

## Title Page

# Modelling the Complex Smoking Exposure History in Assessment of Pan-Cancer Risk

## Author List and Affiliations

Wei Liu<sup>1,2#</sup>, Ya-Ting Chen<sup>1,2#</sup>, Baiwenrui Tao<sup>3#</sup>, Ying Lv<sup>3</sup>, Yan-Xi Zhang<sup>1,2</sup>, Hui-Ying Ren<sup>1,2</sup>, Yu-Ting Zhang<sup>1,2</sup>, Yu-Ping Fan<sup>1,2</sup>, Meng-Han Li<sup>1,2</sup>, Ya-Xin Shi<sup>1,2</sup>, Shi-Yuan Wang<sup>1,2</sup>, Bing-Wei Chen<sup>1,2</sup>, Frits van Osch<sup>4,5</sup>, Maurice P. Zeegers<sup>4,6</sup>, Qi-Rong Qin<sup>3,7\*</sup>, Anke Wesselius<sup>4,6\*</sup>, Evan Yi-Wen Yu<sup>1,2,4\*</sup>

<sup>1</sup> Department of Epidemiology & Biostatistics, School of Public Health, Southeast University, Nanjing 210009, China.

<sup>2</sup> Key Laboratory of Environmental Medicine and Engineering of Ministry of Education, School of Public Health, Southeast University, Nanjing 210009, China.

<sup>3</sup> School of Public Health, Wannan Medical College, Wuhu 241002, China.

<sup>4</sup> Department of Epidemiology, CAPHRI Care and Public Health Research Institute, Maastricht University, Maastricht 6229ER, the Netherlands.

<sup>5</sup> VieCuri Medical Center, Department of Clinical Epidemiology, Venlo, The Netherlands.

<sup>6</sup> School of Nutrition and Translational Research in Metabolism, Maastricht University, Maastricht 6229 ER, the Netherlands.

<sup>7</sup> Ma'anshan Center for Disease Control and Prevention, Ma'anshan 243011, China.

**Running title:** Smoking and Pan-Cancer Risk

<sup>#</sup>These authors contributed equally.

**\*Correspondence to:**

**Evan Yi-Wen Yu, PhD**

Mail address: 87 Dingjiaqiao, Gulou District, Nanjing 210009, China.

Tel: +86 25 83272566; Fax: +86 25 83324322

E-mail address: [evan.yu@maastrichtuniversity.nl](mailto:evan.yu@maastrichtuniversity.nl)

ORCID: <https://orcid.org/0000-0001-7825-5087>

**&**

**Anke Wesselius, PhD**

Mailing address: Universiteitssingel 40 (Room C5.570), 6229 ER, Maastricht, the Netherlands.

Tel: +31 6 39014333

E-mail address: [anke.wesselius@maastrichtuniversity.nl](mailto:anke.wesselius@maastrichtuniversity.nl)

ORCID: <https://orcid.org/0000-0003-4474-9665>

**&**

**Qi-Rong Qin, PhD**

Mail address: 849 JiangDong Road, Yushan District, Ma'anshan 243011, China.

Tel: +86 0555-8367659; Fax: +86 0555-8367682

E-mail address: [qqr2022@163.com](mailto:qqr2022@163.com)

## 1 **Highlights (Key context and significance)**

- 2 • Distinct cancer risk patterns emerge based on smoking exposure beyond equal  
3 pack-years: smoking duration is a stronger risk factor for some cancers, while  
4 smoking intensity dominates for others.
- 5 • Time since smoking cessation (TSC) significantly lowers cancer risk: former smokers  
6 experience substantial reductions in risk for most cancers within the first 20 years  
7 after quitting.
- 8 • Favorable lifestyle behaviors mitigate cancer risks: individuals with regular physical  
9 activity and healthy diets show lower excess relative risks (ERRs) for most cancers,  
10 compared to those with unhealthy habits.
- 11 • Tailored cancer screening based on smoking behavior: cancer screening guidelines  
12 should consider detailed smoking patterns, including intensity, duration, and cessation,  
13 for more precise prevention strategies.

14 **Abstract**

15 Modelling complex smoking histories, with more comprehensive and flexible methods, to show  
16 what profile of smoking behavior is associated with the risk of different cancers remains poorly  
17 understood. This study aims to provide insight into the association between complex smoking  
18 exposure history and pan-cancer risk by modelling both smoking intensity and duration in a  
19 large-scale prospective cohort. Here, we used data including a total of 0.5 million with cancer  
20 incidences of 12 smoking-related cancers. To jointly interpret the effects of intensity and  
21 duration of smoking, we modelled excess relative risks (ERRs)/pack-year isolating the intensity  
22 effects for fixed total pack-years, thus enabling the smoking risk comparison for total exposure  
23 delivered at low intensity (for long duration) and at a high intensity (for short duration). The  
24 pattern observed from the ERR model indicated that for a fixed number of pack-years, low  
25 intensity/long duration or high intensity/short duration is associated with a different greater  
26 increase in cancer risk. Those findings were extended to an increase of time since smoking  
27 cessation (TSC) showing a reduction of ERR/pack-year for most cancers. Moreover,  
28 individuals with favorable lifestyle behaviors, such as regular physical activity and healthy  
29 dietary intakes, were shown to have lower ERRs/pack-year, compared to those with  
30 unfavorable lifestyle behaviors. Overall, this study systematically evaluates and demonstrates  
31 that for pan-cancer risks, smoking patterns are varied, while reducing exposure history to a  
32 single metric such as pack-years was too restrictive. Therefore, cancer screening guidelines  
33 should consider detailed smoking patterns, including intensity, duration, and cessation, for  
34 more precise prevention strategies.

35

36 **Keywords:** Pan-cancer; Complex smoking exposure; Excess relative risk; Prospective cohort



## 37 **Main Document**

### 38 **Introduction**

39 Smoking is an important modifiable risk factor for many cancers but a comprehensive  
40 demonstration of its influence by integration of intensity, duration, and cessation has been  
41 challenged<sup>1-3</sup>. While it is helpful to use a single estimate, such as cumulative exposure (e.g.,  
42 pack-years), to summarize smoking behavior, this approach may not adequately capture the  
43 complex relationship between smoking and cancer risk<sup>4,5</sup>. Although consensus has been  
44 reached that modelling pack-years is for the modifying effect of cigarette formulation or ways  
45 of targeting screening or other public health interventions, it may fail to reveal subtler  
46 phenomena that could shed light on mechanisms<sup>6</sup>. Several researchers have interpreted  
47 pack-years limited in making biologically credible models that provide unbiased information  
48 on complex smoking exposure histories<sup>7,8</sup> and circumvent multicollinearity issues<sup>9</sup>. Other  
49 time-related modifiers suggest that smoking has multiple effects in the carcinogenic process,  
50 thereby, the cooperation of flexible modelling, to capture more complex smoking and enhance  
51 the understanding of the molecular basis of smoking-related cancer risk, is needed.

52 To our knowledge, models that incorporate smoking duration and intensity typically examine  
53 these factors in isolation or various combinations, rather than integrating them  
54 comprehensively, which can be problematic for exploring complex patterns due to changing  
55 total exposures. That is, assessing the smoking risk for either increasing intensity or duration  
56 also incorporates the effect of increasing pack-years, leading to biased interpretation and  
57 comparison for the independent effect of intensity and duration.

