Associations of early tobacco exposure and smoking dose with all-cause and cause-specific cancer risk: A prospective cohort study

Wen Liu[1], Bo-Wu Su[2], Jun-Ying Hou[2], Shu-Qing Xu[2], Xing Yin[2], Ge-Fang Li[2], Jun Deng[1], Jie Zhong[3], Tong-Tong Zhang[4], Yi-Sheng Cai[5], Zilin Zhu[2]

ABSTRACT

INTRODUCTION The long-term association of early tobacco exposure with adult cancer risk remains uncertain, and dose-response associations between cigarette smoking and cancer may be nonlinear.

METHODS We analyzed prospective UK Biobank data through 1 July 2023. All-cause analyses included 467333 participants for pack-years of smoking (PYS) and PYS/(recruitment age − 16) (PYSP), 402849 for maternal smoking during pregnancy (MSDP), and 404584 for age at smoking initiation (ASI). The primary ASI analysis used self-reported never smokers as the reference and excluded ever smokers without initiation age and discordant never smokers with recorded initiation age. Cox models estimated hazard ratios (HRs) with demographic (Model 1) and extended (Model 2) adjustment; ASI Model 2 omitted smoking status. ASI trends were tested among initiators only. Four-knot restricted cubic splines examined exploratory dose–response patterns.

RESULTS In the primary ASI cohort, median follow-up was 14.11 years, with 5263895 person-years, 1037 participants censored at loss to follow-up, and 60160 recorded all-cause cancer events. Compared with never smoking, ASI Model 2 hours were 1.11 (95% CI: 1.09–1.14) for adult, 1.16 (95% CI: 1.13–1.18) for adolescent, and 1.24 (95% CI: 1.15–1.35) for childhood initiation. The initiator-only all-cause trend had nominal p=0.004 and false discovery rate (FDR)=0.038; oropharyngeal cancer showed no corresponding trend (p=0.693). MSDP was associated with all-cause cancer (HR=1.05; 95% CI: 1.03–1.07). Each standard-deviation increase in PYS and PYSP was associated with all-cause cancer (both HR=1.06; 95% CI: 1.05–1.07). Spline analyses indicated nonlinear associations of both dose measures with all-cause and lung cancer (nominal p for nonlinearity ≤0.0013).

CONCLUSIONS These observational associations support continued evaluation of early tobacco exposure and cumulative smoking dose as cancer-risk markers. Comparisons with never smoking and initiation-age trends among smokers address different questions. The results do not establish causal effects or demonstrate the benefit of a specific preventive intervention.

AFFILIATION

1 Department of Pharmacy, Hunan Provincial People's Hospital (The First Affiliated Hospital of Hunan Normal University), Changsha, China
2 School of Pharmacy, Hunan Normal University, Changsha, China
3 Scientific Research Department, Hunan Provincial People's Hospital (The First Affiliated Hospital of Hunan Normal University), Changsha, China
4 The Center of Gastrointestinal and Minimally Invasive Surgery, The Third People's Hospital of Chengdu, Chengdu, China
5 Hunan Key Laboratory of Aging Biology, Xiangya Hospital, Central South University, Changsha, China

CORRESPONDENCE TO

Wen Liu. Department of Pharmacy, Hunan Provincial People's Hospital (The First Affiliated Hospital of Hunan Normal University), Changsha, 410005, China
E-mail: liuwen@hunnu.edu.cn

KEYWORDS

early tobacco exposure, smoking dose, cancer risk, UK Biobank, prospective cohort study

Received: 23 June 2026
Revised: 10 October 2026
Accepted: 17 October 2026

Published by European Publishing. © 2026 Liu W. et al. This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0/).

INTRODUCTION

Epidemiological and experimental studies suggest that transplacental exposure to tobacco-smoke carcinogens may contribute to carcinogenesis. Maternal smoking during pregnancy has been associated with hematological and neurological cancers in childhood[1-3]. Experimental studies show that tobacco-smoke carcinogens can cross the placenta and that developing tissues may be particularly susceptible to irreversible carcinogenic injury[4-10]. Because some tumors induced by transplacental exposure appear only later in life, the long-term association between maternal smoking and cancer in adult offspring warrants investigation.

Early life is a potentially vulnerable period for environmental exposures, including tobacco exposure during fetal development, childhood, and adolescence[11]. Earlier smoking initiation has been associated with lung cancer[12], breast cancer[13], and colorectal cancer[14]. Dose-response associations have also been reported for breast[15], gastric[16], pancreatic[17], lung[18], renal[19], and bladder cancers[20]. However, evidence integrating age at smoking initiation, cumulative smoking dose, and a broad range of incident cancers within a single large prospective cohort remains limited.

Using prospective UK Biobank data, we examined maternal smoking during pregnancy (MSDP), age at smoking initiation (ASI), and cumulative smoking dose in relation to all-cause and cause-specific cancers. The dose metrics were pack-years of smoking (PYS) and pack-years divided by recruitment age minus 16 (PYSP).

METHODS

Study design and population

This study was a secondary analysis of the prospective UK Biobank cohort. The UK Biobank recruited more than 0.5 million participants aged 37–73 years at 22 assessment centers in England, Scotland, and Wales from April 2006 to December 2010, with follow-up through 1 July 2023[21]. Baseline information was collected using touchscreen questionnaires, computer-assisted interviews, physical and functional examinations, biological samples, and linked electronic health records[22]. Of 502236 participants, 34903 with cancer at baseline were excluded from the all-cause base. Exposure-specific all-cause cohorts were the smoking-dose cohort (PYS/PYSP; n=467333), MSDP cohort (n=402849), and primary ASI cohort (n=404584) (Figure 1). After the original exclusion of 1041 participants with unknown/refused ASI responses, the ASI cohort additionally excluded 61691 current/former smokers without initiation age and 17 self-reported never smokers with a discordant recorded initiation age. Cancer-specific models applied the corresponding exposure-data eligibility criteria after excluding the relevant baseline cancer. The UK Biobank study was reported in accordance with the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement[23] and received ethical approval from the North West Multi-Centre Research Ethics Committee (21/NW/0157).

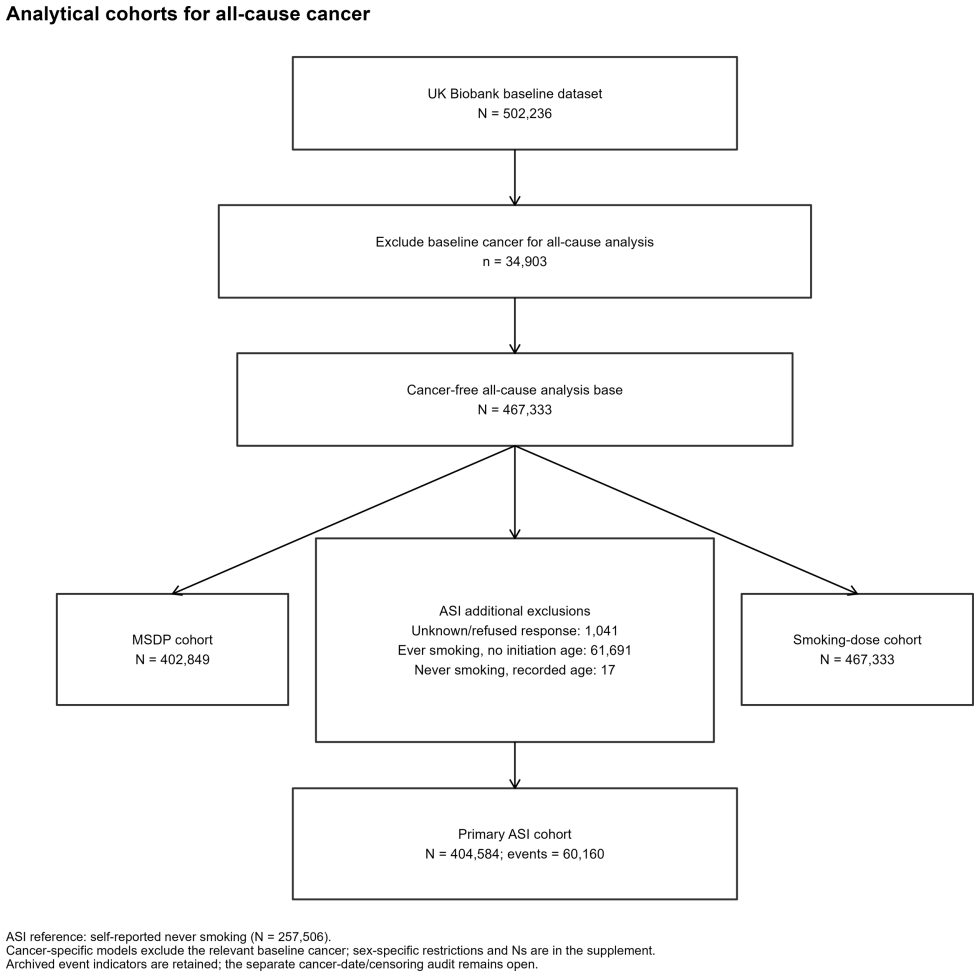
Assessment of tobacco exposure

Early tobacco exposure was assessed using MSDP and ASI, and smoking dose was quantified using PYS and PYSP. Exposure data were collected through touchscreen questionnaires and face-to-face interviews[24].

