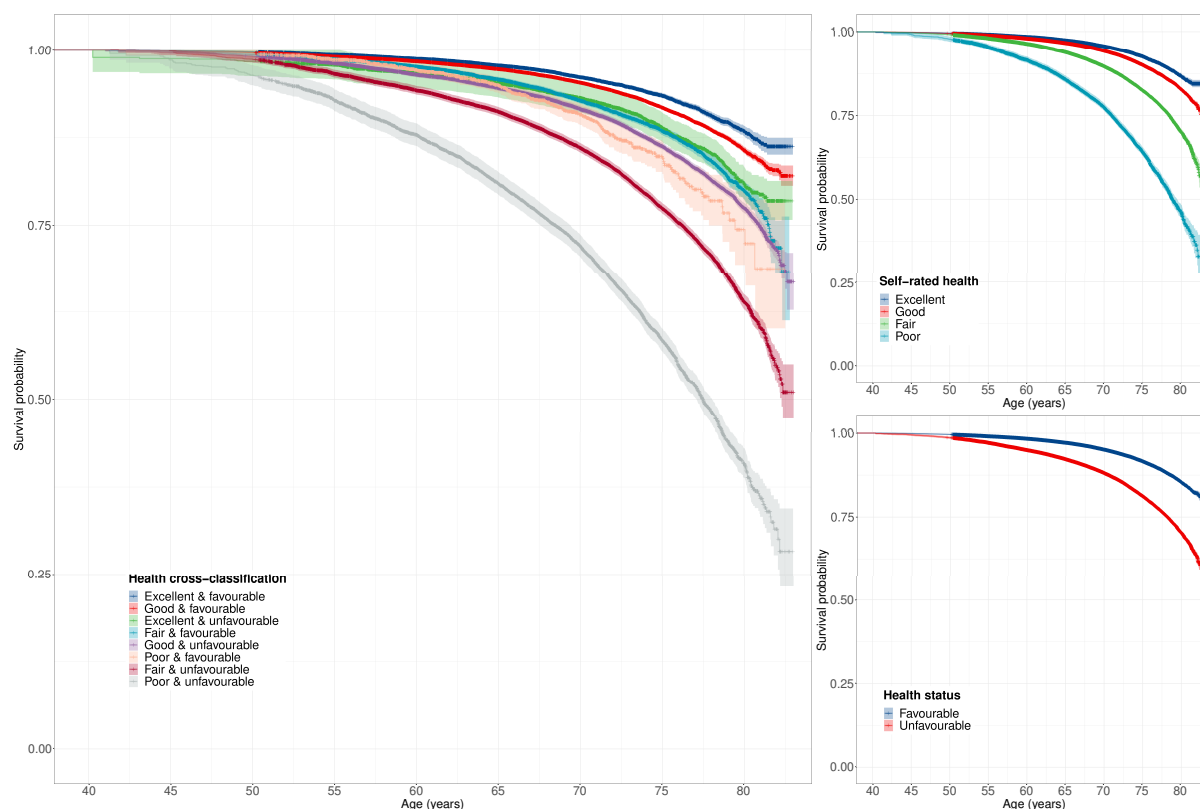
203 (0.11%) for influenza and pneumonia, 701 (0.19%) for dementia and Alzheimer's disease, 699
204 (0.19%) for chronic lower respiratory disease and 11,171 (2.99%) for malignant neoplasms.

205

206 All-cause mortality

207 Kaplan-Meier survival probabilities for all-cause mortality are presented in Figure 3. Favourable
208 health status and better self-rated health were associated with longer survival times (log-rank test p -
209 values < 0.001). Different levels of the cross-classification between health status and self-rated
210 health were associated with varying survival times (log-rank test $p < 0.001$). We found no evidence
211 that survival probabilities differed by year of attending the baseline assessment ($p = 0.15$).

212



213

214

215 **Figure 3.** Kaplan-Meier survival probabilities for all-cause mortality. Full health cross-classification,
216 self-rated health and health status. 89 observations were above graph maximum of age 83.

