Genetic predisposition, modifiable lifestyles, and their joint effects on human lifespan: evidence from multiple cohort studies

Zilong Bian ,[1,2] Lijuan Wang,[1,3] Rong Fan,[1,2] Jing Sun,[1] Lili Yu,[1] Meihong Xu,[4] Paul R H J Timmers,[3,5] Xia Shen,[6] James F Wilson,[3,5] Evropi Theodoratou,[3,7] Xifeng Wu,[1,8] Xue Li[1,3]

10.1136/bmjebm-2023-112583

► Additional supplemental material is published online only. To view, please visit the journal online (https://doi.org/10.1136/bmjebm-2023-112583).

For numbered affiliations see end of article.

Correspondence to:
Professor Xifeng Wu and Dr Xue Li, Department of Big Data in Health Science, School of Public Health and Center of Clinical Big Data and Analytics of The Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China; xifengw@zju.edu.cn, xueli157@zju.edu.cn

ZB, LW and RF are joint first authors.

To cite: Bian Z, Wang L, Fan R, *et al*. *BMJ Evidence-Based Medicine* Epub ahead of print: [please include Day Month Year]. doi:10.1136/bmjebm-2023-112583

Abstract

Objective To investigate the associations across genetic and lifestyle factors with lifespan.

Design A longitudinal cohort study.

Setting UK Biobank.

Participants 353 742 adults of European ancestry, who were recruited from 2006 to 2010 and were followed up until 2021.

Exposures A polygenic risk score for lifespan with long (<lowest quintile), intermediate (quintiles 2 to 4), and short (>highest quintile) risk categories and a weighted healthy lifestyle score, including no current smoking, moderate alcohol consumption, regular physical activity, healthy body shape, adequate sleep duration, and a healthy diet, categorised into favourable, intermediate, and unfavourable lifestyles.

Main outcome measures Lifespan defined as the date of death or the censor date minus the date of birth.

Results Of the included 353 742 participants of European ancestry with a median follow-up of 12.86 years, 24 239 death cases were identified. Participants were grouped into three genetically determined lifespan categories including long (20.1%), intermediate (60.1%), and short (19.8%), and into three lifestyle score categories including favourable (23.1%), intermediate (55.6%), and unfavourable (21.3%). The hazard ratio (HR) of death for individuals with a genetic predisposition to a short lifespan was 1.21 (95% CI 1.16 to 1.26) compared to those with a genetic predisposition to a long lifespan. The HR of death for individuals in the unfavourable lifestyle category was 1.78 (95% CI 1.71 to 1.85), compared with those in the favourable lifestyle category. Participants with a genetic predisposition to a short lifespan and an unfavourable lifestyle had 2.04 times (95% CI 1.87 to 2.22) higher rates of death compared with those with a genetic predisposition to a long lifespan and a favourable lifestyle. No multiplicative interaction was detected between the polygenic risk score of lifespan and the weighted healthy lifestyle score (p=0.10). The optimal combination of healthy lifestyles, including never smoking, regular physical activity, adequate sleep duration, and a healthy diet, was derived to decrease risk of premature death (death before 75 years).

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ It is well established that a shorter lifespan or premature death could be ascribed to modifiable lifestyle factors, specifically tobacco use, alcohol consumption, diet quality, and physical activity. A health-conscious lifestyle might have great potential to assuage the genetic susceptibility towards a shorter lifespan.
⇒ There has been no investigation to probe the joint effects of lifestyle factors and genetic determinants on human lifespan.
⇒ The extent to which a healthy lifestyle could counterbalance the high genetic risk remains elusive.

WHAT THIS STUDY ADDS

⇒ A high genetic risk corresponded to a 21% increased risk of death compared with a low genetic risk independent of lifestyle factors.
⇒ Genetic and lifestyle factors manifested independent associations with lifespan.
⇒ Adherence to healthy lifestyles could largely attenuate the genetic risk of shorter lifespan or premature death.

HOW MIGHT THIS STUDY AFFECT RESEARCH, PRACTICE OR POLICY

⇒ This study elucidates the pivotal role of a healthy lifestyle in mitigating the impact of genetic factors on lifespan reduction.
⇒ Given that our analysis was confined to white-European ancestry, the generalisability of our findings should be further evaluated in more diverse populations.
⇒ Public health policies for improving healthy lifestyles would serve as potent complements to conventional healthcare and mitigate the influence of genetic factors on human lifespan.

Conclusion Genetic and lifestyle factors were independently associated with lifespan. Adherence to healthy lifestyles could largely attenuate the genetic risk of a shorter lifespan or premature death. The optimal combination of healthy lifestyles could convey better benefits for a longer lifespan, regardless of genetic background.

Introduction

Human lifespan is modulated by a combination of genetic and non-genetic factors including lifestyle behaviours.[1] The heritability of lifespan has been estimated to be around 16% according to a study with sufficient global scope conducted on large genealogical trees.[2] Apolipoprotein E (*APOE*) has been perceived as a longevity gene, which was identified as the top associated locus at genome-wide significance and consistently replicated in several studies.[3–7] Other genetic loci such as *CHRNA 3/5*, *LPA*, *CDKN2B-AS1* and *LDLR* are additionally identified to be associated with lifespan from recent genome-wide association study (GWAS) meta-analyses.[7–9] Although single genetic variant accounts for only a small fraction of the variability of human lifespan, the polygenic risk score (PRS) combining multiple loci together provides a measurement of the predisposition for longer lifespan and more potential clinical utility.[8]

It is well established that shorter lifespan or premature death could be attributable to modifiable lifestyle factors, in particular tobacco use, alcohol consumption, diet quality, and physical activity.[10–16] A healthy lifestyle may be able to attenuate the genetic risk of shorter lifespan. Researchers revealed strong negative correlations between body fat, smoking, and susceptibility to coronary artery disease and longer lifespan.[16] However, there is no study to examine the joint effects of lifestyle factors and genetic determinants on the human lifespan. The extent to which individuals with high genetic risk can be offset by a healthy lifestyle remains elusive.

In this study, we incorporated data from three large population-based cohorts (LifeGen, US National Health and Nutrition Examination Survey (NHANES), and UK Biobank) to create a polygenic risk score to capture the genetic susceptibility associated with human lifespan, to assess the influence of common lifestyle factors (ie, smoking, alcohol consumption, diet, physical activity, body shape, and sleep duration), and to investigate the joint effects across genetic and lifestyle factors on human lifespan.