The primary ASI analysis used participants reporting never smoking, with no recorded initiation age, as the reference group. Current or former smokers with recorded initiation age were classified as childhood (aged <12 years), adolescence (aged 12 to <18 years), or adulthood (aged ≥18 years). Ever smokers without initiation age and never smokers with a discordant recorded initiation age were excluded, in addition to the original unknown/refused-response exclusions. Never smoking status was self-reported, not independently validated. The original code had labelled all missing initiation ages as 'Never_smoker', including 61691 current/former smokers in the all-cause cohort. That mixed non-recorded-age reference is retained only as a legacy-reference sensitivity analysis not as the primary reference.

PYS is defined as the number of cigarettes smoked daily, divided by twenty (the typical number of cigarettes in a pack), multiplied by the years spent smoking. The duration of smoking is calculated by subtracting the age at smoking initiation from the age at smoking cessation:

Figure 1. Cohort flow for all-cause cancer analyses in UK Biobank, recruited April 2006–December 2010 and followed through 1 July 2023



The ASI reference comprises 257506 self-reported never smokers. Cancer-specific analyses excluded the relevant baseline cancer rather than every baseline cancer, so their numbers differ, see Supplementary file Table 1. ASI: age at smoking initiation. MSDP: maternal smoking during pregnancy. PYS: pack-years. PYSP: PYS divided by recruitment age minus 16.

$$\text{PYS} = \frac{\text{Number of cigarettes per day}}{20} \times$$

(Age stopped smoking - Age started smoking)

Where PYSP is defined as the pack-years divided by the number of years between the participants' age at recruitment and the age of 16 years:

$$\text{PYSP} = \frac{\text{Pack-years}}{(\text{Age at recruitment} - 16)}$$

Cancer ascertainment

Disease ascertainment in our study was primarily based on hospital records from the Hospital Episode Statistics in England, the Patient Episode Database for Wales, and the Scottish Morbidity Records. These data were supplemented by self-reported diagnoses, primary care records, and death registry data. Cancer diagnoses were systematically coded according to the International Classification of Diseases, 9th and 10th Revisions (ICD-9 and ICD-10), based on the date of diagnosis.

The monitored outcomes included keratinocyte carcinoma (non-melanoma skin cancer), squamous cell carcinoma, basal cell carcinoma, and cutaneous melanoma; oropharyngeal, thyroid, esophageal, gastric, colorectal, pancreatic, liver, gallbladder, lung, renal, bladder, brain, breast, cervical, endometrial, ovarian, and prostate cancers; and Hodgkin disease, multiple myeloma, leukemia, and non-Hodgkin lymphoma. Corresponding ICD-9 and ICD-10 codes are provided below for reproducibility.

Skin cancers comprised keratinocyte carcinoma (non-melanoma skin cancer), including squamous cell carcinoma and basal cell carcinoma (ICD-9: 173; ICD-10: C44), and cutaneous melanoma (ICD-9: 172; ICD-10: C43).

Other solid cancers comprised oropharyngeal (ICD-9: 140-149; ICD-10: C00-C14), thyroid (ICD-9: 193; ICD-10: C73), esophageal (ICD-9: 150; ICD-10: C15), gastric (ICD-9: 151; ICD-10: C16), colorectal (ICD-9: 153-154; ICD-10: C18-C20), pancreatic (ICD-9: 157; ICD-10: C25), liver (ICD-9: 155; ICD-10: C22), gallbladder (ICD-9: 156; ICD-10: C23), lung (ICD-9: 162; ICD-10: C33-C34), renal (ICD-9: 189.0-189.1; ICD-10: C64-C65), bladder (ICD-9: 188 and 189.2; ICD-10: C67), brain (ICD-9: 191; ICD-10: C71), breast (ICD-9: 174; ICD-10: C50), cervical (ICD-9: 180; ICD-10: C53), endometrial (ICD-9: 182; ICD-10: C54), ovarian (ICD-9: 183; ICD-10: C56-C57), and prostate cancers (ICD-9: 185; ICD-10: C61).

Hematological malignancies comprised Hodgkin disease (ICD-9: 201; ICD-10: C81), multiple myeloma (ICD-9: 203; ICD-10: C90), leukemia (ICD-9: 204-208; ICD-10: C91-C95), and non-Hodgkin lymphoma (ICD-9: 200 and 202; ICD-10: C82-C85).

Follow-up began at recruitment and ended at the earliest recorded diagnosis of the outcome, death, loss to follow-up, or 1 July 2023. The original all-cause analysis excluded baseline cancer of any type. Original cancer-specific analyses excluded the relevant baseline cancer rather than all baseline cancers; consequently, their sample sizes differ from the all-cause cohorts and are reported in Supplementary file Table 1. Breast, cervical, endometrial, and ovarian cancers were analyzed in women, and prostate cancer in men.

For MSDP, the numbers for breast, cervical, endometrial, ovarian, and prostate cancer were 229309; 236182; 236239; 236393; and 193090, respectively; the corresponding numbers for primary ASI were 228659; 235539; 235572; 235777; and 195694. Smoking-dose numbers were 264301; 272314; 272379; 272565; and 225964. All first four outcomes included women and prostate cancer included men.

Covariates

Covariates were selected *a priori* as potential confounders because of their plausible associations with tobacco exposure and cancer risk. Data were collected through questionnaires, biochemical assays, and linked medical records. Demographic and socioeconomic covariates included age, sex, education level (certificate of secondary education (CSE) or equivalent, to professional qualifications), annual gross household income (£) (<18000, 18000–30999, 31000–51999, 52000–100000, or >100000), race/ethnicity (White, Asian or Asian British, Black or Black British, Chinese, mixed, or other)[25], and the Townsend deprivation index. The Townsend index incorporates unemployment, non-car ownership, non-home ownership, and household overcrowding and was categorized as low socioeconomic status (highest quintile), medium (second to fourth quintiles), or high (lowest quintile)[26,27].

Health-related covariates included body mass index (BMI; kg/m^2), alcohol consumption, smoking status, total metabolic equivalent of task (MET) minutes per week, the Healthy Eating Score, C-reactive protein (CRP), family history of cancer, cancer screening, cancer-specific covariates, and chronic diseases. Physical activity (MET-min/week) was categorized as low (<600), medium (600–3000), or high (>3000)[28]. The Healthy Eating Score was based on seven food groups (vegetables, fruit, unprocessed red meat, processed meat, fish, whole grains, and refined grains) and was classified as healthy when at least four prespecified dietary criteria were met[29]. CRP was measured from blood samples collected at the baseline assessment.

Cancer-specific covariates included time spent outdoors in summer, use of sunbeds or sunlamps, skin pigmentation, tanning ability, ultraviolet-protection practices, and childhood sunburn for skin cancers; prostate-specific antigen testing and family history for prostate cancer; and hormone replacement therapy, menopausal status, age at menopause, reproductive history, contraceptive use, age at menarche, hysterectomy, and bilateral oophorectomy, as applicable to female-specific cancers. Diabetes was defined from ICD-9 code 250,

ICD-10 codes E10-E14, self-report, or antidiabetic medication use. Hypertension was defined from antihypertensive medication use or ICD-9 codes 401-405 and ICD-10 codes I10-I15. Dyslipidemia was defined from ICD-9 code 272, ICD-10 code E78, self-report, or lipid-lowering medication use.

