nature communications

Article

https://doi.org/10.1038/s41467-025-65963-x

Modifiable influencing factors and their joint effects on early- and late-onset coronary heart disease

Received: 10 December 2024

Accepted: 28 October 2025

Published online: 08 December 2025

Check for updates

Jianhui Guo[1], Petros Koutrakis[2], Carolina L. Zilli Vieira[2], Yi Fan[3], Yaqi Wang[1], Xinyao Lian[1], Zhaokun Wang[3], Ziqi Tang[1], Shaodan Huang[3,4] & Jing Li[1]

Over the past decade, the previously declining trend in coronary heart disease burden has reversed, particularly among younger adults. Using an exposure-wide association study with data from 394,579 UK Biobank participants, we examined the relationship between 213 modifiable factors across eight domains and coronary heart disease. Weighted domain scores were calculated to assess the combined effects of these factors and their interactions with genetic risk. We identified 155 modifiable factors associated with coronary heart disease, 31 of which showed evidence of causality. The major contributors are health and medical history (15.75%), including diabetes mellitus; blood assays (13.87%), including cystatin C; lifestyle factors (10.01%), including time spent watching television; and physical measures (8.70%), including systolic blood pressure. We estimate that 40% to 62% of cases could be prevented by modifying these factors, which have a stronger effect on younger populations. These findings underscore the importance of early and comprehensive prevention strategies.

Cardiovascular disease (CVD) remains the leading cause of premature death globally, with coronary heart disease (CHD) accounting for the largest share of disability-adjusted life years among non-communicable diseases[1,2]. Although the associations between several well-established modifiable risk factors—such as elevated blood pressure, dyslipidemia, obesity, smoking, and diabetes—and CHD have been well documented, the decline in age-standardized CHD incidence has plateaued in recent years[1,3–5]. Alarmingly, the CHD burden is now increasing among younger adults (aged 20–54 years), highlighting the urgent need for more effective preventive strategies targeting modifiable risk factors in this demographic[1,5–8].

As the range of exposures linked to cardiovascular outcomes continues to expand[9], traditional hypothesis-driven studies, while informative, are limited by their constrained focus, susceptibility to inflated effect sizes and type I errors, and potential for selective reporting[10,11]. The exposure-wide association study (EWAS) provides a hypothesis-free framework to systematically evaluate numerous modifiable risk factors simultaneously[12]. By capturing the combined effects across multiple domains, EWAS can reduce bias, validate known associations, and identify novel risk factors with greater precision[10,13–18]. This approach has been successfully applied to studies of depression[10], dementia[13], diabetes[14], and CVD[15–17]. Notably, the use of composite risk scores within the EWAS framework enables the quantification of cumulative exposure burden, allowing for a more accurate estimation of their contribution to CHD risk.

Observed associations between modifiable factors and CHD may arise from non-causal mechanisms, including residual confounding and reverse causation[9]. Integrating Mendelian randomization (MR)

[1]Institute of Child and Adolescent Health, School of Public Health, Peking University and National Health Commission Key Laboratory of Reproductive Health, Beijing, China. [2]Department of Environmental Health, Harvard T.H. Chan School of Public Health, Boston, MA, USA. [3]Department of Occupational and Environmental Health Sciences, School of Public Health, Peking University, Beijing, China. [4]Key Laboratory of Epidemiology of Major Diseases (Peking University), Ministry of Education, Beijing, China. e-mail: jing.li@hsc.pku.edu.cn

 https://doi.org/10.1038/s41467-025-65963-x

into the EWAS framework can strengthen causal inference and help prioritize factors for intervention[19]. Both genetic and environmental factors contribute to the risk of early-onset CHD[20], and prior studies have demonstrated interactions between lifestyle, local environment, and genetic predisposition[21,22]. However, little is known about how modifiable factors from other domains interact with genetic risk. Assessing interactions on an additive scale may help identify individuals who would benefit most from targeted risk factors modification[23].

For the first time, we conducted an EWAS of CHD using phenotypic and genomic data from nearly 400,000 participants in the UK Biobank, with a particular focus on early-onset cases. We assessed the joint effects of modifiable factors across eight domains by constructing composite weighted scores and examined their additive interactions with polygenic risk scores (PRS) for CHD events[24]. Finally, we estimated domain-specific and overall population attributable fractions (PAFs) to quantify the potential impact of preventive interventions, especially among younger populations where early risk modification may yield the most significant long-term benefits (Fig. 1).

Results

This study included 394,579 participants (56.1% females; Supplementary Data 4), with a mean (SD) age of 56.1 (8.1) years at recruitment. After 13.9 (2.3) years of follow-up, 27,592 participants were diagnosed with CHD, with a mean (SD) age of diagnosis at 68.4 (7.6) years. This study identified 155 out of 213 modifiable factors associated with CHD (Fig. 2 and Supplementary Data 5). These factors were categorized into eight domains: socioeconomic status (SES, 7 factors), lifestyles (32 factors), early life factors (2 factors), psychosocial factors (7 factors), local environment (8 factors), blood assays (37 factors), physical measures (10 factors), and health and medical history (52 factors). Among these, 33 factors were identified as protective factors, and 122 were identified as risk factors. According to the variable importance rankings from the random forest model, the top 10 predictors were: diabetes mellitus [hazard ratio (HR) = 2.04, 95% confidence interval (CI): 1.96, 2.13, $p < 2.74 \times 10^{-265}$], Cystatin C (HR = 1.18, 95% CI: 1.17, 1.19, $p < 2.35 \times 10^{-04}$), long-standing illness, disability or infirmity (HR = 1.73, 95% CI: 1.69, 1.78, $p < 2.35 \times 10^{-04}$), Overall health rating (HR = 1.42, 95% CI: 1.36, 1.47, $p = 2.87 \times 10^{-67}$ for "good", HR = 2.37, 95% CI: 2.27, 2.48, $p < 2.35 \times 10^{-04}$ for "fair," HR = 4.01, 95% CI: 3.79, 4.25, $p < 2.35 \times 10^{-04}$ for "poor"; all compared with excellent), systolic blood pressure (HR = 1.20, 95% CI: 1.18, 1.22, $p = 8.90 \times 10^{-118}$), testosterone (HR = 0.84, 95% CI: 0.81, 0.88, $p = 8.43 \times 10^{-16}$), glycated hemoglobin (HR = 1.12, 95% CI :1.12, 1.13, $p < 2.35 \times 10^{-04}$), leg fat percentage (HR = 2.33, 95% CI: 2.23, 2.44, $p = 4.23 \times 10^{-300}$), attendance/disability/mobility allowance (HR = 2.11, 95% CI: 2.03, 2.20, $p = 1.15 \times 10^{-275}$), and HDL cholesterol (HR = 0.72, 95% CI: 0.70, 0.73, $p = 2.30 \times 10^{-252}$).

Further analyses stratified by age, sex, genetic risk for CHD, and follow-up duration revealed that >100 factors remained stably associated with CHD in most subgroups, except for the group with >10 years of follow-up, although the specific associations varied across subgroups (Fig. 3, Supplementary Data 6 and Supplementary