58 To address all these dimensions together and uncover novel associations, modelling complex  
59 smoking exposures has emerged as a powerful tool<sup>8,10-15</sup>, offering insights that are often  
60 missed when examining factors separately, however, evidence specifically linking smoking to  
61 pan-cancers remains limited. Although research has long focused on modelling smoking

62 behavior effects, complex smoking exposure—encompassing the simultaneous consideration  
63 of smoking duration, intensity, and time since smoking cessation (TSC)—has not been  
64 systematically assessed concerning pan-cancers. Additionally, most previous studies have  
65 employed a case-control design, which is prone to recall bias and has limited capacity for  
66 causal inference. Evidence from large-scale prospective cohort studies remains sparse.  
67 Several studies have suggested that with equal pack-year, individuals who had smoked lower  
68 cigarettes/day for a longer duration have different cancer risks compared to those who smoked  
69 higher cigarettes/day for a shorter duration<sup>4,16-19</sup>. These studies have compared excess cancer  
70 risks/pack-year across smoking intensity categories, and recent models have incorporated  
71 TSC or stratified by age to account for exposure timing<sup>8,18,20</sup>. Cancer is a complex trait  
72 influenced by lifestyle behavior such as smoking, the inconsistency in findings from studies  
73 examining the association between cancer risk and complex smoking exposure highlights the  
74 need for further research. Therefore, this study aimed to investigate a more nuanced  
75 understanding between complex smoking exposure and smoking-related pan-cancer risks,  
76 based on our previous study<sup>19</sup> that modelled excess cancer risk per pack-year by intensity and  
77 TSC continuously and to interpret the various smoking effects, in a large-scale prospective  
78 cohort.

79

## 80 **Methods and Materials**

### 81 **Study Population**

82 Participants for this study were sourced from the UK Biobank, a large-scale prospective study  
83 that has been described in detail elsewhere<sup>21</sup>. In summary, the UK Biobank includes over  
84 500,000 individuals from the United Kingdom. Between 2006 and 2010, invitations were  
85 extended to all individuals aged 40 to 69 years living within 25 miles of a study center, from a  
86 total of 9.2 million invitations sent. Ultimately, 503,325 participants were recruited. Baseline

87 data collection involved extensive self-reporting through touchscreen tests, questionnaires, and  
88 nurse-led interviews, as well as anthropometric assessments and biological sample collection.  
89 Health records were supplemented with secondary care data from linked hospital episode  
90 statistics (HES).

91 For our analysis, the UK Biobank database initially included 502,505 participants. We  
92 excluded individuals with incomplete or inaccurate smoking status information and data  
93 regarding smoking intensity, duration, or pack-years (n=11,431) and those who withdraw with  
94 informed consent (n=12), resulting in a final analysis cohort of 491,062 participants (Figure 1).  
95 This study utilized the UK Biobank resource under Application #55889.

#### 96 **Ascertainment of Pan-Cancers**

97 In this study, “pan-cancers” refer to the 12 cancers associated with smoking as identified by the  
98 U.S. CDC<sup>22</sup>. These include: i) mouth and throat; ii) larynx; iii) esophagus; iv) lung, bronchus,  
99 and trachea; v) stomach; vi) liver; vii) kidney and renal pelvis; viii) pancreas; ix) colon and  
100 rectum; x) urinary bladder; xi) uterine cervix; and xii) acute myeloid leukemia. Cancer  
101 definitions within the UK Biobank are detailed in Table S1 and are based on International  
102 Classification of Diseases (ICD) codes (ICD-10 and ICD-9). For each cancer, participants who  
103 were diagnosed i) at or before baseline and ii) by self-report were excluded.

#### 104 **Statistical Analysis and Delivery Rate of Smoking Exposure**

105 We applied a statistical approach based on the models proposed by Vlaanderen et al.<sup>8</sup> and  
106 Lubin et al.<sup>18</sup>. Our smoking data included smoking status (never, current, former), pack-years  
107 (packs smoked per day × years as a smoker), cigarettes/day, and TSC. Participants were  
108 categorized into “never smokers”, “former smokers”, and “current smokers” based on their  
109 responses to questions about their smoking history. Those who answered “Yes” to smoking  
110 were classified as “current smokers”. Those who answered “No” and indicated a history of

111 smoking were categorized as “former smokers”, while those who neither currently smoked nor  
112 had ever smoked were considered “never smokers”<sup>21</sup>.

113 We tested for violations of survival time following a piecewise exponential distribution and  
114 confirmed that the data adhered to this model<sup>23</sup>. Consequently, we utilized a Poisson  
115 regression model for our analyses. The primary analysis adjusted for age (<45, 45–50, 50–60,  
116 60–65, ≥65 years), sex (men and women), body mass index (BMI, <18.5, 18.5–25, 25–30, ≥30,  
117 or unknown, kg/m<sup>2</sup>), ethnicity (White, Asian or Asian British, Black or Black British, Chinese,  
118 Mixed, other ethnic groups, or unknown), alcohol consumption status (never, former, current,  
119 or unknown), and socioeconomic status (SES, low, middle, high, or unknown), as main  
120 adjustment model. Detailed adjustments are provided in the Supplementary materials. In the  
121 UK Biobank, SES was derived from education level, household income, employment status,  
122 and the Townsend deprivation index. Education was categorized into various qualifications,  
123 while income and employment status were grouped accordingly. The Townsend deprivation  
124 index, an area-level SES measure from national census data, was recoded with higher values  
125 indicating higher SES<sup>24</sup>. The overall SES variable was computed by combining these factors  
126 and categorized into three groups based on tertiles (low, middle, high, or unknown).

127 To further examine the impact of lifestyle factors such as physical activity and diet on the  
128 relationship between smoking exposure and pan-cancer risk, we adjusted for regular physical  
129 activity and healthy dietary scores. Regular physical activity was defined as meeting the 2017  
130 UK Physical Activity Guidelines<sup>25</sup>. Dietary intake was assessed using the healthy dietary score  
131 by Lourida et al., based on the Mediterranean Diet and Heart-Healthy Dietary Priorities<sup>26,27</sup>.  
132 The score included factors such as servings of fruit, vegetables, fish, and meat, and was  
133 categorized into healthy (≥4) and unhealthy (<4) diets (Table S2 & S3). We also evaluated  
134 potential interactions between TSC and these lifestyle factors<sup>28,29</sup> to explore effect  
135 modification.

136 We fitted an exponential model to estimate the excess relative risk (ERRs) per pack-year by  
137 smoking intensity to investigate the independent effect of cigarette smoking duration and  
138 intensity of cigarette smoking on each cancer risk. In these models, the low intensity with long  
139 duration smokers was compared with the high intensity with short duration smokers based on  
140 equal pack-years. In addition, TSC was also taken into account, which can expand the scope of  
141 the research subjects from current smokers and never smokers to include all former smokers,  
142 and incorporating the time after smoking cessation into these models was estimated to better fit  
143 the data<sup>8</sup>.

144 We used the model as:

$$145 \quad RR(d) = 1 + \beta d * \exp(g_1(n) + g_2(t)),$$

146 Where  $\beta$  represents the ERR per pack-year at  $g_1=1$  and  $g_2=1$ ,  $g_1$  and  $g_2$  are the distribution of  
147 lifetime average intensity of cigarette smoking of all smokers and TSC of all former smokers,  
148 respectively.

149 We also selected the optimal transformation of  $g_1(n)$  and  $g_2(t)$  by comparing Akaike  
150 information criterion (AIC), with lower AIC indicating a better performance (Table S4).

151 Therefore, we fitted with different  $g_1(n)$  and  $g_2(t)$  transformations per cancer for relatively best  
152 performance as: i)  $G = \exp\{\mu_1 \ln(n) + \mu_2 \ln(t)\}$ ; ii)  $G = \exp\{\mu_1 \ln(n) + \mu_2 \ln(n)^2 + \mu_3 \ln(t) + \mu_4 \ln(t)^2\}$ ; iii)  
153  $G = \exp\{\mu_1 \ln(n) + \mu_2 n + \mu_3 \ln(t) + \mu_4 t\}$ ; iv)  $G = \exp\{\mu_1 n + \mu_2 n^2 + \mu_3 t + \mu_4 t^2\}$  where  $d$  represents the  
154 pack-years,  $n$  represents the cigarettes/day, and  $t$  represents the TSC (i.e., years).