Methods

Study design and data sources

This study adopted a multi-staged design by incorporating data from the LifeGen,[16] US NHANES,[17] and UK Biobank cohorts.[18] The GWAS statistics derived from the LifeGen cohort, which included 26 independent European-heritage population cohorts with data on 606 059 parental lifespans,[16] were used for the construction of lifespan PRS. The healthy lifestyle scores (HLS) associated with human lifespan were generated using data from the 2005–2018 NHANES. The established PRS and HLS were then adopted in the independent UK Biobank cohort to assess the joint effects across genetic and lifestyle factors. Details on the study populations and data preparation are described in the online supplemental methods and our previous studies.[8] Local institutional review board ethics approval was not necessary for this study. All participants provided informed consent at the baseline assessment.

Polygenic risk score of lifespan

Independent genetic variants, captured by the LifeGen GWAS to be associated with human lifespan (p<5e-08) without linkage disequilibrium (r^2<0.001), were used for the construction of PRS (online supplemental table 1).[8] A polygenic risk score for lifespan was constructed for all individuals in the UK Biobank by multiplying the number of lifespan-decreasing alleles for each single nucleotide polymorphism (SNP) by its effect size on lifespan and then summing up this weighted score for all used SNPs. This PRS was then used to categorise the UK Biobank participants into long (<lowest quintile), intermediate (quintiles 2 to 4), and short (>highest quintile) groups to present the genetically determined human lifespan. The genotyped *APOE* SNPs rs429358 and rs7412, which possessed the largest genetic effect on lifespan, were further stratified to examine how the *APOE ε4* status, polygenic risk, and lifestyles interplay together to influence the human lifespan in the UK Biobank participants.

Healthy lifestyle score

We adopted six common lifestyle factors associated with lifespan according to previous evidence,[19 20] that is, smoking, alcohol consumption, physical activity, body shape, sleep duration, and diet. The definitions of lifestyle factors for NHANES and UK Biobank cohorts are shown with full details in online supplemental methods and table 2. Effect estimates (β coefficient) derived from NHANES were used for the construction of a weighted and standardised HLS in the UK Biobank participants. Briefly, we performed the Cox proportional hazards regression model in US NHANES to obtain β coefficients of each lifestyle factor with adjustment for other lifestyle factors and available covariates. Then, a weighted and standardised HLS was constructed in UK Biobank based on the β coefficient of each lifestyle factor derived from US NHANES. This score was then used to categorise UK Biobank participants into unfavourable (<lowest quintile), intermediate (quintiles 2–4), and favourable (>highest quintiles) groups.

Lifespan ascertainment

Lifespan was defined as the date of death minus the date of birth or the sum of age at baseline and follow-up time. In US NHANES, death certificate records were linked by the National Center for Health Statistics through the National Death Index to 31 December 2019. In our analysis, data for the lifespan of survivors in UK Biobank were censored on 31 December 2021. Death event was ascertained using the *International Classification of Diseases, Tenth Revision* (ICD-10) coding system and obtained from data field 40 000 and 40 001. Deaths due to accidents and injuries or COVID-19 were excluded.

Covariates

Information on covariates, including age (continuous in years), sex (men and women), education attainment (college or university degree and above, and high school and below), and socioeconomic status, were collected in the baseline questionnaire. The Townsend deprivation index as a complex indicator of socioeconomic status was constructed using the method mentioned online (https://biobank.ndph.ox.ac.uk/showcase/label.cgi?id=76). The Charlson Comorbidity Index (CCI) was defined using the method developed by Quan *et al* (based on ICD-10 and enhance ICD-9-CM; online supplemental table 3).[21] Missing data were coded as a missing indicator category for categorical variables,[22] using sex-specific means to impute the missing value for continuous variables.

Statistical analysis

Baseline characteristics of included participants were described across their survivorship as frequency (n) and proportion (%) for categorical variables and mean (±SD) for normally distributed

continuous variables. Multivariable logistic regression analysis was used to assess the associations between the polygenic risk score and individual lifestyle factors. We applied the Cox proportional hazard regression model regressed against lifespan and surviving (alive or dead) status to examine the associations of genetic risk categories, lifestyle categories, and genetic risk and lifestyle combined categories. The model was fully adjusted for the covariates mentioned above as well as the first 20 principal components of ancestry.[23] The interactions between the PRS and lifestyle factors were tested using a multiplicative interaction model. The calculation of life expectancy and its confidence interval was carried out for individuals with different genetic and lifestyle risk categories using flexible parametric survival models with age as timescale.[24] The proportionality of hazards assumption was assessed using the Schoenfeld residuals method.

Secondary analysis was performed to derive the 'optimal lifestyle combination', in which we eliminated each lifestyle factor to reconstruct the weighted lifestyle score and rank the importance of the lifestyle according to the size of the coefficient. Several sensitivity analyses were also conducted, including: (1) analysis using the genetic risk quintiles instead of categories; (2) analysis using the number of healthy lifestyle factors instead of categories; (3) analysis using an unweighted lifestyle score; (4) analysis excluding participants with incomplete covariate data (n=2644). To examine the consistency of the association in subpopulations, we conducted stratification analyses by age (≥60 and <60 years), sex (female and male), education attainment (≥college/university and <college/university), and the tertiles of the Townsend deprivation index (from low to high, T1–3). We also stratified the analysis on the associations of the healthy lifestyle categories with death risk by genetic risk. We additionally adjusted self-reported family history of cancer, cardiovascular disease (CVD) or diabetes and depression symptoms, assessed using a two-item depression scale (PHQ-2).[25]

All p values were two-sided, and p<0.05 was considered statistically significant. All statistical analyses were performed using the R version 4.2.0.