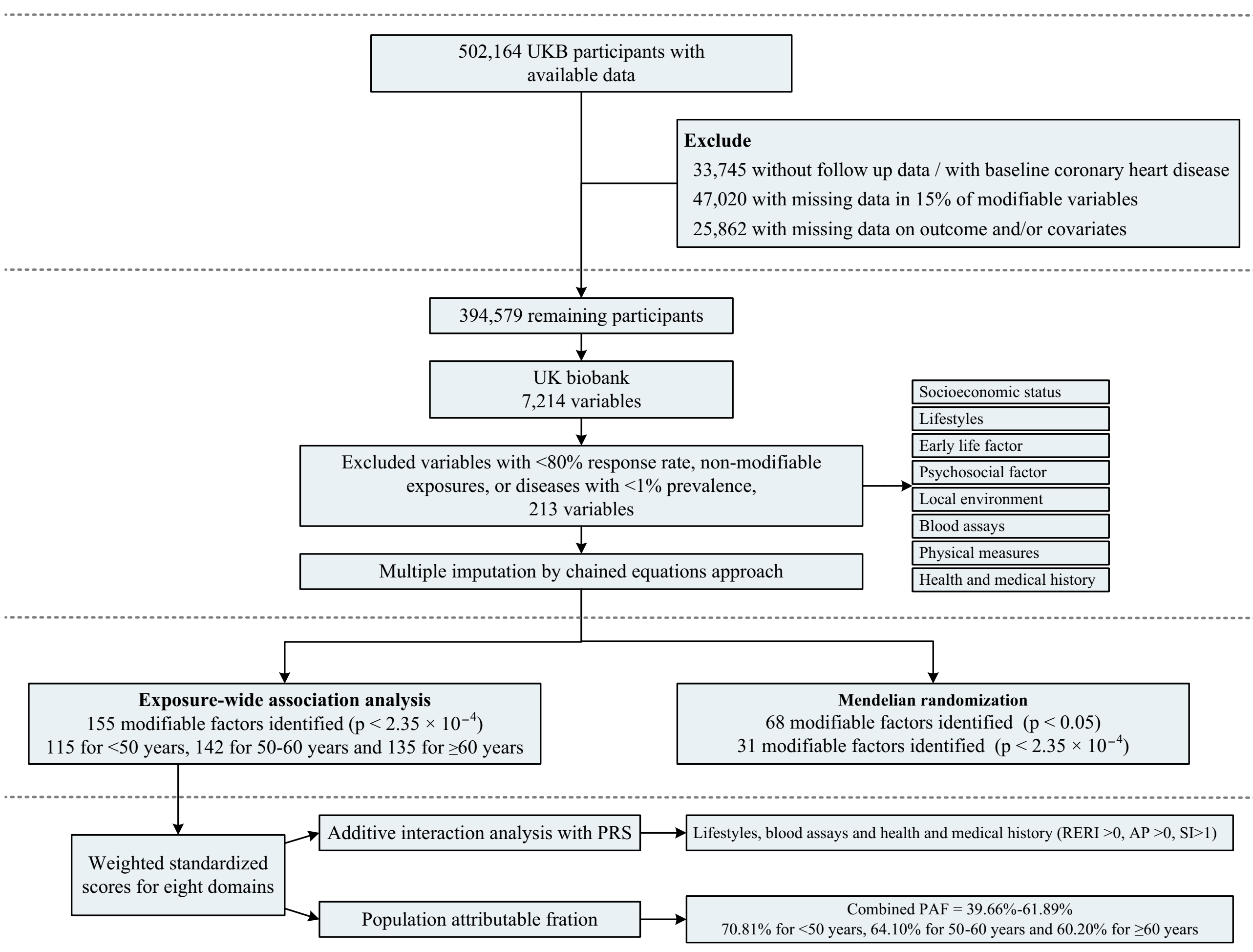
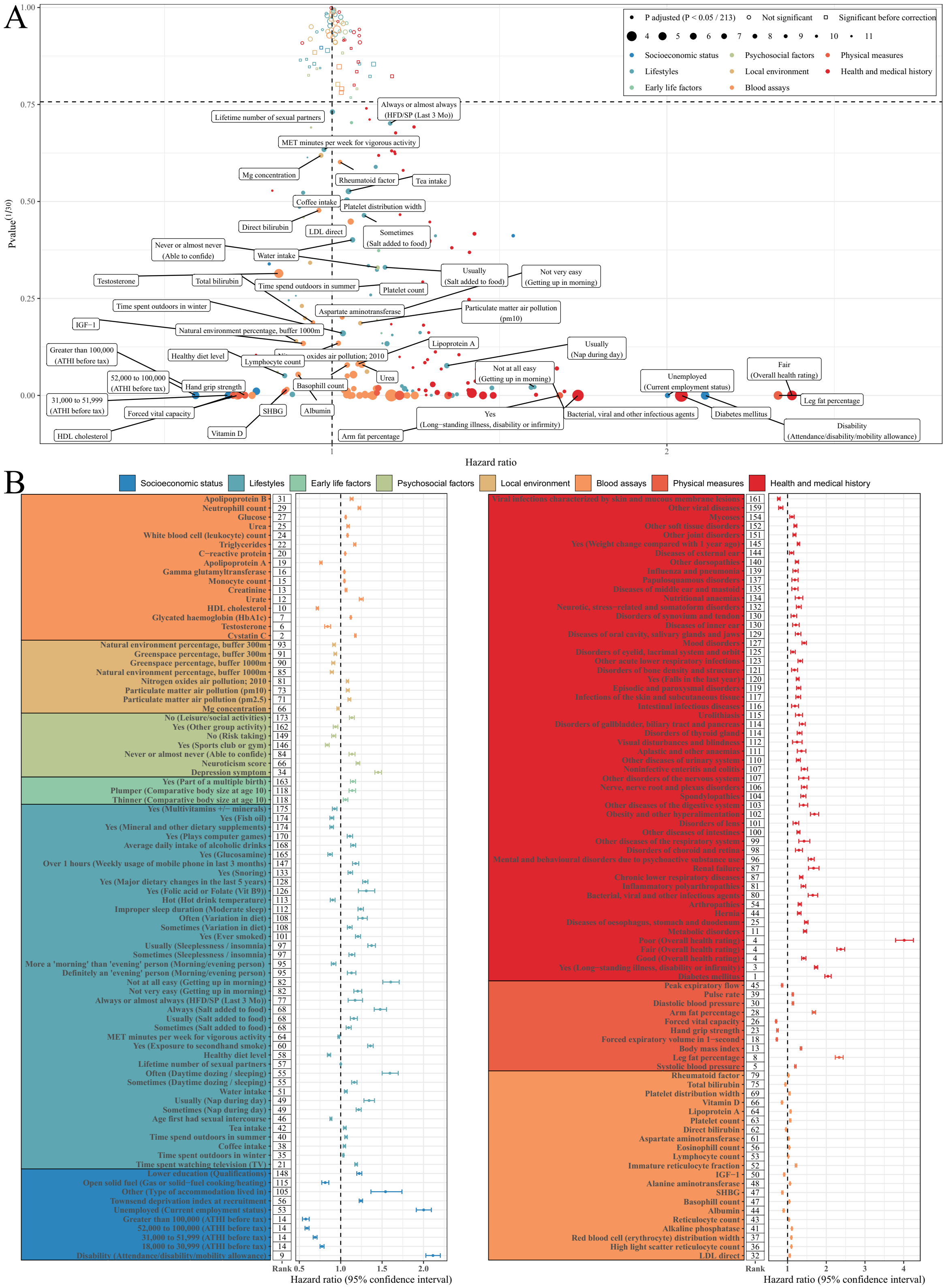


Fig. 1 | Overview of the analytic design. Analytical procedure to identify modifiable risk factors associated with incident coronary heart disease in the UK Biobank. PRS polygenic risk scores, RERI relative excess risk due to interaction, AP attributable proportion due to interaction, SI synergy index.

https://doi.org/10.1038/s41467-025-65963-x



Figs. 1–9). The number of modifiable factors associated with CHD increased with age, yet the effects of most factors decreased. Moreover, older populations showed greater sensitivity to local environmental factors. Variable importance rankings from the random forest model showed broadly consistent results across subgroups, with factors in health and medical history, blood assays, and physical measures consistently ranking high in importance, while the importance of factors in other domains diminished with age.

Inverse variance-weighted MR was employed as the primary analysis after removing outliers (Fig. 4, Supplementary Data 7, and Supplementary Figs. 11–14), given the observed evidence of heterogeneity ($p < 0.05$ for the Q test) and horizontal pleiotropy ($p < 0.05$ for the Egger intercept). By the threshold of $p < 0.05$, 68 factors showed

Fig. 2 | Association between 213 modifiable factors and incident coronary heart disease. A The *x*-axis represents the hazard ratio value, and the *y*-axis represents statistical significance (that is, the 30th root of the *p* value). The dashed line represents the threshold after multiple testing correction (Bonferroni correction, p value < 2.35×10^{-04}). The point size represents minimal depth, which was calculated by simultaneously including 213 modifiable factors and covariates in a random forest model with Cox proportional hazards regression to rank the importance of the variables; smaller minimal depth indicates greater variable importance. A set of the highest-risk factors was annotated. **B** The points represent hazard ratios, and the horizontal lines represent the corresponding 95% confidence interval. The rank indicates the importance of the variable in order of minimal depth. Hazard ratios were calculated using Cox proportional hazards regression analysis, adjusted for baseline age, sex, ethnic background, family history of cardiovascular disease, and assessment center (N = 394,579). Two-sided Z-tests were used to assess statistical significance. ATHI average total household income before tax, HFD/SP hands-free device/speakerphone use with mobile phone, VIT B9 Vitamin B9, HbA1c hemoglobin A1c, HDL high-density lipoprotein, LDL low-density lipoprotein, IGF-1 Insulin-like growth factor 1, MET metabolic equivalent of task, SHBG sex hormone-binding globulin. Source data are provided as a Source data file (Supplementary Data 5).

associations with CHD. At a more stringent threshold of $p < 2.35 \times 10^{-04}$, 31 factors demonstrated causal relationships with CHD, including 22 factors exhibiting positive associations and 9 with negative associations. These factors spanned six domains: blood assays (13 factors, e.g., triglycerides), early life factors (2 factors, e.g., comparative body size at age 10), health and medical history (5 factors, e.g., self-reported hypertension), lifestyle behaviors (3 factors, e.g., alcohol intake frequency), physical measures (6 factors, e.g., waist circumference), and SES (2 factors, e.g., educational attainment). Results from alternative methodologies are presented in Supplementary Data 7.

