

SUPPLEMENTARY METHODS

Definition of Prevalent Disease

Prevalent diabetes was defined as self-reported diagnosis by physicians, use of glucose-lowering medications, measured serum glycated hemoglobin of ≥ 48 mmol/mol, or diagnoses before baseline retrieved through medical records. Prevalent cardiovascular disease including coronary heart disease and stroke was defined as self-reported diagnosis by physicians or diagnoses or related operation procedures before baseline retrieved through medical records. Prevalent cancer except for nonmelanoma was defined as self-reported diagnosis by physicians or diagnoses before baseline retrieved through cancer registry.

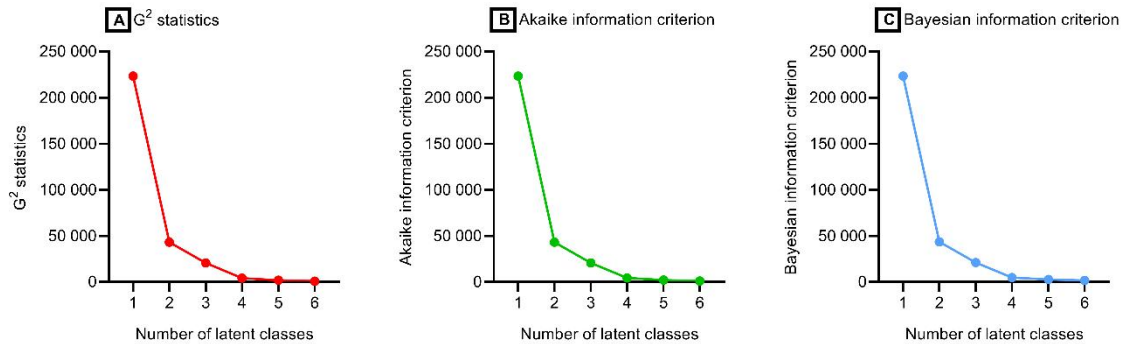
Assessment of Socioeconomic Status (SES)

Average total household income before tax, employment status and job code, and qualifications were used to construct the latent variable SES. The rationality of the classification of each variable is listed below.

- For average total household income before tax, participants could choose an option from “less than £18 000”, “£18,000 to 30,999”, “£31,000 to 51,999”, “£52,000 to 100,000”, “greater than £100,000”, “do not know”, or “prefer not to answer”^[1]. A group with a small sample size will add very limited information for distinguishing participants in the latent class analysis and make the interpretation of the model difficult, thus, we merged “£52,000 to 100,000” and “greater than £100,000” given that only 4.6% of participants’ average total household income before tax was greater than £100,000.
- The occupation was defined by employment status and job code^[1,2], including skilled job (i.e., managers and senior officials, professional occupations, associate professional and technical occupations, administrative and secretarial occupations, and skilled trades occupations), unskilled job (personal service occupations, sales and customer service occupations, process/plant/machine operatives, and elementary occupations), unemployment (looking after home/family, unable to work because of sickness/disability, unemployed, and other employment status), and others (unpaid or voluntary work, retired, and full or part-time student).
- Participants’ educational levels were classified according to qualifications (3), i.e., college or university degree (equal 20 years of education), NVQ or HND or HNC or equivalent (19 years), other professional qualifications (15 years), A levels/AS levels or equivalent (13 years), O levels/GCSEs or CSEs or equivalent (10 years), and none of the above (7 years).

The latent class analysis with different numbers of latent classes was performed, and a reasonable model was selected according to model fit and the practical interpretation of the latent classes. In two consecutive iterations of the estimation procedure during modeling, the maximum absolute deviation between parameter estimates was set to 0.000001, which meant that iterations would end when the difference in parameter estimates between two consecutive iterations was less than 0.000001. For model selection, the likelihood ratio statistic G^2 , Akaike information criterion (AIC), and Bayesian information criterion (BIC) were considered, with smaller values indicating better model fits. The mean posterior probability reflects the uncertainty of posterior classification and was also used for model selection, with a value of 0.7 or greater indicating an acceptable uncertainty. The posterior probability – item-response probability – was used to define latent classes.

Since the model cannot converge when the number of latent classes > 6 , Supplementary Figure S1 shows the G^2 statistics, AIC, and BIC for models with ≤ 6 latent classes, and all statistics continued to go down as more latent classes were added. However, the decrease leveled off when the number of latent classes reached three.



Supplementary Figure S1. Supplementary Methods. Likelihood ratio statistic G², Akaike Information Criterion, and Bayesian Information Criterion.

Supplementary Table S1 shows the mean posterior probabilities in models with ≥ 3 latent classes. All mean posterior probabilities from the three-latent-class solution were all ≥ 0.80 ; mean posterior probabilities of latent classes 2 and 4 from the four-latent-class solution were 0.67 which were less than 0.70; the mean posterior probability of latent classes 2 and 5 from the five-latent-class solution were 0.62 and 0.66, respectively; the mean posterior probability of latent classes 3 to 6 were less than 0.68. Thus, the three-latent-class solution was the best in terms of the uncertainty of the posterior classification.

Supplementary Table S1. Mean posterior probabilities, prevalence of latent classes, and item-response probabilities in latent class models^a

Item	Latent class					
	1	2	3	4	5	6
Three-latent-class solution						
MeanPP	0.84	0.80	0.85	NA	NA	NA
Prevalence	0.31	0.46	0.23	NA	NA	NA
Qualification						
College/university	0.63	0.28	0.12	NA	NA	NA
NVQ/HND/HNC	0.12	0.20	0.13	NA	NA	NA
Other professional	0.10	0.14	0.09	NA	NA	NA
A/AS levels	0.06	0.06	0.04	NA	NA	NA
O levels/GCSEs/CSEs	0.08	0.21	0.19	NA	NA	NA
None of the above	0.01	0.11	0.42	NA	NA	NA
Income, £						
$\geq 52,000$	0.69	0.09	0.01	NA	NA	NA
31,000 to 51,999	0.29	0.37	0.00	NA	NA	NA
18,000 to 30,999	0.01	0.44	0.21	NA	NA	NA
< 18,000	0.01	0.09	0.78	NA	NA	NA
Occupation						
Skilled job	0.87	0.48	0.04	NA	NA	NA
Unskilled job	0.01	0.16	0.14	NA	NA	NA

Item	Latent class					
	1	2	3	4	5	6
Others	0.09	0.33	0.63	NA	NA	NA
Unemployment	0.04	0.03	0.19	NA	NA	NA
Four-latent-class solution						
MeanPP	0.87	0.67	0.91	0.67	NA	NA
Prevalence	0.37	0.19	0.25	0.19	NA	NA
Qualification						
College/university	0.64	0.02	0.13	0.41	NA	NA
NVQ/HND/HNC	0.11	0.34	0.13	0.11	NA	NA
Other professional	0.11	0.08	0.10	0.20	NA	NA
A/AS levels	0.06	0.06	0.04	0.06	NA	NA
O levels/GCSEs/CSEs	0.08	0.32	0.19	0.16	NA	NA
None of the above	0.01	0.18	0.41	0.05	NA	NA
Income, £						
≥ 52,000	0.61	0.11	0.02	0.03	NA	NA
31,000 to 51,999	0.29	0.39	0.03	0.38	NA	NA
18,000 to 30,999	0.06	0.36	0.22	0.58	NA	NA
< 18,000	0.04	0.13	0.74	0.01	NA	NA
Occupation						
Skilled job	0.84	0.62	0.03	0.32	NA	NA
Unskilled job	0.02	0.34	0.12	0.05	NA	NA
Others	0.11	0.00	0.66	0.62	NA	NA
Unemployment	0.04	0.04	0.18	0.01	NA	NA
Five-latent-class solution						
MeanPP	0.77	0.62	0.80	0.72	0.66	NA
Prevalence	0.29	0.15	0.17	0.17	0.22	NA
Qualification						
College/university	0.72	0.07	0.14	0.48	0.12	NA
NVQ/HND/HNC	0.08	0.25	0.11	0.07	0.30	NA
Other professional	0.10	0.06	0.10	0.22	0.12	NA
A/AS levels	0.05	0.03	0.05	0.07	0.08	NA
O levels/GCSEs/CSEs	0.04	0.24	0.18	0.14	0.29	NA
None of the above	0.00	0.35	0.42	0.02	0.09	NA
Income, £						
≥ 52,000	0.64	0.00	0.02	0.10	0.22	NA
31,000 to 51,999	0.26	0.16	0.01	0.37	0.44	NA
18,000 to 30,999	0.06	0.52	0.11	0.51	0.25	NA
< 18,000	0.03	0.32	0.85	0.03	0.08	NA
Occupation						
Skilled job	0.87	0.18	0.02	0.26	0.76	NA
Unskilled job	0.01	0.35	0.07	0.00	0.19	NA
Others	0.08	0.42	0.69	0.71	0.00	NA
Unemployment	0.03	0.04	0.23	0.02	0.05	NA

Item	Latent class					
	1	2	3	4	5	6
Six-latent-class solution						
MeanPP	0.77	0.75	0.68	0.63	0.64	0.55
Prevalence	0.28	0.18	0.11	0.18	0.18	0.06
Qualification						
College/university	0.72	0.08	0.02	0.51	0.11	0.29
NVQ/HND/HNC	0.08	0.26	0.09	0.05	0.31	0.18
Other professional	0.10	0.06	0.09	0.23	0.13	0.06
A/AS levels	0.05	0.03	0.03	0.08	0.08	0.09
O levels/GCSEs/CSEs	0.04	0.23	0.16	0.13	0.28	0.30
None of the above	0.00	0.33	0.61	0.00	0.09	0.09
Income, £						
≥ 52,000	0.66	0.01	0.02	0.08	0.24	0.13
31,000 to 51,999	0.27	0.15	0.02	0.30	0.50	0.13
18,000 to 30,999	0.06	0.48	0.11	0.43	0.26	0.16
< 18,000	0.00	0.36	0.85	0.19	0.00	0.57
Occupation						
Skilled job	0.88	0.21	0.00	0.25	0.80	0.33
Unskilled job	0.01	0.32	0.05	0.00	0.19	0.14
Others	0.08	0.43	0.75	0.73	0.00	0.08
Unemployment	0.03	0.04	0.20	0.02	0.01	0.45

Note. meanPP, mean posterior probability; NA, not available. The leading item-response probabilities for each latent class were marked in bold.

