Journal of Translational Medicine

RESEARCH

Open Access

Modifiable factors affects cancer-specific survival: findings from a large population-based prospective cohort study

Xiong Chen[1,2,3†], Dongxi Zhu[1,2,3†], Chenchen Li[4†], Yi Lin[1,2,3†], Lin Lv[5], Feiling Ai[6], Zhijian Chen[1,2,3], Runnan Shen[1], Xuwen Shen[5], Guochang You[5], Zhige Jiang[5], Wenlong Zhong[1,2,3*] and Tianxin Lin[1,2,3*]

Abstract

Background Modifiable factors affect cancer's survival but literature did not differentiate prior to versus after cancer diagnosis. It is essential to provide references for the intervention prioritized at different stages.

Methods In this prospective cohort study, we analyzed national data from the UK Biobank, including 121,399 participants, to assess the association of modifiable factors with cancer-specific survival (CSS) in two independent cohorts: a pre-cancer cohort ($n=78,027$) and a post-cancer cohort ($n=43,372$). Additionally, a weighted standardized score was derived to evaluate the joint effects across different domains. Interactions between the six domains and age at diagnosis, sex, and cancer site were evaluated using likelihood ratio tests. Subgroup analyses were then performed for factors showing significant effect modification. Population-attributable fractions (PAF) of different domains on 5-year cancer-specific death were calculated.

Results Our study comprehensively presented the differential patterns of modifiable factors' impact on CSS among pre-cancer and post-cancer cohorts, sexes and different cancer sites. In the pre-cancer cohort, CSS were predominantly attributable to smoking/alcohol consumption (PAF 9·2%) and daily activity (PAF 10·6%). Men exhibited a higher risk than women for dietary habits (HR:1·25 versus 1·18), daily activity (HR:1·50 versus 1·29) and living environment (HR:1·13 versus 1·03). The impact of modifiable factors, including daily activity, smoking/alcohol consumption, and physical measures, on CSS varied across different cancer sites. In the post-cancer cohort, 18·6% of 5-year cancer-specific deaths were attributable to unfavourable mental health. In subgroup analysis, the risk of CSS in the domain of smoking/alcohol consumption was higher in men than that in women (HR: 1·58 versus 1·34). The impact of modifiable factors, including smoking/alcohol consumption, mental health and physical measures, on CSS varied across different cancer sites.

†Drs. Xiong Chen, Dongxi Zhu, Chenchen Li and Yi Lin contribute equally and fully.

*Correspondence:
Wenlong Zhong
zhongwlong3@mail.sysu.edu.cn
Tianxin Lin
lintx@mail.sysu.edu.cn

Full list of author information is available at the end of the article

Conclusions Our findings suggested that targeted prevention and early intervention strategies should be implemented to reduce the risk of cancer-related deaths.

Keywords Cancer-specific survival, Modifiable factor, Population-attributable fractions

Introduction

Cancer ranks as the second leading cause of death worldwide, with a rising trend. It is estimated that the cancer burden of 19 million new cases in 2020 is expected to increase to more than 30 million by 2040 [1]. Despite the recent advances in early detection and novel treatment strategies for cancers, the decline in cancer mortality rate (17·9%) is relatively modest compared to the dramatic decrease in mortality from heart diseases (67·5%) [2]. Hence, it is necessary to perform a thorough analysis to identify the possible and preventable risk factors for cancer development in the population level.

Accumulating evidence indicates that modifiable factors affect prognosis in cancer survivors [3–7]. It is estimated that modifiable factors account for 35–50% of all cancer-related deaths, with estimates including 45·2% in China [8], and 45·1% in the USA [9]. However, previous studies have several limitations. First, there is insufficient evidence to ascertain whether pre-cancer factors continue to have a definitive and sustained effect on the long-term survival of cancer survivors. Accordingly, a common problem is the inaccurate estimation of the populational attributable fraction (PAF) using risk ratios on cancer cases but not on cancer death [8]. Second, hypothesis-driven methods have been commonly adopted to explore post-cancer modifiable factors on cancer survival, which are likely to produce type I error and selective reporting error [10]. Third, most previous studies did not differentiate the impact of modifiable factors before and after cancer diagnosis on cancer survivors. Since many patients adjust their lifestyle after diagnosed with cancer, it is essential to provide references for the intervention measures that should be prioritized at different stages. Up to now, no studies have investigated broader factors associated with cancer survival comprehensively in a large population-based design.

Hence, using data from the UK Biobank, a prospective cohort encompassing over 500,000 participants [11], we first employed exposome-wide association study (EWAS), which is a non-hypothesis-driven statistical approach, to systematically identify the impact of pre-cancer and post-cancer modifiable factors on cancer-specific survival (CSS). By leveraging a data-driven approach, our study mitigates selection bias inherent in hypothesis-driven analyses, allowing for a more comprehensive identification of key factors influencing cancer survival. Through the establishment of two distinct sub-cohorts (pre-cancer and post-cancer cohorts), we address the limitations of traditional studies, which often overestimate or underestimate the impact of pre-cancer factors on survival, thereby filling a critical research gap. Moreover, our findings provide stage-specific, personalized intervention strategies, offering empirical support for both early prevention efforts and post-diagnosis cancer management.

Methods

Study design

In this prospective cohort study, we identified modifiable factors before and after the cancer diagnosis categorized as pre-cancer and post-cancer factors, respectively. Modifiable factors were categorized into six domains: dietary habits, daily activity, smoking/alcohol consumption, mental health, living environment, and physical measures. In EWAS, we examined the association of modifiable factors, both before and after cancer diagnosis, with CSS applying Cox proportional hazard regression models. Additionally, a weighted standardized score was derived for each individual, aiming to evaluate the joint effects of modifiable factors across different domains on CSS. Lastly, PAF of different domains on 5-year cancer-specific death were calculated considering the non-independence of different domains. Subgroup analysis was conducted based on sex, age at cancer diagnosis ($>=65$ years, <65 years) and cancer sites (digestive system, respiratory and intrathoracic system, skin, breast, female genital organs, male genital organs, urinary tract, lymphoid or hematopoietic system).

The analysis plans are shown in Fig. 1 and Figure S1. Variable evaluation is detailed in the Supplementary file, with Figure S2 illustrating the processing of continuous variables, which served as the basis for their categorization.