Statistical analysis

Continuous baseline variables are summarized as means and standard deviations (SDs); smoking-dose distributions additionally include medians and interquartile ranges (IQRs). Categorical variables are summarized as frequencies and percentages. Standardized mean differences (SMDs), rather than significance tests, describe baseline imbalance. For ASI, the largest absolute pairwise SMD is reported for each continuous variable or category indicator; for MSDP, the absolute two-group SMD is reported. Missing covariates were imputed using chained equations implemented in mice[30]. Models were fitted to the selected completed covariate dataset. Cox models estimated HRs and 95% CIs.

Model 1 adjusted for recruitment age, sex where applicable, race/ethnicity, Townsend deprivation category, household income, and education level. For MSDP and smoking dose, Model 2 additionally adjusted for BMI, alcohol frequency, smoking status, diet, physical-activity category, CRP, hypertension, heart disease, stroke, and diabetes, with the outcome-specific covariates described above. For the primary ASI analysis, Model 2 used the same extended covariates but omitted smoking status because the never smoking indicator is determined by the ASI reference definition; including both produces redundant adjustment. The extended model is a conditional association model, not an identified total or direct causal effect. Own smoking and adult health characteristics may mediate an association of MSDP with cancer. We added MSDP and dose models without smoking-status adjustment and a hypothesized pathway diagram. Removing smoking status from ASI Model 2 does not exclude possible over-adjustment by the remaining adult-health covariates.

Sensitivity analyses omitted smoking status from the MSDP and dose extended adjustment sets and examined the legacy ASI reference coding using Model 1 and the same no-status extended adjustment as in the primary ASI analysis. ASI trend tests were restricted to current or former smokers with a recorded initiation age. Categories were scored adult=0, adolescence=1, and childhood=2, so that a positive coefficient denotes a trend toward greater risk with earlier initiation. Primary trend tests used Model 1 and ASI Model 2 without smoking status; an additional sensitivity adjusted for current/former status among initiators, where the never smoking reference is absent. The ordinal test assumes equally spaced category scores, not equal differences in initiation age. Sparse category-specific contrasts with fewer than 10 events were not interpreted or displayed.

Analyses used R 4.3.1 and survival::coxph with Efron handling of ties. Original coefficient tables and new sensitivity-model outputs were merged by cancer identifier and coefficient name rather than row position. Two-sided nominal p-values and Benjamini–Hochberg adjusted p-values are provided separately. The multiplicity family comprises 27 cancer outcomes for each exposure coefficient or ASI contrast, model, and analysis; suppressed contrasts have no tested p-value. Initiator-only trend tests form separate 27-outcome families. FDR <0.05 defines multiplicity-adjusted evidence. Unadjusted models contain the exposure only.

Exploratory dose–response analyses used Cox models with four-knot restricted cubic splines fitted with rms::cph and Efron handling of ties. The original smoking-dose cohorts, event definitions, and dose coding were retained. Knots were selected by the rms routine with its handling of tied dose values; exact outcome-specific locations are supplied in RCS_model_statistics.csv. The reference was PYS=0 or PYSP=0. The spline models adjusted for age, sex where applicable, race/ethnicity, BMI, household income, alcohol category, physical activity, deprivation category, education level, smoking status, heart disease, stroke, diabetes, hypertension, CRP, and diet. Prostate models additionally adjusted for prostate-specific antigen testing and family history of prostate cancer. This original spline specification did not include the skin-specific covariates used in the linear dose models. Overall association and nonlinearity were evaluated using Wald tests with 3 and 2 degrees

of freedom, respectively. These exploratory spline p-values were nominal and were not included in the 27-outcome FDR families for the linear models.

RESULTS

Baseline characteristics

The primary ASI all-cause cohort comprised 404584 participants: 257506 self-reported never smokers, 60012 adult initiators, 83506 adolescent initiators, and 3560 childhood initiators. Corresponding recorded cancer events were: 34557; 10341; 14646; and 616. There were 218295 women and 186289 men in this cohort. The MSDP all-cause cohort included 117959 exposed participants (29.3%). Table 1 describes all three cancer-free all-cause analysis cohorts and provides SMDs. In the dose cohort, PYS averaged 6.96 pack-years (SD=14.77; median=0; IQR: 0–7), and PYSP averaged 0.165 pack-years/year (SD=0.341; median=0, IQR: 0–0.170).

For the smoking-dose all-cause analysis, median follow-up was 14.10 years (IQR: 13.18–14.95), totaling 6078382 person-years; 1209 participants were censored at their recorded loss-to-follow-up date. Corresponding MSDP values were 14.11 years (IQR: 13.18–14.95), 5248919 person-years, and 1048 lost. For the primary ASI analysis, median follow-up was 14.11 years (IQR: 13.17–14.95), totaling 5263895 person-years, with 1037 censored at loss to follow-up and 60160 recorded cancer events. Outcome-specific follow-up is provided in Supplementary file Table 1.

Associations of early tobacco exposure with all-cause and cause-specific cancer risk

In Model 2, MSDP was associated with all-cause cancer (HR=1.05; 95% CI: 1.03–1.07), lung cancer (HR=1.29; 95% CI 1.19–1.39), melanoma (HR=1.14; 95% CI: 1.04–1.25), keratinocyte carcinoma (HR=1.06; 95% CI: 1.03–1.10), and basal cell carcinoma (HR=1.08; 95% CI: 1.05–1.12), with FDR <0.05 (Figure 2; and Supplementary file Table 2). Oropharyngeal cancer had HR=1.17 (95% CI: 1.01–1.36; nominal p=0.034), but did not meet the FDR threshold. Renal cancer (HR=0.95; 95% CI: 0.83–1.08), bladder cancer (HR=1.10; 95% CI 0.95–1.28), leukemia (HR=1.02; 95% CI: 0.89–1.18), and multiple myeloma (HR=0.87; 95% CI: 0.73–1.03) were not associated at the multiplicity-adjusted threshold.

In the primary ASI analysis, Model 2 without smoking-status adjustment gave all-cause HRs of 1.11 (95% CI: 1.09–1.14) for adult, 1.16 (95% CI: 1.13–1.18) for adolescent, and 1.24 (95% CI: 1.15–1.35) for childhood initiation relative to self-reported never smoking (all FDR <0.05) (Figure 3). Corresponding Model 1 hours were 1.14 (95% CI: 1.11–1.16), 1.19 (95% CI: 1.17–1.21), and 1.29 (95% CI: 1.19–1.40). These contrasts combine differences in smoking history with initiation age and do not themselves estimate an effect of earlier initiation among smokers. The initiator-only ASI Model 2 all-cause score HR was 1.036 (95% CI: 1.011–1.061; nominal p=0.0042; FDR=0.0379), based on 147078 initiators and 25603 events. Trend tests of earlier initiation met FDR <0.05 for all-cause, lung, and bladder cancer. For oropharyngeal cancer, adult, adolescent, and childhood HRs versus never smoking were 2.35 (95% CI: 1.96–2.81), 2.11 (95% CI: 1.77–2.50), and 3.01 (95% CI: 1.90–4.78), respectively, whereas the initiator-only trend had p=0.693 (FDR=0.970), providing no evidence of a monotonic earlier-initiation gradient. Gastric and liver patterns were not assumed to be monotonic, and their trend tests did not meet the FDR threshold. Renal cancer is not described as associated with earlier initiation (initiator-only p=0.863). Full results are in given in Supplementary file Table 3. Childhood contrasts with fewer than 10 events, including cervical, endometrial, gallbladder, thyroid, and Hodgkin lymphoma outcomes, are suppressed; no CI is plotted for zero cervical events.