217

218 Results from Cox proportional hazards Model 1-3 are presented in Table 3. Compared to excellent
219 self-rated health and favourable health status, all other levels of the full health cross-classification
220 were associated with greater hazards, ranging from HR = 1.22 (95% CI 1.15-1.29, $p_{\text{Bonf.}} < 0.001$)
221 for good self-rated health and favourable health status to HR = 7.14 (95% CI 6.70-7.60, $p_{\text{Bonf.}} <$
222 0.001) for poor self-rated health and unfavourable health status. The order of these levels was
223 consistent across the three models, suggesting that individuals with concordant favourable self-rated
224 health and health status had the lowest hazard and those with concordant unfavourable self-rated
225 health and health status had the highest hazard. Participants with discordant self-rated health and
226 health status had intermediate hazards. Compared to the results from Model 1, we observed
227 attenuation in the hazard ratios from Model 2 and 3, although the overall pattern of results was
228 consistent across these analyses (all Bonferroni-adjusted p -values < 0.001). Survival probabilities
229 and cumulative hazards for Model 3 are presented in Figure 4.

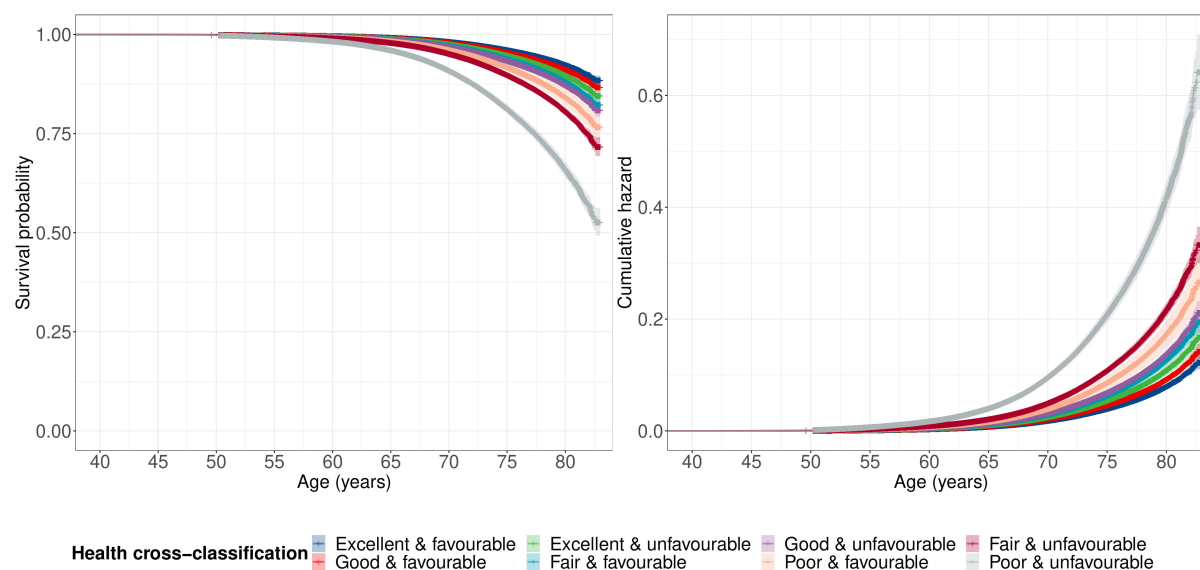
230

Table 3. Cox proportional hazards model: all-cause mortality

Health cross-classification		Model 1			Model 2			Model 3		
		HR	95% CI		HR	95% CI		HR	95% CI	
Excellent	Favourable	Ref	-	-	Ref	-	-	Ref	-	-
Good	Favourable	1.22	1.15	1.29	1.20	1.14	1.27	1.16	1.10	1.23
Excellent	Unfavourable	1.42	1.30	1.55	1.38	1.26	1.51	1.37	1.25	1.50
Fair	Favourable	1.86	1.74	1.99	1.77	1.65	1.89	1.59	1.49	1.70
Good	Unfavourable	1.87	1.77	1.98	1.81	1.71	1.91	1.73	1.63	1.83
Poor	Favourable	2.79	2.42	3.22	2.62	2.27	3.02	2.17	1.88	2.51
Fair	Unfavourable	3.26	3.08	3.45	3.02	2.85	3.19	2.72	2.57	2.88
Poor	Unfavourable	7.14	6.70	7.60	6.34	5.94	6.76	5.24	4.90	5.61

Note: HR = hazard ratio; CI = confidence interval. Model 1 – unadjusted; Model 2 – adjusted for sociodemographic characteristics; Model 3 – additionally adjusted for lifestyle factors. All Bonferroni-adjusted p -values < 0.001.