Study population and outcome definition

Data were retrieved from the UK Biobank, a population-based cohort study that included over 500,000 subjects enrolled at baseline during a period from 2006 to 2010. Overall survival (OS) data were based on the National Death Registries, as of March 2023. Cancer diagnosis was based on the link to the National Cancer Registries System. The cancer site was based on the International Classification of Diseases 10 codes (ICD-10). All participants provided written consent and approval was given by the North West Multicenter Research Ethics Committee (MREC, https://www.ukbiobank.ac.uk/learn-more-about-uk-biobank/about-us/ethics). Participants with complete data on the dates of cancer diagnosis were

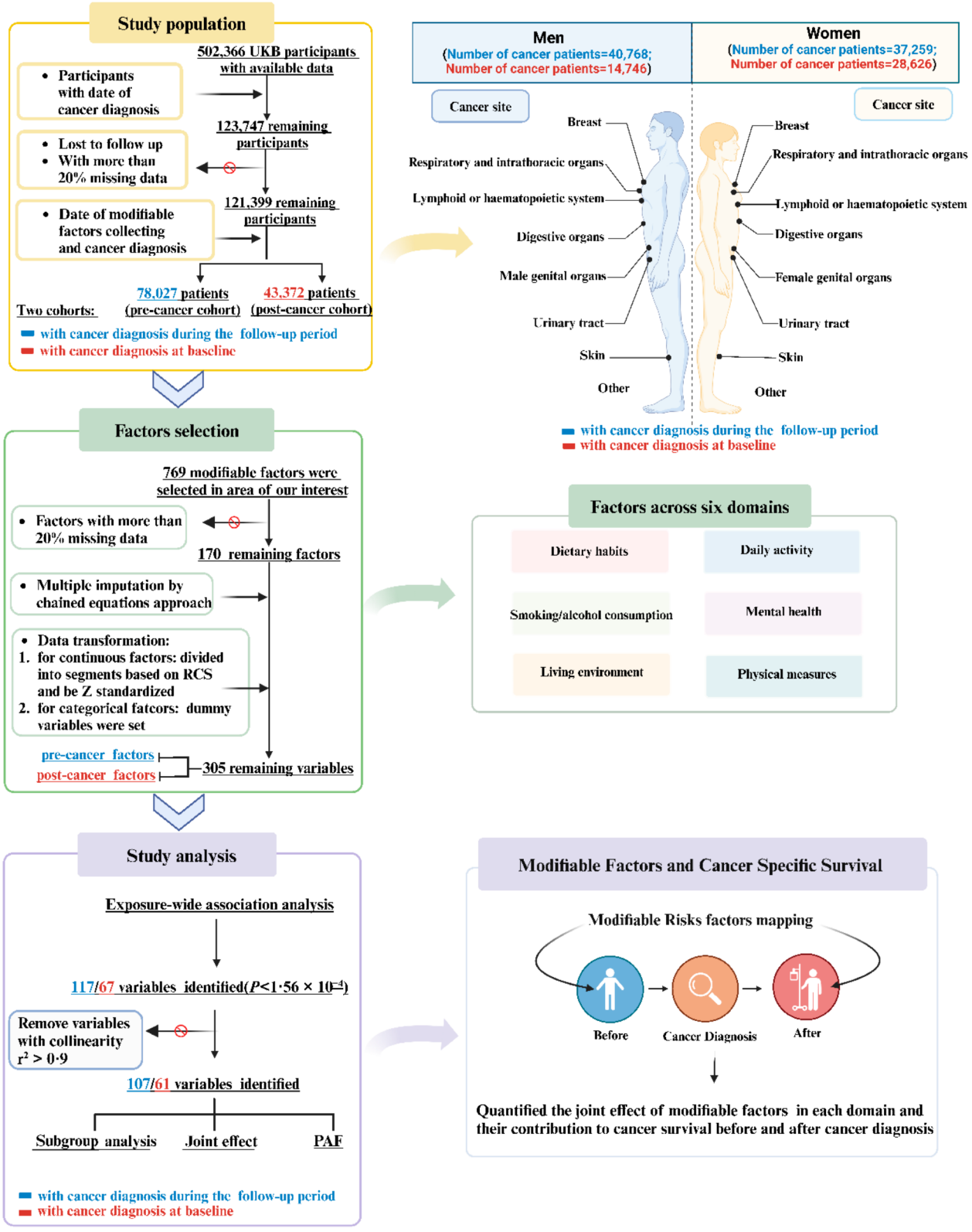

Fig. 1 A schematic overview of the study design. Analyses to examine the potential impact of a diverse array of modifiable risk factors before and after cancer diagnosis on CSS. Red: the pre-cancer cohort (participants diagnosed with cancer at baseline); blue: the post-cancer cohort (participants without a history of cancer at baseline but developed cancer during the follow-up period). Abbreviations: UKB, the UK Biobank; PAF: population attributable fraction; RCS, restricted cubic splines

included while participants with more than 20% missing data of selected modifiable factors, lost to follow-up were excluded from the present study.

CSS was defined as the outcome in pre-cancer and post-cancer cohorts. It was derived by limiting the underlying (primary) cause of death to ICD-10 codes C00-C97, extracted from UKB health outcome datasets death register (Category100093, obtained through linkage to national death registries and updated two or three times per year). In both cohorts, survival time and follow-up time were calculated from the date of cancer diagnosis to the date of death registries or the last date with update information (March 2023), whichever came first.

EWAS

In EWAS, the adjusted covariables included sex, ethnicity, age at cancer diagnosis, cancer stage, number of non-cancer diseases and Townsend deprivation index. For factors that violated the proportional hazard assumption ($P < 0{\cdot}0001$ based on Schoenfeld residuals), an extended Cox model with a time-dependent coefficient was applied [12, 13]. To control false discovery, a conservative Bonferroni-corrected significance threshold (0·05 divided by 319 tests, or $1{\cdot}567 \times 10^{-4}$) was used [10]. Significant factors identified by EWAS were checked for collinearity [14]. In cases where two factors were highly correlated ($r^2 > 0.9$), the factor more strongly correlated (higher correlation coefficient) with other identified factors was excluded.

Joint effects across six domains on CSS

Based on the identified modifiable factors in the EWAS, we constructed risk scores in six domains for each participant. First, we converted continuous variables to binary variables based on the mean. Participants scored 1 point for each detrimental factor (HR > 1), and the sum of points of all modifiable factors in each domain were the total unweighted scores. Weighted standardized scores for each domain were generated based on the absolute value of β coefficients of each variable in the Cox proportional hazard models [15]. A higher score indicated a higher risk of unhealthy modifiable factors in one domain.

The study population was divided into three groups including participants with favourable, intermediate or unfavourable modifiable factors according to the tertiles of their total scores.

Cox proportional hazard regression was conducted to assess the association of the six domains with CSS estimating the joint effects of identified factors using the sum of scores in each domain. In sensitivity analysis, an unweighted score was used to conduct the same analysis. Interactions between six domains and age at cancer diagnosis, sex, and cancer site were assessed by conducting likelihood ratio tests [16]. Subgroup analyses were then performed for factors showing significant effect modification.

PAF calculation

PAF was defined as the proportion by which an outcome occurrence is reduced if the whole population is hypothesized to attain the same risk of disease as the individuals within the lowest consumption category [17]. We performed two models- Model 1 eliminated the unfavourable factors from the population; Model 2 eliminated unfavourable and intermediate factors. PAFs were calculated based on the odds ratio (OR) for 5-year cancer-specific death estimated using logistic regression models [18]. Considering the overlap of six domains and to reduce the overestimation of PAFs caused by the interactions of domains, communality was calculated to compute the weighted PAFs [19]. Additionally, we performed sequential PAF estimation as a sensitivity analysis, calculating PAFs by systematically removing risk factors to assess the robustness of our findings.

Results

Study population and baseline characteristics

Finally, 121,399 participants were enrolled. The pre-cancer cohort included 78,027 participants (diagnosis age, mean [SD]: 66·5 [7·7] years; 52·3% men; 92·6% white). Within the median follow-up of 6·7 years, cancer-specific death occurred in 14,891 (19·1%) participants. The post-cancer cohort included 43,372 participants (diagnosis age, mean [SD]: 50·6 (10·9) years; 34·0% men; 92·97% white). Within the median follow-up of 21·1 years, cancer-specific deaths occurred in 4,060 (10·7%) individuals.