Associations of smoking dose with all-cause and cause-specific cancer risk

In the extended models, one original SD of PYS and PYSP was associated with all-cause cancer (both HR=1.06; 95% CI: 1.05–1.07). The dose scale used in these models was based on the pre-exclusion baseline dataset: 14.869 pack-years for PYS and 0.34118 pack-years/year for PYSP. Figure 4 presents both adjustment models, and Supplementary file Table 4 provides the corresponding estimates and FDR values. Dose coefficients should be read as conditional associations; the

Table 1. Baseline characteristics and recorded follow-up outcomes in the cancer-free all-cause analysis cohorts of UK Biobank, recruited April 2006–December 2010 and followed through 1 July 2023

Characteristic	ASI overall	Never smoking	Adult	Adolescent	Childhood	ASI max SMD	MSDP overall	MSDP no	MSDP yes	MSDP SMD	Dose overall
Total, n	404584	257506	60012	83506	3560	-	402849	284890	117959	-	467333
Sex											
Female	218295 (54.0)	151221 (58.7)	31129 (51.9)	34989 (41.9)	956 (26.9)	0.680	218975 (54.4)	156579 (55.0)	62396 (52.9)	0.041	252245 (54.0)
Male	186289 (46.0)	106285 (41.3)	28883 (48.1)	48517 (58.1)	2604 (73.1)	0.680	183874 (45.6)	128311 (45.0)	55563 (47.1)	0.041	215088 (46.0)
Age at recruitment (years)	56.19 (8.10)	55.57 (8.15)	57.81 (7.98)	57.01 (7.81)	55.10 (8.07)	0.337	56.07 (8.13)	56.22 (8.28)	55.73 (7.75)	0.061	56.24 (8.11)
Race/ethnicity											
White	376958 (93.2)	239002 (92.8)	55063 (91.8)	79525 (95.2)	3368 (94.6)	0.141	373894 (92.8)	260365 (91.4)	113529 (96.2)	0.203	435122 (93.1)
Asian or Asian British	14548 (3.6)	8398 (3.3)	3206 (5.3)	2833 (3.4)	111 (3.1)	0.111	15417 (3.8)	12154 (4.3)	3263 (2.8)	0.082	17437 (3.7)
Black or Black British	2502 (0.6)	1799 (0.7)	391 (0.7)	296 (0.4)	16 (0.4)	0.048	2545 (0.6)	2216 (0.8)	329 (0.3)	0.069	2891 (0.6)
Mixed	5187 (1.3)	4268 (1.7)	529 (0.9)	357 (0.4)	33 (0.9)	0.121	5369 (1.3)	5096 (1.8)	273 (0.2)	0.156	5755 (1.2)
Chinese	1520 (0.4)	1288 (0.5)	143 (0.2)	85 (0.1)	4 (0.1)	0.073	1515 (0.4)	1393 (0.5)	122 (0.1)	0.071	1650 (0.4)
Other	3869 (1.0)	2751 (1.1)	680 (1.1)	410 (0.5)	28 (0.8)	0.072	4109 (1.0)	3666 (1.3)	443 (0.4)	0.100	4478 (1.0)
Annual household income (£)											
<18000	98735 (24.4)	55758 (21.7)	15892 (26.5)	25533 (30.6)	1552 (43.6)	0.481	94784 (23.5)	66795 (23.4)	27989 (23.7)	0.007	113142 (24.2)
18000–30999	103837 (25.7)	64247 (24.9)	15951 (26.6)	22736 (27.2)	903 (25.4)	0.052	102294 (25.4)	72611 (25.5)	29683 (25.2)	0.007	119822 (25.6)
31000–51999	103454 (25.6)	68020 (26.4)	14559 (24.3)	20156 (24.1)	719 (20.2)	0.147	103653 (25.7)	72554 (25.5)	31099 (26.4)	0.020	119366 (25.5)
52000–100000	78288 (19.4)	54715 (21.2)	10710 (17.8)	12521 (15.0)	342 (9.6)	0.327	80511 (20.0)	57203 (20.1)	23308 (19.8)	0.008	91081 (19.5)
>100000	20270 (5.0)	14766 (5.7)	2900 (4.8)	2560 (3.1)	44 (1.2)	0.247	21607 (5.4)	15727 (5.5)	5880 (5.0)	0.024	23922 (5.1)
Deprivation category											
Low	82264 (20.3)	44208 (17.2)	14732 (24.5)	21825 (26.1)	1499 (42.1)	0.568	79995 (19.9)	54542 (19.1)	25453 (21.6)	0.060	94430 (20.2)
Medium	241332 (59.6)	156343 (60.7)	34960 (58.3)	48301 (57.8)	1728 (48.5)	0.246	240976 (59.8)	170850 (60.0)	70126 (59.4)	0.011	279132 (59.7)
High	80988 (20.0)	56955 (22.1)	10320 (17.2)	13380 (16.0)	333 (9.4)	0.356	81878 (20.3)	59498 (20.9)	22380 (19.0)	0.048	93771 (20.1)
Education level											
College or university degree	130503 (32.3)	92844 (36.1)	21645 (36.1)	15587 (18.7)	427 (12.0)	0.587	135230 (33.6)	102052 (35.8)	33178 (28.1)	0.166	152131 (32.6)

Continued

Table 1. **Continued**

Characteristic	ASI overall	Never smoking	Adult	Adolescent	Childhood	ASI max SMD	MSDP overall	MSDP no	MSDP yes	MSDP SMD	Dose overall
A levels/AS levels or equivalent	45329 (11.2)	30225 (11.7)	7614 (12.7)	7261 (8.7)	229 (6.4)	0.214	45491 (11.3)	32715 (11.5)	12776 (10.8)	0.021	52217 (11.2)
O levels/GCSEs or equivalent	85893 (21.2)	53894 (20.9)	11830 (19.7)	19550 (23.4)	619 (17.4)	0.150	84424 (21.0)	58737 (20.6)	25687 (21.8)	0.028	98689 (21.1)
CSEs or equivalent	22583 (5.6)	14190 (5.5)	2383 (4.0)	5710 (6.8)	300 (8.4)	0.186	21944 (5.4)	13941 (4.9)	8003 (6.8)	0.081	25878 (5.5)
NVQ or HND or HNC or equivalent	26853 (6.6)	14865 (5.8)	3653 (6.1)	7949 (9.5)	386 (10.8)	0.184	26171 (6.5)	17103 (6.0)	9068 (7.7)	0.067	30762 (6.6)
Other	20683 (5.1)	13020 (5.1)	3279 (5.5)	4231 (5.1)	153 (4.3)	0.054	20645 (5.1)	14726 (5.2)	5919 (5.0)	0.007	23907 (5.1)
None of the above	68360 (16.9)	35566 (13.8)	9085 (15.1)	22332 (26.7)	1377 (38.7)	0.589	64845 (16.1)	42504 (14.9)	22341 (18.9)	0.107	78611 (16.8)
Prefer not to answer	4380 (1.1)	2902 (1.1)	523 (0.9)	886 (1.1)	69 (1.9)	0.091	4099 (1.0)	3112 (1.1)	987 (0.8)	0.026	5138 (1.1)
BMI (kg/m^2)	27.47 (4.84)	27.21 (4.81)	27.61 (4.83)	28.13 (4.82)	28.96 (5.39)	0.343	27.41 (4.80)	27.18 (4.70)	27.94 (4.99)	0.156	27.44 (4.81)
Alcohol consumption											
Never	34836 (8.6)	24730 (9.6)	3966 (6.6)	5699 (6.8)	441 (12.4)	0.198	33150 (8.2)	25147 (8.8)	8003 (6.8)	0.076	37867 (8.1)
Occasionally	94486 (23.4)	64639 (25.1)	11732 (19.5)	17236 (20.6)	879 (24.7)	0.134	91223 (22.6)	65224 (22.9)	25999 (22.0)	0.020	106073 (22.7)
Sometimes	194898 (48.2)	129022 (50.1)	26311 (43.8)	38173 (45.7)	1392 (39.1)	0.223	197160 (48.9)	138350 (48.6)	58810 (49.9)	0.026	228742 (48.9)
Daily	80364 (19.9)	39115 (15.2)	18003 (30.0)	22398 (26.8)	848 (23.8)	0.360	81316 (20.2)	56169 (19.7)	25147 (21.3)	0.040	94651 (20.3)
Smoking status											
Current	41115 (10.2)	0 (0.0)	16985 (28.3)	22670 (27.1)	1460 (41.0)	1.179	41355 (10.3)	27001 (9.5)	14354 (12.2)	0.087	49954 (10.7)
Former	105963 (26.2)	0 (0.0)	43027 (71.7)	60836 (72.9)	2100 (59.0)	2.317	135248 (33.6)	95233 (33.4)	40015 (33.9)	0.010	159854 (34.2)
Never	257506 (63.6)	257506 (100)	0 (0.0)	0 (0.0)	0 (0.0)	NA	226246 (56.2)	162656 (57.1)	63590 (53.9)	0.064	257525 (55.1)
Physical activity											
Low	74833 (18.5)	46338 (18.0)	11311 (18.8)	16383 (19.6)	801 (22.5)	0.112	73097 (18.1)	51110 (17.9)	21987 (18.6)	0.018	85290 (18.3)
Medium	203521 (50.3)	132181 (51.3)	30579 (51.0)	39300 (47.1)	1461 (41.0)	0.208	203843 (50.6)	145892 (51.2)	57951 (49.1)	0.042	235644 (50.4)
High	126230 (31.2)	78987 (30.7)	18122 (30.2)	27823 (33.3)	1298 (36.5)	0.133	125909 (31.3)	87888 (30.8)	38021 (32.2)	0.030	146399 (31.3)
Healthy diet	157585 (38.9)	104409 (40.5)	23462 (39.1)	28653 (34.3)	1061 (29.8)	0.226	159927 (39.7)	114960 (40.4)	44967 (38.1)	0.046	184314 (39.4)
Health status											
CRP (mg/L)	2.53 (4.20)	2.32 (3.96)	2.74 (4.38)	2.99 (4.68)	3.59 (5.16)	0.276	2.48 (4.15)	2.40 (4.07)	2.67 (4.36)	0.062	2.50 (4.18)
Heart disease	49356 (12.2)	24328 (9.4)	8966 (14.9)	15125 (18.1)	937 (26.3)	0.451	47641 (11.8)	32597 (11.4)	15044 (12.8)	0.040	56363 (12.1)