Next, we focused on explaining each latent class from different models. Generally, the numbers of participants with other professional qualifications, A/AS levels, and unskilled jobs were relatively small, which provided limited information for the model explanation and diluted the effects of adjacent levels. Specifically, for the three-latent-class model, latent class 1 was explicitly characterized by college/university education, high-income level, and skilled job, which could be defined as a “high SES” group. Latent class 2 was characterized by medium-income level, but the proportions of participants with various education levels from O levels/GCSEs/CSEs to college/university were similar, and the participants with skilled jobs and other jobs constituted a large proportion of this group. Although we could define latent class 2 as the “medium SES” group, the participants from latent class 2 had varied educational and occupational characteristics. Latent class 3 was characterized by O levels/GCSEs/CSEs or no qualification, low-income level, and other jobs or unemployment, which could be defined as the “low SES” group.

As for the four-latent-class model, latent classes 1 to 3 were correspondingly similar to those in the three-latent-class model. Particularly, latent class 2 was characterized as both NVQ/HND/HNC and O levels/GCSEs/CSEs (two distinct educational levels), as well as both skilled and unskill jobs, indicating the group’s heterogeneity. While latent class 4 has a clearer profile, i.e., college/university and other professional qualifications, medium-income level, and other jobs, which could be viewed as a subgroup separated from the “medium SES” group and characterized with other jobs.

As for the five-latent-class model, the latent classes 1 to 4 were correspondingly similar to those in the four-latent-class model, while latent class 5 was characterized by NVQ/HND/HNC or O levels/GCSEs/CSEs,

medium income level, and skilled job, which could be viewed as a subgroup separated from “medium SES” group and characterized by having skilled jobs. However, latent classes 2 and 5 have heterogeneous characteristics related to education, i.e., NVQ/HND/HNC and no qualification for latent class 2 and NVQ/HND/HNC and O levels/GCSEs/CSEs for latent class 5. Seemingly, the four- and five-latent-class models regrouped the “medium SES” group according to individuals’ occupations, which made different groups more distinct from others, but there remains some heterogeneity within groups.

As for the six-latent-class model, the latent classes 1 to 5 were correspondingly similar to those in the five-latent-class model; however, it is difficult to give a clear definition to the latent class 6, which is characterized by low-income level, but both college/university education (high education level) and O levels/GCSEs/CSEs (relatively low education level), as well as both skilled job and unemployment.

Above all, the three-latent-class model was the best regarding the certainty of posterior classification, and the sample size for each group was sufficient (all > 20%); however, education and occupation were heterogeneous in latent class 2, and the model fit was inferior to models with more latent classes. Four- and five-latent class models had better model fits but had the following disadvantages: (1) uncertainty of posterior classification; (2) latent class 2 is still heterogeneous; and (3) latent classes 2 and 4 in the four-latent-class model and latent classes 2 to 4 in the five-latent-class model only constituted < 20% of the sample, respectively, which could reduce statistical power. The six-latent class model had large uncertainty of posterior classification, and latent class 6 was too heterogeneous. Seemingly, all models have some disadvantages and cannot confidently explain the participants’ comprehensive SES.

We regrouped the participants due to the small sample sizes of the participants with other professional qualifications, A/AS levels, and unskilled jobs, and better model fits and explanations were expected. Participants were regrouped into average total household income before tax of “less than £18 000”, “£18 000 to 51 999”, and “greater than £52 000”; college/university/NVQ/HND/HNC/other professional qualification, A levels/AS levels/O levels/GCSEs/CSEs, and no degree; skilled job, other jobs (including unskilled job, unpaid or voluntary work, retired, and full or part-time student), and unemployment. Latent class models were fitted by these variables.

The model cannot converge when the number of latent classes > 3, and we compared the three-latent-class model with previous models. Its likelihood ratio statistic G^2 and AIC were between the five- and six-latent-class models’, and its BIC was less than all models’ values. The mean posterior probabilities for all classes were ≥ 0.83 . Latent class 1 was characterized as high education, income, and occupation level, which can be defined as the “high SES” group; latent class 2 was characterized as with qualifications, medium-income level, and having jobs, which can be defined as the “medium SES” group; latent class 3 was characterized as with no qualification or A/AS/O levels/GCSEs/CSEs, low-income level, and unemployment or with other jobs, which can be defined as the “low SES” group. The practical definitions of three latent classes are shown in the following figure.

Average total household income before tax, £		≥ 52,000	31,000 to 51,999	< 18,000
College/university/NVQ/HND /HNC/other professional qualification	Skilled job	17.3	17.6	1.8
	Other jobs	3.1	14.0	5.9
	Unemployment	0.7	1.1	1.4
A levels/AS levels/O levels/GCSEs/CSEs	Skilled job	3.0	6.4	1.0
	Other jobs	0.8	5.7	3.5
	Unemployment	0.3	0.7	1.0
No degree	Skilled job	0.4	1.8	0.5
	Other jobs	0.3	3.8	6.2
	Unemployment	0.1	0.4	1.4
Socioeconomic group		High	Medium	Low

Supplementary Figure S2. Supplementary Methods. Practical definitions of high, medium, and low socioeconomic status in the three-latent-class solution. The values in the cells indicated the proportions of the participants.

Assessment of Risk Factors for Diabetes

Physiological Factors

Most disease factors were defined as diagnoses before baseline according to hospital in-patient records and primary care records. The International Classification of Diseases codes, Tenth Revision (ICD-10) were J45-J46 asthma (self-reported diagnosis by physicians were also used), J40-J44 chronic obstructive pulmonary disease (self-reported diagnosis by physicians were also used), K50-K51 for inflammatory bowel disease, M30-M36 for systemic autoimmune disorders, J30 for allergic rhinitis, M05-M06 for rheumatoid arthritis, L40 and M07 for psoriasis, M15-M19 for osteoarthritis, and L20 for atopic dermatitis. Specifically, severe periodontal disease were defined according to a question which asked whether the participant had mouth ulcers, painful gums, bleeding gums, loose teeth, toothache, or dentures. According to previous studies^[4,5], painful or bleeding gums were defined as mild to moderate periodontal diseases, and loose teeth were defined as severe periodontal diseases, and we set the severe periodontal disease as the risk level.

Measurements of metabolic factors are as follows.

- Serum glycated hemoglobin was measured by a High Performance Liquid Chromatography analysis on five Bio-Rad VARIANT II Turbo analysers (Bio-Rad Laboratories Inc., Hercules, CA, USA).
- Dyslipidemia was defined as self-reported physician-diagnosed high cholesterol, use of lipid-lowering medications, measured total cholesterol of ≥ 6.2 mmol/L, or measured low-density lipoprotein cholesterol of ≥ 4.1 mmol/L. Serum cholesterol was measured by a Beckman Coulter AU5800 (Beckman Coulter Ltd., Buckinghamshire, UK).
- Body mass index was calculated as body weight (kg) divided by height squared (m^2). Body weight was measured using the Tanita BC418MA body composition analyzer (Tanita, Tokyo, Japan) or standard scales. Body height was measured using a Seca 202 device (Seca, Hamburg, Germany) to the nearest 0.1 cm.
- Waist-to-hip ratio was calculated as waist circumference divided by hip circumference.
- Serum uric acid and C-reactive protein were measured using a Beckman Coulter AU5800 (Beckman Coulter Ltd., Buckinghamshire, UK).
- Resting pulse rate was measured using Omron 705 IT electronic blood pressure monitor (OMRON Healthcare Europe, Hoofddorp, Netherlands) during blood pressure measurements.