Results

After excluding individuals who had no genetic data, failed to pass genetic quality control, died of COVID-19, injury, or accidental causes, or had missing data for lifestyle factors, 353 742 European heritage participants from the UK Biobank were included in the main analysis (table 1; online supplemental figure 1). Baseline characteristics of participants are demonstrated by the vital status (dead or alive) in table 1. Over a median follow-up of 12.86 years (IQR 12.14–13.55 years), 24 239 deaths were identified among eligible participants from the UK Biobank. Using the GWAS summary statistics from the LifeGen cohort, we obtained 19 independent SNPs to construct the lifespan PRS among independent UK Biobank participants (online supplemental table 1). The PRS was normally distributed (online supplementary fig 2online supplemental figure 2 2) and was not associated with any lifestyle factor other than healthy diet (OR 1.02, 95% CI 1.01 to 1.03) (online supplemental table 4). To generate the HLS associated with lifespan, we assessed each of the six common lifestyle factors using data from the US NHANES, which included 19 484 eligible adult participants and 1599 death events during a median follow-up of 6.92 years (IQR 3.83–10.42 years). The demographic characteristics of eligible participants are presented in online supplemental table 5. The associations of individual lifestyle factors with lifespan and their weights used for the construction of HLS are presented in online supplemental table 6, in which unfavourable lifestyle was in general associated with reduced length of lifespan for each of the component lifestyle factors.

In the analysis of PRS, the risk of death increased across genetic risk categories (long to short) in a linear way (p_{trend}<0.001) (table 2). Compared to individuals in the genetic category of long lifespan, those in the genetic category of short lifespan had a higher hazard ratio of death (HR 1.21, 95% CI 1.16 to 1.26) (table 2). The associations remained significant after additional adjustment for lifestyle factors. The same pattern of associations was observed in the analysis using the PRS as a continuous variable, instead of categories (online supplemental table 7). The cumulative death rate during the follow-up was also higher in the high genetic risk group compared with the low genetic risk group (log rank p<0.001) (online supplemental figure 3).

In analysis of HLS, the risk of death increased across lifestyle categories (favourable to unfavourable) in a dose-response manner (p_{trend}<0.001) (table 2). The HR of death for individuals in the unfavourable category was 1.78 (95% CI 1.71 to 1.85), compared with those in the favourable category. The associations did not change in sensitivity analysis with further adjustment for genetic risk (table 2) and in the analysis using unweighted HLS (online supplemental table 8). The cumulative death rate of participants during the follow-up was higher in the group with an unfavourable lifestyle compared with the group with a favourable lifestyle (log rank p<0.001) (online supplemental figure 3).

In the analysis of joint categories for genetic and lifestyle risk, the HR of death showed an increasing trend with elevated PRS and HLS (figure 1). Especially compared to individuals with genetic propensity for a long lifespan (low PRS) and a favourable lifestyle (high HLS), those with a genetic propensity for a short lifespan (high PRS) and an unfavourable lifestyle (low HLS) had 104% higher rates of death (HR 2.04, 95% CI 1.87 to 2.22, p<0.001). In contrast, individuals with a genetic propensity for a short lifespan (high PRS) but a favourable lifestyle (high HLS) (HR 1.26, 95% CI 1.14 to 1.39) had 54% lower rates of death than those with a genetic propensity for a short lifespan (high PRS) and an unfavourable lifestyle (low HLS) (HR 1.80, 95% CI 1.64 to 1.96).

Strata analysis confirmed that an unfavourable lifestyle (low HLS) was associated with a higher risk of death across all genetic groups (table 3). We did not detect any multiplicative interaction between the PRS and the HLS ($p_{interaction}$=0.10). There was no statistically significant interaction between a healthy lifestyle and *APOE* ε4 ($p_{interaction}$=0.25). Also, the results remained consistent with the main analysis when stratifying by APOE ε4 carrier status (online supplemental table 9). The observed associations remained statistically significant in a series of sensitivity analyses: (1) using the unweighted HLS (online supplemental table 10); (2) excluding participants with missing data on covariates (online supplemental table 10); (3) additionally adjusted for self-reported family history of cancer, CVD or diabetes and depression symptoms (online supplemental table 11); and (4) stratified by age, sex, education attainment, and Townsend deprivation index (online supplemental table 12).

As for the secondary analysis, we additionally assessed the joint impact of genetic and lifestyle risk on the life expectancy of UK Biobank participants. The life expectancy at 40 years was 52.52 (95% CI 52.00 to 53.01) years for participants with a genetic propensity for long lifespan (low PRS) and a favourable lifestyle (high HLS), and was 45.83 (95% CI 45.35 to 46.32) years for participants with a genetic propensity for short lifespan (high PRS) and an unfavourable lifestyle (low HLS), with a mean difference of 6.69 (95% CI 5.98 to 7.39) years in lifespan (online supplemental table 13). An unfavourable lifestyle has a strong effect on the years of