155 The results from such models describe delivery rate patterns of exposure to smoking in relation  
156 to pan-cancer risk. The delivery rate is described by estimating how increasing intensity within  
157 a fixed number of pack-years influences each cancer risk. For example, an  
158 inverse-exposure-rate effect for intensity would mean that the ERR/pack-year (the strength of  
159 association) decreases with more cigarettes/day (and thus decreased duration) or the  
160 ERR/pack-year increases with lower cigarettes/day (and increased duration). Consequently,

161 for two individuals with equal total pack-years, a greater risk accrues to the individual smoking  
162 for lower intensity at a longer duration, or the individual smoking for higher intensity at a  
163 shorter duration.

164 The 95% confidence intervals (CIs) for ERR models were estimated through bootstrapping  
165 with 1,000 replications of the original data. To avoid overinterpreting regions with sparse data,  
166 we excluded predicted values below the 1st percentile (i.e., <2 cigarettes/day) and above the  
167 99th percentile (i.e., >60 cigarettes/day), applying the same method for TSC predictions.

168 Sensitivity analyses included: i) different adjustment models; ii) restricting the analysis to  
169 participants aged 50–74 years and regrouping former smokers who quit within 5 years as  
170 current smokers according to Lubin et al.<sup>18</sup>; iii) excluding non-White individuals; iv)  
171 excluding cancer cases occurring within the first year of follow-up; v) excluding never smokers  
172 who reported passive smoking; vi) adjusting for alcohol consumption frequency and volume  
173 (available only for current alcohol consumers in UK Biobank).

174 Statistical analyses were conducted using SAS 9.4 and R software (version 4.0.5). All tests  
175 were two-sided, with  $p$  values <0.05 considered statistically significant.

176

## 177 **Results**

### 178 **Population Characteristics**

179 Table 1 summarizes smoking data for 12 types of cancer in the UK Biobank, categorized by  
180 sex. Compared to non-cancer controls, individuals with cancer exhibited worse smoking  
181 behaviors across several metrics, including the proportion of ever smokers and current smokers,  
182 duration, intensity, and pack-years. This suggests that smoking is a significant risk factor for  
183 cancer development. This study includes 491,062 study participants with a median follow-up  
184 of 4.93 years for cancer cases and 11.63 years for non-cancer controls. Among cancer cases,  
185 42.36% were women, while 55.01% participants in the overall cohort were women. The mean

186 (SD) age at recruitment was 61.34 (6.64) years for cancer cases and 56.85 (8.10) for non-cancer  
187 controls (Table S5).

188 Among current smokers with cancer, 1,878 (75.00%) smoked more than 10 cigarettes/day and  
189 2,410 (96.25%) had a smoking duration longer than 20 years. This contrasts with non-cancer  
190 controls, who 22,363 (64.01%) smoked more than 10 cigarettes/day and 32,780 (93.83%) had a  
191 smoking duration longer than 20 years. For former smokers with cancer, 4,304 (82.31%)  
192 smoked more than 10 cigarettes/day, and 3,499 (66.92%) had a smoking duration longer than  
193 20 years. Additionally, 4,283 participants (81.91%) had a TSC greater than 5 years, compared  
194 to 92,874 participants (85.05%) in the non-cancer control group, respectively (Table S6).

### 195 **Complex Smoking Exposure in Relation to Different Patterns of Pan-cancer Risk**

196 Figure 2 illustrates the ERR/pack-year and 95% CI for various smoking intensities, considering  
197 smoking cessation (based on the main adjustment model). The curve shows different smoking  
198 delivery rate patterns with different cancer risks; as daily cigarette consumption increases  
199 (with reduced smoking duration), the ERRs/pack-year exhibit varying changes with either an  
200 increased pattern or a decreased pattern. This indicates that for equal pack-years, smoking for  
201 a longer duration (at lower cigarettes/day) is differentially associated with different cancer  
202 risks than smoking higher cigarettes/day (at a shorter duration). Specifically, we observed the  
203 lower cigarettes/day over a longer duration showing a greater ERR/pack-year of cancers than  
204 smoking more cigarettes/day over a shorter duration (e.g., cancer sites of the stomach, urinary  
205 bladder, pancreas, and kidney), while the higher cigarettes/day over a shorter duration showing  
206 a greater ERR/pack-year of cancers than smoking lower cigarettes/day over a longer duration  
207 (e.g., cancer sites of mouth and throat, esophagus, liver, uterine cervix, and acute myeloid  
208 leukemia). In cancer sites of the lung, bronchus and trachea, larynx, and colon and rectum, we  
209 observed an overall decreased ERRs/pack-year of smoking pattern with cigarettes/day  
210 increment, while an altered point at 40 cigarettes/day with slightly increased ERRs/pack-year

211 of smoking pattern after. The risk estimate remained essentially unchanged regardless of which  
212 other covariates were included in the models (Figure S1–S5).

213 By comparing the mean ERRs/pack-year, we found the association strength was varied across  
214 cancers, where some of them yielded an ERR/pack-year over 10% (e.g., cancer sites of lung,  
215 bronchus and trachea, larynx, and mouth) but others with an ERR/pack-year under 10% for  
216 lifetime smoking behavior when accounting for duration and intensity together (Figure S6).

### 217 **Excess Risk of Pan-Cancer in Relation to Time Since Smoking Cessation (TSC)**

218 We examined the association of smoking cessation with the ERR/pack-year for pan-cancers  
219 considering cumulative pack-years smoked, in an attempt to control for confounding by  
220 pack-years. After partitioning of the qualitative effect of former smoking, TSC was associated  
221 with significant risk reductions for 10 cancers, mostly reaching a relatively small  
222 ERR/pack-year in the first 20 years, in comparison with current and never smokers (Figure 3).

223 Again, the risk estimate remained essentially unchanged regardless of which other dimensions  
224 of smoking or covariates were included in the models (Figure S7–S11). By comparing the  
225 Min-Max ranges of the ERR/pack-year estimates, we found for the cancers showing decreased  
226 ERR/pack-year with TSC, the majority of them yielded a ratio of range of excess  
227 risk/maximum ERR over 50%, indicating quitting smoking could significantly reduce the  
228 ERRs when considering a more complex smoking exposure (Table S8).

229 However, we observed an inconsistent change of TSC pattern with ERRs/pack-year for colon  
230 and rectum cancer, the ERR/pack-year was increased within 7 years but decreased after; and  
231 for acute myeloid leukemia, the ERRs/pack-year still increased. Those findings suggest for  
232 some cancers, the reduction of cancer risk at the beginning of smoking cessation was temporal  
233 with a possible forward increased risk, which could be biased by other residual factors or due to  
234 statistical power. Particularly for colon and rectum cancer in line with a previous study that  
235 observed the cancer risk seems to increase for approximately 7 years after quitting smoking,



236 before gradually decreasing<sup>30</sup>. This may reflect that, in the early stages of smoking cessation,  
237 the health effects caused by prior smoking persist, which is especially true for individuals who  
238 have already sustained severe damage from smoking, referred to as “sick quitters”. Even after  
239 quitting, the damage may lead to a higher short-term cancer risk. As for acute myeloid  
240 leukemia, we hypothesized the harmful effect of smoking on the increase of cancer risk might  
241 not alleviate or diminish once quitting smoking, which is partially supported by a review  
242 showing quitting smoking did not stop the increased risks of acute myeloid leukemia at once<sup>31</sup>.

### 243 **Effect Modification by Major Lifestyle Factors**

244 In recent years, smoking, physical activity, and diet have been recognized as key lifestyle  
245 factors influencing human health, often contributing to the formulation of a healthy lifestyle  
246 score. In this study, we examined the impact of physical activity and diet—both separately and  
247 in combination—on the relationship between complex smoking exposure and pan-cancer risk.  
248 Our analysis revealed that neither physical activity nor diet significantly altered the  
249 associations between smoking intensity-duration patterns, TSC, and ERR/pack-year for  
250 various cancers (see Figure S12–S16 for smoking intensity and Figure S17–S21 for TSC).