- Sex hormone binding globulin was measured by two step sandwich immunoassay analysis on a Beckman Coulter Unicel Dxl 800 (Beckman Coulter Ltd., Buckinghamshire, UK).
- Platelet count was measured by the Coulter LH 750 system (Beckman Coulter Ltd., Buckinghamshire, UK).
- Hypertension was defined by self-reported physician-diagnosis of hypertension, use of antihypertensive medications, diagnoses before baseline according to hospital in-patient records and primary care records (ICD-10 codes: I10-I15), or measured systolic/diastolic blood pressure of $\geq 140/90$ mmHg. Blood pressure was measured using Omron 705 IT electronic blood pressure monitor (OMRON Healthcare Europe, Hoofddorp, Netherlands) or a manual sphygmometer when the standard automated device could not be employed. Two measures of blood pressure were taken a few moments apart, and the mean values of the two measurements were calculated.
- We considered both biomarkers related to liver function and non-alcoholic fatty liver disease to evaluate liver function. Alanine aminotransferase of > 52 U/L, aspartate aminotransferase of > 39 U/L, gamma glutamyltransferase of > 64 U/L, direct bilirubin of > 0.18 mg/dL, or total bilirubin of > 1.0 mg/dL were defined as liver dysfunction^[6]. All these biomarkers were measured using a Beckman Coulter AU5800 (Beckman Coulter Ltd., Buckinghamshire, UK). Non-alcoholic fatty liver disease was defined by diagnoses before baseline according to hospital in-patient records and primary care records (ICD-10 codes: K76.0, K74.0, K74.1, K74.2, and K75.8).
- Chronic kidney disease was defined according to measured albuminuria of > 30 mg/L, estimated glomerular filtration rate (eGFR) of < 60 mL/min per 1.73 m², diagnoses before baseline according to hospital in-patient records and primary care records (ICD-10 codes: N03, N06, N08, N11-N16, and N18-N21), and indications of renal replacement therapy (ICD-10 codes: E85.3, N16.5, Q60.1, T82.4, T86.1, Y60.2, Y61.2, Y62.2, Y84.1, Z49.0, Z49.1, Z49.2, Z94.0, and Z99.2; Office of Population Censuses and Surveys Classification of Interventions and Procedures codes, version 4: L74.1, L74.2, L74.3, L74.4, L74.5, L74.6, L74.8, L74.9, M01.2, M01.3, M01.4, M01.5, M01.8, M01.9, M02.3, M08.4, M17.2, M17.4, M17.8, M17.9, X40.1, X40.2, X40.3, X40.4, X40.5, X40.6, X40.7, X40.8, X40.9, X41.1, X41.2, X41.8, X41.9, X42.1, X42.8, X42.9, and X43.1)^[7]. Chronic Kidney Disease Epidemiology Collaboration equation was used for eGFR calculation, i.e., $eGFR = 141 \times \min(\text{serum creatinine}/\kappa, 1)^\alpha \times \max(\text{serum creatinine}/\kappa, 1)^{-1.209} \times 0.993^{\text{Age}} \times 1.018$ [if female], where κ is 0.7 for females and 0.9 for males, and α is -0.329 for females and -0.411 for males^[8].
- Hyperthyroidism and hypothyroidism reflected thyroid function and were defined as diagnoses before baseline according to hospital in-patient records and primary care records (ICD-10 codes: E01-E07).

Behavioral Factors

All information on behavioral factors was collected through questionnaires.

- Dietary quality was evaluated through a recent dietary recommendation for cardiovascular health which considered ten food items, i.e., consuming ≥ 3 servings/d of vegetables, ≥ 3 servings/d of fruit, ≥ 3 servings/d of whole grains, ≥ 2 servings/wk of (shell)fish, ≥ 2 servings/d of dairy products, ≥ 2 servings/d vegetable oils, ≤ 2 servings/d of refined grains, ≤ 2 servings/wk of unprocessed meat, ≤ 1 serving/wk of processed meat, and no sugar-sweetened beverages. The methods were widely used and detailed in publications related to diabetes and cardiovascular disease from UK Biobank, and meeting ≥ 5 items of the recommendations was considered a healthy diet^[1,9].
- The unhealthy sleep score was constructed according to previous publications and considered inadequate sleep (< 7 or > 8 h/d), late chronotype (evening or more an evening than morning person), snoring, usually having insomnia, and frequent excessive daytime sleepiness (often or all the time)^[7].

The participants had 1 point if meeting each unhealthy sleep behavior or 0 points otherwise, and the unhealthy sleep score ranged between 0 and 5, with higher scores indicating less healthy sleep pattern.

- Alcohol drinking was assessed according to the frequency of drinking alcohol and the volume of alcoholic drinks consumed each month or week.
- Leisure-time physical activity was the sum of walking for pleasure (metabolic equivalent score = 3.3), strenuous sports (metabolic equivalent score = 8.0), and other exercises (e.g., swimming, cycling, keeping fit, and bowling; metabolic equivalent score = 4.0), which was calculated according to the frequency and duration for each type of activity.
- Sedentary behavior was the sum of the time spent on driving, using the computer (not including workplace usage), and watching TV.
- Active smoking was evaluated according to the current and past tobacco smoking status, time when quitting smoking, and the number of cigarettes smoked daily.
- Passive smoking included exposure at home and outside of the home.

Environmental Factors

Area-level data were provided by the UK Biobank and included environmental noise, air pollution, land used as the natural environment, and the index of multiple deprivation scores.

- Noise estimates in 2009 were modeled using a version of the Common NOise aSSessment methOdS in Europe noise model, and the methods have been detailed previously^[10].
- Air pollution estimates including PM_{2.5}, PM₁₀, and nitrogen dioxide in 2010 were modeled for each address using a Land Use Regression model developed as part of the European Study of Cohorts for Air Pollution Effects, and details were reported previously^[11,12].
- The percentage of the home location buffer (buffered at 1000 m) classed as “natural environment” in the Land Cover Map (LCM) 2007 was provided by the UK Biobank. Land use data were obtained from the Land Cover Map, 2007, and the 23 land cover classes were reclassified into two groups, i.e., “natural environment” and “built environment”^[13,14].
- The UK government, Welsh government, and Scottish government routinely released the index of multiple deprivation scores for every small area in the UK, and the UK Biobank could link the participants to the index through the Lower-layer Super Output Area code. Health score was included in the current analysis to reflect the area-level health resources and health level. The health score in England was calculated according to years of potential life lost, comparative illness and disability ratio, measures of acute morbidity (an age and sex standardised rate of emergency admissions to hospital), and proportion of adults under 60 suffering from mood or anxiety disorders. The health score in Scotland was calculated according to the standardized mortality ratio, hospital episodes related to alcohol use, hospital episodes related to drug use, comparative illness factor, emergency admissions to hospital, proportion of population being prescribed drugs for anxiety, depression or psychosis, and proportion of live singleton births of low birth weight were used to calculate the health score. The health score in Wales was calculated according to limiting long-term illness, standardised all-cause death rate, standardised cancer incidence rate, and singleton low birth weights. Details have been reported previously^[15].

Psychological Factors

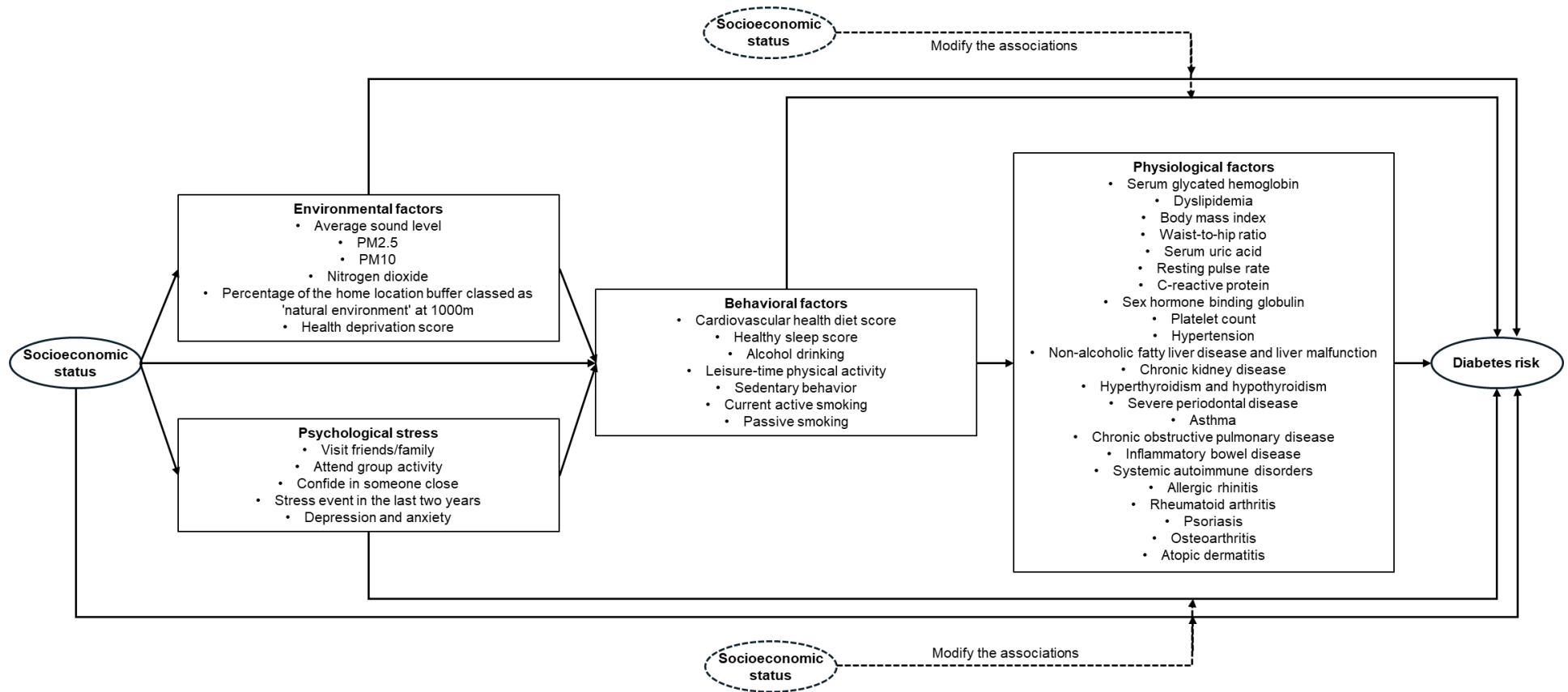
Frequency of friend/family visits, whether living with family, group activity attended once a week or more often, frequency of being able to confide in someone close, and whether experiencing serious illness/injury/assault to themselves or close relatives, death of close relatives/spouse/partner, marital

separation/divorce, or financial difficulties in the last 2 years were documented via touchscreen questionnaires.