Table 1 Baseline characteristics of eligible participants in UK Biobank

	All (number (%)*)			Women (number (%)*)			Men (number (%)*)		
Characteristic	Alive at end of follow-up (n=329 503)	Dead (n=24 239)	P value	Alive at end of follow-up (n=178 986)	Dead (n=9618)	P value	Alive at end of follow-up (n=150 517)	Dead (n=14 621)	P value
Age (mean (SD))	56.85 (8.00)	62.39 (6.22)	<0.001	56.70 (7.92)	61.93 (6.43)	<0.001	57.02 (8.08)	62.70 (6.06)	<0.001
CCI† (mean (SD))	0.19 (0.85)	1.91 (4.86)	<0.001	0.18 (0.85)	2.10 (5.47)	<0.001	0.20 (0.86)	1.79 (4.40)	<0.001
Education (%)									
Higher	125 955 (38.2)	7266 (30.0)	<0.001	66 523 (37.2)	2917 (30.3)	<0.001	59 432 (39.5)	4349 (29.7)	<0.001
Upper secondary	113 204 (34.4)	7137 (29.4)		67 187 (37.5)	3215 (33.4)		46 017 (30.6)	3922 (26.8)	
Lower secondary	18 207 (5.5)	848 (3.5)		9792 (5.5)	345 (3.6)		8415 (5.6)	503 (3.4)	
Vocational	21 646 (6.6)	2067 (8.5)		7620 (4.3)	455 (4.7)		14 026 (9.3)	1612 (11.0)	
Others	48 362 (14.7)	6678 (27.6)		26 726 (14.9)	2593 (27.0)		21 636 (14.4)	4085 (27.9)	
Socioeconomic status quintile‡ (%)									
1 (least deprived)	66 541 (20.2)	4212 (17.4)	<0.001	35 608 (19.9)	1681 (17.5)	<0.001	30 933 (20.6)	2531 (17.3)	<0.001
2–4	198 585 (60.3)	13 641 (56.3)		108 484 (60.6)	5520 (57.4)		90 101 (59.9)	8121 (55.5)	
5 (most deprived)	64 377 (19.5)	6386 (26.3)		34 894 (19.5)	2417 (25.1)		29 483 (19.6)	3969 (27.1)	
Healthy lifestyle factors									
No current smoking	184 786 (56.1)	9494 (39.2)	<0.001	107 692 (60.2)	4630 (48.1)	<0.001	77 094 (51.2)	4864 (33.3)	<0.001
Moderate alcohol consumption	241 264 (73.2)	17 732 (73.2)	0.829	131 847 (73.7)	7406 (77.0)	<0.001	109 417 (72.7)	10 326 (70.6)	<0.001
Healthy body shape§	110 787 (33.6)	6513 (26.9)	<0.001	72 542 (40.5)	3225 (33.5)	<0.001	38 245 (25.4)	3288 (22.5)	<0.001
Adequate sleep duration (7–8 hours)	229 492 (69.6)	15 223 (62.8)	<0.001	124 553 (69.6)	6033 (62.7)	<0.001	104 939 (69.7)	9190 (62.9)	<0.001
Healthy diet	232 662 (70.6)	15 685 (64.7)	<0.001	142 017 (79.3)	7444 (77.4)	<0.001	90 645 (60.2)	8241 (56.4)	<0.001
Regular physical activity	193 824 (58.8)	13 099 (54.0)	<0.001	102 930 (57.5)	5140 (53.4)	<0.001	90 894 (60.4)	7959 (54.4)	<0.001
Number of healthy lifestyle factors (%)									
0	2014 (0.6)	354 (1.5)	<0.001	590 (0.3)	59 (0.6)	<0.001	1424 (0.9)	295 (2.0)	<0.001
1	14 941 (4.5)	1949 (8.0)		5542 (3.1)	479 (5.0)		9399 (6.2)	1470 (10.1)	
2	46 248 (14.0)	4901 (20.2)		20 268 (11.3)	1516 (15.8)		25 980 (17.3)	3385 (23.2)	
3	85 128 (25.8)	6946 (28.7)		43 116 (24.1)	2580 (26.8)		42 012 (27.9)	4366 (29.9)	
4	96 195 (29.2)	6076 (25.1)		54 979 (30.7)	2780 (28.9)		41 216 (27.4)	3296 (22.5)	
5	64 648 (19.6)	3225 (13.3)		40 707 (22.7)	1717 (17.9)		23 941 (15.9)	1508 (10.3)	
6	20 329 (6.2)	788 (3.3)		13 784 (7.7)	487 (5.1)		6545 (4.3)	301 (2.1)	
Genetic risk category¶									
Long	66 239 (20.1)	4510 (18.6)	<0.001	35 796 (20.0)	1809 (18.8)	<0.001	30 443 (20.2)	2701 (18.5)	<0.001
Intermediate	197 911 (60.1)	14 333 (59.1)		107 605 (60.1)	5686 (59.1)		90 306 (60.0)	8647 (59.1)	

Continued

Table 1 Continued

Characteristic	All (number (%)*)			Women (number (%)*)			Men (number (%)*)		
	Alive at end of follow-up (n=329 503)	Dead (n=24 239)	P value	Alive at end of follow-up (n=178 986)	Dead (n=9618)	P value	Alive at end of follow-up (n=150 517)	Dead (n=14 621)	P value
Short	65 353 (19.8)	5396 (22.3)		35 585 (19.9)	2123 (22.1)		29 768 (19.8)	3273 (22.4)	
APOE ε4 carrier	137 359 (41.7)	10 559 (43.6)		74 652 (41.7)	4193 (43.6)		62 707 (41.7)	6366 (43.5)	

*Percentages may not sum to 100 because of rounding.
†Charlson Comorbidity Index.
‡Socioeconomic status was assessed using the Townsend deprivation index, which combines information on social class, employment, car availability, and housing.
§Healthy body shape was defined as body mass index: 18.5–30 kg/m^2.
¶Genetic risk categories were classified as long (<lowest quintile), intermediate (2–4 quintiles), and short (>highest quintile) according to polygenic risk scores.

life lost regardless of which lifespan group an individual is at, and the subgroup of long lifespan and unfavourable lifestyle has the most years of life lost (figure 2). Among individuals with a genetic propensity for short lifespan (high PRS), those with a favourable lifestyle (high HLS) would have 5.22 (95% CI 5.18 to 5.24) years longer of lifespan than those with an unfavourable lifestyle (low HLS) (online supplemental table 13). Given that the largest proportion of participants had four healthy lifestyle factors (28.91%), the 'optimal lifestyle combination' for a prolonged lifespan were derived to be never smoking, regular physical activity, adequate sleep duration, and healthy diet, according to the rank of the size of the effect estimates (online supplemental table 14).

Discussion

In this study, we comprehensively investigated the associations between genetic risk and lifestyle risk factors regressed against lifespan in 353 742 participants of the UK Biobank cohort. Our results indicated that a high genetic risk was associated with a 21% increased risk of death compared with a low genetic risk, independent of lifestyle factors. In contrast, an unfavourable lifestyle was associated with an approximately 78% increased risk of death compared with a favourable lifestyle within and across genetic risk categories. Furthermore, the genetic risk of a shorter lifespan or premature death might be offset by a favourable lifestyle by approximately 62%. Participants with a genetic predisposition to a short lifespan and an unfavourable lifestyle had a 2.04 times higher death risk compared with those with a genetic predisposition to a long lifespan and a favourable lifestyle. Our study also indicated that adherence to healthy lifestyles could substantially attenuate the loss of lifespan for individuals with genetic susceptibility to a shorter lifespan. In addition, we constituted the optimal lifestyle combination containing four lifestyle factors, including no current smoking, regular physical activity, adequate sleep duration, and a healthy diet, to bring better benefits for prolonging human lifespan.