Compared to favorable profiles (Table 1), both unfavorable early life and psychosocial factors, as well as intermediate or unfavorable profiles across SES, lifestyle, local environment, blood assays, physical measurements, and health and medical history, were significantly associated with higher CHD risk. A significant increasing trend in CHD risk was observed across all eight domains (p for trend <0.001). Stratified analysis by age (<50 years, 50–60 years, and ≥60 years) revealed that the adverse effects of intermediate and unfavorable profiles across SES, lifestyles, early life factors, psychosocial factors, blood assays, as well as health and medical history diminished with increasing age (Fig. 5 and Supplementary Data 8). Additionally, a multiplicative interaction between age and the weighted score for health and medical history demonstrated that the effect of having a moderate or unfavorable health and medical history score on CHD risk gradually diminished with increasing age (p for interaction <0.05, Supplementary Data 8).

Meanwhile, participants were divided into nine groups based on tertiles of PRS and weighted scores for each domain, respectively. As shown in Fig. 6, Supplementary Data 9 and Supplementary Figs. 15–17, individuals with the highest genetic risk scores and unfavorable weighted scores had substantially increased CHD risks compared with those with the lowest genetic risk scores and favorable weighted scores: 134% higher for SES, 157% for lifestyles, 94% for early life factors, 97% for psychosocial factors, 96% for local environment, 192% for blood assays, 162% for physical measures, and 240% for health and medical history. Statistically significant interactions were observed between weighted scores for lifestyles, blood assays, and health and medical history with PRS, which diminished with age (Supplementary Data 10).

When shifting all unfavorable profiles to intermediate and favorable ones (Model 1), PAF estimation suggested that 39.66% of CHD cases could be prevented (Table 2), increasing to 61.89% when shifting all factors to the favorable tertile. In the conservative assessment (Model 1), the estimated maximum preventive effect was attributed to health and medical history, resulting in an 11.37% reduction in CHD incidence. The contributions of remaining domains were described as follows: blood assays (9.54%), lifestyles (5.97%), physical measures (5.70%), SES (3.70%), psychosocial factors (2.06%), early life factors (0.99%), and local environment (0.33%). In the context of a feasible implementation of comprehensive elimination of these factors, health and medical history would still account for the highest proportion of prevented CHD cases (15.75%), followed by blood assays (13.87%) and lifestyles (10.01%). Notably, implementing preventive measures in younger populations (under the age of 50), including blood indicator monitoring, healthy lifestyles, and disease prevention, could yield substantial benefits, potentially preventing 49.34–70.81% of CHD cases (Supplementary Data 11–13). Two sensitivity analyses were conducted, calculating PAFs by including only factors significantly associated with CHD ($p < 0.05$), and further adjusting for PRS when estimating PAFs. Both approaches yielded results consistent with the main findings (Supplementary Data 14 and 15). Additionally, considering the potential temporal ordering among different domains, PAFs were also calculated using models without mutual adjustment between domains, suggesting the possibility of potential prevention of an even larger proportion (59.77–77.65%) of CHD cases (Supplementary Data 16).

Discussion

Despite substantial progress in CHD prevention in recent decades, the overall decline in CHD burden has plateaued, accompanied by a concerning rise in incidence among younger populations[1,25]. Here, we systematically investigated 155 modifiable risk factors across eight domains. In addition to well-established factors such as obesity, systolic blood pressure, non-high-density lipoprotein cholesterol, smoking, and diabetes, several additional contributors, including socioeconomic indicators (e.g., income, disability benefits), infections, and sex hormones, were also identified[3,4,15,17]. Our findings demonstrate that addressing a broader range of risk factors beyond traditional biomedical targets is essential for reducing health disparities and preventing early-onset CHD effectively.

A wide range of modifiable factors was associated with CHD risk across multiple domains. Health and medical history contributed the most to CHD incidence (PAF: 11.37–15.75%), particularly endocrine and metabolic disorders such as diabetes, obesity, and thyroid disease[26–28]. Other disease categories, including genitourinary, neurological, and psychiatric conditions, were also strongly linked to CHD, potentially through mechanisms such as metabolic dysregulation, vascular dysfunction, and autonomic imbalance[29–31]. Lifestyle factors—such as smoking, physical inactivity, poor sleep, and unhealthy dietary patterns—were associated with a 5.97–10.01% reduction in CHD risk when improved[32,33]. Additionally, biomarkers (e.g., cystatin C, testosterone), physical function (e.g., lung function, grip strength), and environmental exposures (e.g., air pollution) were essential contributors, likely operating through inflammatory, oxidative stress, and cardiometabolic pathways[34,35]. SES, psychosocial factors, and early-life exposures together accounted for 6.75–11.05% of the preventable CHD burden. These upstream determinants influence health behaviors, healthcare access, and biological vulnerability throughout the life course[36–38].

We also incorporated MR within the framework of EWAS. However, MR identified only 31 variables that were causally associated with CHD, in contrast to the larger number of associations observed in EWAS. Several factors may explain this discrepancy. First, some instrumental variables lacked sufficient strength, which may have limited causal inference and made it difficult to rule out reverse causation[39]. Second, MR estimates can indicate the lifelong impact of genetic variation on disease, while longitudinal studies cover shorter timeframes, potentially leading to underestimation of results[11]. Third, due to the lack of suitable genetic instruments, several exposures, such as natural environment percentage, water percentage, and magnesium

https://doi.org/10.1038/s41467-025-65963-x

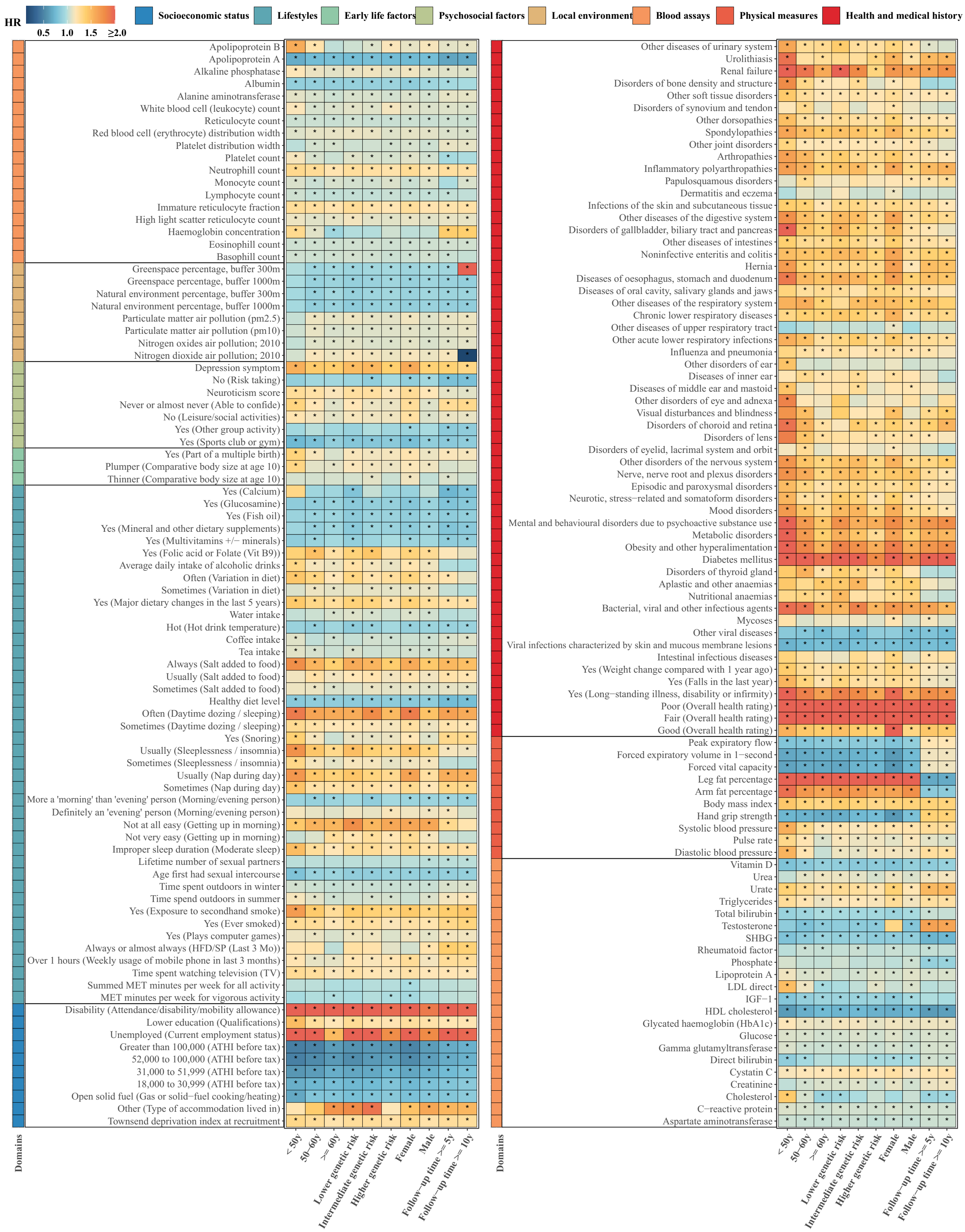


Fig. 3 | Summary heat map for significant factors in EWAS analysis across the subgroups. Models were estimated using Cox proportional hazards regression, adjusted for baseline age, sex, ethnic background, family history of cardiovascular disease, and assessment center ($N = 394{,}579$). Two-sided Z-tests were used to assess statistical significance. The color of cells indicates the effect sizes between each risk factor and incident coronary heart disease. Asterisks in cells represent significant associations after correction for multiple testing (Bonferroni-corrected, $p < 2.35 \times 10^{-04}$). ATHI average total household income before tax, HFD/SP hands-free device/speakerphone use with mobile phone, VIT B9 Vitamin B9, HbA1c hemoglobin A1c, HDL high-density lipoprotein, LDL low-density lipoprotein, IGF-1 Insulin-like growth factor 1, MET metabolic equivalent of task, SHBG sex hormone-binding globulin. Source data are provided as a Source data file (Supplementary Data 6).