231



232

233 **Figure 4.** Cox proportional hazard survival probabilities and cumulative hazards for all-cause
 234 mortality. Estimates from Model 3 adjusted for sex, ethnicity, highest qualification, household
 235 income, smoking status, alcohol intake and physical activity. 89 observations were above graph
 236 maximum of age 83. Figures show continuous covariates fixed at their mean value and categorical
 237 covariates fixed at the largest group.

238

239 Cause-specific mortality

240 Survival probabilities for specific causes of death are presented in [Supplement Figure 3](#). Results from
 241 Cox proportional hazards Model 1-3 are presented in [Table 4](#) and [Supplement Table 2](#). Compared to
 242 participants with excellent self-rated health and favourable health status, all other levels of the health
 243 cross-classification were associated with higher hazards. For example, HRs for ischaemic heart
 244 disease from Model 3 ranged from HR = 1.14 (95% CI 0.97-1.35, $p_{BH} = 0.12$) for good self-rated
 245 health and favourable health status to HR = 5.17 (95% CI 4.17-6.26, $p_{Bonf.} < 0.001$) for poor self-
 246 rated health and unfavourable health status. We generally observed attenuation in the hazard ratios
 247 from Model 2 and 3, compared to those from Model 1. For example, the highest hazard ratio for
 248 malignant neoplasm was HR = 8.10 (95% CI 5.51-11.9, $p_{Bonf.} < 0.001$) for poor self-rated health
 249 and unfavourable health status in Model 1, HR = 6.96 (95% CI 4.71-10.31, $p_{Bonf.} < 0.001$) in Model
 250 2 and HR = 5.80 (95% CI 3.87-8.68, $p_{Bonf.} < 0.001$) in Model 3. The overall pattern of results was
 251 similar to the results for all-cause mortality, although several comparisons were not statistically
 252 significant after multiple testing correction, typically for causes of death with lower numbers of
 253 observed deaths (influenza and pneumonia, dementia and Alzheimer's disease, and chronic lower

254 respiratory disease) and levels of the health cross-classification that were associated with longer
 255 survival times than good self-rated health and unfavourable health status.
 256