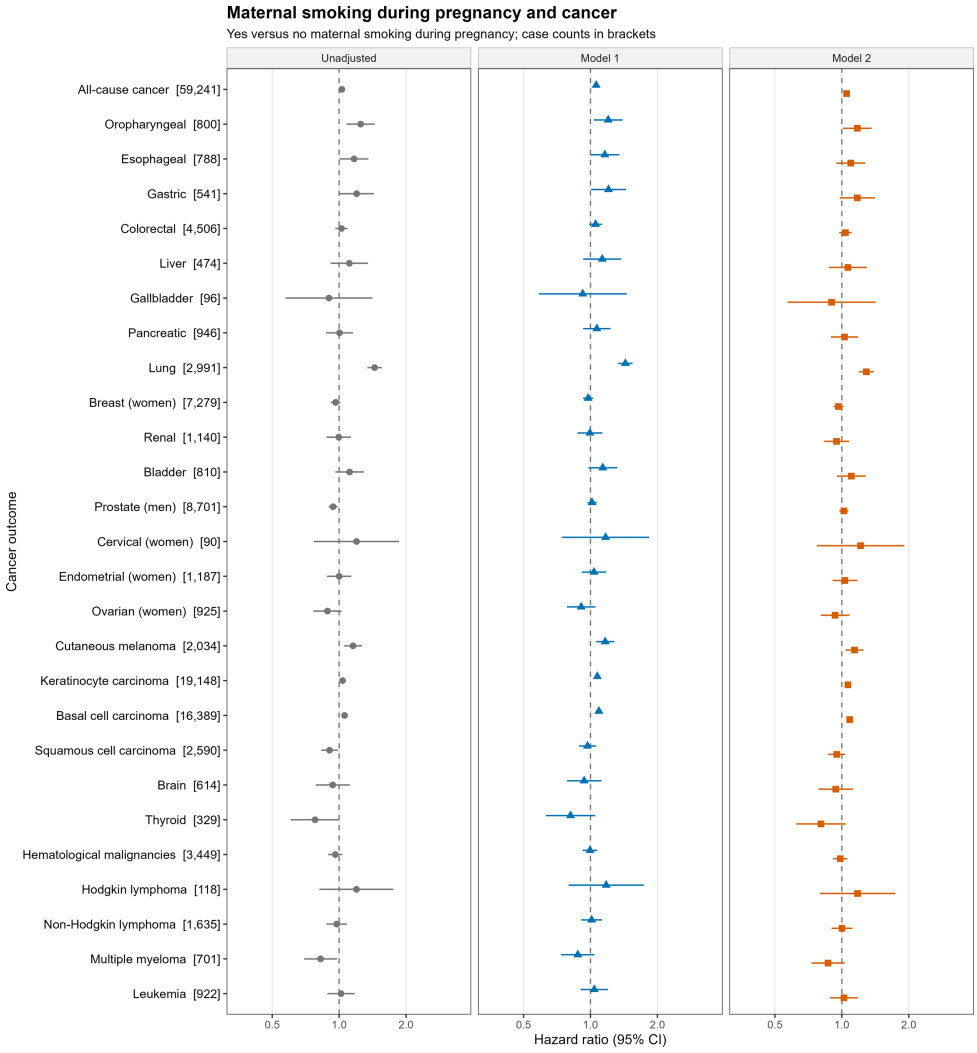
Continued

Table 1. **Continued**

Characteristic	ASI overall	Never smoking	Adult	Adolescent	Childhood	ASI max SMD	MSDP overall	MSDP no	MSDP yes	MSDP SMD	Dose overall
Stroke	12345 (3.1)	6365 (2.5)	2311 (3.9)	3473 (4.2)	196 (5.5)	0.155	11936 (3.0)	8351 (2.9)	3585 (3.0)	0.006	14175 (3.0)
Diabetes mellitus	38014 (9.4)	19640 (7.6)	6670 (11.1)	11005 (13.2)	699 (19.6)	0.355	36349 (9.0)	25038 (8.8)	11311 (9.6)	0.028	43058 (9.2)
Hypertension	128747 (31.8)	72364 (28.1)	21751 (36.2)	33049 (39.6)	1583 (44.5)	0.345	125498 (31.2)	87007 (30.5)	38491 (32.6)	0.045	148089 (31.7)
Pack-years											
PYS	8.03 (15.60)	0.00 (0.00)	17.55 (15.43)	24.61 (19.61)	40.03 (31.18)	1.816	6.72 (14.53)	5.89 (13.43)	8.72 (16.71)	0.186	6.96 (14.77)
PYSP (pack-years/year)	0.19 (0.36)	0.00 (0.00)	0.42 (0.35)	0.58 (0.44)	0.85 (0.65)	1.859	0.16 (0.34)	0.14 (0.31)	0.21 (0.39)	0.198	0.16 (0.34)
All-cause cancer during follow-up	60160 (14.9)	34557 (13.4)	10341 (17.2)	14646 (17.5)	616 (17.3)	0.114	59241 (14.7)	41573 (14.6)	17668 (15.0)	0.011	69647 (14.9)

Values are n (%) or mean (SD) consistently across all cohorts. Primary ASI (N=404584; self-reported never smoking reference), MSDP (N=402849), and smoking dose (N=467333) are described with standardized differences, not baseline hypothesis-test p-values. ASI maximum SMD is the largest absolute pairwise standardized difference for the displayed continuous variable or category indicator; MSDP SMD is the absolute two-group difference. Counts are recomputed from the saved analysis datasets. Primary ASI uses self-reported never smoking as the reference and excludes ever smokers with missing initiation age and discordant never smokers with recorded age. For the never smoking indicator, the SMD is undefined because the groups are deterministically separated with zero within-group variance; this cell is shown as NA, not as zero. PYS/PYSP rows are descriptive, and stored zero values may include originally missing dose data. Dose medians (IQR): PYS 0 (0–7) pack-years; PYSP 0 (0–0.170) pack-years/year. The original exposure SDs used in regression (14.869 and 0.34118) were derived before baseline-cancer exclusion and differ slightly from the SDs displayed here. Alcohol consumption retains the original four categories: never, occasionally, sometimes, and daily. The recurring values were checked against the analysis data: CSE/MSDP yes = 8003; alcohol never/MSDP yes = 8003; alcohol never/MSDP no = 25147; alcohol daily/MSDP yes = 25147.