https://doi.org/10.1038/s41467-025-65963-x

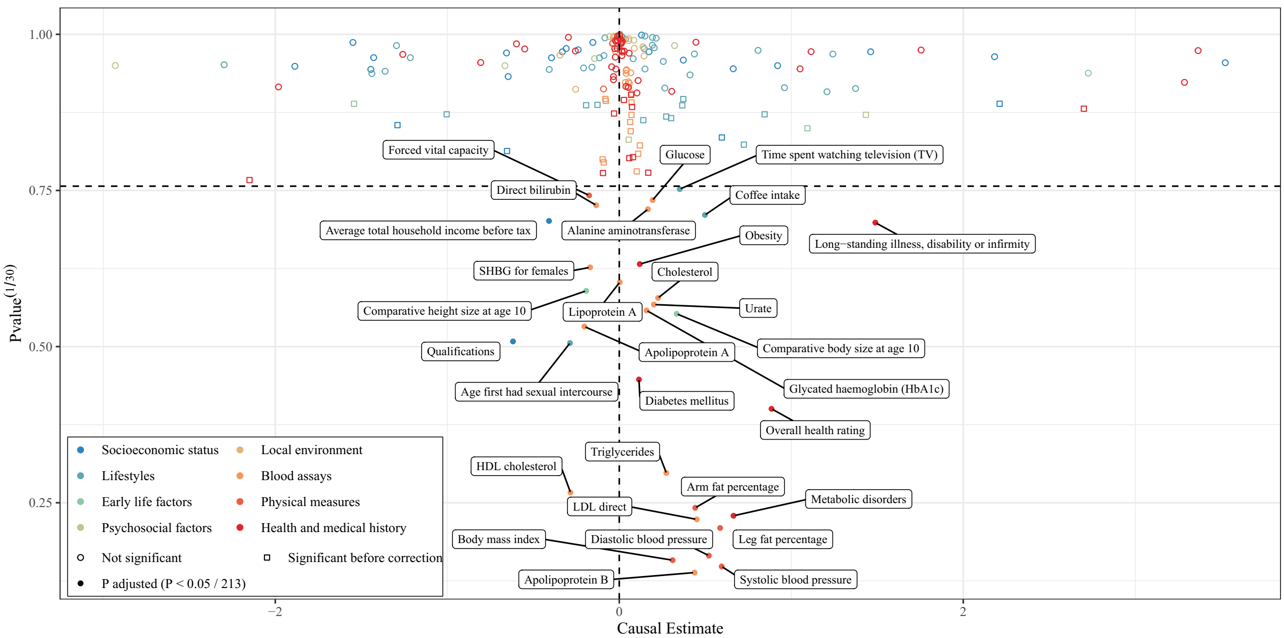


Fig. 4 | Mendelian randomization estimates of factors in relation to coronary heart disease risk. Causal estimates were generated using the inverse variance-weighted method after removing outliers. The dashed line represents the threshold after multiple testing correction (Bonferroni correction, p value < 2.35×10^{-04}). Two-sided statistical tests were performed. Dots represent ln (odds ratios). A set of the highest-risk factors was annotated. Source data are provided as a Source data file (Supplementary Data 7).

concentration, could not be robustly assessed within the MR framework. Because suitable instrumental variables for certain exposures were unavailable in the UKB, the present study utilized exposure genome-wide association studies (GWAS) data from the FinnGen consortium as an alternative. Potential bias arising from sample overlap between the exposure GWAS and the outcome dataset should be considered when interpreting corresponding results[40]. Our further examination of the interaction between genetic and modifiable factors revealed an additive interaction between the PRS for CHD and the weighted scores of the eight domains, with lifestyle, blood assays, as well as health and medical history being the most predominant ones. This addresses a limitation of previous studies, which have primarily focused on the interaction between lifestyles[41], local environment[42], and genetic risk. Notably, there was a gradual reduction in the effect size of this interaction with age, which may be related to survival bias, as individuals with high genetic risk and unfavorable lifestyles may be more susceptible to both CVD and premature mortality[43].

In this study, we found that the strength of association between multiple modifiable factors and CHD varied significantly with age, showing an overall decreasing trend as age increased. This observation is biologically plausible and aligns with findings from prior studies[44]. Our analysis further extended and quantified this phenomenon from multiple perspectives. First, using an EWAS framework, we systematically compared the exposure profiles and their explanatory power for CHD risk across different age groups. Although older individuals were exposed to more complex and diverse risk factors, the independent explanatory effects of each domain appeared to be attenuated. Second, age-specific comparisons of PAFs revealed distinct priorities for preventive interventions across the lifespan. In younger populations, lifestyle factors emerged as one of the main contributors to CHD risk, underscoring substantial opportunities for lifestyle-based prevention strategies. Conversely, among older adults, the contribution of lifestyle diminished, shifting the emphasis towards integrated management of chronic diseases, particularly within the context of multimorbidity, and the preservation of functional capacity. Furthermore, we observed that the additive interaction between genetic susceptibility and environmental exposures was more pronounced in younger individuals, suggesting that gene-environment synergy in early life may play a more substantial role in CHD development[22,45]. Taken together, our findings highlight the considerable heterogeneity in CHD risk profiles across different age groups, underscoring the importance of developing age-specific, stratified prevention strategies from a life course perspective. For individuals with high genetic risk, targeted interventions should focus on the accumulation of modifiable risk factors. Our study not only quantified the PAFs for each domain and overall modifiable burden, but also ranked individual factors by their importance. These results suggest that effective prevention should prioritize key factors within the most influential domains to achieve precision and efficiency in reducing the burden of CHD, especially in efforts to curb the rising trend of early-onset CHD among younger and middle-aged populations.

Using a data-driven EWAS framework and large-scale data from nearly 400,000 participants, this study systematically integrated environmental, behavioral, phenotypic, and genetic risk factors. It comprehensively evaluated the PAFs of modifiable risk factors, explored their additive interactions with PRS, and examined age-related trends in these interactions. These findings provide valuable insights into the multifactorial pathogenesis of CHD and offer practical guidance for developing age-specific precision prevention strategies. However, several limitations should be acknowledged. First, the generalizability of our findings is limited by volunteer bias and the age range of UK Biobank participants (40–69 years), highlighting the need to study younger populations. Second, although most modifiable factors were objectively measured or derived from validated tools, some domains (e.g., early life factors) were sparsely represented, which may underestimate their contribution. Third, the use of Bonferroni correction, while conservative, may have masked true associations. Residual confounding may also remain despite adjustments for collinearity and domain-specific factors. Fourth, this study did not systematically assess the impact of interactions between modifiable factors or gene-environment interactions on the estimated attributable risks. Multimorbidity was only partially reflected within the "health and medical history" domain, without detailed analysis of specific disease combinations or their synergistic effects. Fifth, the PAF

Table 1 | Risk of incident coronary heart disease according to the categories of the eight domains

Domains	*n*/*N*	HR	*p* value	*p* for trend
Socioeconomic status				
Favorable	6905/136,714	1.00 (Reference)	-	
Intermediate	9399/140,380	1.06 (1.03, 1.09)	3.57×10^{-04}	1.04×10^{-32}
Unfavorable	11,288/117,485	1.21 (1.17, 1.26)	1.52×10^{-29}	
Lifestyles				
Favorable	6828/131,527	1.00 (Reference)	-	
Intermediate	8717/131,526	1.11 (1.07, 1.15)	2.35×10^{-10}	1.80×10^{-72}
Unfavorable	12,047/131,526	1.33 (1.29, 1.37)	6.93×10^{-66}	
Early life factors				
Favorable	9013/132,366	1.00 (Reference)	-	
Intermediate	9267/132,742	1.00 (0.97, 1.03)	9.96×10^{-01}	6.61×10^{-03}
Unfavorable	9312 /129,471	1.04 (1.01, 1.07)	1.05×10^{-02}	
Psychosocial factors				
Favorable	7701/131,538	1.00 (Reference)	-	
Intermediate	9056/131,668	0.98 (0.95, 1.01)	1.25×10^{-01}	5.47×10^{-06}
Unfavorable	10,835/131,373	1.06 (1.03, 1.10)	8.43×10^{-05}	
Local environment				
Favorable	8899/132,918	1.00 (Reference)	-	
Intermediate	9515/130,137	1.05 (1.01, 1.08)	3.58×10^{-03}	1.66×10^{-02}
Unfavorable	9178/131,524	1.04 (1.00, 1.07)	2.96×10^{-02}	
Blood assays				
Favorable	5010/131,527	1.00 (Reference)	-	
Intermediate	8213/131,526	1.20 (1.16, 1.25)	2.68×10^{-24}	1.53×10^{-205}
Unfavorable	14,369/131,526	1.65 (1.59, 1.70)	2.41×10^{-177}	
Physical measures				
Favorable	7210/131,527	1.00 (Reference)	-	
Intermediate	8686/132,932	1.11 (1.07, 1.14)	1.81×10^{-09}	4.89×10^{-62}
Unfavorable	11,696/130,120	1.34 (1.29, 1.39)	1.81×10^{-58}	
Health and medical history				
Favorable	5512/131,836	1.00 (Reference)	-	
Intermediate	7858/131,217	1.24 (1.20, 1.28)	6.89×10^{-33}	1.03×10^{-267}
Unfavorable	14,222/131,526	1.78 (1.72, 1.84)	7.78×10^{-239}	