Table 4. Cox proportional hazards model: cause-specific mortality

Health cross-classification		Model 1			Model 2			Model 3		
		HR	95% CI		HR	95% CI		HR	95% CI	
Ischaemic heart disease										
Excellent	Favourable	Ref	-	-	Ref	-	-	Ref	-	-
Good	Favourable	1.21	1.02	1.43	1.19	1.01	1.41	1.14	0.97	1.35
Excellent	Unfavourable	1.51	1.16	1.98	1.47	1.12	1.92	1.45	1.11	1.90
Fair	Favourable	2.16	1.77	2.63	2.03	1.67	2.48	1.81	1.48	2.21
Good	Unfavourable	1.84	1.55	2.18	1.77	1.49	2.10	1.68	1.42	2.00
Poor	Favourable	3.71	2.53	5.43	3.44	2.35	5.05	2.81	1.91	4.14
Fair	Unfavourable	3.38	2.85	4.01	3.10	2.60	3.68	2.76	2.32	3.29
Poor	Unfavourable	7.14	5.90	8.66	6.28	5.16	7.64	5.11	4.17	6.26
Influenza/pneumonia										
Excellent	Favourable	Ref	-	-	Ref	-	-	Ref	-	-
Good	Favourable	1.25	0.93	1.66	1.25	0.94	1.67	1.22	0.91	1.62
Excellent	Unfavourable	1.67	1.06	2.61	1.63	1.04	2.55	1.61	1.03	2.53
Fair	Favourable	2.13	1.52	3.00	2.09	1.48	2.95	1.90	1.34	2.69
Good	Unfavourable	2.20	1.65	2.94	2.16	1.62	2.89	2.09	1.56	2.79
Poor	Favourable	3.24	1.61	6.53	3.16	1.56	6.39	2.64	1.30	5.36
Fair	Unfavourable	3.91	2.92	5.22	3.73	2.78	5.00	3.41	2.53	4.59
Poor	Unfavourable	9.49	6.93	13.01	8.75	6.34	12.07	7.19	5.15	10.03
Dementia/Alzheimer's disease										
Excellent	Favourable	Ref	-	-	Ref	-	-	Ref	-	-
Good	Favourable	1.22	0.82	1.82	1.17	0.79	1.75	1.11	0.75	1.66
Excellent	Unfavourable	1.13	0.55	2.3	1.09	0.53	2.22	1.07	0.53	2.19
Fair	Favourable	1.32	0.77	2.25	1.19	0.70	2.04	1.05	0.61	1.80
Good	Unfavourable	2.17	1.46	3.24	2.02	1.35	3.01	1.90	1.27	2.84
Poor	Favourable	4.13	1.72	9.91	3.63	1.51	8.74	2.96	1.22	7.18
Fair	Unfavourable	2.74	1.81	4.17	2.40	1.57	3.66	2.11	1.38	3.25
Poor	Unfavourable	6.92	4.36	10.98	5.71	3.56	9.16	4.66	2.86	7.58
Chronic lower respiratory disease										
Excellent	Favourable	Ref	-	-	Ref	-	-	Ref	-	-
Good	Favourable	1.25	0.92	1.70	1.24	0.91	1.68	1.19	0.87	1.62
Excellent	Unfavourable	1.28	0.76	2.17	1.24	0.73	2.10	1.23	0.72	2.07
Fair	Favourable	1.44	0.96	2.15	1.38	0.92	2.07	1.23	0.82	1.85
Good	Unfavourable	2.02	1.48	2.75	1.95	1.43	2.66	1.85	1.35	2.53
Poor	Favourable	2.87	1.30	6.32	2.79	1.26	6.16	2.29	1.03	5.09
Fair	Unfavourable	3.22	2.34	4.41	3.00	2.18	4.13	2.67	1.93	3.70
Poor	Unfavourable	7.00	4.91	9.99	6.35	4.42	9.14	5.16	3.54	7.51
Malignant neoplasm										
Excellent	Favourable	Ref	-	-	Ref	-	-	Ref	-	-
Good	Favourable	1.50	1.07	2.09	1.46	1.04	2.04	1.42	1.01	1.98
Excellent	Unfavourable	2.52	1.58	4.03	2.46	1.54	3.94	2.41	1.51	3.86
Fair	Favourable	2.38	1.6	3.54	2.19	1.47	3.26	1.99	1.33	2.97
Good	Unfavourable	2.35	1.67	3.30	2.24	1.59	3.16	2.14	1.52	3.01
Poor	Favourable	4.01	1.88	8.55	3.58	1.67	7.65	2.99	1.39	6.41
Fair	Unfavourable	4.09	2.90	5.76	3.68	2.60	5.21	3.30	2.33	4.69
Poor	Unfavourable	8.10	5.51	11.9	6.96	4.71	10.31	5.80	3.87	8.68