The demographic and baseline characteristics of the two cohorts are shown in Table S1-S3.

Modifiable factors in pre-cancer cohort

In the EWAS of the pre-cancer cohort, CSS was associated with 117 modifiable factors (Fig. 2 and Table S4). After excluding variables due to multicollinearity (Figure S3), 107 factors remained (20 dietary habits, 38 daily activities, 12 smoking/alcohol consumption, 9 mental health, 3 living environment, and 25 physical measures). Among these factors, current smoking most of the day had the highest HR (2·55, 95%CI: 2·25 to 2·88, $P < 1{\cdot}567 \times 10^{-4}$). Among the protective factors, the brisk walking pace had the lowest HR = 0·63 (95%CI: 0·58 to 0·65, $P < 1{\cdot}567 \times 10^{-4}$). Sex-stratified analysis indicated that the association between CSS and modifiable factors varied substantially between sexes. CSS was associated with a broader range of modifiable factors in men than in women. Notably, more differentiated performance had been observed in the sex-specific patterns of modifiable factors related to mental health and daily

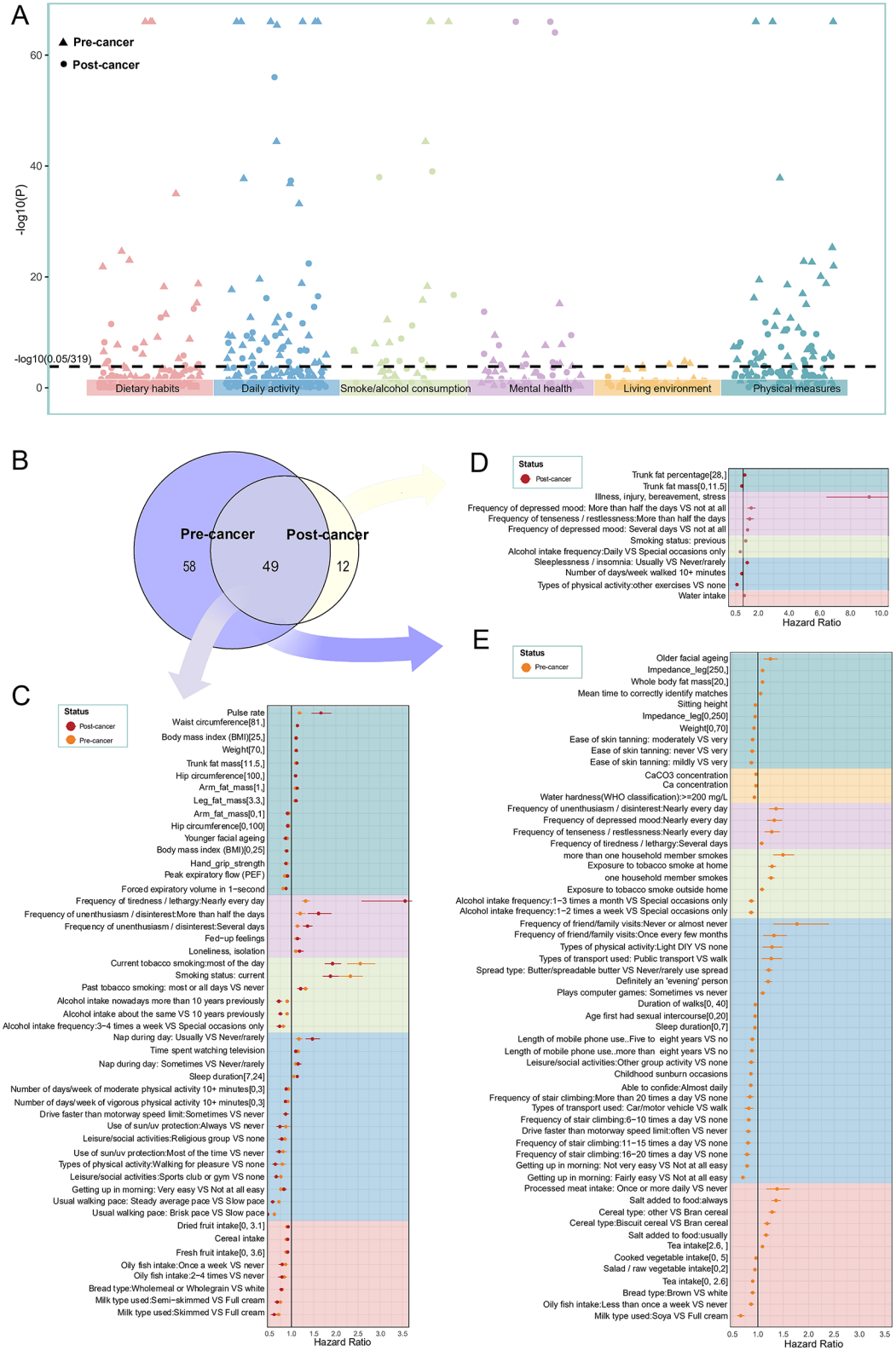

Fig. 2 (See legend on next page.)

(See figure on previous page.)

Fig. 2 Associations between modifiable factors across six domains and cancer-specific survival in pan-cancer using Exposure-wide association analysis. (**A**) Results of EWAS in full sample. X-axis: category domains; y-axis: statistical significance as − log10 of p-value; dotted horizontal line: significance threshold corrected for multiple testing (Bonferroni correction, $P < 1{\cdot}567 \times 10^{-4}$). (**B**). Venn diagram showing the similarities and differences between pre-cancer and post-cancer risk factors. (**C**) Hazard ratios of significant factors identified in both cohorts by EWAS in the full sample. (**D**) Hazard ratios of significant factors identified in the post-cancer cohort but not in the pre-cancer cohort. (**E**) Hazard ratio of significant factors identified in the pre-cancer cohort but not in the post-cancer cohort. Dots: hazard ratios; horizontal lines: corresponding 95% confidence intervals. The association was estimated by applying the Cox hazard regression model or an extended cox model with a time-dependent coefficient), adjusted for age at cancer diagnosis, sex, ethnicity, cancer stage, number of non-cancer diseases and Townsend deprivation index

activities. Age-stratified analysis showed that CSS was associated with more modifiable factors in the elderly versus younger participants, especially in the domain of mental health. Cancer site-stratified analysis showed a distinct pattern of association, except for in the domain of smoking/alcohol consumption (Fig. 3 and Table S4).

CSS was associated with all six domains (P for trend < 0·05, Fig. 4A). The top three domains with the highest joint effect were respectively unfavourable profiles of smoking/alcohol consumption (HR = 1·49, 95%CI: 1·43 to 1·56, $P < 0{\cdot}0001$), physical measures (HR = 1·45, 95%CI: 1·38 to 1·52, $P < 0{\cdot}0001$) and daily activity (HR = 1·40, 95%CI = 1·34 to 1·47, $P < 0{\cdot}0001$). The pattern of results was nearly identical in the sensitivity analysis (Table S5). Domains interacted with sex and cancer site, but not age at cancer diagnosis (Fig. 4B and Table S6). Men exhibited higher HRs for unfavorable dietary habits (HR = 1·25, 95%CI:1·18 to 1·33 vs. HR = 1·18, 95%CI:1·11 to 1·26), daily activity (HR = 1·50, 95%CI = 1·40 to 1·60 vs. HR = 1·29, 95%CI:1·20 to 1·37), and living environment (HR = 1·13, 95%CI = 1·07 to 1·20 vs. HR = 1·03, 95%CI:0·97 to 1·09) than women (Fig. 4C and Table S7). Cancer site-stratified analysis showed the highest HRs for poor daily activity and physical measures in male genital cancers, while long-term smoking/alcohol consumption had the strongest impact on respiratory and intrathoracic cancers (Fig. 4D, Table S7).