The favorable profile was set as a reference in each domain. The associations were estimated using a Cox proportional hazards regression model that included all eight domains, mutually adjusted, and with adjustments for baseline age, sex, ethnic background, family history of cardiovascular disease, and assessment center. Two-sided statistical tests were used.
N number of individuals at risk, *n* number of CHD cases.

may overestimate the reduction in risk when multiple overlapping factors are considered. Therefore, we applied conservative models and weighting procedures, as well as multiple sensitivity analyses, to reduce potential bias. However, the estimates should still be interpreted with caution. Finally, as PAFs are population-specific, our estimates may not generalize to other countries or ethnic groups, although many UKB-based risk estimates align with those from more representative cohorts[46,47]. Therefore, further validation in diverse populations is needed to ensure the effectiveness and appropriateness of prevention strategies.

In summary, our comprehensive analysis demonstrates that increased risk of CHD is associated with SES, lifestyles, early life factors, psychosocial factors, local environment, blood assays, physical measures, as well as health and medical history. Genetic risk amplifies the association between modifiable factors and CHD, particularly among younger populations. Importantly, approximately 40%–62% of CHD cases may be preventable through modification of identified risk factors. The proportion of CHD cases attributable to modifiable factors declines with age, from 70.81% in individuals under 50 years to 60.20% in those aged 60 years and older. These findings underscore the need to integrate a more comprehensive range of modifiable factors into public health strategies to reduce the incidence of CHD, particularly by mitigating the accumulation of modifiable factors during early life stages. A comprehensive approach may substantially reduce the overall CHD burden and help reverse the rising trend of early-onset CHD.

Method

Study design

We employed Cox proportional hazards regression models in an EWAS framework to examine the association between 213 modifiable factors and CHD. Two-sample MR analysis was then conducted for each of the 213 factors to identify causal relationships between modifiable factors and CHD. Subsequently, weighted standardized scores were constructed to assess the combined effect of risk factors within each domain, with an exploration of variations across different age groups (<50, 50–60, and ≥60 years). Additionally, we assessed additive interactions between weighted standardized scores and genetic risk on CHD. Finally, we estimated PAFs and weighted PAFs for each domain.

Study population and CHD diagnosis

We drew our study population from the UKB study, acquiring baseline data on 502,164 participants between 2006 and 2010 who were

 https://doi.org/10.1038/s41467-025-65963-x

Domains	<50 years n / N	<50 years HR (95% CI)	<50 years p−value	50−60 years n / N	50−60 years HR (95% CI)	50−60 years p−value	≥60 years n / N	≥60 years HR (95% CI)	≥60 years p−value
Socioeconomic status									
Favourable	995 / 43073	1.00 (Reference)	−	3042 / 61550	1.00 (Reference)	−	2868 / 32091	1.00 (Reference)	−
Intermediate	969 / 31916	1.08 (0.98, 1.18)	1.05×10^{-01}	3123 / 53370	1.07 (1.01, 1.12)	1.54×10^{-02}	5307 / 55094	1.03 (0.98, 1.08)	2.02×10^{-01}
Unfavorable	1039 / 23132	1.21 (1.10, 1.34)	1.34×10^{-04}	3336 / 39139	1.27 (1.20, 1.34)	3.97×10^{-17}	6913 / 55214	1.17 (1.11, 1.22)	2.51×10^{-10}
Lifestyles									
Favourable	508 / 30973	1.00 (Reference)	−	2153 / 49747	1.00 (Reference)	−	4167 / 50807	1.00 (Reference)	−
Intermediate	815 / 31192	1.27 (1.13, 1.42)	4.07×10^{-05}	2853 / 51464	1.10 (1.04, 1.16)	1.60×10^{-03}	5049 / 48870	1.11 (1.06, 1.15)	2.88×10^{-06}
Unfavorable	1680 / 35956	1.64 (1.47, 1.83)	6.64×10^{-19}	4495 / 52848	1.36 (1.28, 1.44)	4.52×10^{-26}	5872 / 42722	1.27 (1.21, 1.32)	5.82×10^{-27}
Early life factors									
Favourable	829 / 30325	1.00 (Reference)	−	2780 / 48854	1.00 (Reference)	−	5404 / 53187	1.00 (Reference)	−
Intermediate	903 / 33400	0.92 (0.84, 1.01)	8.26×10^{-02}	3049 / 50111	1.01 (0.96, 1.06)	7.25×10^{-01}	5315 / 49231	1.01 (0.98, 1.05)	5.02×10^{-01}
Unfavorable	1271 / 34396	1.10 (1.00, 1.20)	4.50×10^{-02}	3672 / 55094	1.03 (0.98, 1.08)	2.86×10^{-01}	4369 / 39981	1.03 (0.99, 1.07)	1.59×10^{-01}
Psychosocial factors									
Favourable	665 / 30901	1.00 (Reference)	−	2359 / 48809	1.00 (Reference)	−	4677 / 51828	1.00 (Reference)	−
Intermediate	801 / 29702	1.00 (0.90, 1.11)	9.35×10^{-01}	2963 / 51292	0.97 (0.91, 1.02)	2.26×10^{-01}	5292 / 50674	0.98 (0.94, 1.02)	3.34×10^{-01}
Unfavorable	1537 / 37518	1.11 (1.01, 1.23)	2.80×10^{-02}	4179 / 53958	1.05 (0.99, 1.11)	7.68×10^{-02}	5119 / 39897	1.06 (1.01, 1.10)	8.32×10^{-03}
Local environment									
Favourable	808 / 28701	1.00 (Reference)	−	2862 / 51797	1.00 (Reference)	−	5229 / 52420	1.00 (Reference)	−
Intermediate	1059 / 32297	1.03 (0.94, 1.13)	5.68×10^{-01}	3320 / 50920	1.07 (1.02, 1.13)	6.98×10−03	5136 / 46920	1.03 (0.99, 1.07)	1.11×10^{-01}
Unfavorable	1136 / 37123	0.95 (0.86, 1.06)	3.70×10^{-01}	3319 / 51342	1.06 (1.00, 1.12)	4.31×10^{-02}	4723 / 43059	1.04 (1.00, 1.09)	6.35×10^{-02}
Blood assays									
Favourable	653 / 43252	1.00 (Reference)	−	1838 / 52145	1.00 (Reference)	−	2519 / 36130	1.00 (Reference)	−
Intermediate	858 / 30049	1.34 (1.21, 1.49)	2.58×10^{-08}	2841 / 52263	1.22 (1.15, 1.29)	8.46×10^{-11}	4514 / 49214	1.15 (1.09, 1.20)	5.88×10^{-08}
Unfavorable	1492 / 24820	2.14 (1.93, 2.37)	6.10×10^{-49}	4822 / 49651	1.72 (1.62, 1.82)	1.04×10^{-74}	8055 / 57055	1.50 (1.43, 1.57)	1.94×10^{-64}
Physical measures									
Favourable	1421 / 50027	1.00 (Reference)	−	3034 / 52973	1.00 (Reference)	−	2755 / 28527	1.00 (Reference)	−
Intermediate	856 / 30929	1.11 (1.01, 1.22)	2.32×10^{-02}	3056 / 54030	1.13 (1.08, 1.20)	2.69×10^{-06}	4774 / 47973	1.07 (1.02, 1.12)	9.05×10^{-03}
Unfavorable	726 / 17165	1.26 (1.13, 1.39)	2.15×10^{-05}	3411 / 47056	1.33 (1.25, 1.41)	1.19×10^{-21}	7559 / 65899	1.33 (1.26, 1.40)	4.65×10^{-29}
Health and medical history									
Favourable	654 / 38543	1.00 (Reference)	−	2056 / 52685	1.00 (Reference)	−	2802 / 40608	1.00 (Reference)	−
Intermediate	884 / 32547	1.32 (1.19, 1.46)	2.11×10^{-07}	2621 / 51005	1.18 (1.11, 1.25)	4.79×10^{-08}	4353 / 47665	1.25 (1.19, 1.31)	8.63×10^{-20}
Unfavorable	1465 / 27031	1.91 (1.72, 2.11)	5.30×10^{-34}	4824 / 50369	1.74 (1.64, 1.84)	6.61×10^{-79}	7933 / 54126	1.77 (1.69, 1.85)	7.70×10^{-127}

Fig. 5 | Risk of incident coronary heart disease according to categories of eight domains within each age group. The favorable profile was set as a reference in each domain. The associations were estimated using a Cox proportional hazards regression model that included all eight domains, mutually adjusted, and with adjustments for baseline age, sex, ethnic background, family history of cardiovascular disease, and assessment center. The points represent hazard ratios (HR), and the horizontal lines represent the corresponding 95% confidence interval (CI). Two-sided statistical tests were used. *N* number of individuals at risk, *n* number of CHD cases. Source data are provided as a Source data file (Supplementary Data 8).