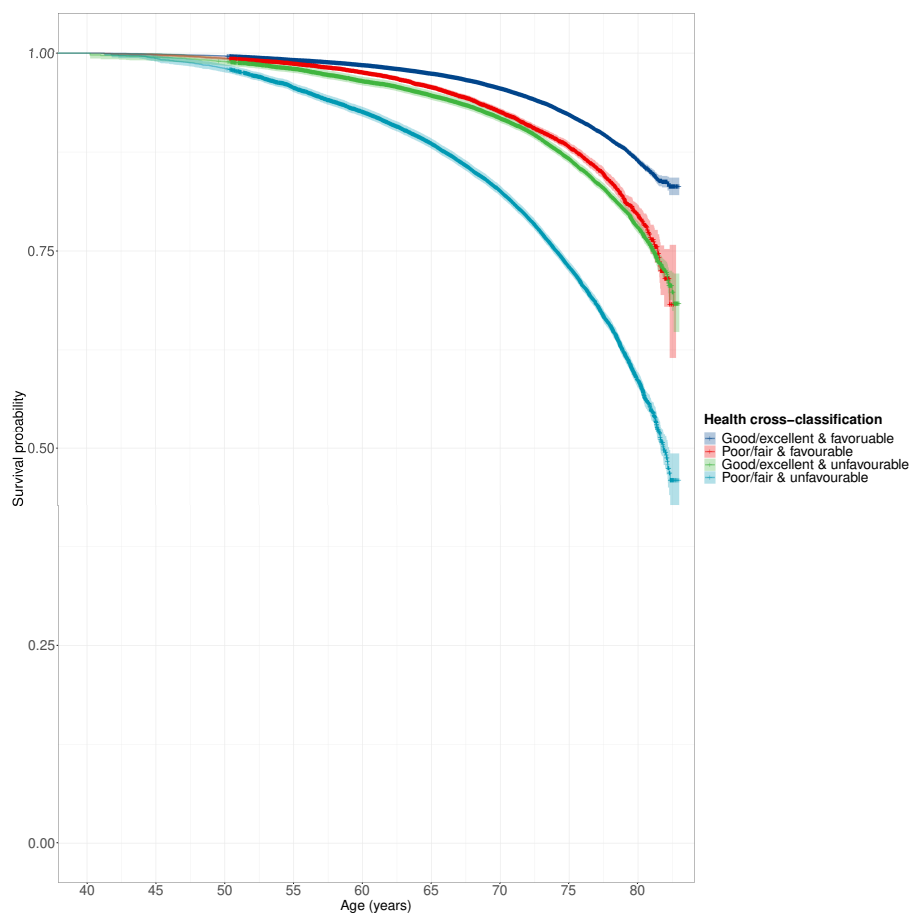
Note: HR = hazard ratio; CI = confidence interval; Bonferroni and Benjamini & Hochberg adjusted *p*-values are presented in [Supplement Table 2](#). Model 1 – unadjusted; Model 2 – adjusted for sociodemographic characteristics; Model 3 – additionally adjusted for lifestyle factors.

257
 258 **Sensitivity analysis**

259 The average number of illnesses was 1.07 (SD = 1.22) in individuals with good/excellent self-rated
 260 health and favourable health status, 2.78 (SD = 1.67) in individuals with good/excellent self-rated
 261 health and unfavourable health status, 1.86 (SD = 1.59) in individuals with poor/fair self-rated health
 262 and favourable health status and 4.00 (SD = 2.30) in individuals with poor/fair self-rated health and
 263 unfavourable health status.

264
 265 Kaplan-Meier survival probabilities from our sensitivity analysis are presented in [Figure 5](#),
 266 suggesting that different levels of the simplified health cross-classification were associated with
 267 varying survival times (log-rank test *p* < 0.001). We found that participants with favourable health
 268 status and good or excellent self-rated health had the longest survival times, while participants with

269 unfavourable health status and poor or fair self-rated health had the shortest survival times.
 270 Participants with discordant health status and self-rated health had intermediate survival times.
 271 Individuals with good or excellent self-rated health but unfavourable health status, i.e., for whom their
 272 objective assessment of health was worse, had only slightly shorter survival times than participants
 273 with favourable health status but fair or poor self-rated health.
 274



275
 276 **Figure 5.** Kaplan-Meier survival probabilities for all-cause mortality. Simplified health cross-
 277 classification. 89 observations were above graph maximum of age 83.
 278

279 Results from Cox proportional hazards Model 1-3 are presented in [Table 5](#). Compared to favourable
 280 health status and good or excellent self-rated health, all other levels of the simplified health cross-
 281 classification were associated with increased hazards. Across Model 1 and 2, individuals with better
 282 self-rated health than health status had lower hazards than individuals with worse self-rated health
 283 than health status, compared to the reference group, although in Model 3 those with worse self-rated
 284 health than health status had lower hazards, similar to the Kaplan-Meier estimates. Individuals with
 285 unfavourable health status and self-rated health had the highest hazards across all models.
 286