251 However, given the modifiable impact of physical activity on smoking cessation, we further  
252 compared smoking patterns across stratified analyses of regular versus irregular physical  
253 activity. The patterns remained largely consistent with those observed in the main adjustment  
254 model, but we noted that for most cancers, ERR/pack-year with TSC was lower among  
255 individuals engaging in regular physical activity (Figure 4). Similarly, when stratifying by the  
256 healthy dietary score—healthy versus unhealthy—we found that smoking patterns were  
257 comparable to those in the main adjustment model. Notably, a healthy dietary score was  
258 associated with lower ERRs with increasing TSC compared to an unhealthy dietary score for  
259 most cancers (Figure 5). The differences between favorable and unfavorable lifestyle behaviors  
260 were statistically significant ( $p < 0.05$ ) as shown in Table S7. These findings suggest that

261 maintaining a healthy lifestyle significantly modifies the relationship between smoking and  
262 cancer risk. While most cancers showed reduced ERRs/pack-year with favorable lifestyle  
263 behaviors, a few cancers exhibited either similar or partially reduced ERRs/pack-year among  
264 those with less favorable behaviors.

### 265 **Sensitivity Analysis**

266 Wide confidence intervals observed in some patterns indicate that certain cancers or cigarette  
267 consumption levels had limited statistical power, which may account for the unexpected and  
268 inconsistent trends in ERRs/pack-year for some cancers. These confidence intervals should be  
269 interpreted with caution. For many cancers where a pattern could be discerned, there was  
270 considerable uncertainty in effect modification by smoking intensity per pack-year,  
271 particularly at very low or very high extremes, despite excluding <1% and >99% of extreme  
272 intensity values. Additionally, the cumulative effect of heavy smoking, such as exceeding 40  
273 cigarettes/day, should be considered, as there might be a potential plateau effect in smoking  
274 risk<sup>32</sup>. The bootstrapped 95% CIs show that the most reliable data is for individuals smoking  
275 between 10 and 40 cigarettes/day, as this range included the majority of smokers in the UK  
276 Biobank, making the shape of the curve most dependable within this interval.

277 Sensitivity analyses, incorporating physical activity and diet into the fully adjusted model,  
278 yielded consistent results regarding the association between complex smoking exposure and  
279 pan-cancer risks for both intensity-duration combinations per fixed pack-year and TSC (see  
280 Figure S12–S16 for smoking intensity and Figure S17–S21 for TSC). Comparisons across  
281 different adjusted models showed that the range of ERR estimates and patterns was consistent  
282 for most cancer sites, except for uterine cervix cancer and liver. This consistency reinforces the  
283 validity of our flexible ERR model<sup>32</sup>.

284

### 285 **Discussion**

## 286 **Complexity of Smoking Exposure and Cancer Risk**

287 Our study explored the intricate relationships between smoking exposure—encompassing  
288 duration, intensity, and time since cessation (TSC)—and cancer risk across various cancer  
289 types. We observed that different smoking patterns are associated with distinct cancer risks.  
290 Specifically, smoking lower cigarettes/day over a prolonged period was linked to a higher  
291 ERR/pack-year for cancers such as the stomach, urinary bladder, pancreas, and kidney.  
292 Conversely, higher daily smoking intensity over a shorter duration was more strongly  
293 associated with cancers like mouth and throat, esophagus, liver, uterine cervix, and acute  
294 myeloid leukemia. Furthermore, the impact of TSC on cancer risk varied: generally, longer  
295 cessation periods were associated with reduced risk, though some exceptions were noted.  
296 These findings underscore the complex interplay between smoking exposure patterns,  
297 cessation, and cancer risk.

## 298 **Challenges in Measuring Smoking Exposure and the Role of Duration**

299 Accurately measuring smoking intensity presents significant challenges due to fluctuations in  
300 the number of cigarettes smoked/day, variability in cigarette formulations, and differences in  
301 smoking behavior. Factors such as the depth of inhalation also influence exposure  
302 assessments, leading to potential misclassification biases and constrained reporting ranges,  
303 which can result in underestimation or overestimation of risk.  
304 These challenges are further complicated by the need to understand how smoking intensity,  
305 beyond merely counting cigarettes, affects carcinogenic processes through various molecular  
306 and biological pathways<sup>33</sup>. Increasing smoking intensity may activate or alter specific  
307 molecular mechanisms that influence cancer initiation and progression. Despite substantial  
308 research into smoking-induced carcinogenesis, the precise mechanisms linking complex  
309 exposure patterns—such as duration, intensity, and TSC—remain inadequately explored.

310 Current models that incorporate smoking duration, intensity, and TSC, while accounting for  
311 pack-years, offer a more comprehensive understanding of smoking's multifaceted nature.  
312 Our findings highlight the critical role of smoking duration in cancer risk. Extended smoking  
313 periods are associated with a greater accumulation of genetic and epigenetic changes<sup>34,35</sup>,  
314 particularly for cancers of the urinary bladder, kidney, pancreas, and stomach. While personal  
315 monitoring and biochemical assays provide valuable insights into smoking burden<sup>33</sup>, they  
316 face limitations related to feasibility and cost, especially for cancers with long latency periods.  
317 Reviews indicate that smoking-related cancer risks involve complex mechanisms, including  
318 nicotine stimulation, increased inflammation, and tobacco smoke particulates<sup>36</sup>, emphasizing  
319 the need for comprehensive exposure models to fully elucidate the impact of smoking  
320 duration, intensity, and overall exposure.

### 321 **The Role of Smoking Duration and Age in Cancer Risk**

322 Our study, consistent with other research (Table S9), confirms that smoking duration is a key  
323 determinant of cancer risk, particularly due to its nonlinear effects on long-term exposure to  
324 tobacco carcinogens. Longer smoking durations are associated with increased accumulation  
325 of genetic and epigenetic changes, which heighten cancer risk<sup>34,35</sup>. The role of age in  
326 carcinogenesis, particularly age at smoking onset, complicates the analysis of smoking  
327 duration's effects. Smoking duration and age of onset are strongly correlated (Spearman's  
328 correlation test  $p < 0.001$ ), and adjusting for age at enrolment and age of smoking onset did  
329 not significantly alter the relationship between smoking duration and cancer risk (data not  
330 shown). Moreover, the model incorporating smoking duration and age onset of smoking might  
331 cause colinear and overfitting, therefore, the close correlation of age at initiation, attained age,  
332 and the duration of smoking effectively prevents us from studying the effects of age at  
333 initiation in this study<sup>9,20</sup>. This suggests that while age at initiation and smoking duration are  
334 interrelated, their independent effects are challenging to disentangle.

335 Our findings reinforce the importance of smoking duration as a critical factor in cancer risk  
336 and support public health initiatives aimed at delaying smoking initiation among adolescents.  
337 Measures such as excise taxes and counter-advertising are crucial, though their effectiveness  
338 is limited by cancer's long latency period and the challenges in measuring exposure time<sup>20</sup>.  
339 Therefore, more effective smoking cessation programs are needed to encourage early  
340 quitting.

341 The inclusion of TSC in our study further emphasizes that quitting smoking reduces cancer  
342 risk, highlighting that it is never too late to quit. TSC is highly sensitive to smoking duration  
343<sup>37</sup>, underscoring the need to consider both duration and intensity in cancer risk assessments.  
344 While pack-years—combining cigarettes/day and smoking duration—provide a measure of  
345 total exposure, the relative contributions of smoking intensity versus duration to cancer  
346 development are not fully understood. Our study suggests that smoking duration is more  
347 reliably recalled and measured than daily cigarette intake, which is subject to fluctuation and  
348 less accurately quantified.