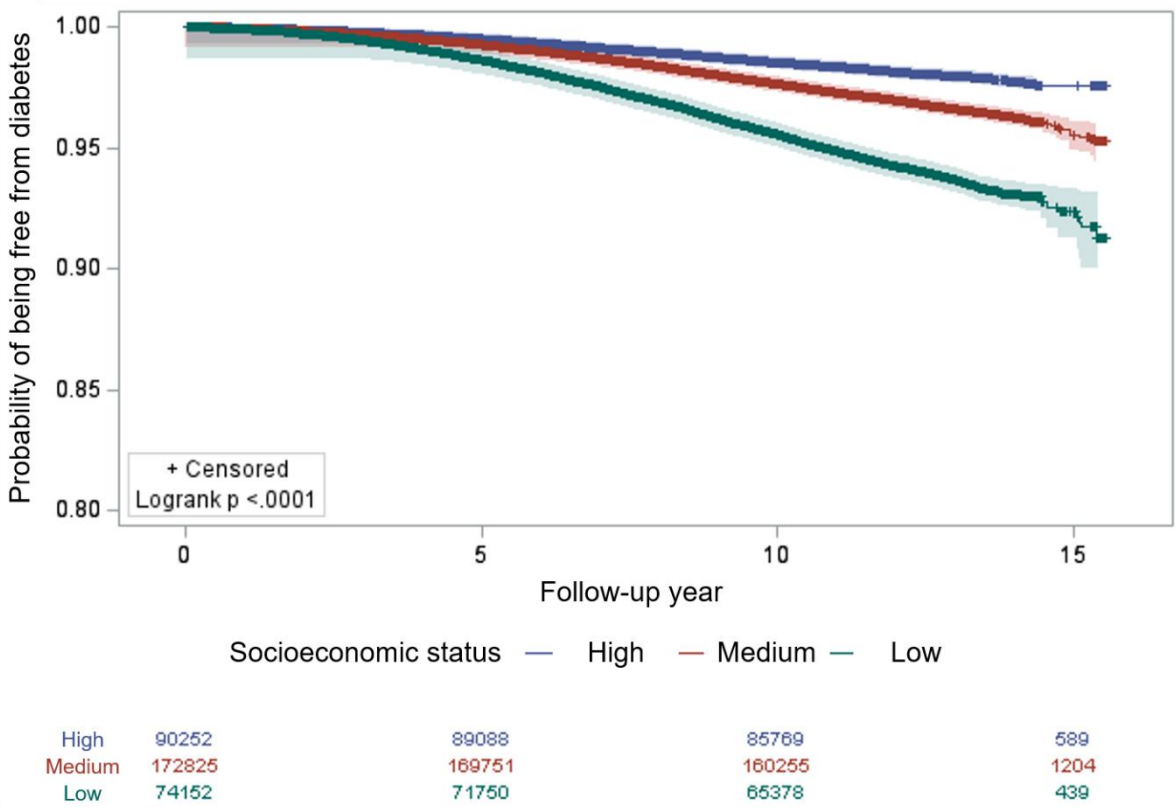
Anxiety and depression were defined according to self-reports, Patient Health Questionnaire-4, and linkages to hospital in-patient records and primary care records based on previous publications from the UK Biobank^[16,17]. The participants were asked whether ever saw a general practitioner or psychiatrist for nerves, anxiety, tension, or depression, as well as noncancer disease. Besides, the UK Biobank applied a 4-term Patient Health Questionnaire, which asked participants about the frequency of (1) feeling down, depressed, or hopeless, (2) having little interest or pleasure in doing things, (3) feeling tense, fidgety, or restless, and (4) feeling tired or having little energy over the past two weeks. The first two questions were related to depression, and the latter two questions were related to anxiety. The response options for each question included “not at all”, “several days”, “more than half the days”, and “nearly every day”, which were assigned a score of 0, 1, 2, and 3, respectively. The participants would be considered to have depression if the sum of the scores for the first two questions was > 2 points, and the participants would be considered to have anxiety if the sum of the scores for the latter two questions was > 2 points^[17]. Besides, the International Classification of Diseases, Tenth Revision codes were F32, F33, F34, F38, and F39 for depression and F40 and F41 for anxiety^[16].

Sensitivity Analysis

First, incident diabetes cases occurring within the first five years of follow-up were excluded to reduce the possibility of reverse causation according to a previous study^[18]. Second, incident type 2 diabetes was examined as a secondary outcome. Third, all analyses were repeated by substituting SES with separate socioeconomic factors (i.e., income, occupation, and education; each factor had three levels^[1], and these three factors were mutually adjusted. For average total household income before tax, less than £18,000, £18,000 to 51,999, and £52,000 or more were defined as the high, medium, and low levels, respectively. For education levels, college or university degree, NVQ, HND, HNC, or other professional qualifications (e.g., nursing, teaching, etc) were equivalent to college or above and were defined as the high level; A/AS levels, O levels, GCSEs, or CSEs were equivalent to high school and were defined as the medium level; and no education qualification was equivalent to less than high school and was defined as the low level. For occupation, skilled jobs were defined as the high level; unskilled jobs, students, and retirees were defined as the medium level; unemployment was defined as the low level. Fourth, serum glycated hemoglobin was excluded from the physiological score due to its strong predictive value for incident diabetes. Fifth, risk factors were examined separately instead of being summed as scores. The last two analyses were only conducted for mediation analyses.



Supplementary Figure S3. Directed acyclic graph of the hypothesized relationships among socioeconomic status, modifiable risk factors, and diabetes risk.



Supplementary Figure S4. Survival plots for participants with different socioeconomic status. The x axis denotes the follow-up year, and the y axis denotes the probability of being free from diabetes.

Supplementary Table S2. Definitions of mediators and their associations with the risk of diabetes

Variable	Levels	HR (95% CI) for the risk of diabetes ^a	Missingness, <i>n</i>
Physiological factors			
Serum glycated hemoglobin	≥ 39 mmol/mol	1.00	23,020 (6.8%)
	< 39 mmol/mol	0.16 (0.15 to 0.16)	
Dyslipidemia	Yes	1.00	0 (0.0%)
	No	0.70 (0.67 to 0.73)	
Body mass index(19)	≥ 30 kg/m ²	1.00	1,243 (0.4%)
	< 30 kg/m ²	0.23 (0.23 to 0.24)	
Waist-to-hip ratio	≥ 0.90/0.85 (for men/women)	1.00	859 (0.3%)
	< 0.90/0.85	0.26 (0.24 to 0.27)	
Serum uric acid	> 7.6/6.6 mg/dL (for men/women)	1.00	21,187 (6.3%)
	≤ 7.6/6.6 mg/dL	0.38 (0.36 to 0.40)	
Resting pulse rate	< 60 or > 100 bpm	1.00	329 (0.1%)
	60-100 bpm	1.55 (1.47 to 1.63)	
C-reactive protein	≥ 5 mg/L	1.00	21,467 (6.4%)
	< 5 mg/L	0.37 (0.36 to 0.39)	
Sex hormone binding globulin	< 13.3/18.2/16.8 nmol/L for men/premenopausal women/postmenopausal women	1.00	50,493 (15.0%)
	≥ 13.3/18.2/16.8 nmol/L	0.81 (0.77 to 0.84)	
	< 169.06 or > 397.10*10 ³ cells/μL	1.00	
Platelet count	169.06-397.10*10 ³ cells/μL	0.82 (0.76 to 0.87)	15,063 (4.5%)
Hypertension	Yes	1.00	0 (0.0%)
	No	0.42 (0.40 to 0.44)	
Non-alcoholic fatty liver disease and liver malfunction	Yes	1.00	0 (0.0%)
	No	0.53 (0.51 to 0.55)	
Chronic kidney disease	Yes	1.00	0 (0.0%)
	No	0.59 (0.56 to 0.62)	
Hyperthyroidism and hypothyroidism	Yes	1.00	0 (0.0%)
	No	0.68 (0.64 to 0.73)	
Severe periodontal disease(4,5)	Yes	1.00	0 (0.0%)
	No	0.70 (0.64 to 0.75)	
Asthma	Yes	1.00	0 (0.0%)
	No	0.68 (0.65 to 0.72)	
Chronic obstructive pulmonary disease	Yes	1.00	0 (0.0%)
	No	0.55 (0.51 to 0.60)	