		Model 1		Model 2		Model 3	
Term		HR	95% CI	HR	95% CI	HR	95% CI
Good/excellent	Favourable	Ref	-	Ref	-	Ref	-
Good/excellent	Unfavourable	1.55	1.50 1.61	1.51	1.46 1.56	1.49	1.44 1.54
Poor/fair	Favourable	1.66	1.58 1.75	1.58	1.50 1.66	1.43	1.36 1.50
Poor/fair	Unfavourable	3.43	3.32 3.54	3.14	3.04 3.25	2.81	2.72 2.91

Note: HR = hazard ratio; CI = confidence interval. Model 1 – unadjusted; Model 2 – adjusted for sociodemographic characteristics; Model 3 – additionally adjusted for lifestyle factors. All Bonferroni-adjusted *p*-values < 0.001.

287 **Discussion**

288

289 Using data from more than 370,000 UK Biobank participants with a median follow-up period of
290 almost 12 years, we found that individuals with favourable health status or good to excellent self-rated
291 health at baseline had a substantially lower risk of mortality than individuals with unfavourable health
292 status or poor to fair self-rated health, respectively. Examining the cross-classification between health
293 status and self-rated health provided additional granularity to differentiate survival times, confirming
294 that self-rated health captures additional health-related information and highlighting the potential
295 value of combining these measures for risk stratification.

296

297 **Principal findings**

298 As hypothesised, individuals with favourable health status and good to excellent self-rated health had
299 the longest survival times, while individuals with unfavourable health status and poor to fair self-rated
300 health had the shortest survival times. Individuals with discordant health status and self-rated health
301 had intermediate survival times. For example, individuals with favourable health status based on their
302 medical history but poor self-rated health had substantially shorter survival times than individuals
303 with favourable health status and good to excellent self-rated health.

304

305 We observed similar results for leading causes of death in the UK⁶. However, there were some
306 inconsistencies for causes of death with fewer observed deaths during the follow-up period. For
307 dementia and Alzheimer's disease, we did not find evidence of differences between the four health
308 cross-classification levels with the longest survival times: (i) good self-rated health and favourable
309 health status, (ii) excellent self-rated health and unfavourable health status and (iii) fair self-rated
310 health and favourable health status, compared to (iv) excellent self-rated health and favourable health
311 status.

312

313 In the sensitivity analysis in which we merged good and excellent self-rated health and poor and fair
314 self-rated health prior to analysis, we observed that individuals with discordant health status and self-
315 rated health had intermediate survival times, showing that both measures have a key contribution to
316 predicting all-cause mortality.

317

318 **Findings in context**

319 A previous analysis of 5-year mortality in the UK Biobank identified self-rated health as the strongest
320 predictor of all-cause mortality in men, out of 655 variables, and showed that self-rated health was
321 consistently associated with cause-specific mortality¹². Our findings support most previous research
322 that found a graded association between self-rated health levels and mortality¹. Findings from a
323 Brazilian cohort study suggested that the 10-year mortality risk was similar for participants who
324 reported fair or good to excellent self-rated health, compared to poor self-rated health⁴. However, the
325 sample size was limited and the difference in result might be due to linguistic factors; 'fair' self-rated
326 health might reflect average or normal health in Spanish or Portuguese¹⁷, while it might reflect less
327 favourable health in English. Although several previous studies have examined mortality outcomes
328 associated with objective and subjective assessments of health^{3,4}, no studies have, to our knowledge,
329 examined the cross-classification between health status and self-rated health for potential risk
330 stratification.

331 **Strengths and limitations**

332 Strengths of this study include its large sample size (>370,00 participants) and a median follow-up of
333 almost 12 years. Risk indices are often developed in high-risk populations (e.g., in older individuals),
334 focus on a single health outcome and studies are often limited by small sample sizes. The health
335 cross-classification examined in this study might be applicable for risk stratification for a wide range
336 of health outcomes.