To our knowledge, our study is the first to investigate the joint association of genetic risk and lifestyle factors with human lifespan. Previous epidemiological studies have established the critical role of healthy lifestyles in prolonging lifespan. Adherence to a healthy lifestyle is associated with a gain in life expectancy of 8.8 years (for men) and 8.1 years (for women) at age 30 in China, 12.2 years (for men) and 14.0 years (for women) at age 50 in the USA, and 17 years (for men) and 13.9 years (for women) at age 40 in Germany.[26–28] Also, the life expectancy free of diabetes, cardiovascular diseases, and cancer at age 50 was 34.4 years for women and 31.1 years for men who adopted a healthy lifestyle.[29] A study of middle-aged adults based on data from the UK Biobank found that engaging in a healthier lifestyle was associated with up to 6.3 years longer life for men and 7.6 years for women, regardless of the presence of multimorbidity.[19] Zhang *et al* revealed that both in American and British adults, a minor portion of the disparity in health outcomes resulting from socioeconomic factors was mediated by unhealthy lifestyles.[30] Another study in UK Biobank found that unfavourable lifestyles portended a higher risk of all-cause mortality and CVD mortality, independent of the genetic risk score constructed by 300 CVD-related SNPs.[31] The general conclusions of these studies are consistent with our study. However, those studies only considered four or five healthy lifestyles and did not incorporate genetic factors for lifespan. Also, there have been studies constructing PRSs to evaluate the association with human lifespan.[32 33] Researchers have explored the relationship between some of the lifestyles (eg, regular exercise, smoking status, etc) and lifespan individually. The joint relationship of multiple lifestyles

Table 2 Risk of death according to genetic risk and lifestyle categories in UK Biobank

Category	Events/person-years	Model 1*		Model 2†	
		HR (95% CI)	P value	HR (95% CI)	P value
Genetic propensity					
Long	4510/4944075	1 (Reference)		1 (Reference)	
Intermediate	14333/14821419	1.05 (1.02 to 1.09)	1.90E-03	1.06 (1.02 to 1.1)	7.09E-04
Short	5396/4933241	1.21 (1.16 to 1.26)	5.05E-21	1.21 (1.17 to 1.26)	1.42E-21
P value for trend‡			1.53E-29		4.64E-30
Healthy lifestyle§					
Favourable	4039/5718655	1 (Reference)		1 (Reference)	
Intermediate	12370/13716591	1.19 (1.15 to 1.23)	1.60E-21	1.19 (1.15 to 1.23)	1.44E-21
Unfavourable	7830/5263490	1.78 (1.71 to 1.85)	7.05E-186	1.78 (1.71 to 1.85)	4.65E-186
P value for trend‡			2.41E-282		8.88E-283

*Adjusted for age, age-square, sex, socioeconomic status, education, Charlson Comorbidity Index, and first 20 principal components of ancestry.
†Adjusted for model 1 and weighted lifestyle category or genetic risk category.
‡The p value for trend was calculated using genetic risk or healthy lifestyle scores as continuous variables.
§Weighted healthy lifestyle categories were classified as favourable (23.07%), intermediate (55.63%), and unfavourable (21.29%) in UK Biobank.

with PRS and lifespan has not been discussed. In our study, we found that a healthy lifestyle could lower overall risk within and between genetic risk groups, and the genetic predisposition to a shorter lifespan can be substantially compensated by having a healthy lifestyle. Participants with high genetic risk could prolong approximately 5.22 years of life expectancy at age 40 with a favourable lifestyle. Given that lifestyle behavioural habits are usually developed before middle age, taking effective public health interventions is quite crucial for those at high genetic risk to extend their lifespan before the formation of a fixed lifestyle.

Strengths and limitations

The major strengths of our study include the prospective design, a large sample size from two well-established cohorts from the USA and the UK, and the availability of genotype and lifestyle information that enabled us to examine their joint effect comprehensively. We constructed the healthy lifestyle score derived from NHANES and applied it in UK Biobank to avoid the inflation of the weight. Also, we leveraged genetic loci associated with lifespan from LifeGen's GWAS independent of UK Biobank to avoid overfitting.[8] Notably, the adoption of GWAS data distinct from UK Biobank participants might mitigate the inherent heterogeneity in PRS computation. In addition, to enhance the robustness of the results, we included comprehensive sensitivity analyses by incorporating variables such as family histories of non-communicable diseases, known to exert significant influence on longevity, and the evaluation of depression symptoms. Furthermore, most people have shown poor adherence to healthy lifestyles in modern society. Evidence from the Nurses' Health Study and the Health Professionals Follow-up Study showed that <2% of participants had five or more healthy lifestyle factors simultaneously,[28] and only 6.2% of the UK Biobank population had six healthy lifestyle factors in our analysis. Therefore, our study brought up the concept of 'optimal lifestyle combination' for the first time. The combination of the listed four lifestyles could convey better benefits for a longer lifespan than any other combination of four healthy lifestyle factors, offering people health recommendations with strong practical implications.