Inflammatory bowel disease	Yes	1.00	0 (0.0%)
	No	0.67 (0.58 to 0.78)	
Systemic autoimmune disorders	Yes	1.00	0 (0.0%)
	No	0.63 (0.52 to 0.76)	
Allergic rhinitis	Yes	1.00	0 (0.0%)
	No	1.09 (1.02 to 1.17)	
Rheumatoid arthritis	Yes	1.00	0 (0.0%)
	No	0.58 (0.51 to 0.66)	
Psoriasis	Yes	1.00	0 (0.0%)
	No	0.74 (0.66 to 0.82)	
Osteoarthritis	Yes	1.00	0 (0.0%)
	No	0.58 (0.56 to 0.61)	
Atopic dermatitis	Yes	1.00	0 (0.0%)
	No	0.92 (0.81 to 1.05)	
Behavioral factors			
Cardiovascular health diet score	0-4 points	1.00	3,257 (1.0%)
	5-10 points	0.80 (0.76 to 0.84)	
Healthy sleep score(7)	0-3 points	1.00	56,244 (16.7%)
	4-5 points	0.69 (0.66 to 0.71)	
Alcohol drinking(2,20)	Current heavy drinking (> 2/1 unit for men/women)	1.00	1,740 (0.5%)
	Current none or moderate drinking	1.47 (1.42 to 1.53)	
Leisure-time physical activity(21)	< 150 min/w of moderate-to-vigorous leisure-time physical activity	1.00	3,328 (1.0%)
	≥ 150 min/w of moderate-to-vigorous leisure-time physical activity	0.58 (0.56 to 0.61)	
Sedentary behavior	Top three quintiles	1.00	8,183 (2.4%)
	Bottom two quintiles	0.59 (0.57 to 0.62)	
Current active smoking	Yes	1.00	143 (0.0%)
	No	0.53 (0.50 to 0.56)	
Passive smoking	Yes	1.00	0 (0.0%)
	No	0.77 (0.74 to 0.80)	
Environmental factors			
Average sound level	Top three quintiles	1.00	4,979 (1.5%)
	Bottom two quintiles	0.95 (0.91 to 0.98)	
PM2.5	Top three quintiles	1.00	28,530 (8.5%)
	Bottom two quintiles	0.80 (0.77 to 0.83)	
PM10	Top three quintiles	1.00	28,530 (8.5%)
	Bottom two quintiles	0.92 (0.88 to 0.96)	
Nitrogen dioxide	Top three quintiles	1.00	4,979 (1.5%)

	Bottom two quintiles	0.78 (0.75 to 0.81)	
Percentage of the home location buffer classed as 'natural environment' at 1000m(22)	Top three quintiles	1.00	3,515 (1.0%)
	Bottom two quintiles	0.85 (0.82 to 0.89)	
Health deprivation score	Top three quintiles	1.00	8,611 (2.6%)
	Bottom two quintiles	0.56 (0.54 to 0.58)	
Psychological factors			
Visit friends/family(23)	Not live with relatives or visit friends/family once a week or less	1.00	17 (0.0%)
	Live with relatives or visit friends/family more than once a week	0.80 (0.75 to 0.84)	
Attend group activity(23)	Less than once a week	1.00	532 (0.2%)
	Once a week or more often	0.78 (0.75 to 0.82)	
Confide in someone close(23)	Once a week or less	1.00	8,953 (2.7%)
	More than once a week	0.86 (0.83 to 0.90)	
Stress event in the last two years	Yes	1.00	3,695 (1.1%)
	No	0.73 (0.70 to 0.76)	
Depression and anxiety	Yes	1.00	0 (0.0%)
	No	0.65 (0.63 to 0.68)	

Note. *CI*, confidence interval; *HR*, hazard ratio. ^a*HRs* for the risk of diabetes were estimated by Cox regression models, with adjustment of age, sex, and race. Factors will not be included in the physiological, behavioral, environmental, and psychological scores if their healthy levels are associated with an *HR* of > 1 for diabetes risk.

Supplementary Table S3. Odds ratios (95% confidence intervals) associated with socioeconomic status for the healthy levels of mediators

Mediators	Socioeconomic status		
	High	Medium	Low
Physiological factors			
Serum glycated hemoglobin	1.00	0.84 (0.82 to 0.86)	0.71 (0.69 to 0.73)
Dyslipidemia	1.00	0.99 (0.97 to 1.01)	0.91 (0.89 to 0.93)
Body mass index	1.00	0.78 (0.77 to 0.80)	0.57 (0.55 to 0.58)
Waist-to-hip ratio	1.00	0.81 (0.79 to 0.82)	0.60 (0.58 to 0.61)
Serum uric acid	1.00	1.00 (0.96 to 1.04)	0.82 (0.79 to 0.86)
Resting pulse rate ^a	1.00	1.17 (1.15 to 1.19)	1.33 (1.29 to 1.36)
C-reactive protein	1.00	0.74 (0.72 to 0.76)	0.49 (0.47 to 0.51)
Sex hormone binding globulin	1.00	0.98 (0.96 to 1.00)	0.95 (0.92 to 0.97)
Platelet count	1.00	0.96 (0.93 to 0.99)	0.87 (0.84 to 0.91)
Hypertension	1.00	0.85 (0.83 to 0.86)	0.70 (0.69 to 0.72)
Non-alcoholic fatty liver disease and liver malfunction	1.00	1.02 (1.00 to 1.05)	0.91 (0.89 to 0.94)
Chronic kidney disease	1.00	0.90 (0.88 to 0.93)	0.75 (0.73 to 0.78)
Hyperthyroidism and hypothyroidism	1.00	0.90 (0.87 to 0.94)	0.78 (0.75 to 0.82)
Severe periodontal disease	1.00	0.68 (0.64 to 0.71)	0.45 (0.43 to 0.48)
Asthma	1.00	1.02 (1.00 to 1.05)	0.91 (0.88 to 0.94)
Chronic obstructive pulmonary disease	1.00	0.76 (0.72 to 0.81)	0.43 (0.40 to 0.46)
Inflammatory bowel disease	1.00	0.92 (0.85 to 1.00)	0.77 (0.70 to 0.85)
Systemic autoimmune disorders	1.00	0.77 (0.68 to 0.88)	0.65 (0.57 to 0.75)
Allergic rhinitis ^a	1.00	1.11 (1.08 to 1.14)	1.41 (1.36 to 1.47)
Rheumatoid arthritis	1.00	0.67 (0.61 to 0.74)	0.40 (0.36 to 0.45)
Psoriasis	1.00	0.89 (0.84 to 0.94)	0.78 (0.73 to 0.84)
Osteoarthritis	1.00	0.75 (0.72 to 0.77)	0.51 (0.50 to 0.53)
Atopic dermatitis	1.00	0.94 (0.89 to 1.00)	0.87 (0.81 to 0.94)
Behavioral factors			
Cardiovascular health diet score	1.00	1.01 (0.98 to 1.03)	0.93 (0.90 to 0.96)
Healthy sleep score	1.00	0.95 (0.94 to 0.97)	0.85 (0.83 to 0.86)
Alcohol drinking ^a	1.00	1.55 (1.52 to 1.57)	2.40 (2.35 to 2.45)
Leisure-time physical activity	1.00	0.78 (0.76 to 0.79)	0.66 (0.64 to 0.67)
Sedentary behavior	1.00	0.70 (0.68 to 0.71)	0.47 (0.46 to 0.48)
Current active smoking	1.00	0.48 (0.46 to 0.50)	0.20 (0.20 to 0.21)
Passive smoking	1.00	0.73 (0.72 to 0.75)	0.60 (0.59 to 0.61)
Environmental factors			
Average sound level	1.00	0.94 (0.92 to 0.95)	0.88 (0.86 to 0.90)
PM2.5	1.00	0.93 (0.91 to 0.94)	0.78 (0.76 to 0.80)
PM10	1.00	0.93 (0.91 to 0.94)	0.78 (0.76 to 0.80)
Nitrogen dioxide	1.00	0.85 (0.83 to 0.86)	0.59 (0.58 to 0.60)
Natural environment	1.00	0.90 (0.88 to 0.92)	0.69 (0.67 to 0.70)
Health deprivation score	1.00	0.38 (0.37 to 0.38)	0.15 (0.14 to 0.15)

Psychological factors			
Visit friends/family	1.00	0.30 (0.28 to 0.31)	0.17 (0.17 to 0.18)
Attend group activity	1.00	0.89 (0.88 to 0.91)	0.65 (0.63 to 0.66)
Confide in someone close	1.00	0.73 (0.72 to 0.74)	0.54 (0.53 to 0.56)
Stress event in two years	1.00	0.74 (0.72 to 0.75)	0.58 (0.56 to 0.59)
Depression and anxiety	1.00	0.69 (0.68 to 0.70)	0.48 (0.47 to 0.49)

Note. Logistic regression models with adjustment of age, sex, and race were used. Each mediator was included as the dependent variable, and the probability of the healthy level of the mediator was modeled.