Modifiable factor	Genetic risk	Socioeconomic status n / N	Socioeconomic status HR (95% CI)	Lifestyles n / N	Lifestyles HR (95% CI)	Early life factors n / N	Early life factors HR (95% CI)	Psychosocial factors n / N	Psychosocial factors HR (95% CI)
Favourable	Lower	1570 / 46907	1.00 (Reference)	1567 / 45290	1.00 (Reference)	2159 / 44795	1.00 (Reference)	1828 / 44528	1.00 (Reference)
Favourable	Intermediate	2284 / 45645	1.50 (1.41, 1.60)	2232 / 44175	1.47 (1.37, 1.56)	2905 / 43944	1.37 (1.29, 1.44)	2473 / 43737	1.38 (1.30, 1.46)
Favourable	Higher	3051 / 44162	2.09 (1.96, 2.22)	3029 / 42062	2.13 (2.00, 2.27)	3949 / 43627	1.90 (1.80, 2.00)	3400 / 43273	1.93 (1.83, 2.05)
Intermediate	Lower	2263 / 46549	1.16 (1.09, 1.24)	2120 / 44186	1.21 (1.14, 1.30)	2262 / 44859	1.03 (0.97, 1.09)	2203 / 43906	1.02 (0.96, 1.09)
Intermediate	Intermediate	2984 / 46877	1.52 (1.43, 1.61)	2813 / 43761	1.63 (1.53, 1.74)	3073 / 44308	1.40 (1.32, 1.48)	2913 / 44152	1.33 (1.26, 1.41)
Intermediate	Higher	4152 / 46954	2.16 (2.03, 2.29)	3784 / 43579	2.24 (2.11, 2.38)	3932 / 43575	1.84 (1.75, 1.94)	3940 / 43610	1.86 (1.76, 1.96)
Unfavorable	Lower	2758 / 38071	1.38 (1.30, 1.47)	2904 / 42051	1.52 (1.43, 1.62)	2170 / 41873	1.05 (0.99, 1.12)	2560 / 43093	1.10 (1.03, 1.17)
Unfavorable	Intermediate	3711 / 39004	1.82 (1.71, 1.94)	3934 / 43590	1.99 (1.87, 2.11)	3001 / 43274	1.42 (1.35, 1.51)	3593 / 43637	1.52 (1.44, 1.61)
Unfavorable	Higher	4819 / 40410	2.34 (2.20, 2.48)	5209 / 45885	2.57 (2.43, 2.73)	4141 / 44324	1.94 (1.84, 2.05)	4682 / 44643	1.97 (1.87, 2.09)

Modifiable factor	Genetic risk	Local environment n / N	Local environment HR (95% CI)	Blood assays n / N	Blood assays HR (95% CI)	Physical measures n / N	Physical measures HR (95% CI)	Health and medical history n / N	Health and medical history HR (95% CI)
Favourable	Lower	2085 / 44759	1.00 (Reference)	1301 / 46981	1.00 (Reference)	1630 / 45938	1.00 (Reference)	1293 / 46261	1.00 (Reference)
Favourable	Intermediate	2919 / 44148	1.43 (1.36, 1.52)	1646 / 43759	1.37 (1.27, 1.47)	2386 / 44038	1.53 (1.44, 1.63)	1759 / 43964	1.44 (1.34, 1.55)
Favourable	Higher	3895 / 44011	1.93 (1.83, 2.03)	2063 / 40787	1.87 (1.75, 2.01)	3194 / 41551	2.20 (2.07, 2.33)	2460 / 41611	2.19 (2.05, 2.35)
Intermediate	Lower	2311 / 42763	1.12 (1.05, 1.19)	1979 / 44290	1.16 (1.08, 1.25)	2139 / 44686	1.25 (1.17, 1.33)	1856 / 43681	1.32 (1.23, 1.42)
Intermediate	Intermediate	3026 / 43733	1.43 (1.35, 1.51)	2638 / 44024	1.60 (1.49, 1.71)	2787 / 44178	1.65 (1.55, 1.76)	2513 / 43888	1.79 (1.68, 1.92)
Intermediate	Higher	4178 / 43641	2.01 (1.91, 2.12)	3596 / 43212	2.30 (2.16, 2.45)	3760 / 44068	2.29 (2.16, 2.43)	3489 / 43648	2.56 (2.40, 2.74)
Unfavorable	Lower	2195 / 44005	1.09 (1.02, 1.16)	3311 / 40256	1.66 (1.56, 1.77)	2822 / 40903	1.59 (1.49, 1.69)	3442 / 41585	2.04 (1.91, 2.18)
Unfavorable	Intermediate	3034 / 43645	1.49 (1.40, 1.58)	4695 / 43743	2.23 (2.10, 2.38)	3806 / 43310	2.03 (1.91, 2.16)	4707 / 43674	2.70 (2.53, 2.88)
Unfavorable	Higher	3949 / 43874	1.96 (1.85, 2.07)	6363 / 47527	2.92 (2.75, 3.10)	5068 / 45907	2.62 (2.47, 2.78)	6073 / 46267	3.40 (3.20, 3.62)

Fig. 6 | Risk of incident coronary heart disease according to categories of eight domains within each genetic risk category. Participants were categorized into nine groups based on the tertiles of the PRS, combined with domain-weighted scores. Hazard ratios were generated using a Cox model that included all eight domains, mutually adjusted, and with adjustments for baseline age, sex, ethnic background, family history of cardiovascular disease, and assessment center. The points represent hazard ratios (HR), and the horizontal lines represent the corresponding 95% confidence interval (CI). Two-sided statistical tests were used. The additive interaction results of the PRS and the weighted scores for each domain are available in Supplementary Data 9 and 10. *N* number of individuals at risk, *n* number of CHD cases. Source data are provided as a Source data file.

followed up until the earliest occurrence of a first diagnosis of CHD, death, loss to follow-up, or the last available information (up to 1 September 2023). CHD was identified using relevant International Classification of Diseases codes (I20, I21, I22, I23, I24, and I25) from the Health Outcomes dataset in UKB, comprising primary care cases, hospital records, death registrations, and algorithmically defined outcomes. We excluded participants with pre-existing CHD or stroke at baseline, those with >15% missing data regarding the included variables, and individuals with missing outcomes or covariates. Finally, 394,579 participants (aged 40–69 years at recruitment, including both men and women) who provided informed consent were included in the present study; participation in the UK Biobank is entirely voluntary and without financial compensation. The North West Multi-center Research Ethics Committee approved this study. The analysis was conducted under UKB application number 199688. We followed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) reporting guideline for cohort studies.

Modifiable factors

After excluding variables with >20% missing data and those classified as unmodifiable (https://ukbiobank.dnanexus.com/panx/projects), 213 modifiable factors, either measured or derived at baseline, were included for subsequent analysis. These factors spanned eight domains, involving SES (e.g., household income and accommodation type), lifestyle (e.g., physical activity), early life factors (e.g., breast-feeding), psychosocial factors (e.g., neuroticism score), local environment (e.g., distance to major roads), blood assays (e.g., platelet

Table 2 | Weighted and unweighted PAFs for eight domains

Domains	Model 1			Model 2		
	Unweighted PAF (%)	Communality	Weighted PAF (%)	Unweighted PAF (%)	Communality	Weighted PAF (%)
Socioeconomic status	6.09	0.18	3.70	11.69	0.19	6.15
Lifestyles	9.84	0.24	5.97	19.02	0.24	10.01
Early life factors	1.64	0.07	0.99	3.39	0.25	1.79
Psychosocial factors	3.40	0.31	2.06	5.90	0.27	3.11
Local environment	0.54	0.28	0.33	4.80	0.20	2.52
Blood assays	15.72	0.27	9.54	26.35	0.20	13.87
Physical measures	9.39	0.41	5.70	16.53	0.42	8.70
Health and medical history	18.74	0.23	11.37	29.93	0.23	15.75
Overall PAF						
All participants			39.66			61.89
<50 years			49.34			70.81
50–60 years			42.27			64.10
≥60 years			38.92			60.20

Weighted PAFs were calculated after considering the overlap between risk factors.
In Model 1, individuals in the unfavorable group (representing the highest-risk tertile) were shifted to either the intermediate or favorable group, corresponding to the elimination of approximately one-third of the most adverse risk profiles.
In Model 2, all individuals were shifted to the favorable group (representing the lowest-risk tertile), corresponding to the elimination of two-thirds of modifiable risk factors.

count), physical measurements (e.g., blood pressure), and medical history (e.g., overall health rating). Data cleaning and processing for these variables are detailed in Supplementary Methods and Supplementary Data 1–3. We employed multiple imputation using chained equations, generating five imputed datasets over ten iterations. For each variable, an imputation method was automatically assigned according to its data type, incorporating the 10 most relevant predictors, with age and sex included by default[48].