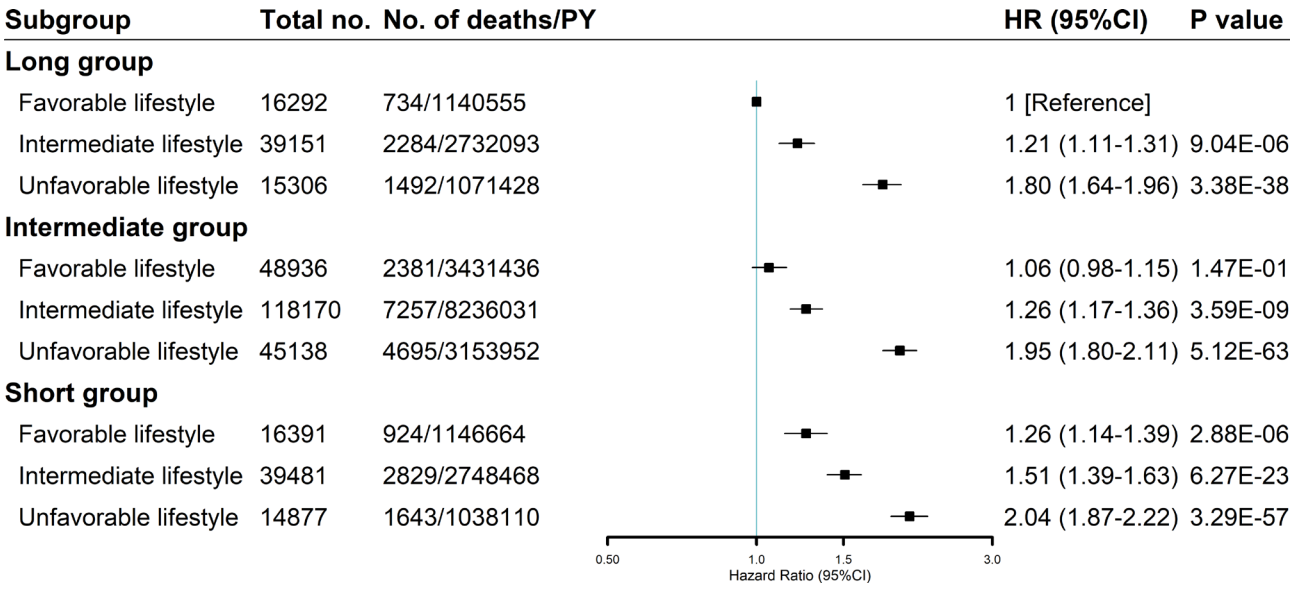

Figure 1 Risk of death by joint categorisation for genetic risk and healthy lifestyle score in UK Biobank. Adjusted for age, age-square, sex, socioeconomic status, education, Charlson Comorbidity Index, and first 20 principal components of ancestry.

Table 3 Association of risk of death with lifestyle categories by genetic risk level

Subgroup	Events/person-years	HR (95% CI)	P value
Long group			
Favourable lifestyle	734/1140555	1 (Reference)	
Intermediate lifestyle	2284/2732093	1.20 (1.11 to 1.31)	1.47E-05
Unfavourable lifestyle	1492/1071428	1.79 (1.64 to 1.96)	1.80E-36
P value for trend*			3.89E-55
Intermediate group			
Favourable lifestyle	2381/3431436	1 (Reference)	
Intermediate lifestyle	7257/8236031	1.18 (1.13 to 1.24)	9.77E-13
Unfavourable lifestyle	4695/3153952	1.83 (1.74 to 1.93)	2.26E-123
P value for trend*			1.06E-185
Short group			
Favourable lifestyle	924/1146664	1 (Reference)	
Intermediate lifestyle	2829/2748468	1.19 (1.11 to 1.29)	3.51E-06
Unfavourable lifestyle	1643/1038110	1.61 (1.48 to 1.75)	1.20E-29
P value for trend*			3.10E-47

*Association of risk of death with lifestyle categories by genetic risk level.

Our study also has several limitations. First, there is still abundant room for further progress in determining the genetic variants associated with lifespan. While the polygenic risk score has included 19 validated SNPs, it explains only limited proportions of the genetic risk of a shorter lifespan. Second, the life expectancy at birth in the UK has approached 79.0 years for males and 82.9 years for females from 2018 to 2020.[34] Nevertheless, the follow-up period of UK Biobank in our study is confined to about 12.86 years, and the longest lifespan observed is 87 years, which may engender an underestimate of our findings. Third, it is inevitable that self-reported lifestyle factors could lead to an incorrect assessment of healthy lifestyle scores. Fourth, lifestyle factors were measured only once at baseline. However, people may have made changes in their lifestyle during the follow-up for many reasons, such as the diagnosis of diseases, thus affecting risk estimates. Fifth, there were varied definitions of healthy lifestyle factors in the two cohorts we employed. Thus, we defined healthy lifestyle factors based on previous studies. Furthermore, previous research has shown significant variability in lifestyle choices across different age groups, especially among the young and elderly subgroups.[27] Therefore, the age disparity between participants in NHANES and UK Biobank cohorts introduces complexity. In light of this, weighted lifestyle scores may not be able to capture effectively the nuanced impact of lifestyle choices

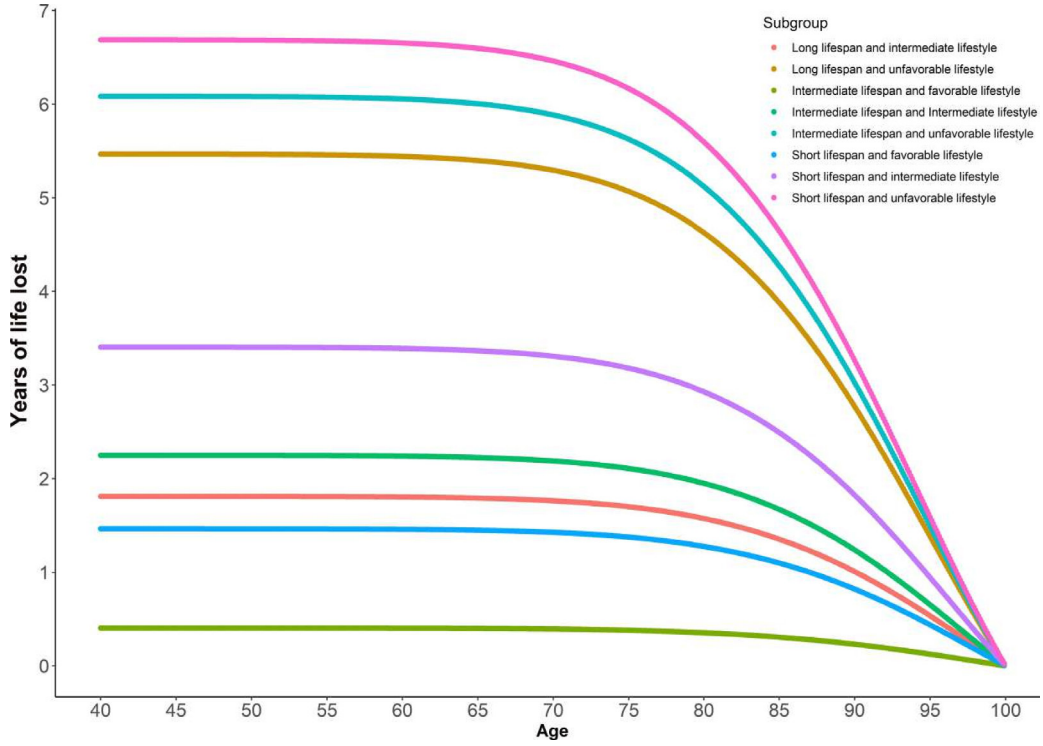

Figure 2 Years of life lost of other subgroups versus long lifespan and favourable lifestyle group by joint categorisation for genetic risk and healthy lifestyle score in UK Biobank.

within the UK Biobank cohort. Finally, previous research has suggested that the UK Biobank cohort is not fully representative of the general UK population given the 'healthy volunteer' selection bias and low participation rate.[35] Also, our analysis was limited to white-European ancestry, making the findings less generalisable to the general population and other ethnic groups. The generalisability of our findings should be further evaluated in more diverse populations.