Figure 2. MSDP and cancer in UK Biobank. The exposure contrast is reported maternal smoking during pregnancy versus no maternal smoking (reference)

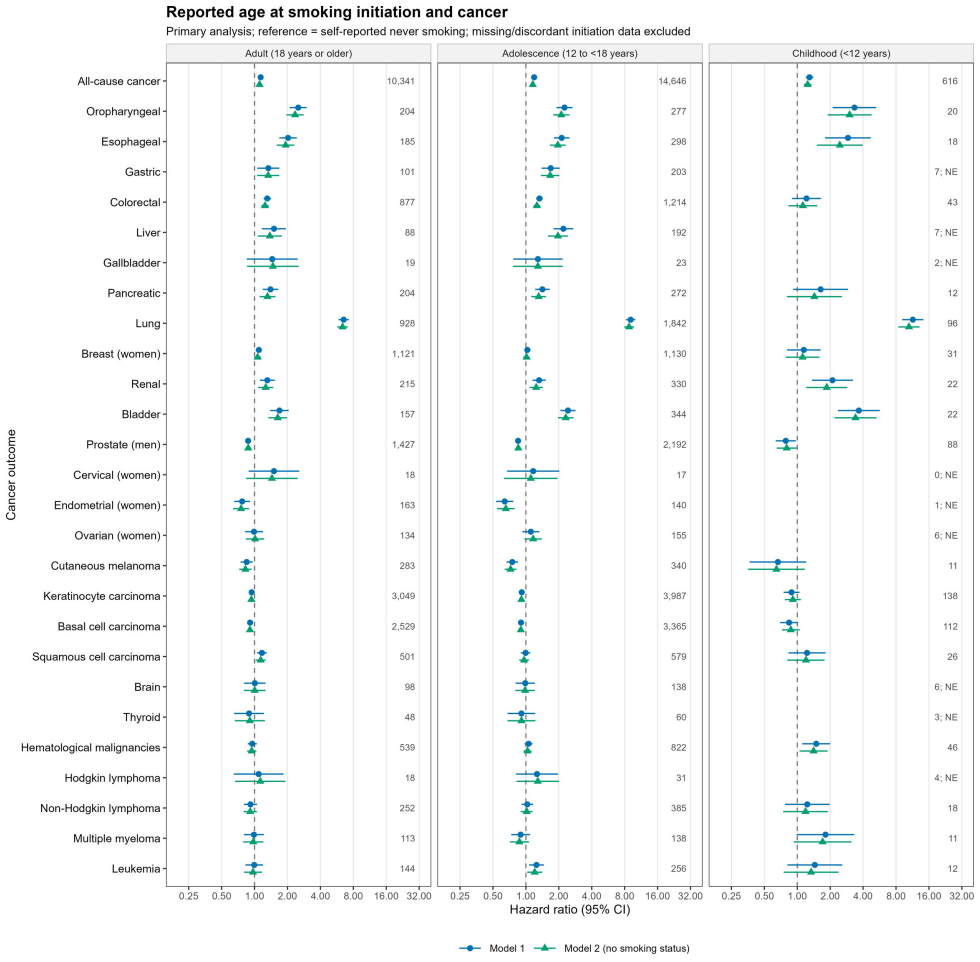


Brackets beside outcome labels give event counts, including 59241 all-cause events among 402849 participants. Outcome-specific numbers, sex restrictions, follow-up, and event counts are given in Supplementary file Table 1. Unadjusted models include MSDP only. Model 1 adjusts for recruitment age, sex where applicable, race/ethnicity, deprivation category, income, and education level. Model 2 additionally includes BMI, alcohol frequency, own smoking status, physical activity, and diet. Points and bars show HRs and 95% CIs; the dashed vertical line marks HR=1. The axis is logarithmic. Nominal and Benjamini–Hochberg adjusted p-values are reported in Supplementary file Table 2, across 27 outcomes within each model. Model 2 is not interpreted as a total causal effect. CRP: chronic diseases, and relevant outcome-specific covariates, as detailed in the Methods. CI: confidence interval. HR: hazard ratio. MSDP: maternal smoking during pregnancy.

supplementary models omitting smoking status assess sensitivity to this overlapping adjustment variable.

Restricted cubic spline analyses showed departures from linearity for all-cause cancer with both PYS (p for nonlinearity=0.0002) and PYSP (p=0.0012), and for lung cancer with both measures (both p<0.0001; Supplementary file Figure 8). Nominal evidence of nonlinearity was also observed for gastric cancer (PYS, p=0.0240; PYSP, p=0.0069), prostate cancer (PYS, p=0.0296; PYSP, p=0.0204), and leukemia (PYS, p=0.0179; PYSP, p=0.0249). For PYSP, liver cancer (p=0.0371) and basal cell carcinoma (p=0.0219) also showed nominal nonlinearity. The remaining curves did not meet nominal p<0.05 for nonlinearity. Complete results for 23 outcomes and both dose measures are provided in Supplementary file Table 7 and Supplementary file Figures 6–7. These exploratory findings describe curvature rather than exposure thresholds or evidence of a protective dose range.

Figure 3. Primary ASI analysis using self-reported never smoking as the reference, showing Models 1 and 2 together



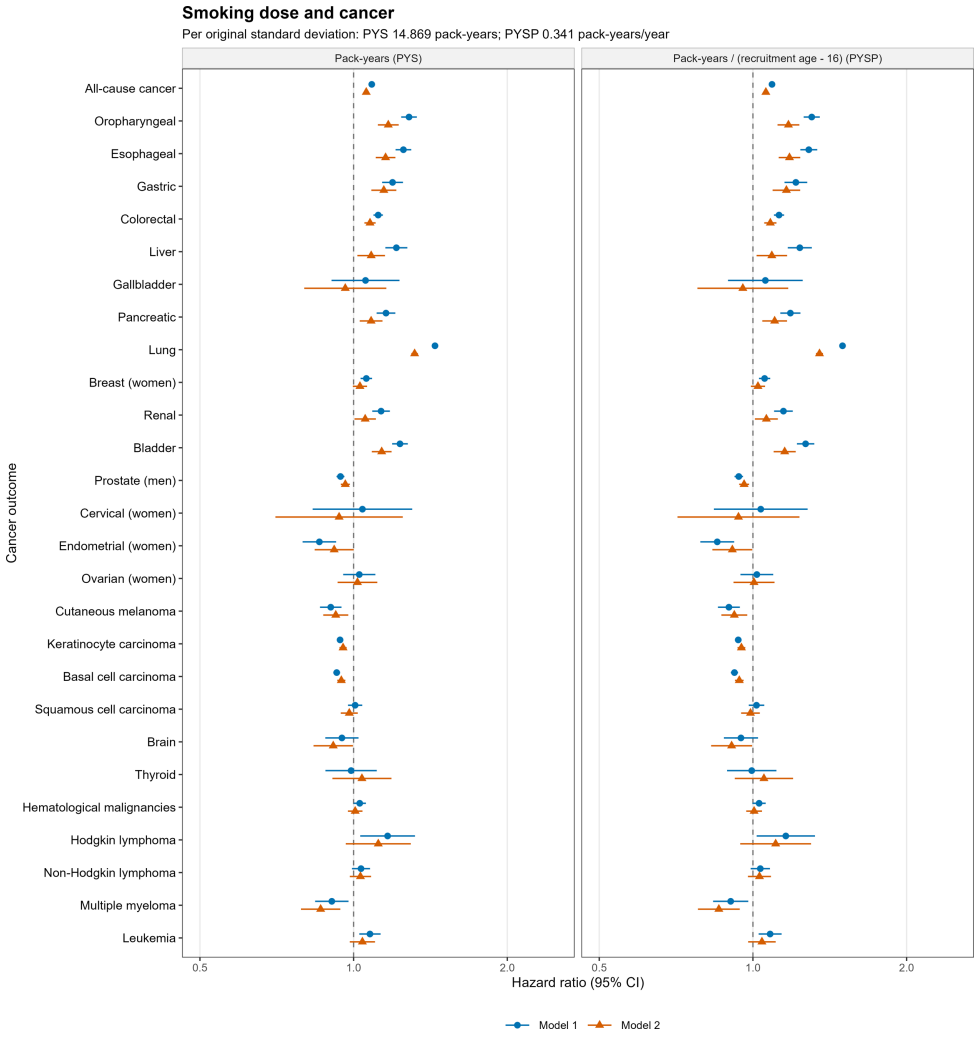
Numbers at the right of each panel are events. NE: estimate suppressed because fewer than 10 events.
Primary ASI Model 2 omits smoking status. Initiator-only trend tests are in Supplementary Table 3 and eFigure 2.

Current/former smokers without initiation age and discordant never smokers with recorded initiation age are excluded. Initiation categories are adult (≥18 years), adolescent (12 to <18 years), and childhood (<12 years). Model 1 adjusts for age, sex where applicable, race/ethnicity, deprivation, income, and education. ASI Model 2 adds the extended covariates described in the Methods but omits smoking status to avoid deterministic overlap with the reference category. All panels share a logarithmic HR scale; the dashed line denotes HR=1. Numbers at the right of each panel are category-specific events. NE denotes a suppressed estimate with fewer than 10 events; zero-event categories have no plotted CI. Each contrast and model has a separate 27-outcome FDR family. Full estimates and FDR values are in Supplementary file Table 3. Trend tests in that table include only current/former smokers with recorded initiation age, scored adult=0, adolescent=1, childhood=2; each model has a separate 27-outcome trend family. Sex-specific restrictions and numbers are given in Supplementary file Table 1. ASI: age at smoking initiation. CI: confidence interval. FDR: false discovery rate. HR: hazard ratio.