#### 349 **Smoking Intensity, Time Since Cessation (TSC), and Cancer Risk**

350 Our study underscores the significant impact of smoking intensity on cancer risk, particularly  
351 within the range of 10–40 cigarettes/day. This relationship is largely consistent across various  
352 cancer sites, reflecting how smoking intensity influences site-specific disease risk. Tobacco  
353 metabolites, including nicotine and cotinine, may contribute to local immunosuppressive  
354 effects and DNA damage, potentially explaining these patterns<sup>8,38</sup>. Increased  
355 ERRs/pack-year with higher smoking intensity could result from mechanisms such as  
356 metabolic saturation or enhanced DNA repair capacity<sup>8,38</sup>.

357 However, the relationship between smoking intensity and cancer risk is complex.  
358 High-intensity smoking (e.g., 40 or more cigarettes/day) is associated with higher risks for  
359 cancers such as those of the larynx and lung, but this relationship may diminish at very high

360 intensities due to misclassification or variations in inhalation patterns. Our study found that  
361 after excluding extreme values, intensity effects remained consistent but showed an increase  
362 above 40 cigarettes/day. Results for very low and high intensities should be interpreted with  
363 caution, necessitating further research.

364 TSC also plays a critical role in modifying cancer risk. Longer cessation periods generally  
365 reduce cancer risk, although exceptions exist. For example, the ERRs/pack-year of colon and  
366 rectal cancer increases within the first 7 years of quitting but may decrease afterward, while  
367 ERRs/pack-year of acute myeloid leukemia initially rises with longer cessation. This suggests  
368 that TSC interacts with other factors like age and genetic predisposition, affecting the overall  
369 cancer risk profile.

370 Previous research indicates that TSC is a major driver of cancer mortality, with varying  
371 sensitivity to smoking duration across different cancers<sup>8,39</sup>. For instance, lung cancer risk is  
372 more strongly associated with smoking duration than with average daily cigarette  
373 consumption. Our study highlights that TSC, in conjunction with smoking duration,  
374 significantly influences cancer risk profiles and has implications for tobacco cessation  
375 strategies. The pack-years measure, which combines cigarettes/day and smoking duration,  
376 may not fully capture the differential impacts of these factors on cancer risk.

377 Our analysis suggests that smoking intensity impacts cancer risk through biological processes  
378 such as reduced DNA repair capacity at lower intensities or increased detoxification enzyme  
379 induction at higher intensities<sup>4</sup>. However, patterns of intensity effects may also reflect  
380 nicotine satiation, where carcinogenic yield per cigarette decreases as smokers increase their  
381 daily intake to maintain addiction-sufficient nicotine levels. This makes the number of  
382 cigarettes/day an overestimate of internal exposure rates.

383 The use of pack-years has been criticized for assuming equal importance of smoking intensity  
384 and duration in determining cancer risk. Our study demonstrates that smoking duration or

385 intensity alone may be more strongly associated with cancer risk than the other. Relying  
386 solely on pack-years for screening may exclude high-risk individuals who smoke less  
387 intensely. Thus, a paradigm shift is needed in assessing and reporting smoking history,  
388 incorporating both smoking duration and intensity into more complex exposure models for  
389 cancer susceptibility screening.

390 Additionally, our study explored interactions between smoking and other factors such as  
391 physical activity and diet. These interactions significantly modify the association between  
392 smoking and cancer risk, emphasizing the need for integrated models that account for these  
393 variables.

#### 394 **Strengths and Limitations of the Excess Relative Risk (ERR) Model**

395 We applied a model similar to the ERR model described by Vlaanderen et al.<sup>8</sup>, building on  
396 earlier approaches such as the Lubin et al.<sup>18</sup> model. Our model effectively addresses  
397 differences in risk between low intensity/long duration and high intensity/short duration  
398 smokers with equal pack-years.

399 Major strengths of our study include a large-sample-size and well-characterized cohort of  
400 current and former smokers with diverse ethnicity and sex, and valid information on smoking  
401 and potential confounding factors. The prospective design minimized recall bias and allowed  
402 us to examine multiple smoking exposures within the same cohort. Our consistent findings  
403 across different cancer types, with multiple adjustments and sensitivity analyses, strengthen  
404 our inferences and reduce the likelihood of chance or differential reporting.

405 However, our study also has limitations. While the ERR model provides detailed insights,  
406 some factors remain unaccounted for. For example, more vigorous inhalation patterns, often  
407 associated with lighter smokers, could confound risk estimates between heavy and light  
408 smokers. Discrepancies between self-reported and objective information were more likely  
409 among long-term heavy smokers, potentially leading to inaccurate reporting of intensity and

410 duration. Variations in smoking intensity over time and by age of exposure, along with  
411 reliance on self-reported averages, may introduce errors in exposure measurements. To  
412 mitigate information bias and residual confounding, we excluded extreme values of cigarette  
413 intensity and duration from our analysis. Due to some missing data and the criteria for the  
414 definition of favorable/unfavorable on physical activity and dietary intakes, the stratified  
415 findings and interpretations upon the impact of lifestyle behaviors on excess risks when  
416 quitting smoking should be with caution. Even though competing risks could influence results  
417 since smokers were more likely to be selectively removed from follow-up due to other diseases,  
418 a previous study<sup>14</sup> has demonstrated when assessing complex smoking history, competing  
419 risks had no appreciable effect on the results which suggested that any potential bias from  
420 competing risk considerations was minimal. Finally, the observational nature of our study  
421 limits our ability to uncover molecular or biological pathways linking smoking exposure to  
422 cancer development, though it highlights important associations that warrant further research.

423

## 424 **Conclusions**

425 In summary, our study reveals qualitative and quantitative differences in the association  
426 between dimensions of smoking exposure and cancer risk. Smoking duration was a strong  
427 determinant of cancers such as the urinary bladder, kidney, pancreas, and stomach, while  
428 intensity was a major determinant for cancers including the esophagus, liver, and acute  
429 myeloid leukemia, with significant nonlinear dose effects observed. TSC emerged as an  
430 independent and dominant predictor for most cancers, emphasizing the substantial benefits of  
431 quitting smoking in reducing cancer risk, regardless of the duration or intensity of smoking.



432 **Conflicts of Interest Statement**

433 The authors declare no conflicts of interest.

434 **Ethics Approval**

435 The studies involving human participants were reviewed and approved by NHS National  
436 Research Ethics Service North West (11/NW/0382). The patients/participants provided their  
437 written informed consent to participate in this study.

438 **Author Contributions**

439 Conceptualization, E.Y.W.Y. and A.W.; Data curation, W.L., Y.C. and B.T.; Formal analysis,  
440 W.L. and Y.C.; Funding acquisition, E.Y.W.Y. and Q.Q.; Investigation, W.L. and Y.C.;  
441 Methodology, W.L., Y.C. and E.Y.W.Y.; Project administration, E.Y.W.Y. and A.W.;  
442 Resources, E.Y.W.Y. and A.W.; Supervision, E.Y.W.Y. and A.W.; Writing-original draft,  
443 W.L., A.W. and E.Y.W.Y.; and all other authors Y.L., Y.X.Z., H.Y.R., Y.T.Z., Y.P.F.,  
444 M.H.L., Y.X.S., S.Y.W., B.W.C., F.O., MP.Z. Q.Q. reviewed, edited the writing, and  
445 approved the final manuscript.