337
338 Our research inevitably has limitations. There might be some misclassification in the reporting of
339 medical illnesses that were used to determine health status. However, participants were asked to
340 report illnesses that had been diagnosed by a doctor and these diagnoses were confirmed during a
341 nurse-led interview. Nevertheless, there is the possibility of recall bias in the reporting of long-term
342 and past conditions. Regarding self-rated health, there could be differences by native language in
343 evaluating fair self-rated health as positive or negative. However, most participants were born in the
344 UK and such differences, if present, would likely be minimal. Mortality data from death certificates
345 might have some misclassification in causes of death, especially between similar diseases, and we
346 observed fewer deaths during follow-up for several specific causes of death. However, all-cause
347 mortality is a more robust endpoint than cause-specific mortality and we found similar results across
348 most outcomes.

349
350 **Generalisability**

351 Compared with non-responders, UK Biobank participants were older, more likely to be female and
352 more likely of higher socioeconomic status. They were also less likely to engage in unhealthy lifestyle
353 behaviours, reported fewer medical illnesses compared with data from a nationally representative
354 survey and all-cause mortality was 46.2% lower for 70-74-year-olds¹⁸. Although these findings show
355 that the UK Biobank is not fully representative of the UK population, a recent empirical investigation
356 comparing the UK Biobank with data from 18 prospective cohort studies with conventional response
357 rates showed that the direction of risk factor associations were similar, although with differences in
358 magnitude¹⁹. It is also worth noting that most participants (79.6%) in our sample reported good to
359 excellent self-rated health, which is comparable to estimates from the Office for National Statistics
360 suggesting that >72% of people in England and Wales rated their health as good or very good²⁰.

361
362 **Implications**

363 Risk stratification is a key public health priority that is central to clinical decision making and
364 resource allocation. Self-rated health and health status can be obtained through verbal interview, self-
365 report or from medical records. The cross-classification between health status and self-rated health
366 represents a straightforward metric for initial risk stratification, with applications to population health,
367 clinical decision making and resource allocation. Our findings also highlight that self-rated health
368 captures additional health-related information and should therefore be more widely assessed across
369 settings.

370 **Authorship contributions**

371 CML acquired the studentship funding, interpreted the findings and critically reviewed the
372 manuscript. JM conceived the idea of the study, acquired the data, carried out the statistical analysis,
373 interpreted the findings, wrote the manuscript and revised the manuscript for final submission. Both
374 authors read and approved the final manuscript. JM had full access to all data used in this study and
375 takes responsibility for the integrity of the data and the accuracy of the data analysis.

376

377 **Funding and disclosure**

378 JM acknowledges studentship funding from the Biotechnology and Biological Sciences Research
379 Council (BBSRC) (ref: 2050702) and Eli Lilly and Company Limited. CML is part-funded by the
380 National Institute for Health Research (NIHR) Biomedical Research Centre at South London and
381 Maudsley NHS Foundation Trust and King's College London. The views expressed are those of the
382 authors and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care.
383 CML is a member of the Scientific Advisory Board of Myriad Neuroscience.

384

385 **Acknowledgments**

386 This research has been conducted using data from UK Biobank, a major biomedical database. This
387 project made use of time on Rosalind HPC, funded by Guy's & St Thomas' Hospital NHS Trust
388 Biomedical Research Centre (GSTT-BRC), South London & Maudsley NHS Trust Biomedical
389 Research Centre (SLAM-BRC), and Faculty of Natural Mathematics & Science (NMS) at King's
390 College London.

391

392 **Ethics**

393 Ethical approval for the UK Biobank study has been granted by the National Information Governance
394 Board for Health and Social Care and the NHS North West Multicentre Research Ethics Committee
395 (11/NW/0382). No project-specific ethical approval is needed. Data access permission has been
396 granted under UK Biobank application 45514.