Sensitivity analysis

Supplementary file Table 5 and Supplementary file Figures 1 and 3 report the newly fitted MSDP and dose extended models without smoking-status adjustment. Supplementary file Table 3 and Supplementary file Figure 2 report initiator-only ASI trend analyses using the primary adjustment sets; Supplementary file Table 3 additionally shows the current/former-status-adjusted sensitivity. The all-cause trend remained positive with that additional adjustment (HR=1.042; 95% CI: 1.017–1.068; p=0.00085; FDR=0.0077), while the oropharyngeal trend remained unsupported (p=0.847). Supplementary file Table 6 and Supplementary file Figure 4 retain the original mixed no-recorded-age reference as a sensitivity analysis, with Model 1 and the same no-status extended covariates as the primary ASI analysis. This comparison changes both reference composition and eligibility and is not a pure reference relabeling. Differences between analyses are not interpreted as evidence of a causal pathway.

Figure 4. PYS and PYSP associations with cancer, showing Models 1 and 2 together



HRs compare values differing by one original baseline SD: 14.869 pack-years for PYS and 0.34118 pack-years/year for PYSP. PYSP is PYS divided by recruitment age minus 16, not the proportion of lifespan exposed. Both panels use a common logarithmic HR scale; the dashed line marks HR=1. Points and bars are HRs and 95% CIs. Models are defined in Figure 2. Each exposure and model has a separate 27-outcome Benjamini–Hochberg family; exact nominal and adjusted p-values are in given Supplementary file Table 4. Sex-specific numbers are in given Supplementary file Table 1. CI: confidence interval. HR: hazard ratio. PYS: pack-years of smoking. PYSP: scaled pack-years.

DISCUSSION

The results support an association of MSDP with all-cause cancer and selected outcomes, including lung cancer, melanoma, and keratinocyte carcinoma. MSDP associations with renal cancer, bladder cancer, leukemia, and multiple myeloma did not meet the multiplicity-adjusted threshold. Greater smoking dose was associated with all-cause cancer, but the extent of adjustment materially affects interpretation. Initiation-age contrasts and trends among initiators answer different questions, and neither establishes causality.

Previous epidemiological studies have focused mainly on MSDP and cancer in young offspring, including testicular cancer[31], central nervous system tumors[2,32], and hematological malignancies[1]. Experimental evidence indicates that fetal tissues may be especially susceptible to carcinogenic injury and that tumors induced by transplacental exposure may manifest later in life[7,8,10]. Our analysis extends this literature by examining cancer risk in adult offspring within a large prospective resource. The absence of a significant association with brain cancer differs from some childhood findings. Maternal smoking has also been associated with childhood obesity

after adjustment for major confounders[33], which may be relevant to obesity-related cancers. Mechanistic and replication studies are needed before these associations can be interpreted causally.

Early life may be a vulnerable period for tobacco exposure[11]. Previous studies have associated childhood or adolescent smoking with lung[12,34,35], breast[13], and colorectal cancers[14]. In this analysis, all-cause cancer HRs relative to self-reported never smoking were higher in earlier initiation categories. However, those contrasts are distinct from a gradient among smokers. The initiator-only tests supported an earlier initiation trend for all-cause, lung, and bladder cancer after multiplicity correction, but not for oropharyngeal, gastric, or liver cancer; category patterns should not be assumed to be monotonic. Renal cancer is therefore excluded from the list of cancers associated with earlier ASI. Excluding smokers with missing initiation age improves reference classification but may introduce selection bias. Rare childhood-initiation contrasts are not interpreted when they contain fewer than 10 events. Tobacco exposure may interact with immune-related genetic pathways[36], but this requires targeted investigation.

Prior studies have reported smoking-dose associations with breast[15], gastric[16], pancreatic[17], lung[18], renal[19], and bladder cancers[20]. PYS and PYSP quantify related aspects of accumulated exposure, with PYSP scaling pack-years by recruitment age minus 16. Their coefficients are not independent biological effects and do not identify risk thresholds. Inverse estimates for some skin cancers, prostate cancer, or multiple myeloma should not be interpreted as protective effects. The spline models additionally suggested that a single linear coefficient may not fully describe the dose–response association for all-cause and lung cancer, with exploratory evidence of curvature for selected other outcomes. These nominal findings require cautious interpretation given multiple comparisons, wider uncertainty at high doses, and the shared information captured by PYS and PYSP. A crossing of HR=1 is not a biological or clinical safety threshold.

Strengths and limitations

Strengths include the large cohort, long follow-up, linked outcomes, and evaluation of early life and cumulative tobacco exposure. Several limitations affect interpretation. Exposures were assessed once at baseline, so subsequent initiation, cessation, relapse, or changes in intensity could not be represented as time-varying exposure. Maternal smoking and initiation age were retrospectively self-reported, with possible recall error and misclassification; secondhand tobacco exposure was not comprehensively measured. The primary ASI reference now excludes ever smokers with missing initiation age and discordant never smokers, but self-reported never smoking status is not independently validated and these exclusions may introduce selection bias. The legacy-reference sensitivity does not remove this concern. Residual familial, social, and behavioral confounding remains possible. Smoking, BMI, CRP, and chronic disease may be mediators of MSDP, and extended adjustment may introduce over-adjustment or collider bias. Smoking status was omitted from primary ASI models to avoid deterministic overlap, but remaining extended covariates may still be downstream of tobacco exposure. Death was treated as censoring rather than modelled as a competing event, so the cause-specific HRs should not be interpreted as cumulative-incidence risk ratios. Analyses of a single completed covariate dataset do not propagate between-imputation uncertainty. Sparse contrasts were suppressed rather than interpreted. Spline tests were exploratory and not adjusted for multiple comparisons, and estimates at extreme doses were less precise. UK Biobank volunteer selection and the predominance of White participants limit generalizability. These observational estimates do not establish causality.

CONCLUSIONS

The present observational associations support continued evaluation of early tobacco exposure and cumulative smoking dose as cancer-risk markers. The primary ASI analysis uses self-reported never smoking as the reference, with separate initiator-only tests to assess an earlier initiation gradient among smokers. The study does not establish

causal effects or demonstrate the effectiveness of a particular preventive intervention.