PRS of CHD

A subset of 337,151 individuals from the UKB, known as the White British Unrelated (WBU) group, was created by Bycroft et al.[49] to include individuals of British European ancestry while excluding closely related participants. This WBU subset, referred to as the "training subset," was used to generate summary statistics from GWAS and, subsequently, meta-analyzed with external GWAS datasets to develop the Enhanced PRS set. Further details can be found in a previous publication[50]. Using this approach, the PRS for CHD was derived for over 480,000 individuals in the UKB (field ID: 26227).

Statistical analyses

We performed the EWAS using Cox proportional hazard regression models to test the association between each of the 213 baseline exposures and newly diagnosed CHD, using Bonferroni-corrected significance thresholds (0.05 divided by 213 tests, resulting in 2.35×10^{-4}). We made further adjustments for age, sex, ethnic background, family history of CVD, and assessment center. An interaction term with the time of follow-up was added in case of violated proportional hazards assumption ($p < 0.0001$ via a test using Schoenfeld's residuals)[51]. We conducted stratified analyses based on the baseline age, sex, PRS, and time of follow-up, identifying significance by the Bonferroni correction method. We employed random forest models combined with Cox proportional hazards regression to rank the importance of 213 modifiable variables by calculating minimal depth—a measure of the depth at which a variable splits—to prioritize modifiable factors that may serve as effective targets for intervention[52].

To explore the causal relationship between modifiable factors and CHD, we incorporated a two-sample MR into the EWAS analytical framework. We utilized GWAS data for CHD (phenocode: I9_IHD) from the FinnGen database, including 75,592 cases and 378,141 controls. We prioritized GWAS data from the FinnGen database, where medical history-related factors were available. For the remaining variables, data from the MRC IEU OpenGWAS database (https://gwas.mrcieu.ac.uk/) were sourced to access publicly available UKB-based summary statistics. Supplementary Data 1–3 summarizes the GWAS IDs for each variable in detail. We identified significant single-nucleotide polymorphisms (SNPs) for instrument variables ($p < 5 \times 10^{-08}$). For traits with ≤3 SNPs after outlier removal, we either relaxed the threshold ($p < 5 \times 10^{-06}$) or excluded the trait. Inverse variance weighting was chosen as the primary method, with outliers removed. Subsequently, potential heterogeneity and horizontal pleiotropy were assessed using Cochran's Q test, MR Egger intercept, and MR-PRESSO global test. Moreover, other methods for MR, such as weighted median and MR Egger regression, were applied to ensure robust conclusions.

Following previous methodology[11], weighted standardized scores for each domain were also calculated by reversing the coding for protective factors (HR < 1) identified in the EWAS, normalizing continuous variables, and dummy-coding categorical variables. The weighted standardized scores were determined based on adjustments for variables within the same domain using Cox proportional hazards regression models, accounting for age, sex, ethnic background, family history of CVD, and assessment center. These scores were obtained by multiplying the raw value of each variable by its corresponding β coefficient and dividing by the total sum of β coefficients. Higher weighted scores indicated greater exposure to risk factors. According to these scores, three tertiles were established, including unfavorable (higher risk), moderate, and favorable (lower risk). Then, the Cox proportional hazards regression model was also used to assess the association between domain-specific weighted scores and CHD risk, with covariate adjustment as described above. Subsequently, we conducted stratified analysis by age (<50, 50–60, and ≥60 years) to explore the age-related heterogeneity in the relationship between domain-specific scores and CHD. To construct the CHD genetic risk score, we classified participants into nine groups based on tertiles of the PRS combined with the domain-weighted scores. We also examined CHD risk across different combinations of genetic risk and modifiable factors using a Cox proportional hazards regression model, with the group having low genetic risk and a favorable weighted score as the reference. Additionally, we assessed the additive interaction effect between genetic risk and weighted scores across eight domains of CHD risk.

Finally, the PAF was calculated for each domain, representing the proportion of reduction in specific disease risk that would result from replacing a given risk factor with a more favorable profile. We employed two models: Model 1 eliminated the most adverse one-third of modifiable factors by integrating intermediate and unfavorable profiles across the eight domains to produce a more conservative outcome; Model 2 eliminated modifiable factors by 2/3 to estimate the proportion of coronary events that could be reduced, indicating a more complete elimination of unfavorable factors. The PAFs for each domain were calculated using the "graphPAF" package, employing a Cox proportional hazards regression model with adjustments for sex, age, assessment center, ethnic background, and family history of CVD[53]. To account for correlations among the eight domains, the common factor variance was calculated through principal component analysis to estimate the weights of each PAF. These weights were then used to calculate both the combination-weighted and individual-weighted PAFs, accounting for the coexistence of risk factors within individuals, which resulted in reduced overestimation of PAFs caused by factor interactions[54]. Additionally, we conducted three sensitivity analyses to assess the robustness of the PAF estimates as follows: (1) analysis restricted to factors associated significantly with CHD ($p < 0.05$); (2) additional adjustment for the PRS; and (3) omitting mutual adjustment between domains.

All *P* values were two-sided, and analyses were conducted using R v4.3.2 (R Core Team, 2023).

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The data used in the present study are available from UKB with restrictions applied. Data were used under license and are therefore not publicly available. Access can be obtained by applying to the UK Biobank through the standard protocol (https://www.ukbiobank.ac.uk/register-apply/). The processed data generated in this study are available in the Supplementary Information and the Source data file. Publicly available GWAS summary statistics for risk factors can be obtained from the MRC IEU OpenGWAS database (https://gwas.mrcieu.ac.uk/). Summary statistics for CHD and other diseases (e.g., diabetes mellitus and obesity) are available from the FinnGen consortium (https://www.finngen.fi/en/access_results). Source data are provided with this paper.

Code availability

Scripts used to perform the analyses are available at https://github.com/gjh0828/UKB_CHD_EWAS.git.

References

1. GBD 2021 Diseases and Injuries Collaborators Global incidence, prevalence, years lived with disability (YLDs), disability-adjusted life-years (DALYs), and healthy life expectancy (HALE) for 371 diseases and injuries in 204 countries and territories and 811 subnational locations, 1990-2021: a systematic analysis for the Global Burden of Disease Study 2021. *Lancet* **403**, 2133–2161 (2024).
2. Mensah, G. A., Roth, G. A. & Fuster, V. The global burden of cardiovascular diseases and risk factors: 2020 and beyond. *J. Am. Coll. Cardiol.* **74**, 2529–2532 (2019).
3. The Global Cardiovascular Risk Consortium Global effect of modifiable risk factors on cardiovascular disease and mortality. *N. Engl. J. Med.* **389**, 1273–1285 (2023).
4. Pencina, M. J. et al. Quantifying importance of major risk factors for coronary heart disease. *Circulation* **139**, 1603–1611 (2019).
5. Tzoulaki, I., Elliott, P., Kontis, V. & Ezzati, M. Worldwide exposures to cardiovascular risk factors and associated health effects. *Circulation* **133**, 2314–2333 (2016).
6. GBD 2019 Diseases and Injuries Collaborators Global burden of 369 diseases and injuries in 204 countries and territories, 1990-2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet* **396**, 1204–1222 (2020).
7. Roth, G. A. et al. Global burden of cardiovascular diseases and risk factors, 1990–2019: update from the GBD 2019 study. *J. Am. Coll. Cardiol.* **76**, 2982–3021 (2020).
8. Ezzati, M. et al. Contributions of risk factors and medical care to cardiovascular mortality trends. *Nat. Rev. Cardiol.* **12**, 508–530 (2015).
9. Georgiou, A. N. et al. Appraising the causal role of risk factors in coronary artery disease and stroke: a systematic review of mendelian randomization studies. *J. Am. Heart Assoc.* **12**, e029040 (2023).
10. Choi, K. W. et al. An exposure-wide and Mendelian randomization approach to identifying modifiable factors for the prevention of depression. *Am. J. Psychiatry* **177**, 944–954 (2020).
11. Zhang, Y. et al. Identifying modifiable factors and their joint effect on dementia risk in the UK Biobank. *Nat. Hum. Behav.* **7**, 1185–1195 (2023).
12. Vermeulen, R., Schymanski, E. L., Barabási, A. L. & Miller, G. W. The exposome and health: Where chemistry meets biology. *Science* **367**, 392–396 (2020).
13. Licher, S. et al. Genetic predisposition, modifiable-risk-factor profile and long-term dementia risk in the general population. *Nat. Med.* **25**, 1364–1369 (2019).
14. Sheehan, A., Freni Sterrantino, A., Fecht, D., Elliott, P. & Hodgson, S. Childhood type 1 diabetes: an environment-wide association study across England. *Diabetologia* **63**, 964–976 (2020).
15. Wang, T. et al. Age-specific modifiable risk factor profiles for cardiovascular disease and all-cause mortality: a nationwide, population-based, prospective cohort study. *Lancet Reg. Health West. Pac.* **17**, 100277 (2021).
16. Li, S. et al. Modifiable risk factors associated with cardiovascular disease and mortality in China: a PURE substudy. *Eur. Heart J.* **43**, 2852–2863 (2022).
17. Tian, F. et al. Ranking age-specific modifiable risk factors for cardiovascular disease and mortality: evidence from a population-based longitudinal study. *eClinicalMedicine* **64**, 102230 (2023).
18. Manrai, A. K. et al. Informatics and data analytics to support exposome-based discovery for public health. *Annu. Rev. Public Health* **38**, 279–294 (2017).
19. Lawlor, D. A., Harbord, R. M., Sterne, J. A., Timpson, N. & Davey Smith, G. Mendelian randomization: using genes as instruments for making causal inferences in epidemiology. *Stat. Med.* **27**, 1133–1163 (2008).
20. Wang, H. et al. Pathogenesis of premature coronary artery disease: Focus on risk factors and genetic variants. *Genes Dis.* **9**, 370–380 (2022).
21. Wang, N. et al. Long-term night shift work is associated with the risk of atrial fibrillation and coronary heart disease. *Eur. Heart J.* **42**, 4180–4188 (2021).
22. Sun, D. et al. Joint impact of polygenic risk score and lifestyles on early- and late-onset cardiovascular diseases. *Nat. Hum. Behav.* **8**, 1810–1818 (2024).
23. Rothman, K. J., Greenland, S. & Walker, A. M. Concepts of interaction. *Am. J. Epidemiol.* **112**, 467–470 (1980).