^a These factors were not included in the final scores because their healthy levels were not more prevalent among participants with high socioeconomic status than those with low socioeconomic status.

Supplementary Table S4. Baseline characteristics among participants with missing socioeconomic factors and those included in the analyses

Socioeconomic status	Excluded	Included	SMD
Total, <i>n</i> (%)	57,281	337,229	-
Mean age (<i>SD</i>), <i>y</i>	57.8 (8.0)	55.3 (8.1)	31.7
Men, <i>n</i> (%)	19,194 (33.5)	156,297 (46.3)	-26.4
Non-white, <i>n</i> (%)	4,949 (8.6)	15,590 (4.6)	16.2
Family history of diabetes, <i>n</i> (%)	12,313 (21.5)	70,703 (21.0)	1.3
Annual income of ≥ £31,000, <i>n</i> (%)	640 (32.4)	185,990 (55.2)	-47.2
Skilled job, <i>n</i> (%)	16,248 (29.6)	180,903 (53.6)	-50.3
College or university degree, <i>n</i> (%)	10,325 (19.6)	123,720 (36.7)	-38.7
Serum glycated hemoglobin of < 39 mmol/mol, <i>n</i> (%)	44,100 (77.0)	276,139 (81.9)	-12.1
No prevalent dyslipidemia, <i>n</i> (%)	30,646 (53.5)	195,092 (57.9)	-8.8
Body mass index of < 30 kg/m ² , <i>n</i> (%)	44,051 (76.9)	264,956 (78.6)	-4.0
Waist-to-hip ratio of < 0.90/0.85 for men/women, <i>n</i> (%)	31,648 (55.3)	181,193 (53.7)	3.1
Serum uric acid of ≤ 7.6/6.6 mg/dL, <i>n</i> (%)	53,974 (94.2)	317,327 (94.1)	0.5
C-reactive protein < 5 mg/L, <i>n</i> (%)	50,747 (88.6)	304,791 (90.4)	-5.8
Normal sex hormone binding globulin level, <i>n</i> (%)	47,596 (83.1)	283,294 (84.0)	-2.5
Platelet count of 169.06-397.10*10 ³ cells/μL, <i>n</i> (%)	53,716 (93.8)	316,186 (93.8)	0.1
No prevalent hypertension, <i>n</i> (%)	22,999 (40.2)	153,575 (45.5)	-10.9
No prevalent non-alcoholic fatty liver disease or liver malfunction, <i>n</i> (%)	48,008 (83.8)	278,091 (82.5)	3.6
No prevalent chronic kidney disease, <i>n</i> (%)	52,887 (92.3)	295,480 (87.6)	15.7
No prevalent hyperthyroidism or hypothyroidism, <i>n</i> (%)	52,965 (92.5)	318,064 (94.3)	-7.5
No severe periodontal disease, <i>n</i> (%)	54,758 (95.6)	324,181 (96.1)	-2.7
No prevalent asthma, <i>n</i> (%)	50,453 (88.1)	296,328 (87.9)	0.6
No prevalent chronic obstructive pulmonary disease, <i>n</i> (%)	55,097 (96.2)	327,158 (97.0)	-4.6
No prevalent inflammatory bowel disease, <i>n</i> (%)	56,635 (98.9)	333,619 (98.9)	-0.5
No prevalent systemic autoimmune disorders, <i>n</i> (%)	56,801 (99.2)	335,184 (99.4)	-2.7
No prevalent rheumatoid arthritis, <i>n</i> (%)	56,341 (98.4)	333,517 (98.9)	-4.6
No prevalent psoriasis, <i>n</i> (%)	56,182 (98.1)	330,205 (97.9)	1.2
No prevalent osteoarthritis, <i>n</i> (%)	48,637 (84.9)	300,074 (89.0)	-12.1
No prevalent atopic dermatitis, <i>n</i> (%)	56,153 (98.0)	330,494 (98.0)	0.2
Cardiovascular health diet score of ≥ 5 points, <i>n</i> (%)	22,701 (39.6)	132,925 (39.4)	0.4
Healthy sleep score of 4-5 points, <i>n</i> (%)	21,946 (38.3)	134,907 (40.0)	-3.5
≥ 150 minutes/week of moderate to vigorous leisure-time physical activity, <i>n</i> (%)	20,780 (36.3)	124,341 (36.9)	-1.2
Bottom two quintiles of sedentary time, <i>n</i> (%)	21,187 (37.0)	125,467 (37.2)	-0.5
No current active smoking, <i>n</i> (%)	22,465 (39.2)	134,774 (40.0)	-1.5
No current passive smoking, <i>n</i> (%)	20,712 (36.2)	137,666 (40.8)	-9.6
Current no or moderate alcohol drinking, <i>n</i> (%)	10,204 (17.8)	59,573 (17.7)	0.4
Bottom two quintiles of average sound level, <i>n</i> (%)	39,509 (69.0)	224,202 (66.5)	5.3
Bottom two quintiles of nitrogen dioxide, <i>n</i> (%)	19,428 (33.9)	127,476 (37.8)	-8.1
Bottom two quintiles of PM10, <i>n</i> (%)	17,489 (30.5)	117,505 (34.8)	-9.2

Bottom two quintiles of PM2.5, <i>n</i> (%)	52,791 (92.2)	311,865 (92.5)	-1.2
Top two quintiles of the percentage of the home location buffer at 1 km classed as natural environment, <i>n</i> (%)	37,128 (64.8)	238,016 (70.6)	-12.3
Bottom two quintiles of health deprivation score, <i>n</i> (%)	34,607 (60.4)	168,726 (50.0)	21.0
Live with relatives or visit friends/family more than once a week, <i>n</i> (%)	50,920 (88.9)	303,396 (90.0)	-3.5
Attend social group activity once a week or more often, <i>n</i> (%)	39,481 (68.9)	237,853 (70.5)	-3.5
Confide in someone close to you more than once a week, <i>n</i> (%)	37,682 (65.8)	218,946 (64.9)	1.8
No stress events in last 2 years, <i>n</i> (%)	34,518 (60.3)	193,070 (57.3)	6.1
No depression, <i>n</i> (%)	34,912 (60.9)	205,139 (60.8)	0.2

Note. *CI*, confidence interval; *HR*, hazard ratio. ^aFour scores were mutually adjusted.

Supplementary Table S5. Associations between physiological, behavioral, environmental, and psychological scores with incident diabetes

Scores ^a	Bottom tertile	Middle tertile	Top tertile
Physiological score			
Score range	9 to 17	18 to 19	20 to 21
<i>HR</i> (95% <i>CI</i>)	1.00	0.25 (0.24 to 0.26)	0.08 (0.07 to 0.09)
Behavioral score			
Score range	1 to 2	3	4 to 6
<i>HR</i> (95% <i>CI</i>)	1.00	0.79 (0.76 to 0.82)	0.60 (0.58 to 0.63)
Environmental score			
Score range	0 to 1	2 to 3	4 to 6
<i>HR</i> (95% <i>CI</i>)	1.00	0.93 (0.89 to 0.97)	0.85 (0.81 to 0.89)
Psychological score			
Score range	0 to 2	3 to 4	5
<i>HR</i> (95% <i>CI</i>)	1.00	0.83 (0.79 to 0.86)	0.71 (0.67 to 0.76)

Note. *CI*, confidence interval; *HR*, hazard ratio. ^aFour scores were mutually adjusted.

Supplementary Table S6. Proportion of the association between low (vs. high) socioeconomic status and diabetes risk mediated by the mediators (product method)

Mediator	Mediation (%; 95% <i>CI</i>)
Physiological score	35.3 (33.1 to 38.1)
Behavioral score	18.8 (17.3 to 20.6)
Environmental score	6.6 (5.5 to 7.7)
Psychological score	11.7 (10.5 to 13.0)
Four scores jointly	51.9 (48.0 to 56.1)

Note. Base model adjusted for age, sex, race/ethnicity, alcohol drinking, and family history of diabetes. To calculate the mediation proportion for each score, first, a linear regression model was fitted for each mediator (each score) regressing on the socioeconomic status, with adjustment of the covariates in the base model. α_2 is the coefficient for low vs. high socioeconomic status from the model. Second, the direct effect of socioeconomic status was estimated by constructing the Cox regression model which included socioeconomic status, the mediator, and all covariates in the base model. *b* is the coefficient for the mediator. Third, the total effect of socioeconomic status was estimated by constructing the Cox regression

model which included socioeconomic status and all covariates in the base model. c_2 is the coefficient for low vs. high SES, representing the total effect on the log hazard ratio scale. The indirect (mediated) effect on the log hazard ratio scale is $IE = \alpha_2 * b$, representing the log hazard ratio for socioeconomic status that operates through the mediator. Mediation proportion is $IE/c_2 = \alpha_2 * b / c_2$. To calculate the median proportion for four scores jointly, first, a linear regression model was fitted for each mediator (each score) regressing on the socioeconomic status, with adjustment of the covariates in the base model. α_{2k} is the coefficient for low vs. high socioeconomic status for score_k. Second, the direct effect of socioeconomic status was estimated by constructing the Cox regression model which included socioeconomic status, all mediators (four scores), and all covariates in the base model. b_k is the coefficient for score_k. Third, the total effect of socioeconomic status was estimated by constructing the Cox regression model which included socioeconomic status and all covariates in the base model. c_2 is the coefficient for low vs. high SES, representing the total effect on the log hazard ratio scale. The indirect (mediated) effect on the log hazard ratio scale is $IE = \sum_{k=1}^4 (a_{2k} * b_k)$, representing the log hazard ratio for socioeconomic status that operates through the mediator. Mediation proportion is $IE/c_2 = \sum_{k=1}^4 (a_{2k} * b_k) / c_2$. All confidence intervals were estimated by bootstrap resampling 1000 times.