PAF analysis indicated a significant impact of modifiable factors on 5-year cancer-specific death (Fig. 5A). Shifting unfavourable profiles to intermediate/favourable ones (Model 1), the most significant survival benefit was associated with addressing smoking/alcohol consumption, with 9·2% of 5-year cancer-specific deaths attributable to unfavourable smoking/alcohol consumption. PAF was 8·7% for daily activity, 5·8% for physical measures, 5·7% for dietary habits, 2·0% for mental health, and 1·5% for living environment. Shifting all unfavourable/intermediate profiles to favourable ones (Model 2), PAF was 10·6% for daily activity, 10·0% for physical measures, 8·7% for smoking/alcohol consumption, 6·9% for dietary habits, 3·2% for mental health, 2·1% for the living environment. Table S8 showed the results of sensitivity analysis of PAF. PAF by cancer site (Fig. 5B, Table S9) showed the greatest benefit for male genital cancers (PAF = 30.7–42.0%), across all six domains. Furthermore, in the comparison of the two models, PAF was minimal in Model for female genital cancers, but improved in Model 2, where dietary habits played a key role.

These findings highlight the critical role of pre-cancer modifiable factors in cancer-specific survival, particularly smoking/alcohol consumption, daily activity, and physical measures. The impact of pre-cancer modifiable factors varies across sex, emphasizing the importance of targeted interventions tailored to different populations. Addressing modifiable risk factors could yield substantial survival benefits, particularly for cancers of the male genital organs.

Modifiable factors in post-cancer cohort

In the post-cancer cohort, after excluding variables due to multicollinearity (Figure S4), 61 pre-cancer variables was associated with CSS: 9 dietary habits, 18 daily activities, 8 smoking/alcohol consumption, 9 mental health, 17 physical measures, and none in the living environment domain (Fig. 2 and Table S10). Among these factors, illness/injury/bereavement/stress had the highest HR (9·21, 95%CI = 6·43 to 13·21, $P < 1{\cdot}567 \times 10^{-4}$). Among the protective factors, brisk pace had the lowest HR (0·48, 95%CI = 036 to 0·63, $P < 1{\cdot}567 \times 10^{-4}$). Men had more CSS-associated modifiable factors than women (47 versus 41), primarily in daily activity (16 versus 10). Noteworthy, the type of significant factors also varied across sexes. For instance, within the domain of physical measures, CSS was associated with the change of mass of leg in women but not in men, whereas blood pressure and Forced Expiratory Volume in the first second (FEV1) were significant in men but not in women. Concerning diet, fruit intake had significant association with CSS only in men, whereas fish oil consumption was associated with CSS in women but not in men. In participants ≥ 65 years, only eight modifiable factors were significantly associated with CSS: four in physical measures, one in mental health, and three in daily activities. The pattern of association differed significantly across cancer sites, but was nearly identical in the domain of smoking/alcohol consumption (Fig. 3 and Table S10).

The joint effects of post-cancer modifiable factors in each domain are shown in Fig. 4A. Of note, post-cancer mental health had greater e impact on CSS in the post- versus pre-cancer cohort (intermediate profile: HR = 1·43, 95%CI = 1·33 to 1·55, $P < 0{\cdot}0001$ vs. HR = 1·03, 95%CI: 0·99 to 1·07, $P = 0{\cdot}098$; unfavorable profile:

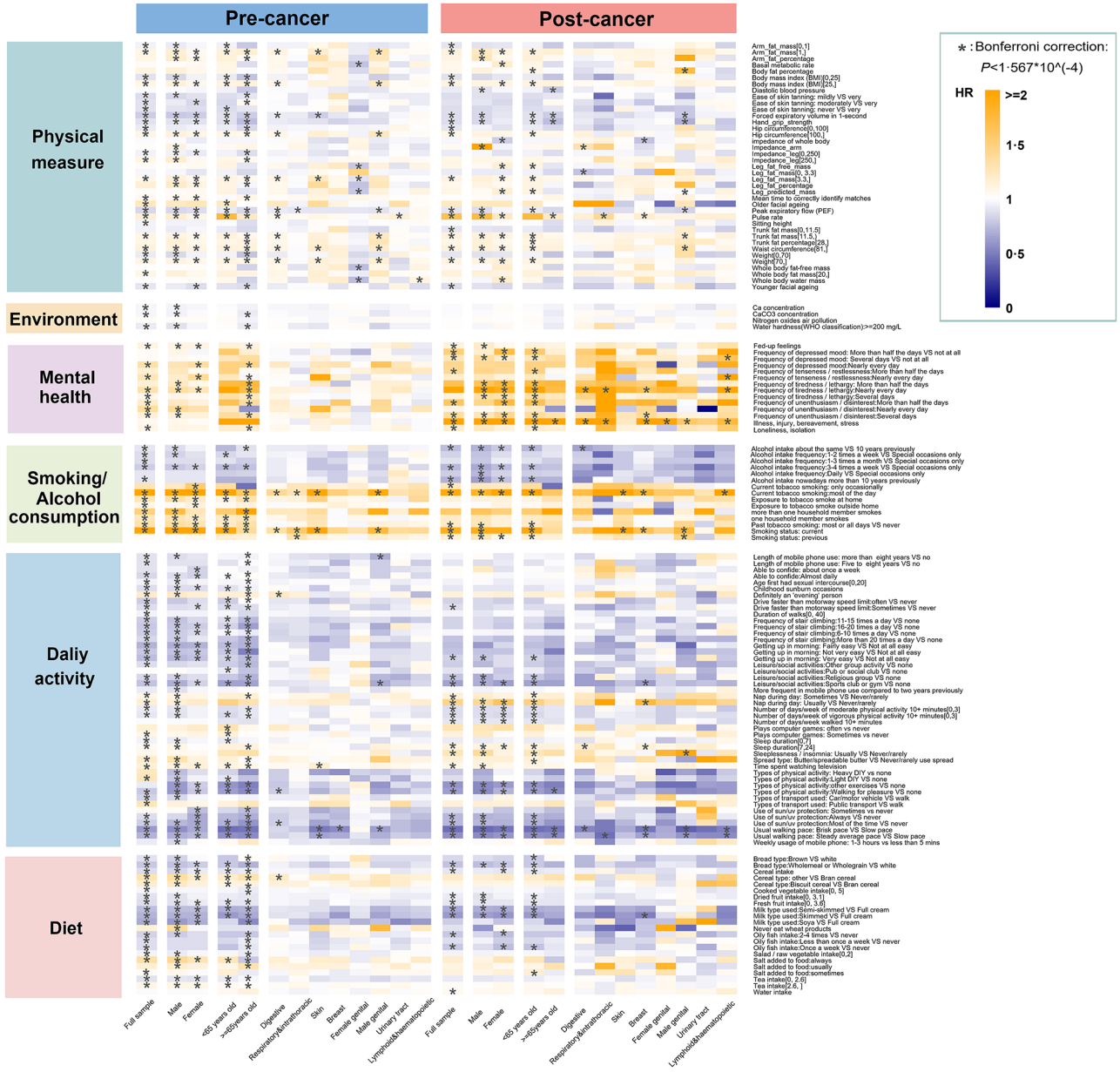

Fig. 3 Summary heatmap for all significant factors identified by EWAS across all analytic sets (subgroup analysis). The color of cells indicates the effect sizes (HR) between each risk factor and all-cause mortality in pan-cancer. Asterisks in cells represent significant associations after correction for multiple testing (Bonferroni correction, $P < 1{\cdot}567 \times 10^{-4}$). Only factors with significant association in at least one analytic set are shown