446 **Acknowledgments**

447 This study was supported by the: National Natural Science Foundation of China (NSFC,  
448 82204033); the Natural Science Foundation of Jiangsu grant (BK2022020826); Fundamental  
449 Research Funds for the Central Universities of China (2242022R10062/3225002202A1),  
450 Medical Foundation of Southeast University (4060692202/021), Zhishan Young Scholar  
451 Award at the Southeast University (2242023R40031); The Scientific Research Project for  
452 Health Commission of Anhui Province (AHWJ2023A20172; AHWJ2023BAa20055). The  
453 funders had no role in the study design, data collection, decision to publish, or preparation of  
454 the manuscript. The authors acknowledge certain figures (Figure 1) were created, adapted, and  
455 exported from BioRender.com (2024).  
456 Retrieved from <https://app.biorender.com/biorender-templates>.

457 **Data and Code Availability Statement**

458 This work has been conducted using the UK Biobank Resource under Application Number  
459 55889. The UK Biobank is an open-access resource and bona fide researchers can apply to  
460 use the UK Biobank dataset by registering and applying at  
461 <http://ukbiobank.ac.uk/register-apply/>. Further information and the key codes for analysis in  
462 this study are available from the corresponding author upon request.

463 **Abbreviations:** ERR, excess relative risk; BMI, body mass index; TSC, time since smoking  
464 cessation.

## References

1. Siegel RL, Giaquinto AN, Jemal A. Cancer statistics, 2024. *CA Cancer J Clin* 2024; **74**(1): 12-49.
2. Thomson B, Emberson J, Lacey B, Lewington S, Peto R, Islami F. Association of Smoking Initiation and Cessation Across the Life Course and Cancer Mortality: Prospective Study of 410 000 US Adults. *JAMA Oncol* 2021; **7**(12): 1901-3.
3. Tu H, Ye Y, Huang M, et al. Smoking, smoking cessation, and survival after cancer diagnosis in 128,423 patients across cancer types. *Cancer Commun (Lond)* 2022; **42**(12): 1421-4.
4. Lubin JH, Alavanja MC, Caporaso N, et al. Cigarette smoking and cancer risk: modeling total exposure and intensity. *Am J Epidemiol* 2007; **166**(4): 479-89.
5. Potter AL, Xu NN, Senthil P, et al. Pack-Year Smoking History: An Inadequate and Biased Measure to Determine Lung Cancer Screening Eligibility. *J Clin Oncol* 2024; **42**(17): 2026-37.
6. de Vocht F, Burstyn I, Sanguanchaiyakrit N. Rethinking cumulative exposure in epidemiology, again. *Journal of exposure science & environmental epidemiology* 2015; **25**(5): 467-73.
7. Thomas DC. Invited commentary: is it time to retire the "pack-years" variable? Maybe not! *Am J Epidemiol* 2014; **179**(3): 299-302.
8. Vlaanderen J, Portengen L, Schüz J, et al. Effect modification of the association of cumulative exposure and cancer risk by intensity of exposure and time since exposure cessation: a flexible method applied to cigarette smoking and lung cancer in the SYNERGY Study. *Am J Epidemiol* 2014; **179**(3): 290-8.
9. Leffondré K, Abrahamowicz M, Siemiatycki J, Rachet B. Modeling smoking history: a comparison of different approaches. *Am J Epidemiol* 2002; **156**(9): 813-23.
10. Baris D, Karagas MR, Verrill C, et al. A Case–Control Study of Smoking and Bladder Cancer Risk: Emergent Patterns Over Time. *JNCI: Journal of the National Cancer Institute* 2009; **101**(22): 1553-61.
11. Lubin JH, Virtamo J, Weinstein SJ, Albanes D. Cigarette Smoking and Cancer: Intensity Patterns in the Alpha-Tocopherol, Beta-Carotene Cancer Prevention Study in Finnish Men. *American Journal of Epidemiology* 2008; **167**(8): 970-5.
12. Lubin JH, Caporaso NE. Cigarette Smoking and Lung Cancer: Modeling Total Exposure and Intensity. *Cancer Epidemiology, Biomarkers & Prevention* 2006; **15**(3): 517-23.
13. Lubin JH, Gaudet MM, Olshan AF, et al. Body mass index, cigarette smoking, and alcohol consumption and cancers of the oral cavity, pharynx, and larynx: modeling odds ratios in pooled case-control data. *Am J Epidemiol* 2010; **171**(12): 1250-61.
14. Lubin JH, Couper D, Lutsey PL, Woodward M, Yatsuya H, Huxley RR. Risk of Cardiovascular Disease from Cumulative Cigarette Use and the Impact of Smoking Intensity. *Epidemiology* 2016; **27**(3): 395-404.

15. Lubin JH, Caporaso N, Wichmann HE, Schaffrath-Rosario A, Alavanja MC. Cigarette smoking and lung cancer: modeling effect modification of total exposure and intensity. *Epidemiology* 2007; **18**(5): 639-48.
16. Baris D, Karagas MR, Verrill C, et al. A case-control study of smoking and bladder cancer risk: emergent patterns over time. *J Natl Cancer Inst* 2009; **101**(22): 1553-61.
17. Lubin JH, Kogevinas M, Silverman D, et al. Evidence for an intensity-dependent interaction of NAT2 acetylation genotype and cigarette smoking in the Spanish Bladder Cancer Study. *Int J Epidemiol* 2007; **36**(1): 236-41.
18. Lubin JH, Caporaso NE. Cigarette smoking and lung cancer: modeling total exposure and intensity. *Cancer Epidemiol Biomarkers Prev* 2006; **15**(3): 517-23.
19. van Osch FHM, Vlaanderen J, Jochems SHJ, et al. Modeling the Complex Exposure History of Smoking in Predicting Bladder Cancer: A Pooled Analysis of 15 Case-Control Studies. *Epidemiology* 2019; **30**(3): 458-65.
20. Flanders WD, Lally CA, Zhu BP, Henley SJ, Thun MJ. Lung cancer mortality in relation to age, duration of smoking, and daily cigarette consumption: results from Cancer Prevention Study II. *Cancer Res* 2003; **63**(19): 6556-62.
21. Sudlow C, Gallacher J, Allen N, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS medicine* 2015; **12**(3): e1001779.
22. PREVENTION USCFDCA. Health Effects of Cigarettes: Cancer. 2024. <https://www.cdc.gov/tobacco/about/cigarettes-and-cancer.html>.
23. Sedgwick P. Poisson regression. *BMJ (Clinical research ed)* 2014; **349**: g6150.
24. Townsend P, Phillimore P, Beattie A. Health and deprivation: inequality and the North: Routledge; 2023.
25. biobank U. 2017 UK Physical activity guidelines. 2017. <https://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=22036>.
26. Mozaffarian D. Dietary and Policy Priorities for Cardiovascular Disease, Diabetes, and Obesity: A Comprehensive Review. *Circulation* 2016; **133**(2): 187-225.
27. Lourida I, Hannon E, Littlejohns TJ, et al. Association of Lifestyle and Genetic Risk With Incidence of Dementia. *Jama* 2019; **322**(5): 430-7.
28. Ussher MH, Faulkner GEJ, Angus K, Hartmann-Boyce J, Taylor AH. Exercise interventions for smoking cessation. *Cochrane Database Syst Rev* 2019; **2019**(10).
29. Patriota P, Guessous I, Marques-Vidal P. No changes in dietary intake after quitting smoking; a prospective study in Switzerland. *BMC Nutr* 2021; **7**(1): 34.
30. Park E, Kang HY, Lim MK, Kim B, Oh JK. Cancer Risk Following Smoking Cessation in Korea. *JAMA Netw Open* 2024; **7**(2): e2354958.