397

398 **Data sharing statement**

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

402

403 **Supplementary material**

404 Supplementary information is available online.

405 References

- 406
407 1 DeSalvo, K. B., Bloser, N., Reynolds, K., He, J. & Muntner, P. Mortality prediction with a single
408 general self-rated health question: A meta-analysis. *Journal of general internal medicine* **21**, 267-275
409 (2006).
- 410 2 Ferraro, K. F. Self-ratings of health among the old and the old-old. *Journal of Health and Social*
411 *Behavior*, 377-383 (1980).
- 412 3 Wu, S. *et al.* The relationship between self-rated health and objective health status: a population-based
413 study. *BMC public health* **13**, 320 (2013).
- 414 4 Lima-Costa, M. F., Cesar, C. C., Chor, D. & Proietti, F. A. Self-rated health compared with objectively
415 measured health status as a tool for mortality risk screening in older adults: 10-year follow-up of the
416 Bambuí Cohort Study of Aging. *American journal of epidemiology* **175**, 228-235 (2012).
- 417 5 ONS. *How do people rate their general health? An analysis of general health by disability and*
418 *deprivation*
419 <https://webarchive.nationalarchives.gov.uk/20160109203542/http://www.ons.gov.uk/ons/dcp171776_353238.pdf> (2014).
- 420
421 6 ONS. *Leading causes of death, UK*,
422 <<https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/causesofdeath/articles/leadingcausesofdeathuk/2001to2018>> (2020).
- 423
424 7 Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203
425 (2018).
- 426 8 Maxwell, J. M. *et al.* Multifactorial disorders and polygenic risk scores: predicting common diseases
427 and the possibility of adverse selection in life and protection insurance. *Annals of Actuarial Science*, 1-
428 16 (2020).
- 429 9 Mutz, J., Roscoe, C. J. & Lewis, C. M. Exploring health in the UK Biobank: associations with
430 sociodemographic characteristics, psychosocial factors, lifestyle and environmental exposures.
431 *medRxiv* (2020).
- 432 10 Hartzband, P. & Groopman, J. There is more to life than death. *N Engl J Med* **367**, 987-989 (2012).
- 433 11 Guggenheim, J. A. & Williams, C. Childhood febrile illness and the risk of myopia in UK Biobank
434 participants. *Eye* **30**, 608 (2016).
- 435 12 Ganna, A. & Ingelsson, E. 5 year mortality predictors in 498 103 UK Biobank participants: a
436 prospective population-based study. *The Lancet* **386**, 533-540 (2015).
- 437 13 Kaplan, E. L. & Meier, P. Nonparametric estimation from incomplete observations. *Journal of the*
438 *American statistical association* **53**, 457-481 (1958).
- 439 14 Cox, D. R. Regression models and life-tables. *Journal of the Royal Statistical Society: Series B*
440 *(Methodological)* **34**, 187-202 (1972).
- 441 15 Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach
442 to multiple testing. *Journal of the Royal statistical society: series B (Methodological)* **57**, 289-300
443 (1995).
- 444 16 Svedberg, P., Bardage, C., Sandin, S. & Pedersen, N. L. A prospective study of health, life-style and
445 psychosocial predictors of self-rated health. *European journal of epidemiology* **21**, 767-776 (2006).
- 446 17 Bzostek, S., Goldman, N. & Pebley, A. Why do Hispanics in the USA report poor health? *Social*
447 *Science & Medicine* **65**, 990-1003 (2007).
- 448 18 Fry, A. *et al.* Comparison of sociodemographic and health-related characteristics of UK Biobank
449 participants with those of the general population. *American Journal of Epidemiology* **186**, 1026-1034
450 (2017).
- 451 19 Batty, G. D., Galec, C. R., Kivimäki, M., Deary, I. J. & Belle, S. Generalisability of Results from
452 UK Biobank: Comparison With a Pooling of 18 Cohort Studies.
- 453 20 ONS. *General Health in England and Wales: 2011 and comparison with 2001*,
454 <[https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/healthandwellbeing/arti](https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/healthandwellbeing/articles/generalhealthinenglandandwales/2013-01-30)
455 [cles/generalhealthinenglandandwales/2013-01-30](https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/healthandwellbeing/articles/generalhealthinenglandandwales/2013-01-30)> (2013).
- 456