Conclusions

Our study reveals that genetic and lifestyle factors were independently associated with lifespan. Adherence to healthy lifestyles could significantly offset the genetic risk of a shorter lifespan or premature death. Accordingly, lifespan could be further extended with public intervention for healthy lifestyles across entire populations. Successes in several regions have set good examples that healthy lifestyle promotion policies would contribute significantly to increased life expectancy and reduced mortality.[36–39] Public health policies for improving healthy lifestyles would be a potent complement to standard healthcare and diminish the impact of genetic factors on human lifespan.

Author affiliations
[1]Department of Big Data in Health Science, School of Public Health and Center of Clinical Big Data and Analytics of The Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, China
[2]Department of Biostatistics, Center for Global Health, School of Public Health, Nanjing Medical University, Nanjing, China
[3]Centre for Global Health, Usher Institute, University of Edinburgh, Edinburgh, UK
[4]Department of Nutrition and Food Hygiene, School of Public Health, Peking University, Beijing, China
[5]MRC Human Genetics Unit, MRC Institute of Genetics and Cancer, University of Edinburgh, Edinburgh, UK
[6]Department of Biostatistics, School of Public Health and The Key Laboratory of Public Health Safety of Ministry of Education, Fudan University, Shanghai, China
[7]Cancer Research UK Edinburgh Centre, Medical Research Council Institute of Genetics and Cancer, University of Edinburgh, Edinburgh, UK
[8]The Key Laboratory of Intelligent Preventive Medicine of Zhejiang Province, Hangzhou, China

Acknowledgements Thanks are due to the technical support provided by the Core Facilities, Zhejiang University Medical Center/Liangzhu Laboratory.

Contributors ZB and XL conceived and designed the study. All authors contributed to acquisition, analysis, or interpretation of data. ZB, LW and RF drafted the manuscript. All authors critically revised the manuscript for important intellectual content and approved the final version of the manuscript. XW and XL contributed to supervision and administrative, technical, or material support. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted. XW and XL are joint last authors. XL is the guarantor for the full content.

Funding XL is supported by the Natural Science Fund for Distinguished Young Scholars of Zhejiang Province (LR22H260001) and the National Nature Science Foundation of China (82204019). XFW is supported by the Key Laboratory of Intelligent Preventive Medicine of Zhejiang Province (2020E10004), the Leading Innovative and Entrepreneur Team Introduction Program of Zhejiang (2019R01007), the Key Research and Development Program of Zhejiang Province (2020C03002), and Healthy Zhejiang One Million People Cohort (K-20230085). ZLB is supported by China Scholarship Council (201909067018). ET is supported by a CRUK Career Development Fellowship (C31250/A22804). PRHJT and JFW are supported by the MRC Human Genetics Unit program grant, Quantitative traits in health and disease (U. MC_UU_00007/10). The funders/sponsors had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Competing interests PRHJT is an employee of BioAge Labs, Inc; there are no other relationships or activities that could appear to have influenced the submitted work.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval The study protocol for the US NHANES was approved by the US NHANES institutional review board and National Center for Health Statistics Research ethics review board. The North West Multi-Centre Research Ethics Committee approved the collection and use of UK Biobank data. All participants provided written informed consent. Institutional review board approval was waived for this analysis because of the publicly available and de-identified data.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available in a public, open access repository. NHANES data are available at http://www.cdc.gov/nchs/nhis/index.htm. UK Biobank study was under Application Number 66354. The UK Biobank is an open access resource and bona fide researchers can apply to use the UK Biobank dataset by registering and applying at http://ukbiobank.ac.uk/register-apply/. Further information is available from the corresponding author upon request.

ORCID iD
Zilong Bian http://orcid.org/0000-0002-2628-7535

References

1 **Shadyab AH**, LaCroix AZ. Genetic factors associated with longevity: a review of recent findings. *Ageing Res Rev* 2015;19:1–7.
2 **Kaplanis J**, Gordon A, Shor T, *et al*. Quantitative analysis of population-scale family trees with millions of relatives. *Science* 2018;360:171–5.
3 **Deelen J**, Beekman M, Uh H-W, *et al*. Genome-wide association study identifies a single major locus contributing to survival into old age; the APOE locus revisited. *Aging Cell* 2011;10:686–98.
4 **Nebel A**, Kleindorp R, Caliebe A, *et al*. A genome-wide association study confirms APOE as the major gene influencing survival in long-lived individuals. *Mech Ageing Dev* 2011;132:324–30.