HR = 1·78, 95%CI = 1·65 to 1·94, $P < 0{\cdot}0001$ vs. HR = 1·06, 95%CI = 1·01 to 1·10, $P = 0{\cdot}009$). The pattern of results was nearly identical in the sensitivity analysis (Table S11). No interactions were found between domains and age (Fig. 4B and Table S12). Unfavorable smoking/alcohol consumption was associated with higher risk in men (HR = 1·58, 95%CI = 1·40 to 1·78, $P < 0{\cdot}0001$) than in women (HR = 1·34, 95%CI = 1·21 to 1·48, $P < 0{\cdot}0001$). Among different cancer sites, unfavourable profiles of mental health and physical measures were associated with the highest risk of cancer of the respiratory and intrathoracic organs, whereas unfavourable smoking/alcohol consumption was associated with the highest risk of cancer of the urinary tract. Details of interaction and subgroup analysis are shown in Fig. 4C-D and Table S13.

PAF analysis highlighted mental health prevention as the most impactful intervention (PAF = 18·4–30·6% across two models, Fig. 5A). The estimated PAFs ranged from 15·7%-17·2% for daily activity, 12·6%-17·4% for physical measures, 11·8%-14·8% for dietary habits, and 8·2%-12·0% for smoking/alcohol consumption. Table S14 showed the results of sensitivity analysis of PAF. In

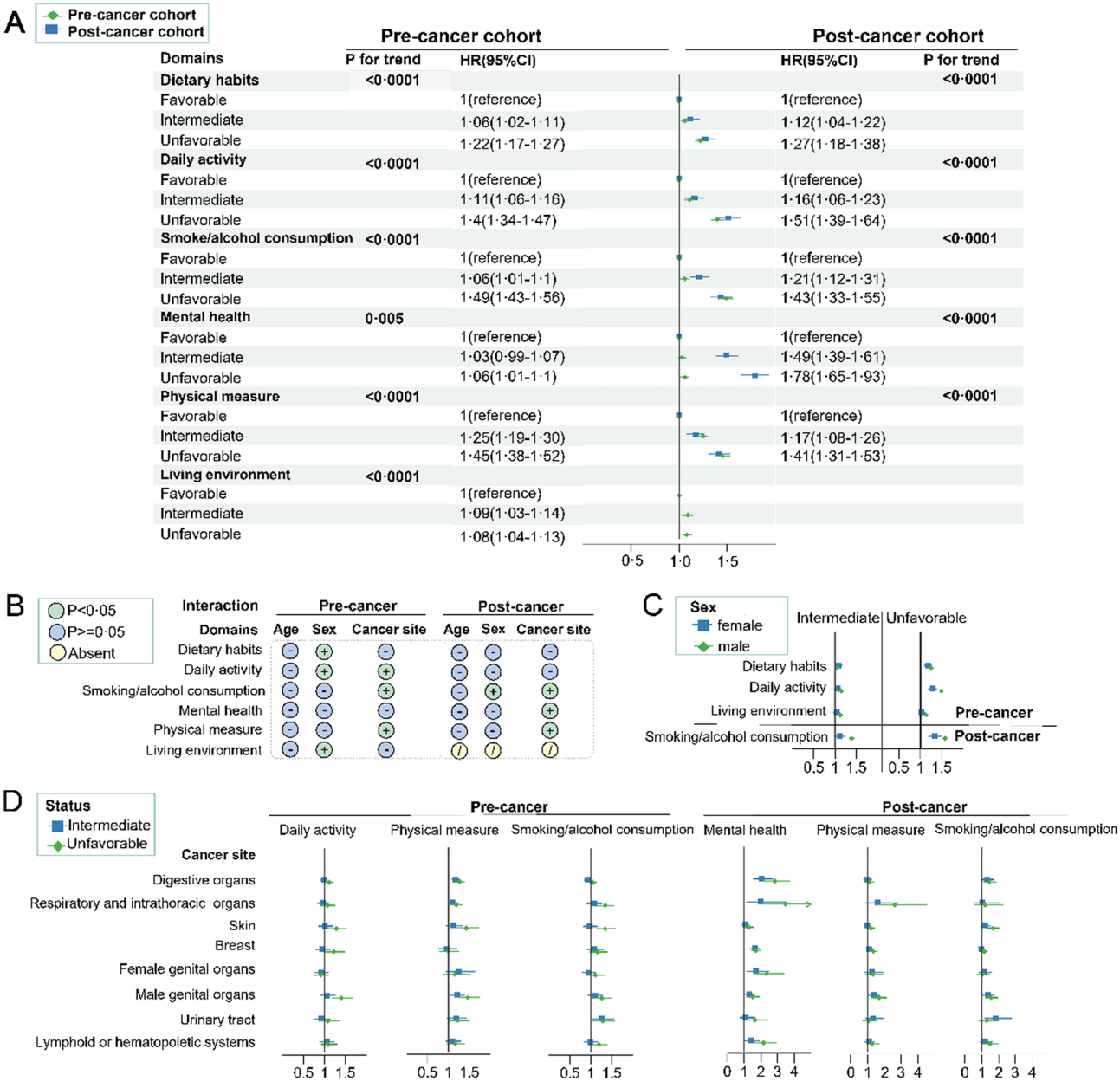

Fig. 4 Joint effects of identified factors across six domains on cancer-specific survival. (**A**) Associations of six domains with cancer-specific survival. Blocks represent hazard ratios and horizontal lines indicate corresponding 95% confidence intervals. The favourable profile of each domain was used as the reference panel. The hazard ratio was generated by a Cox model including six domains and adjusted for age at cancer diagnosis, sex, ethnicity, cancer stage, number of non-cancer diseases and Townsend deprivation index. (**B**) Interaction of six domains and age at cancer diagnosis, sex and cancer site, was assessed by conducting a likelihood ratio test. The red dot indicates the interaction. (**C**) Sex-specific analyses of domains indicating interactions with sex based on Cox regression models. (**D**) Cancer site-specific analyses of domains showing interaction with cancer site based on Cox regression models

analyses stratified by cancer sites, adjustments for mental health conferred the greatest survival benefit across most cancers in both models, except for cancer of the male genital organs, urinary tract, and skin (Fig. 5C and Table S15). Improving daily activity performance offered the most significant survival benefits in patients with cancer of the male genital organs; dietary enhancements offered the most significant survival benefits in patients with urinary tract cancer; smoking/alcohol consumption showed the greatest importance for patients with skin cancer.