Supplementary Table S7. Associations of socioeconomic status with the risk of diabetes (sensitivity analyses)

Analyses and models ^a	Socioeconomic status		
	High	Medium	Low
Excluding incident diabetes cases occurring within the first five years of follow-up			
Base model			
<i>HR (95% CI)</i>	1.00	1.55 (1.46 to 1.65)	2.60 (2.43 to 2.78)
IRD, cases per 1000 PYs (95% CI)	1.00	0.6 (0.5 to 0.7)	2.1 (2.0 to 2.3)
Base model + all mediators			
<i>HR (95% CI)</i>	1.00	1.29 (1.21 to 1.38)	1.61 (1.50 to 1.73)
Mediation, % (95% CI)	-	41.5 (35.1 to 48.3)	50.0 (46.5 to 53.6)
Type 2 diabetes			
Base model			
<i>HR (95% CI)</i>	1.00	1.53 (1.45 to 1.62)	2.57 (2.42 to 2.74)
IRD, cases per 1000 PYs (95% CI)	1.00	0.7 (0.6 to 0.8)	2.6 (2.4 to 2.7)
Base model + all mediators			
<i>HR (95% CI)</i>	1.00	1.27 (1.19 to 1.34)	1.55 (1.45 to 1.65)
Mediation, % (95% CI)	-	44.9 (38.2 to 51.7)	53.7 (50.1 to 57.2)
Excluding serum glycosylated hemoglobin level from physiological score			
Base model <i>HR (95% CI)</i>	1.00	1.41 (1.33 to 1.49)	2.22 (2.10 to 2.36)
Base model + all mediators			
<i>HR (95% CI)</i>	1.00	1.22 (1.15 to 1.29)	1.52 (1.43 to 1.61)
Mediation, % (95% CI)	-	42.7 (35.3 to 50.4)	47.6 (44.1 to 51.2)
Using separate risk factors			
Base model <i>HR (95% CI)</i>	1.00	1.49 (1.41 to 1.57)	2.47 (2.33 to 2.62)
Base model + all mediators			
<i>HR (95% CI)</i>	1.00	1.17 (1.11 to 1.24)	1.43 (1.34 to 1.52)
Mediation, % (95% CI)	-	60.8 (51.2 to 69.6)	60.8 (56.7 to 64.8)

Income			
Base model			
<i>HR (95% CI)</i>	1.00	1.42 (1.34 to 1.50)	1.87 (1.76 to 2.00)
IRD, cases per 1000 PYs (95% CI)	1.00	0.6 (0.5 to 0.7)	1.8 (1.6 to 2.0)
Base model + all mediators			
<i>HR (95% CI)</i>	1.00	1.21 (1.14 to 1.28)	1.35 (1.26 to 1.44)
Mediation, % (95% CI)	-	61.6 (50.6 to 71.6)	62.8 (56.6 to 68.6)
Education			
Base model			
<i>HR (95% CI)</i>	1.00	1.15 (1.10 to 1.21)	1.49 (1.41 to 1.56)
IRD, cases per 1000 PYs (95% CI)	1.00	0.3 (0.2 to 0.5)	1.6 (1.4 to 1.8)
Base model + all mediators			
<i>HR (95% CI)</i>	1.00	1.08 (1.03 to 1.13)	1.19 (1.13 to 1.25)
Mediation, % (95% CI)	-	57.1 (36.4 to 75.6)	64.3 (55.8 to 71.9)
Occupation			
Base model			
<i>HR (95% CI)</i>	1.00	1.03 (0.98 to 1.08)	1.63 (1.52 to 1.74)
IRD, cases per 1000 PYs (95% CI)	1.00	0.0 (-0.1 to 0.2)	1.7 (1.4 to 2.0)
Base model + all mediators			
<i>HR (95% CI)</i>	1.00	1.02 (0.97 to 1.07)	1.18 (1.10 to 1.26)
Mediation, % (95% CI)	-	No mediation	53.2 (47.6 to 58.8)

Note. *CI*, confidence interval; *HR*, hazard ratio; *IRD*, incidence rate difference; *PY*, person-year. ^aBase model adjusted for age, sex, race, alcohol drinking, and family history of diabetes. When the physiological score does not contain serum glycated hemoglobin level, serum glycated hemoglobin level was additionally adjusted. Mediators included physiological, behavioral, environmental, and psychological factors. Multiplicative interactions compared the HRs comparing high versus low socioeconomic status between subgroups, and additive interactions compared the IRDs comparing high versus low socioeconomic status between subgroups.

Supplementary Table S8. HRs and IRDs comparing top (healthiest) versus bottom (least-healthy) risk factor scores by socioeconomic status (sensitivity analyses)

Sensitivity analyses ^a	High SES	Medium SES	Low SES	<i>P</i> for interaction
Sensitivity analyses 1				
Physical score				
IRD (95% CI)	-2.6 (-2.8 to -2.4)	-3.7 (-3.8 to -3.5)	-5.7 (-6.0 to -5.4)	< 0.001
HR (95% CI)	0.07 (0.05 to 0.09)	0.09 (0.08 to 0.11)	0.09 (0.07 to 0.12)	0.035
Behavioral score				
IRD (95% CI)	-0.7 (-0.9 to -0.5)	-0.9 (-1.1 to -0.8)	-1.7 (-2.0 to -1.4)	< 0.001
HR (95% CI)	0.56 (0.49 to 0.65)	0.62 (0.58 to 0.68)	0.63 (0.57 to 0.69)	0.22
Environmental score				
IRD (95% CI)	-0.2 (-0.3 to 0.0)	-0.3 (-0.5 to -0.2)	-0.4 (-0.7 to -0.1)	< 0.001
HR (95% CI)	0.87 (0.76 to 0.99)	0.83 (0.77 to 0.90)	0.87 (0.80 to 0.96)	0.84
Psychological score				
IRD (95% CI)	-0.5 (-0.7 to -0.2)	-0.7 (-0.9 to -0.5)	-1.0 (-1.4 to -0.6)	< 0.001
HR (95% CI)	0.69 (0.57 to 0.83)	0.70 (0.63 to 0.78)	0.75 (0.66 to 0.85)	0.55
Sensitivity analyses 2				
Physical score				
IRD (95% CI)	-3.3 (-3.6 to -3.1)	-4.7 (-4.8 to -4.5)	-7.2 (-7.5 to -6.8)	< 0.001
HR (95% CI)	0.06 (0.05 to 0.08)	0.08 (0.07 to 0.09)	0.08 (0.06 to 0.09)	0.055
Behavioral score				
IRD (95% CI)	-0.9 (-1.1 to -0.7)	-1.2 (-1.4 to -1.0)	-2.1 (-2.5 to -1.8)	< 0.001
HR (95% CI)	0.54 (0.48 to 0.61)	0.60 (0.56 to 0.65)	0.61 (0.56 to 0.66)	0.18
Environmental score				
IRD (95% CI)	-0.2 (-0.4 to 0.0)	-0.5 (-0.6 to -0.3)	-0.5 (-0.9 to -0.2)	< 0.001
HR (95% CI)	0.87 (0.77 to 0.98)	0.80 (0.75 to 0.86)	0.88 (0.81 to 0.95)	0.86
Psychological score				
IRD (95% CI)	-0.5 (-0.8 to -0.2)	-0.8 (-1.0 to -0.6)	-1.4 (-1.9 to -0.9)	< 0.001
HR (95% CI)	0.73 (0.62 to 0.86)	0.70 (0.64 to 0.77)	0.72 (0.64 to 0.80)	0.63
Sensitivity analyses 3				
Physical score				
IRD (95% CI)	-3.6 (-3.9 to -3.4)	-5.0 (-5.2 to -4.8)	-7.2 (-7.6 to -6.9)	<0.001
HR (95% CI)	0.07 (0.06 to 0.09)	0.09 (0.08 to 0.10)	0.08 (0.07 to 0.11)	0.12
Behavioral score				
IRD (95% CI)	-0.8 (-1.0 to -0.6)	-1.4 (-1.5 to -1.2)	-1.9 (-2.3 to -1.5)	<0.001
HR (95% CI)	0.61 (0.54 to 0.68)	0.60 (0.56 to 0.64)	0.65 (0.60 to 0.71)	0.22
Environmental score				
IRD (95% CI)	-0.2 (-0.4 to -0.1)	-0.6 (-0.7 to -0.4)	-0.2 (-0.6 to 0.1)	0.10
HR (95% CI)	0.87 (0.78 to 0.98)	0.80 (0.75 to 0.86)	0.94 (0.86 to 1.03)	0.47
Psychological score				
IRD (95% CI)	-0.7 (-1.0 to -0.4)	-0.8 (-1.1 to -0.6)	-1.3 (-1.8 to -0.8)	< 0.001
HR (95% CI)	0.68 (0.58 to 0.80)	0.74 (0.68 to 0.81)	0.74 (0.66 to 0.84)	0.082
Sensitivity analyses 4				