31. Colamesta V, D'Aguanno S, Breccia M, Bruffa S, Cartoni C, La Torre G. Do the smoking intensity and duration, the years since quitting, the methodological quality and the year of publication of the studies affect the results of the meta-analysis on cigarette smoking and Acute Myeloid Leukemia (AML) in adults? *Crit Rev Oncol Hematol* 2016; **99**: 376-88.
32. Krebs NM, Chen A, Zhu J, et al. Comparison of Puff Volume With Cigarettes per Day in Predicting Nicotine Uptake Among Daily Smokers. *Am J Epidemiol* 2016; **184**(1): 48-57.
33. Etter JF, Perneger TV. Measurement of self reported active exposure to cigarette smoke. *J Epidemiol Community Health* 2001; **55**(9): 674-80.
34. Sugden K, Hannon EJ, Arseneault L, et al. Establishing a generalized polyepigenetic biomarker for tobacco smoking. *Transl Psychiatry* 2019; **9**(1): 92.
35. Hecht SS. Tobacco carcinogens, their biomarkers and tobacco-induced cancer. *Nat Rev Cancer* 2003; **3**(10): 733-44.
36. Liu Y, Lu L, Yang H, et al. Dysregulation of immunity by cigarette smoking promotes inflammation and cancer: A review. *Environ Pollut* 2023; **339**: 122730.
37. Lee JJW, Kunaratnam V, Kim CJH, et al. Cigarette smoking cessation, duration of smoking abstinence, and head and neck squamous cell carcinoma prognosis. *Cancer* 2023; **129**(6): 867-77.
38. Hashibe M, Brennan P, Chuang S-c, et al. Interaction between tobacco and alcohol use and the risk of head and neck cancer: pooled analysis in the International Head and Neck Cancer Epidemiology Consortium. *Cancer Epidemiology Biomarkers & Prevention* 2009; **18**(2): 541-50.
39. Lai H, Liu Q, Ye Q, et al. Impact of smoking cessation duration on lung cancer mortality: A systematic review and meta-analysis. *Crit Rev Oncol Hematol* 2024; **196**: 104323.

## Figure Legends

### Figure 1. Study overview and workflow

This figure illustrates the model formulas and flow chart of participants used in the study, aimed at analyzing the relationship between different smoking patterns and 12 smoking related cancers in the UK Biobank, which includes 491,062 participants into analysis. The diagram in the middle shows two scenarios: high-intensity, low-duration smoking, and low-intensity, high-duration smoking under the same total pack-years. Therefore, we applied a flexible excess relative risk (ERR) model to explore the association of exposure intensity/duration and cancer risk simultaneously. This model isolates the intensity effects while holding total pack-years constant, allowing for the comparison of ERRs for total exposure delivered at lower intensity (for longer duration) and at higher intensity (for shorter duration) across the 12 cancers. Additionally, the model adjusts for the ERR per pack-year after smoking cessation. The figure on the right presents the 12 cancers included in the analysis for which smoking is explicitly identified as a risk factor according to the WHO and US CDC.

**Abbreviations:** ERR, excess relative risk.

### Figure 2. The excess relative risk (ERR) for pan-cancers per pack-year of smoking by smoking intensity.

Bootstrapped 95% confidence intervals are based on 1,000 replications. The blue solid line represents the prediction from the function  $g_1(n)$  in the model, with the area showing 95% confidence interval obtained from bootstrap analysis.

**Model adjustments:** age (<45, 45–50, 50–60, 60–65,  $\geq 65$ , years), sex (men or women), ethnicity (White, Asian or Asian British, Black or Black British, Chinese, Mixed, other ethnic group, or unknown), BMI (<18.5, 18.5–25, 25–30,  $\geq 30$ , or unknown,  $\text{kg}/\text{m}^2$ ), socioeconomic status (low, medium, high, or unknown), and alcohol consumption status (never, previous, current, or unknown).

**Abbreviations:** ERR, excess relative risk; BMI, body mass index.

### Figure 3. The excess relative risk (ERR) for pan-cancers per pack-year of smoking by time since smoking cessation.

Bootstrapped 95% confidence intervals are based on 1,000 replications. The blue solid line represents the prediction from the function  $g_2(t)$  in the model, with the area showing 95% confidence interval obtained from bootstrap analysis.

**Model adjustments:** age (<45, 45–50, 50–60, 60–65,  $\geq 65$ , years), sex (men or women), ethnicity (White, Asian or Asian British, Black or Black British, Chinese, Mixed, other ethnic group, or unknown), BMI (<18.5, 18.5–25, 25–30,  $\geq 30$ , or unknown,  $\text{kg}/\text{m}^2$ ), socioeconomic status (low, medium, high, or unknown), and alcohol consumption status (never, previous, current, or unknown).

**Abbreviations:** ERR, excess relative risk; BMI, body mass index.

**Figure 4.** The excess relative risk (ERR) for pan-cancers per pack-year of smoking by time since smoking cessation within categories of physical activity.

Bootstrapped 95% confidence intervals are based on 1,000 replications. The blue solid line represents the prediction from the function  $g_2(t)$  in the model for individuals who engage in regular physical activity, with the area showing 95% confidence interval obtained from bootstrap analysis.

**Model adjustments:** age (<45, 45–50, 50–60, 60–65,  $\geq 65$ , years), sex (men or women), ethnicity (White, Asian or Asian British, Black or Black British, Chinese, Mixed, other ethnic group, or unknown), BMI (<18.5, 18.5–25, 25–30,  $\geq 30$ , or unknown,  $\text{kg}/\text{m}^2$ ), socioeconomic status (low, medium, high, or unknown), alcohol consumption status (never, previous, current, or unknown).