24. Inouye, M. et al. Genomic risk prediction of coronary artery disease in 480,000 adults: implications for primary prevention. *J. Am. Coll. Cardiol.* **72**, 1883–1893 (2018).
25. McClellan, M., Brown, N., Califf, R. M. & Warner, J. J. Call to Action: urgent challenges in cardiovascular disease: a presidential advisory from the American Heart Association. *Circulation* **139**, e44–e54 (2019).
26. Goodarzi, M. O. & Rotter, J. I. Genetics insights in the relationship between type 2 diabetes and coronary heart disease. *Circ. Res.* **126**, 1526–1548 (2020).
27. Pi-Sunyer, F. X. The obesity epidemic: pathophysiology and consequences of obesity. *Obes. Res.* **10**, 97S–104S (2002).
28. Gobal, F., Deshmukh, A., Shah, S. & Mehta, J. L. Triad of metabolic syndrome, chronic kidney disease, and coronary heart disease with a focus on microalbuminuria death by overeating. *J. Am. Coll. Cardiol.* **57**, 2303–2308 (2011).
29. Hage, F. G. et al. The scope of coronary heart disease in patients with chronic kidney disease. *J. Am. Coll. Cardiol.* **53**, 2129–2140 (2009).
30. Rugulies, R. Depression as a predictor for coronary heart disease. a review and meta-analysis. *Am. J. Prev. Med.* **23**, 51–61 (2002).
31. Osborn, D. P. J. et al. Relative risk of diabetes, dyslipidaemia, hypertension and the metabol ic syndrome in people with severe mental illnesses: systematic review and metaanalysis. *BMC Psychiatry* **8**, 84 (2008).
32. Hasbani, N. R. et al. American Heart Association's Life's Simple 7: lifestyle recommendations, polygenic risk, and lifetime risk of coronary heart disease. *Circulation* **145**, 808–818 (2022).
33. Fan, M. et al. Sleep patterns, genetic susceptibility, and incident cardiovascular disease: a prospective study of 385 292 UK Biobank participants. *Eur. Heart J.* **41**, 1182–1189 (2020).
34. Wallentin, L. et al. Plasma proteins associated with cardiovascular death in patients with chronic coronary heart disease: a retrospective study. *PLOS Med.* **18**, e1003513 (2021).
35. Wan, E. S. et al. Association between preserved ratio impaired spirometry and clinical outcomes in US adults. *JAMA* **326**, 2287–2298 (2021).
36. Hamad, R. et al. Association of low socioeconomic status with premature coronary heart disease in US adults. *JAMA Cardiol.* **5**, 899–908 (2020).
37. Peterson, P. N. JAHA spotlight on psychosocial factors and cardiovascular disease. *J. Am. Heart Assoc.* **9**, e017112 (2020).
38. Pedamallu, H., Zmora, R., Perak, A. M. & Allen, N. B. Life course cardiovascular health: risk factors, outcomes, and interventions. *Circ. Res.* **132**, 1570–1583 (2023).
39. Burgess, S., Thompson, S. G. & Collaboration, C. C. G. Avoiding bias from weak instruments in Mendelian randomization studies. *Int. J. Epidemiol.* **40**, 755–764 (2011).
40. Lawlor, D. A. Commentary: two-sample Mendelian randomization: opportunities and challenges. *Int. J. Epidemiol.* **45**, 908–915 (2016).
41. Said, M. A., 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.* **3**, 693–702 (2018).
42. Fu, Z. et al. Association of air pollution exposure and increased coronary artery disease risk: the modifying effect of genetic susceptibility. *Environ. Health* **22**, 85 (2023).
43. Surakka, I. et al. Sex-specific survival bias and interaction modeling in coronary artery disease risk prediction. *Circ. Genom. Precis. Med.* **16**, e003542 (2023).
44. Dugani, S. B. et al. Association of lipid, inflammatory, and metabolic biomarkers with age at onset for incident coronary heart disease in women. *JAMA Cardiol.* **6**, 437–447 (2021).
45. Jukarainen, S. et al. Genetic risk factors have a substantial impact on healthy life years. *Nat. Med.* **28**, 1893–1901 (2022).
46. Lee, M. et al. Variation in population attributable fraction of dementia associated with potentially modifiable risk factors by race and ethnicity in the US. *JAMA Netw. Open* **5**, e2219672 (2022).
47. Batty, G. D., Gale, C. R., Kivimäki, M., Deary, I. J. & Bell, S. Comparison of risk factor associations in UK Biobank against representative, general population based studies with conventional response rates: prospective cohort study and individual participant meta-analysis. *BMJ* **368**, m131 (2020).
48. van Buuren, S., Boshuizen, H. C. & Knook, D. L. Multiple imputation of missing blood pressure covariates in survival analysis. *Stat. Med.* **18**, 681–694 (1999).
49. Bycroft, C. et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
50. Thompson, D. J. et al. A systematic evaluation of the performance and properties of the UK Biobank Polygenic Risk Score (PRS) Release. *PLOS ONE* **19**, e0307270 (2024).
51. Ganna, A. & Ingelsson, E. 5 year mortality predictors in 498,103 UK Biobank participants: a prospective population-based study. *Lancet* **386**, 533–540 (2015).
52. Ishwaran, H., Kogalur, U. B., Gorodeski, E. Z., Minn, A. J. & Lauer, M. S. High-dimensional variable selection for survival data. *J. Am. Stat. Assoc.* **105**, 205–217 (2010).
53. Ferguson, J. & O'Connell, M. Estimating and displaying population attributable fractions using the R package: graphPAF. *Eur. J. Epidemiol.* **39**, 715–742 (2024).
54. Mukadam, N., Sommerlad, A., Huntley, J. & Livingston, G. Population attributable fractions for risk factors for dementia in low-income and middle-income countries: an analysis using cross-sectional survey data. *Lancet Glob. Health* **7**, e596–e603 (2019).

Acknowledgements
J.L. was supported by the National Key Research and Development Program of China (Grant number: 2023YFC3709302), the National Natural Science Foundation of China (Grant numbers: 42307133), the Special Fund of the State Key Joint Laboratory of Environmental Simulation and Pollution Control (Grant number: 23K02ESPCP), and the Young Elite Scientists Sponsorship Program by CAST (Grant number: 2023QNRC001). S.H. were supported by the National Natural Science Foundation of China (Grant numbers: 52478073). This study was also supported by the High-performance Computing Platform of Peking University.

Author contributions
J.G., S.H., and J.L. had full access to all data in the study and took responsibility for the integrity of the data and the accuracy of the analysis. J.G., S.H., and J.L. conceived and designed the project. All authors contributed to data acquisition, analysis, or interpretation. J.G., S.H., and J.L. performed the statistical analyses. S.H. and J.L. obtained funding. J.G., P.K., C.L.Z.V., Z.T., S.H., and J.L. drafted the manuscript. Y.F., Y.W., X.L., Z.W., and J.L. critically revised the manuscript. All authors reviewed and approved the final manuscript.

Competing interests
The authors declare no competing interests.

https://doi.org/10.1038/s41467-025-65963-x

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-025-65963-x.

Correspondence and requests for materials should be addressed to Jing Li.

Peer review information *Nature Communications* thanks Harold Snieder, who co-reviewed with Rima Triatin, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.