Physical score				
IRD (95% CI)	-4.5 (-4.7 to -4.4)	-5.4 (-5.7 to -5.1)	-7.0 (-7.4 to -6.6)	< 0.001
HR (95% CI)	0.08 (0.07 to 0.10)	0.07 (0.05 to 0.08)	0.12 (0.09 to 0.15)	0.027
Behavioral score				
IRD (95% CI)	-1.3 (-1.4 to -1.1)	-1.1 (-1.3 to -0.8)	-2.3 (-2.8 to -1.8)	< 0.001
HR (95% CI)	0.59 (0.56 to 0.63)	0.67 (0.61 to 0.74)	0.61 (0.55 to 0.68)	0.51
Environmental score				
IRD (95% CI)	-0.5 (-0.6 to -0.3)	-0.3 (-0.5 to 0.0)	-0.4 (-0.9 to 0.1)	0.26
HR (95% CI)	0.82 (0.77 to 0.87)	0.89 (0.81 to 0.98)	0.92 (0.83 to 1.01)	0.062
Psychological score				
IRD (95% CI)	-0.7 (-0.9 to -0.5)	-1.2 (-1.5 to -0.8)	-1.0 (-1.7 to -0.4)	0.024
HR (95% CI)	0.73 (0.67 to 0.80)	0.64 (0.55 to 0.73)	0.82 (0.72 to 0.94)	0.18
Sensitivity analyses 5				
Physical score				
IRD (95% CI)	-4.5 (-4.7 to -4.3)	-5.3 (-5.6 to -5.1)	-7.9 (-8.6 to -7.3)	< 0.001
HR (95% CI)	0.07 (0.06 to 0.08)	0.11 (0.09 to 0.13)	0.06 (0.04 to 0.09)	0.65
Behavioral score				
IRD (95% CI)	-1.0 (-1.2 to -0.9)	-1.6 (-1.9 to -1.4)	-1.7 (-2.4 to -1.0)	< 0.001
HR (95% CI)	0.61 (0.56 to 0.65)	0.62 (0.58 to 0.66)	0.64 (0.54 to 0.76)	0.90
Environmental score				
IRD (95% CI)	-0.3 (-0.5 to -0.2)	-0.5 (-0.7 to -0.3)	-0.6 (-1.2 to 0.1)	0.007
HR (95% CI)	0.85 (0.79 to 0.91)	0.85 (0.80 to 0.91)	0.85 (0.73 to 1.00)	0.73
Psychological score				
IRD (95% CI)	-0.7 (-0.9 to -0.5)	-0.9 (-1.2 to -0.6)	-0.9 (-1.8 to -0.1)	0.002
HR (95% CI)	0.70 (0.64 to 0.78)	0.76 (0.69 to 0.83)	0.73 (0.56 to 0.95)	0.77

Note. *CI*, confidence interval; *HR*, hazard ratio; *IRD*, incidence rate difference. ^aAdditive interactions compared the incidence rate differences comparing top versus bottom tertile of risk factor scores between high and low socioeconomic status groups, which were evaluated by additive hazard models. Multiplicative interactions compared the hazard ratios comparing top versus bottom tertile of risk factor scores between high and low socioeconomic status groups, which were evaluated by Cox regression models. Sensitivity analysis 1 excluded incident diabetes cases occurring within the first five years of follow-up. Sensitivity analysis 2 treated type 2 diabetes as the outcome. Sensitivity analyses 3 to 5 substituted socioeconomic status with income, education, and occupation, and these three individual socioeconomic factors were mutually adjusted.

REFERENCES

1. Zhang YB, Chen C, Pan XF, et al. Associations of healthy lifestyle and socioeconomic status with mortality and incident cardiovascular disease: two prospective cohort studies. *BMJ*, 2021; 373, n604
2. Harrison S, Davies AR, Dickson M, et al. The causal effects of health conditions and risk factors on social and socioeconomic outcomes: Mendelian randomization in UK Biobank. *Int J Epidemiol*, 2020; 49, 1661-81.
3. Lee JJ, Wedow R, Okbay A, et al. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat Genet*, 2018; 50, 1112-21.
4. Larvin H, Wilmott S, Kang J, et al. Additive Effect of Periodontal Disease and Obesity on COVID-19 Outcomes. *J Dent Res*, 2021; 100, 1228-35.
5. Watson S, Woodside JV, Winning L, et al. Associations between self-reported periodontal disease and nutrient intakes and nutrient-based dietary patterns in the UK Biobank. *J Clin Periodontol*, 2022; 49, 428-38.
6. Beckman Coulter. Technical Documents for Beckman Coulter Diagnostics. 2023. <https://www.beckmancoulter.com/en/support/tech-docs>. [2023-5-27]
7. Geng T, Li X, Ma H, Heianza Y, Qi L. Adherence to a Healthy Sleep Pattern and Risk of Chronic Kidney Disease: The UK Biobank Study. *Mayo Clin Proc*, 2022; 97, 68-77.
8. Levey AS, Stevens LA, Schmid CH, et al. A new equation to estimate glomerular filtration rate. *Ann Intern Med*, 2009; 150, 604-12.
9. Said MA, Verweij N, van der Harst P. Associations of combined genetic and lifestyle risks with incident cardiovascular disease and diabetes in the UK Biobank study. *JAMA Cardiol*, 2018; 3, 693-702.
10. Kephelopoulos S, Paviotti M, Anfosso-Ledee F, et al. Advances in the development of common noise assessment methods in Europe: The CNOSSOS-EU framework for strategic environmental noise mapping. *Sci Total Environ*, 2014; 482-483, 400-410.
11. Beelen R, Hoek G, Vienneau D, et al. Development of NO₂ and NO_x land use regression models for estimating air pollution exposure in 36 study areas in Europe – The ESCAPE project. *Atmos Environ*, 2013; 72, 10-23.
12. Eeftens M, Beelen R, de Hoogh K, et al. Development of Land Use Regression models for PM_{2.5}, PM_{2.5} absorbance, PM₁₀ and PM_{coarse} in 20 European study areas; results of the ESCAPE project. *Environ Sci Technol*, 2012; 46, 11195-205.
13. Wheeler B. UK Biobank Project 15374: Documentation of environmental indicators attributed to participants based on home location grid references. 2017. <https://biobank.ndph.ox.ac.uk/showcase/ukb/docs/App15374Docs.pdf>. [2022-8-18].
14. Morton D, Rowland C, Wood C, et al. Final Report for LCM2007 – the new UK Land Cover Map. 2011. <https://www.ceh.ac.uk/sites/default/files/LCM2007%20Final%20Report.pdf>. [2022-8-18].
15. UK Biobank. Index of Multiple Deprivation Scores (Version 1.0). 2018. https://biobank.ndph.ox.ac.uk/showcase/ukb/docs/imd_baseline.pdf. [2022-8-4].
16. Cai N, Revez JA, Adams MJ, et al. Minimal phenotyping yields genome-wide association signals of low specificity for major depression. *Nat Genet*, 2020; 52, 437-47.
17. Hu W, Chu J, Chen X, et al. The role of depression and physical activity in the association

- of between sleep quality, and duration with and health-related quality of life among the elderly: a UK Biobank cross-sectional study. *BMC Geriatr* 2022;22:338
18. Odegaard AO, Koh WP, Gross MD, Yuan JM, Pereira MA. Combined lifestyle factors and cardiovascular disease mortality in Chinese men and women: the Singapore Chinese health study. *Circulation* 2011;124:2847-2854
19. WHO Consultation on Obesity, World Health Organization. Obesity: preventing and managing the global epidemic: report of a WHO consultation. 2000. <https://apps.who.int/iris/handle/10665/42330>. [2021-7-21].
20. National Health Service. The risks of drinking too much. 2019. <https://www.nhs.uk/live-well/alcohol-support/the-risks-of-drinking-too-much/>. [2021-12-9].
21. National Health Service. Benefits of exercise. 2022. <https://www.nhs.uk/live-well/exercise/exercise-health-benefits/>. [2022-8-4].
22. Zhao Y, Li Y, Zhuang Z, et al. Associations of polysocial risk score, lifestyle and genetic factors with incident type 2 diabetes: a prospective cohort study. *Diabetologia*, 2022; 65, 2056-65.
23. Hakulinen C, Pulkki-Raback L, Virtanen M, et al. Social isolation and loneliness as risk factors for myocardial infarction, stroke and mortality: UK Biobank cohort study of 479054 men and women. *Heart*, 2018; 104, 1536-42.