REFERENCES

1. Stjernfeldt M, Berglund K, Lindsten J, Ludvigsson J. Maternal smoking during pregnancy and risk of childhood cancer. Lancet. 1986;1(8494):1350–1352. doi:10.1016/s0140-6736(86)91664-8
2. Heck JE, Contreras ZA, Park AS, Davidson TB, Cockburn M, Ritz B. Smoking in pregnancy and risk of cancer among young children: a population-based study. Int J Cancer. 2016;139(3):613–616. doi:10.1002/ijc.30111
3. Mucci LA, Granath F, Cnattingius S. Maternal smoking and childhood leukemia and lymphoma risk among 1,440,542 Swedish children. Cancer Epidemiol Biomarkers Prev. 2004;13(9):1528–1533. doi:10.1158/1055-9965.1528.13.9
4. Nicolov IG, Chernozemsky IN. Tumors and hyperplastic lesions in Syrian hamsters following transplacental and neonatal treatment with cigarette smoke condensate. J Cancer Res Clin Oncol. 1979;94(3):249–256. doi:10.1007/BF00419284
5. Wynder EL, Hoffmann D. Tobacco and health: a societal challenge. N Engl J Med. 1979;300(16):894–903. doi:10.1056/NEJM197904193001605
6. Everson RB. Individuals transplacentally exposed to maternal smoking may be at increased cancer risk in adult life. Lancet. 1980;2(8186):123–127. doi:10.1016/s0140-6736(80)90007-0
7. Rice JM. Carcinogenesis: a late effect of irreversible toxic damage during development. Environ Health Perspect. 1976;18:133–139. doi:10.1289/ehp.7618133
8. Rice JM. Perinatal period and pregnancy: intervals of high risk for chemical carcinogens. Environ Health Perspect. 1979;29:23–27. doi:10.1289/ehp.792923
9. Armuth V, Berenblum I. Tritiated thymidine as a broad spectrum initiator in transplacental two-stage carcinogenesis, with phorbol as promoter. Int J Cancer. 1979;24(3):355–358. doi:10.1002/ijc.2910240314
10. Goerttler K, Löhrke H. Diaplacental carcinogenesis: tumor localization and tumor incidence in NMRI mice after diaplacental initiation with DMBA and urethane and postnatal promotion and the phorbol ester TPA in a modified 2-stage Berenblum/Mottram experiment. Virchows Arch A Pathol Anat Histol. 1977;376(2):117–132. doi:10.1007/BF00432583
11. Clarke MA, Joshu CE. Early life exposures and adult cancer risk. Epidemiol Rev. 2017;39(1):11–27. doi:10.1093/epirev/mxx004
12. He H, He MM, Wang H, et al. In utero and childhood/adolescence exposure to tobacco smoke, genetic risk, and lung cancer incidence and mortality in adulthood. Am J Respir Crit Care Med. 2023;207(2):173–182. doi:10.1164/rccm.202112-2758OC
13. Gram IT, Braaten T, Terry PD, et al. Breast cancer risk among women who start smoking as teenagers. Cancer Epidemiol Biomarkers Prev. 2005;14(1):61–66. doi:10.1158/1055-9965.61.14.1
14. Liang PS, Chen TY, Giovannucci E. Cigarette smoking and colorectal cancer incidence and mortality: systematic review and meta-analysis. Int J Cancer. 2009;124(10):2406–2415. doi:10.1002/ijc.24191
15. Scala M, Bosetti C, Bagnardi V, et al. Dose-response relationships between cigarette smoking and breast cancer risk: a systematic review and meta-analysis. J Epidemiol. 2023;33(12):640–648. doi:10.2188/jea.JE20220206
16. Rota M, Possenti I, Valsassina V, et al. Dose-response association between cigarette smoking and gastric cancer risk: a systematic review and meta-analysis. Gastric Cancer. 2024;27(2):197–209. doi:10.1007/s10120-023-01459-1
17. Zou L, Zhong R, Shen N, et al. Non-linear dose-response relationship between cigarette smoking and pancreatic cancer risk: evidence from a meta-analysis of 42 observational studies. Eur J Cancer. 2014;50(1):193–203. doi:10.1016/j.ejca.2013.08.014
18. Law MR, Morris JK, Watt HC, Wald NJ. The dose-response relationship between cigarette consumption, biochemical markers and risk of lung cancer. Br J Cancer. 1997;75(11):1690–1693. doi:10.1038/bjc.1997.287
19. Liu X, Peveri G, Bosetti C, et al. Dose-response relationships between cigarette smoking and kidney cancer: a systematic review and meta-analysis. Crit Rev Oncol Hematol. 2019;142:86–93. doi:10.1016/j.critrevonc.2019.07.019
20. Zhao X, Wang Y, Liang C. Cigarette smoking and risk of bladder cancer: a dose-response meta-analysis. Int Urol Nephrol. 2022;54(6):1169–1185. doi:10.1007/s11255-022-03173-w
21. Cox N. UK Biobank shares the promise of big data. Nature. 2018;562(7726):194–195. doi:10.1038/d41586-018-06948-3
22. Mao R, Wang F, Zhong Y, Meng X, Zhang T, Li J. Association of biological age acceleration with cardiac morphology, function, and incident heart failure: insights from UK Biobank participants. Eur Heart J Cardiovasc Imaging. 2024;25(9):1315–1323. doi:10.1093/ehjci/jeae126
23. von Elm E, Altman DG, Egger M, et al. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. Ann Intern Med. 2007;147(8):573–577. doi:10.7326/0003-4819-147-8-200710160-00010
24. UK Biobank. UK Biobank: Protocol for a large-scale prospective epidemiological resource. Protocol No:

UKBB-PROT-09-06 (Main Phase); November 21, 2006. Accessed October 17, 2026. https://www.ukbiobank.ac.uk/wp-content/uploads/2025/01/Main-study-protocol.pdf
25. Siegel DA, O'Neil ME, Richards TB, Dowling NF, Weir HK. Prostate cancer incidence and survival, by stage and race/ethnicity - United States, 2001–2017. MMWR Morb Mortal Wkly Rep. 2020;69(41):1473–1480. doi:10.15585/mmwr.mm6941a1
26. Guillaume E, Pornet C, Dejardin O, et al. Development of a cross-cultural deprivation index in five European countries. J Epidemiol Community Health. 2016;70(5):493–499. doi:10.1136/jech-2015-205729
27. Lourida I, Hannon E, Littlejohns TJ, et al. Association of lifestyle and genetic risk with incidence of dementia. JAMA. 2019;322(5):430–437. doi:10.1001/jama.2019.9879
28. Pearce M, Garcia L, Abbas A, et al. Association between physical activity and risk of depression: a systematic review and meta-analysis. JAMA Psychiatry. 2022;79(6):550–559. doi:10.1001/jamapsychiatry.2022.0609
29. Mozaffarian D. Dietary and policy priorities for cardiovascular disease, diabetes, and obesity: a comprehensive review. Circulation. 2016;133(2):187–225. doi:10.1161/CIRCULATIONAHA.115.018585
30. Zhang Z. Multiple imputation with multivariate imputation by chained equation (MICE) package. Ann Transl Med. 2016;4(2):30. doi:10.3978/j.issn.2305-5839.2015.12.63
31. Tuomisto J, Holl K, Rantakokko P, et al. Maternal smoking during pregnancy and testicular cancer in the sons: a nested case-control study and a meta-analysis. Eur J Cancer. 2009;45(9):1640–1648. doi:10.1016/j.ejca.2009.01.017
32. Rumrich IK, Viluksela M, Vähäkangas K, Gissler M, Surcel HM, Hänninen O. Maternal smoking and the risk of cancer in early life - a meta-analysis. PLoS One. 2016;11(11):e0165040. doi:10.1371/journal.pone.0165040
33. Rayfield S, Plugge E. Systematic review and meta-analysis of the association between maternal smoking in pregnancy and childhood overweight and obesity. J Epidemiol Community Health. 2017;71(2):162–173. doi:10.1136/jech-2016-207376
34. Olivo-Marston SE, Yang P, Mechanic LE, et al. Childhood exposure to secondhand smoke and functional mannose binding lectin polymorphisms are associated with increased lung cancer risk. Cancer Epidemiol Biomarkers Prev. 2009;18(12):3375–3383. doi:10.1158/1055-9965.EPI-09-0986
35. Robles AI, Yang P, Jen J, et al. A DRD1 polymorphism predisposes to lung cancer among those exposed to secondhand smoke during childhood. Cancer Prev Res (Phila). 2014;7(12):1210–1218. doi:10.1158/1940-6207.CAPR-14-0158
36. Simonds NI, Ghazarian AA, Pimentel CB, et al. Review of the gene-environment interaction literature in cancer: what do we know?. Genet Epidemiol. 2016;40(5):356–365. doi:10.1002/gepi.21967

CONFLICTS OF INTEREST
The authors have completed and submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest and none was reported.

FUNDING
This work was supported by Hunan Provincial Natural Science Foundation (2026JJ80935), Changsha Natural Science Foundation (kzd2501018, kq2502151) and Hunan Health Commission High-level Talent Project (20230609-1014).

ETHICAL APPROVAL AND INFORMED CONSENT
Ethical approval and informed consent were not required for this study as it is a secondary analysis of existing data.

DATA AVAILABILITY
The data supporting this research are available from the authors on reasonable request.

AUTHORS' CONTRIBUTIONS
WL, BWS, and ZZ: conceptualization.
WL, BWS, JYH, JZ, and ZZ: methodology.
WL, BWS, JYH, XY, GFL, JZ, TTZ, and YSC: investigation. BWS, JYH, SQX, XY, GFL, and TTZ: data curation. BWS, JYH, SQX, and JD: formal analysis. WL, XY, JZ, and TTZ: resources. SQX and JD: software. SQX, GFL, JD, and YSC: validation. SQX and JD: visualization. WL and ZZ: supervision. WL and ZZ: project administration. WL and BWS: writing – original draft. WL, BWS, JYH, SQX, XY, GFL, JD, JZ, TTZ, YSC, and ZZ: writing – review and editing. All authors: contribution to the critical revision of the manuscript. All authors read and approved the final version of the manuscript.

PROVENANCE AND PEER REVIEW
Not commissioned; externally peer reviewed.

SUPPLEMENTARY FILE DISCLAIMER
The content has been provided by the author(s) and has not been reviewed, verified, or endorsed by European Publishing. It may not have undergone peer review. The views, opinions, and recommendations expressed are solely those of the author(s) and do not necessarily reflect the position of European Publishing. European Publishing accepts no responsibility or liability for any consequences arising from the use of, or reliance on, this content.