These findings underscore the crucial role of post-cancer modifiable factors in influencing cancer-specific survival, particularly mental health, which had a significantly stronger impact after cancer diagnosis compared to before. The effect of modifiable factors varied across

A

Domain	Model1: PAF for unfavorable profiles			Model2: PAF for intermediate and unfavorable profiles		
	Unweighted PAF	Communality	Weighted PAF	Unweighted PAF	Communality	Weighted PAF
Pre-cancer						
Dietary habit	0·100	0·358	0·057	0·152	0·412	0·069
Daily activity	0·153	0·332	0·087	0·233	0·360	0·106
Smoking/Alcohol consumption	0·162	0·287	0·092	0·192	0·388	0·087
Mental health	0·034	0·242	0·020	0·071	0·382	0·032
Living environment	0·026	0·316	0·015	0·047	0·929	0·021
Physical measures	0·103	0·466	0·058	0·220	0·528	0·100
Post-cancer						
Dietary habit	0·217	0·155	0·118	0·370	0·156	0·148
Daily activity	0·289	0·332	0·157	0·432	0·350	0·172
Smoking/Alcohol consumption	0·150	0·226	0·082	0·301	0·231	0·120
Mental health	0·337	0·162	0·184	0·766	0·140	0·306
Physical measure	0·231	0·124	0·126	0·435	0·123	0·174

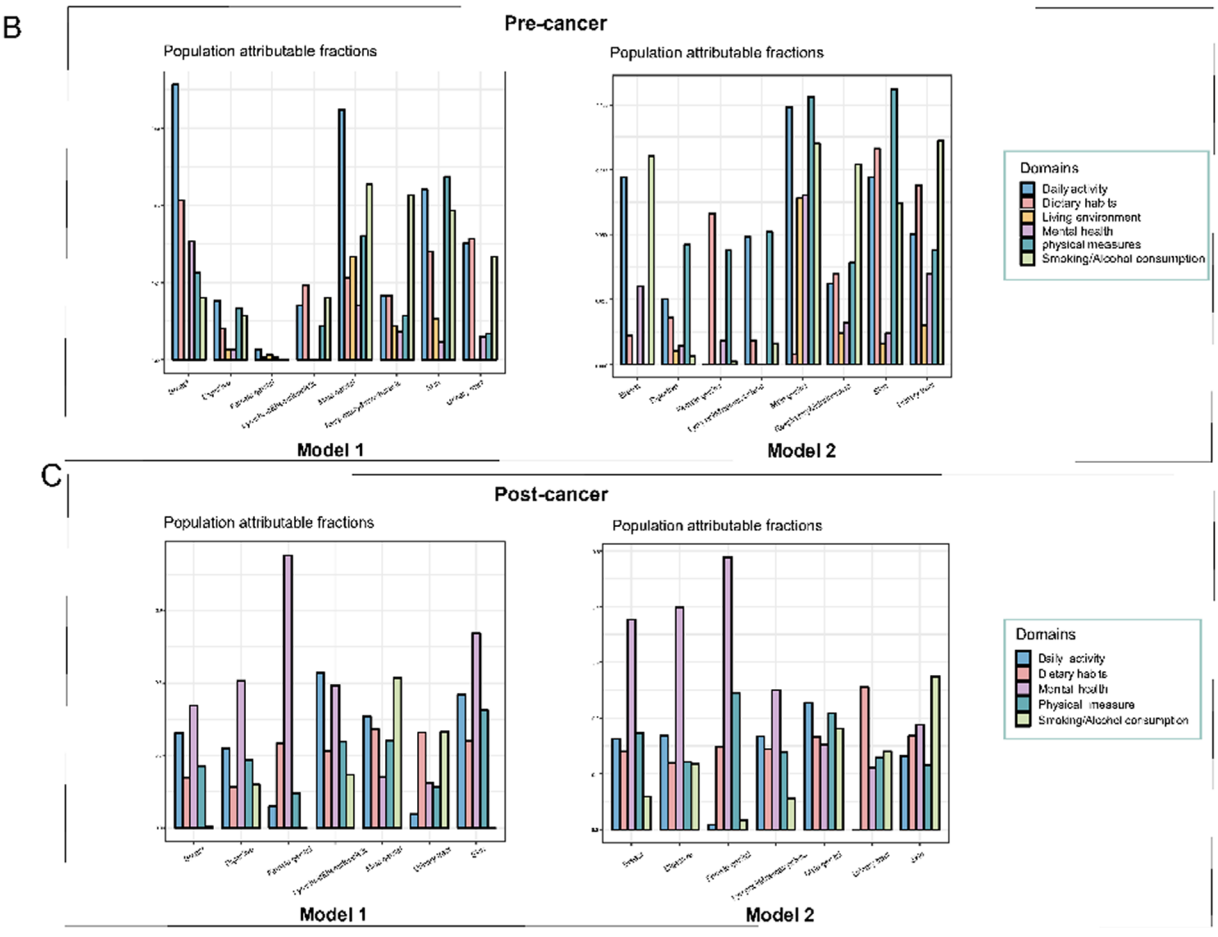

Fig. 5 Population attributable fractions for the six domains on 5-year cancer-specific death. Abbreviation: PAF, population attributable fraction. (**A**) PAFs were calculated based on univariate logistic regression models. Communality represents the overlap of the domain with others, which was calculated using principal component analysis. Weighted PAFs were calculated after considering the overlap among risk factors. In Model 1, we shifted the unfavourable profiles to intermediate and favourable ones. In Model 2, we shifted all factors to the favourable ones. (**B**) Weighted PAFs for six post-cancer domains across cancer sites. (**C**) Weighted PAFs for six pre-cancer domains across cancer sites

sex and cancer sites, emphasizing the importance of tailored interventions. Addressing these factors could provide substantial survival benefits, with mental health interventions emerging as the most impactful strategy.

Discussion

The current study identified a broad array of pre-cancer modifiable factors significantly associated with CSS, being observed in both younger and elderly populations.

It indicated the impact of pre-cancer modifiable factors on CSS persists throughout the entire lifespan, even into gerontic. Modifiable factors are known to be associated with the risk of developing cancers [20, 21]. Accordingly, primary prevention based on these modifiable factors could prevent cancer as well as reduce cancer-specific death. Fewer post-cancer modifiable factors were associated with CSS in the elderly than that in the younger population. Several factors may have contributed to this finding. First, such a discrepancy may reflect a decline in the ability or will to modify unhealthy or maintain healthy lifestyles after the cancer diagnosis [22]. Decreased physical function, less social and psychological support, and higher levels of cancer-related fatigue in elderly subjects are potential reasons for poorer self-management compared to younger patients [23]. Additionally, susceptibility to side effects of medications in the elderly may disrupt the estimation of modifiable factors' effect on the CSS.

Previous studies showed that cigarette smoking accounted for the highest proportion of cancer cases and cancer deaths [9, 24, 25]. Interventions through taxation and regulatory policies have achieved significant results over the past decades in developed countries [26]. Our study also suggests that public health initiatives should prioritize smoking cessation programs and alcohol reduction strategies among cancer patients. Moreover, this study suggests that pre-cancer daily activity is equally important as smoking in reducing cancer-specific mortality. Motivating individuals to participate in more daily activities, such as physical exercise and social interaction, should be advocated at the policy level to mitigate the growing cancer burden. Recent literature suggests that the exposome shapes distinct patterns of disease and mortality risk, irrespective of polygenic disease risk [27]. Our study further supports the necessity of actively intervening in pre-cancer modifiable factors, which not only reduces the risk of tumor development but may also extend patient survival and improve quality of life.