5 **Broer L**, Buchman AS, Deelen J, *et al*. GWAS of longevity in charge consortium confirms APOE and FOXO3 candidacy. *J Gerontol A Biol Sci Med Sci* 2015;70:110–8.
6 **Joshi PK**, Fischer K, Schraut KE, *et al*. Variants near CHRNA3/5 and APOE have age- and sex-related effects on human lifespan. *Nat Commun* 2016;7:11174.
7 **McDaid AF**, Joshi PK, Porcu E, *et al*. Bayesian association scan reveals loci associated with human lifespan and linked biomarkers. *Nat Commun* 2017;8:15842.
8 **Timmers PR**, Mounier N, Lall K, *et al*. Genomics of 1 million parent lifespans Implicates novel pathways and common diseases and distinguishes survival chances. *Elife* 2019;8:e39856.
9 **Wright KM**, Rand KA, Kermany A, *et al*. A prospective analysis of genetic variants associated with human lifespan. *G3 (Bethesda)* 2019;9:2863–78.
10 **Loef M**, Walach H. The combined effects of healthy lifestyle behaviors on all cause mortality: a systematic review and meta-analysis. *Prev Med* 2012;55:163–70.
11 **Berstad P**, Botteri E, Larsen IK, *et al*. Lifestyle changes at middle age and mortality: a population-based prospective cohort study. *J Epidemiol Community Health* 2017;71:59–66.
12 **Colpani V**, Baena CP, Jaspers L, *et al*. Lifestyle factors, cardiovascular disease and all-cause mortality in middle-aged and elderly women: a systematic review and meta-analysis. *Eur J Epidemiol* 2018;33:831–45.
13 **Ibsen DB**, Søgaard K, Sørensen LH, *et al*. Modifiable lifestyle recommendations and mortality in Denmark: a cohort study. *Am J Prev Med* 2021;60:792–801.
14 **Zhang Y-B**, Pan X-F, Chen J, *et al*. Combined lifestyle factors, all-cause mortality and cardiovascular disease: a systematic review and meta-analysis of prospective cohort studies. *J Epidemiol Community Health* 2021;75:92–9.
15 **Lopez-Jaramillo P**, Joseph P, Lopez-Lopez JP, *et al*. Risk factors, cardiovascular disease, and mortality in South America: a PURE substudy. *Eur Heart J* 2022;43:2841–51.
16 **Joshi PK**, Pirastu N, Kentistou KA, *et al*. Genome-wide meta-analysis associates HLA-DQA1/DRB1 and LPA and lifestyle factors with human longevity. *Nat Commun* 2017;8:910.
17 **Centers for Disease Control and Prevention/National Center for Health Statistics**. About the National Health and Nutrition Examination Survey. 2023. Available: https://www.cdc.gov/nchs/nhanes/about_nhanes.htm
18 **Sudlow C**, Gallacher J, Allen N, *et al*. UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med* 2015;12:e1001779.
19 **Chudasama YV**, Khunti K, Gillies CL, *et al*. Healthy lifestyle and life expectancy in people with multimorbidity in the UK Biobank: a longitudinal cohort study. *PLoS Med* 2020;17:e1003332.
20 **Han H**, Cao Y, Feng C, *et al*. Association of a healthy lifestyle with all-cause and cause-specific mortality among individuals with type 2 diabetes: a prospective study in UK Biobank. *Diabetes Care* 2022;45:319–29.
21 **Quan H**, Sundararajan V, Halfon P, *et al*. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. *Med Care* 2005;43:1130–9.
22 **Song M**, Zhou X, Pazaris M, *et al*. The missing covariate indicator method is nearly valid almost always. 2021. Available: https://ui.adsabs.harvard.edu/abs/2021arXiv211100138S [Accessed 01 Oct 2022].
23 **Bycroft C**, Freeman C, Petkova D, *et al*. The UK Biobank resource with deep phenotyping and genomic data. *Nature* 2018;562:203–9.
24 **Dehbi HM**, Royston P, Hackshaw A. Life expectancy difference and life expectancy ratio: two measures of treatment effects in randomised trials with non-proportional hazards. *BMJ* 2017;357:j2250.
25 **Kroenke K**, Spitzer RL, Williams JBW, *et al*. Anxiety disorders in primary care: prevalence, impairment, comorbidity, and detection. *Ann Intern Med* 2007;146:317–25.
26 **Li K**, Hüsing A, Kaaks R. Lifestyle risk factors and residual life expectancy at age 40: a German cohort study. *BMC Med* 2014;12:59.
27 **Sun Q**, Yu D, Fan J, *et al*. Healthy lifestyle and life expectancy at age 30 years in the Chinese population: an observational study. *Lancet Public Health* 2022;7:e994–1004.
28 **Li Y**, Pan A, Wang DD, *et al*. Impact of healthy lifestyle factors on life expectancies in the US population. *Circulation* 2018;138:345–55.
29 **Li Y**, Schoufour J, Wang DD, *et al*. Healthy lifestyle and life expectancy free of cancer, cardiovascular disease, and type 2 diabetes: prospective cohort study. *BMJ* 2020;368:l6669–l69.
30 **Zhang Y-B**, Chen C, Pan X-F, *et al*. Associations of healthy lifestyle and socioeconomic status with mortality and incident cardiovascular disease: two prospective cohort studies. *BMJ* 2021;373:n604.
31 **Livingstone KM**, Abbott G, Ward J, *et al*. Unhealthy lifestyle, genetics and risk of cardiovascular disease and mortality in 76,958 individuals from the UK Biobank cohort study. *Nutrients* 2021;13:4283.
32 **Sakaue S**, Kanai M, Karjalainen J, *et al*. Trans-biobank analysis with 676,000 individuals elucidates the association of polygenic risk scores of complex traits with human lifespan. *Nat Med* 2020;26:542–8.
33 **Meisner A**, Kundu P, Zhang YD, *et al*. Combined utility of 25 disease and risk factor polygenic risk scores for stratifying risk of all-cause mortality. *Am J Hum Genet* 2020;107:418–31.
34 **Office for National Statistics**. National life tables – life expectancy in the UK: 2018 to 2020. 2023. Available: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/lifeexpectancies/bulletins/nationallifetablesunitedkingdom/2018to2020 [Accessed 14 Nov 2023].
35 **Fry A**, Littlejohns TJ, Sudlow C, *et al*. Comparison of sociodemographic and health-related characteristics of UK Biobank participants with those of the general population. *Am J Epidemiol* 2017;186:1026–34.
36 **Ni MY**, Canudas-Romo V, Shi J, *et al*. Understanding longevity in Hong Kong: a comparative study with long-living, high-income countries. *Lancet Public Health* 2021;6:e919–31.
37 **Wagenaar AC**, Tobler AL, Komro KA. Effects of alcohol tax and price policies on morbidity and mortality: a systematic review. *Am J Public Health* 2010;100:2270–8.
38 **Sallis JF**, Cerin E, Kerr J, *et al*. Built environment, physical activity, and obesity: findings from the International Physical Activity and Environment Network (IPEN) adult study. *Annu Rev Public Health* 2020;41:119–39.
39 **Wang YC**, Coxson P, Shen Y-M, *et al*. A penny-per-ounce tax on sugar-sweetened beverages would cut health and cost burdens of diabetes. *Health Aff (Millwood)* 2012;31:199–207.