The current study highlighted the importance of post-cancer mental health in CSS. Several potential mechanisms may explain this association. First, patients with poor mental health were less likely to adhere to treatments and follow-up. A meta-analysis indicated that patients with depression have lower medication adherence and are more likely to delay or discontinue medication, particularly in the long-term management of cancer, diabetes, and cardiovascular diseases [28]. Kielsholm et al. [29] conducted a population-based cohort analysis of 2,036,704 Danish people and reported that individuals with mental disorders were less likely to engage in colorectal cancer screening, illustrating the impact of mental health on healthcare utilization. Second, patients with mental disorders often require treatment with certain medications, which in turn may produce an adverse impact on CSS. Jari Tiihonen et al. [30] found that long-term use of the atypical antipsychotic agent clozapine was associated with a higher risk of mortality from lymphoma and leukaemia compared to the risk of agranulocytosis. Third, mental health may influence CSS through immune system modulation. Depression and anxiety have also been linked to dysregulation of T-cell function and reduced natural killer cell activity, impairing immune surveillance against tumor progression [31]. Notably, Zeng et al. recently demonstrated that pretreatment emotional distress correlated with diminished clinical responses to immune checkpoint inhibitors in non-small-cell lung cancer patients, underscoring the role of psychological factors in shaping treatment efficacy [7]. However, caution must be exercised due to the possibility of reverse causation, which may lead to an overestimation of the causal effect of mental health on CSS. Patients with malignancies or low socioeconomic status often experience higher levels of worry, distress, and other mental health issues. The analysis adjusting for the cancer stage and Townsend index in the current study indicated that mental health remained significantly associated with CSS, suggesting that its influence extends beyond disease severity and social support. Of course, we cannot entirely rule out the possibility that our conclusions overestimate the causal effect of mental health on CSS. Future research with more rigorous designs, such as time-series analyses, is needed to further clarify the temporal sequence and establish the causality of this relationship. All in all, the inclusion of psychiatrists in multidisciplinary treatment teams for cancer patients might prolong survival. Furthermore, a recent review provides evidence that web-based platforms and smartphone applications can effectively alleviate depression, anxiety, and distress in cancer patients [32]. This suggests that integrating digital health interventions could be a scalable and cost-effective strategy to enhance mental well-being and overall survivorship outcomes.

The current study revealed distinct patterns of association between CSS and modifiable factors across different sexes and cancer sites. For individuals with cancer, they could get specific suggestions from our subgroup EWAS results (Fig. 3). Similar to the previous study, the current study indicated that unfavourable modifiable factors, particularly daily activities, dietary habits, living environment before cancer, smoking/alcohol consumption after cancer, carry a higher risk of cancer survival in men versus in women. Also, specific risk patterns in men versus women suggest different measures. Regarding differences across cancer sites, emphasis should be placed on improving mental health and physical measures in patients with cancer of the respiratory and intrathoracic organs. In contrast, emphasis in patients with cancer of

the urinary tract should be placed on avoiding unfavourable smoking/alcohol consumption.

Our study has several limitations. First, most subjects were of European ancestry. Whether the findings are generalizable to other ancestry is unknown. Future research should include more diverse cohorts to validate these associations across different genetic and environmental backgrounds. Second, people who have a poor prognosis of cancer may not participate in UK Biobank; accordingly, cancer-specific death in the post-cancer cohort may have been under-estimated and PAFs might have been over-estimated. Third, our study did not account for treatment modalities (e.g., chemotherapy, immunotherapy), which could introduce confounding bias. Future studies should incorporate detailed treatment data to better account for its potential impact and improve the robustness of the findings.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12967-025-06372-y.

Supplementary Material 1: Figure S1. Flow chart of statistical analysis. PAF, Population attributable fraction

Supplementary Material 2: Figure S2. Restricted cubic spline assisted in finding truncation values for continuous variables. Restricted cubic spline to primarily explore the linear relationship of continuous variables and outcome. The histogram represents the density distribution. The curve represents the value of hazard ratio and red dot indicates the cutoff values

Supplementary Material 3: Figure S3. Correlation heatmap of identified factors in pre-cancer cohort. The color of cells indicates the correlation coefficient which was calculated by Spearman rank-order correlations. The grey box represents the factors excluded in the scrutiny of multicollinearity

Supplementary Material 4: Figure S4. Correlation heatmap of identified factors in post-cancer cohort. The color of cells indicates the correlation coefficient which was calculated by Spearman rank-order correlations. The grey box represents the factors excluded in the scrutiny of multicollinearity

Supplementary Material 5

Supplementary Material 6

Acknowledgements
We thank the participants in the study and the members of the survey teams of the UK Biobank.

Author contributions
XC, DXZ designed the study. XC, DXZ, LCC, YL and FLA conducted statistical analyses. DXZ, LL participated in data interpretation. XC wrote the first draft of the manuscript and verified the underlying data. DXZ, LCC and FLA also involved in manuscript writing and verified the data. ZJC, RNS, XWS, GCY, ZGJ played roles in acquisition of the data and analyses. All authors revised and approved the final manuscript. The corresponding author (WLZ, and TXL) attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Funding
The research was funded by the National Natural Science Foundation of China (#U21A20383, #81961128027, #82373254, and #82202251), Guangdong Provincial Clinical Research Centre for Urological Diseases (#2020B1111170006), Sun Yat-Sen Memorial Hospital Clinical Research 5010 Program (#SYS-5010Z-202401).

Data availability
Details of how to access UKBiobank data are available from https://www.ukbiobank.ac.uk/enable-your-research/about-our-data. Codes for analysis can be found on Github (https://github.com/XiongChen-SYSU/Modifiable-Risk-Factors-and-Cancer-specific-Survival-Evidence-from-a-large-prospective-cohort-study).

Declarations

Ethics approval and consent to participate
Ethics approval and participate have been obtained by the UKBiobank. All data were applicated under application number 100739.

Consent for publication
All authors read and approved the publication.

Competing interests
We declare no competing interests.

Author details
[1]Department of Urology, Sun Yat-sen Memorial Hospital, Sun Yat-sen University, Guangzhou, PR China
[2]Guangdong Provincial Key Laboratory of Malignant Tumor Epigenetics and Gene Regulation, Guangdong-Hong Kong Joint Laboratory for RNA Medicine, Sun Yat-sen Memorial Hospital, Sun Yat-sen University, Guangzhou, PR China
[3]Guangdong Provincial Clinical Research Center for Urological Diseases, Guangzhou, PR China
[4]Department of Medical Oncology, The Sixth Affiliated Hospital, Sun Yat-sen University, Guangzhou, PR China
[5]School of Medicine, Sun Yat-sen University, Guangzhou, PR China
[6]Faculty of Medicine, Ludwig Maximilian University of Munich, Munich, Germany

Received: 29 January 2025 / Accepted: 8 March 2025
Published online: 29 April 2025

References

1. Brennan P, Davey-Smith G. Identifying novel causes of cancers to enhance cancer prevention: new strategies are needed. J Natl Cancer Inst. 2022;114(3):353–60.
2. Song M, Giovannucci E. Preventable incidence and mortality of carcinoma associated with lifestyle factors among white adults in the united States. JAMA Oncol. 2016;2(9):1154–61.
3. Richman EL, Stampfer MJ, Paciorek A, Broering JM, Carroll PR, Chan JM. Intakes of meat, fish, poultry, and eggs and risk of prostate cancer progression. Am J Clin Nutr. 2010;91(3):712–21.
4. Liu VN, Van Blarigan EL, Zhang L, Graff RE, Loeb S, Langlais CS, et al. Plant-Based diets and disease progression in men with prostate cancer. JAMA Netw Open. 2024;7(5):e249053.
5. Langlais CS, Graff RE, Van Blarigan EL, Palmer NR, Washington SL, Chan JM, et al. Post-Diagnostic dietary and lifestyle factors and prostate cancer recurrence, progression, and mortality. Curr Oncol Rep. 2021;23(3):37.
6. Brookman-May SD, Campi R, Henríquez JDS, Klatte T, Langenhuijsen JF, Brausi M, et al. Latest evidence on the impact of smoking, sports, and sexual activity as modifiable lifestyle risk factors for prostate cancer incidence, recurrence, and progression: A systematic review of the literature by the European association of urology section of oncological urology (ESOU). Eur Urol Focus. 2019;5(5):756–87.
7. Lan A, Li H, Shen M, Li D, Shu D, Liu Y, et al. Association of depressive symptoms and sleep disturbances with survival among US adult cancer survivors. BMC Med. 2024;22(1):225.
8. Chen W, Xia C, Zheng R, Zhou M, Lin C, Zeng H, et al. Disparities by Province, age, and sex in site-specific cancer burden attributable to 23 potentially modifiable risk factors in China: a comparative risk assessment. Lancet Glob Health. 2019;7(2):e257–69.

9. Islami F, Goding Sauer A, Miller KD, Siegel RL, Fedewa SA, Jacobs EJ, et al. Proportion and number of cancer cases and deaths attributable to potentially modifiable risk factors in the united States. CA Cancer J Clin. 2018;68(1):31–54.
10. Manrai AK, Cui Y, Bushel PR, Hall M, Karakitsios S, Mattingly CJ et al. Informatics and data analytics to support exposome-based discovery for public health. Annu Rev Public Health. 2017;38:279–94.
11. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK biobank resource with deep phenotyping and genomic data. Nature. 2018;562(7726):203–9.
12. Ganna A, Ingelsson E. 5 Year mortality predictors in 498,103 UK biobank participants: a prospective population-based study. Lancet Lond Engl. 2015;386(9993):533–40.
13. Zhang Z, Reinikainen J, Adeleke KA, Pieterse ME, Groothuis-Oudshoorn CGM. Time-varying covariates and coefficients in Cox regression models. Ann Transl Med. 2018;6(7):121.
14. Lin BD, Pries LK, Sarac HS, van Os J, Rutten BPF, Luykx J, et al. Nongenetic factors associated with psychotic experiences among UK biobank participants: Exposome-Wide analysis and Mendelian randomization analysis. JAMA Psychiatry. 2022;79(9):857–68.
15. Lourida I, Hannon E, Littlejohns TJ, Langa KM, Hyppönen E, Kuzma E, et al. Association of lifestyle and genetic risk with incidence of dementia. JAMA. 2019;322(5):430–7.
16. Wang YX, Mínguez-Alarcón L, Gaskins AJ, Missmer SA, Rich-Edwards JW, Manson JE, et al. Association of spontaneous abortion with all cause and cause specific premature mortality: prospective cohort study. BMJ. 2021;372:n530.
17. Eide GE, Gefeller O. Sequential and average attributable fractions as aids in the selection of preventive strategies. J Clin Epidemiol. 1995;48(5):645–55.
18. Sjölander A. Estimation of causal effect measures with the R-package StdReg. Eur J Epidemiol. 2018;33(9):847–58.
19. Livingston G, Huntley J, Sommerlad A, Ames D, Ballard C, Banerjee S, et al. Dementia prevention, intervention, and care: 2020 report of the lancet commission. Lancet Lond Engl. 2020;396(10248):413–46.
20. Swanton C, Bernard E, Abbosh C, André F, Auwerx J, Balmain A, et al. Embracing cancer complexity: hallmarks of systemic disease. Cell. 2024;187(7):1589–616.
21. Clinton SK, Giovannucci EL, Hursting SD, The World Cancer Research Fund/American Institute for Cancer Research. Third expert report on diet, nutrition, physical activity, and cancer: impact and future directions. J Nutr. 2020;150(4):663–71.
22. Corbett T, Cummings A, Calman L, Farrington N, Fenerty V, Foster C, et al. Self-management in older people living with cancer and multi-morbidity: A systematic review and synthesis of qualitative studies. Psychooncology. 2020;29(10):1452–63.
23. Butt Z, Rao AV, Lai JS, Abernethy AP, Rosenbloom SK, Cella D. Age-associated differences in fatigue among patients with cancer. J Pain Symptom Manage. 2010;40(2):217–23.
24. GBD 2019 Tobacco Collaborators. Spatial, Temporal, and demographic patterns in prevalence of smoking tobacco use and attributable disease burden in 204 countries and territories, 1990–2019: a systematic analysis from the global burden of disease study 2019. Lancet Lond Engl. 2021;397(10292):2337–60.
25. GBD 2019 Cancer Risk Factors Collaborators. The global burden of cancer attributable to risk factors, 2010-19: a systematic analysis for the global burden of disease study 2019. Lancet Lond Engl. 2022;400(10352):563–91.
26. Flor LS, Reitsma MB, Gupta V, Ng M, Gakidou E. The effects of tobacco control policies on global smoking prevalence. Nat Med. 2021;27(2):239–43.
27. Argentieri MA, Amin N, Nevado-Holgado AJ, Sproviero W, Collister JA, Keestra SM et al. Integrating the environmental and genetic architectures of aging and mortality. Nat Med. 2025:1–10.
28. Grenard JL, Munjas BA, Adams JL, Suttorp M, Maglione M, McGlynn EA, et al. Depression and medication adherence in the treatment of chronic diseases in the united States: A Meta-Analysis. J Gen Intern Med. 2011;26(10):1175–82.
29. Thomsen MK, Jørgensen MD, Pedersen L, Erichsen R, Sørensen HT, Mikkelsen EM. Mental disorders, participation, and trajectories in the Danish colorectal cancer programme: a population-based cohort study. Lancet Psychiatry. 2023;10(7):518–27.
30. Tiihonen J, Tanskanen A, Bell JS, Dawson JL, Kataja V, Taipale H. Long-term treatment with clozapine and other antipsychotic drugs and the risk of haematological malignancies in people with schizophrenia: a nationwide case-control and cohort study in Finland. Lancet Psychiatry. 2022;9(5):353–62.
31. Antoni MH, Lutgendorf SK, Cole SW, Dhabhar FS, Sephton SE, McDonald PG, et al. The influence of bio-behavioural factors on tumour biology: pathways and mechanisms. Nat Rev Cancer. 2006;6(3):240–8.
32. Zhong C, Luo X, Tan M, Chi J, Guo B, Tang J, et al. Digital health interventions to improve mental health in patients with cancer: umbrella review. J Med Internet Res. 2025;27(1):e69621.

Publisher's note