**Abbreviations:** ERR, excess relative risk; BMI, body mass index.

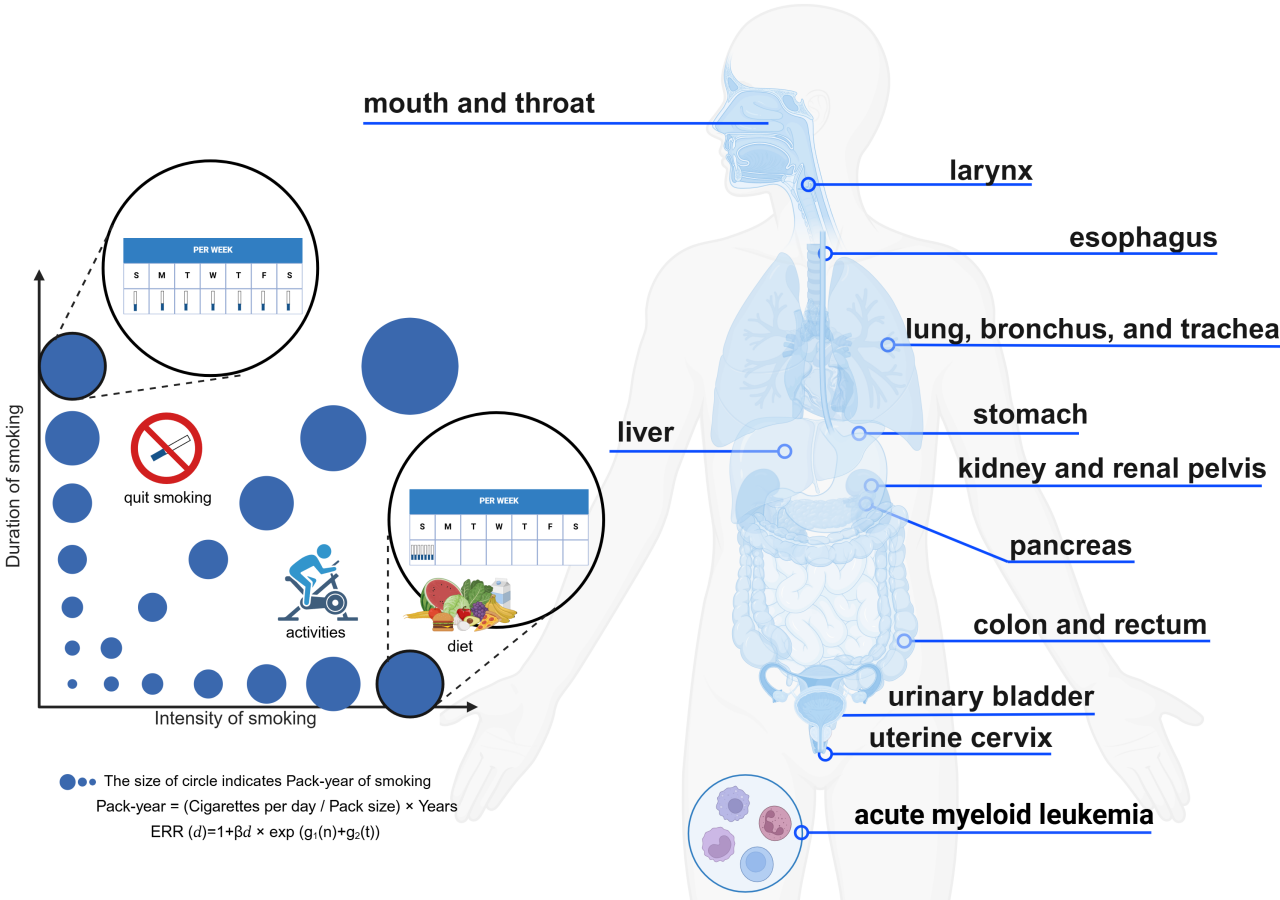
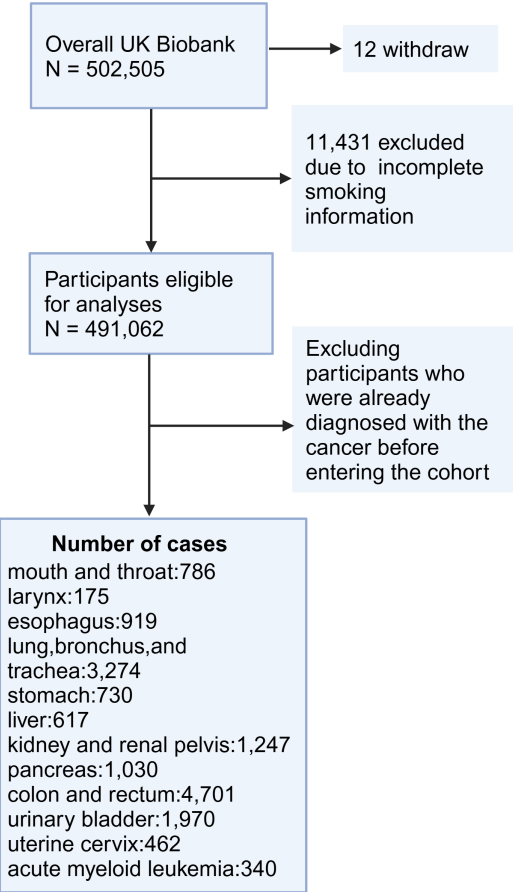
**Figure 5.** The excess relative risk (ERR) for pan-cancers per pack-year of smoking by time since smoking cessation within categories of a healthy diet.

Bootstrapped 95% confidence intervals are based on 1,000 replications. The blue solid line represents the prediction from the function  $g_2(t)$  in the model for individuals who engage in a healthy diet, while the dashed black line corresponds to those with an unhealthy diet.

**Model adjustments:** age (<45, 45–50, 50–60, 60–65,  $\geq 65$ , years), sex (men or women), ethnicity (White, Asian or Asian British, Black or Black British, Chinese, Mixed, other ethnic group, or unknown), BMI (<18.5, 18.5–25, 25–30,  $\geq 30$ , or unknown,  $\text{kg}/\text{m}^2$ ), socioeconomic status (low, medium, high, or unknown), alcohol consumption status (never, previous, current, or unknown).

**Abbreviations:** ERR, excess relative risk; BMI, body mass index.

# Different smoking pattern related to different cancer risk pattern

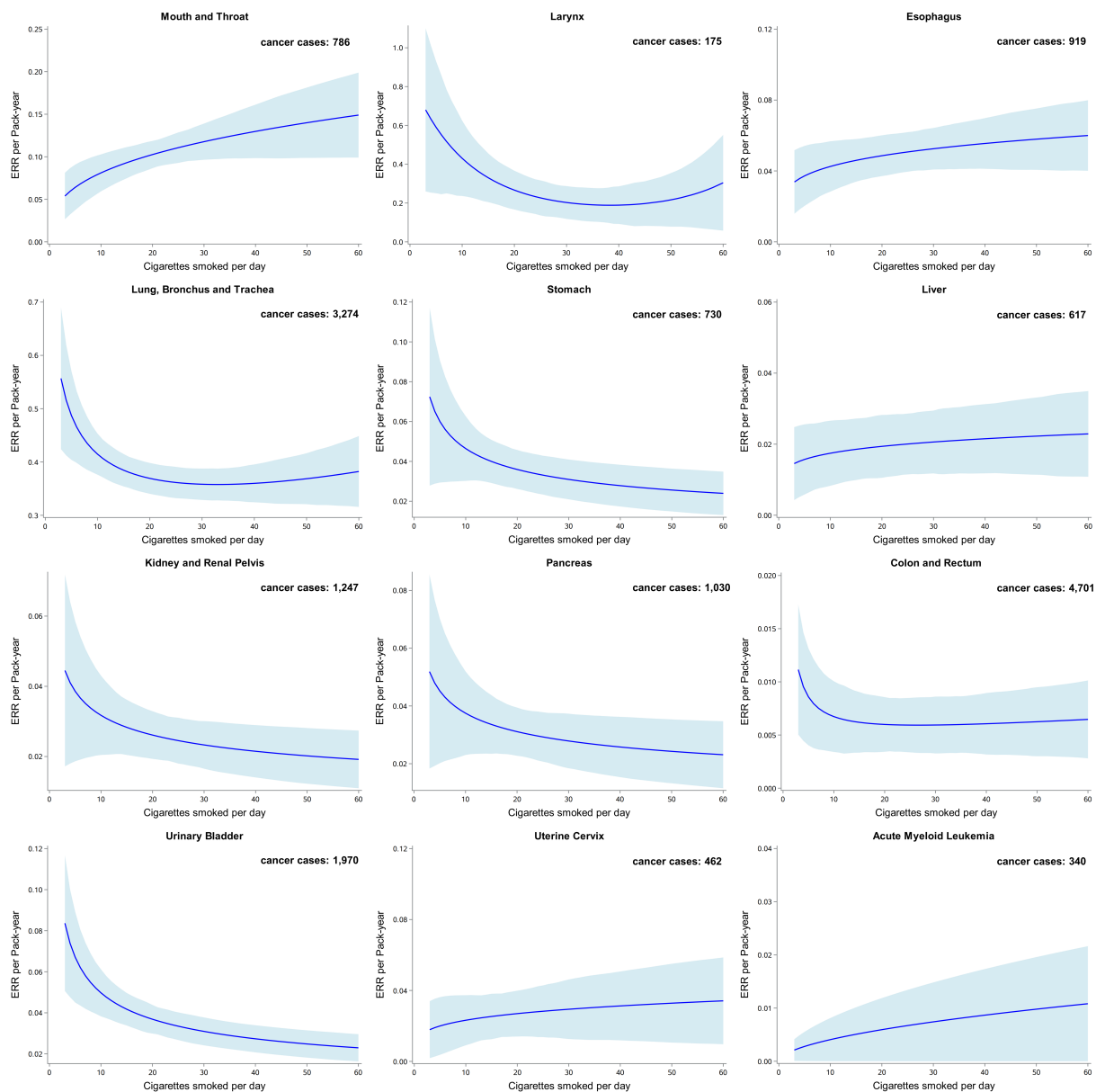




**Figure 1.** Study overview and workflow

This figure illustrates the model formulas and flow chart of participants used in the study, aimed at analyzing the relationship between different smoking patterns and 12 smoking related cancers in the UK Biobank, which includes 491,062 participants into analysis. The diagram in the middle shows two scenarios: high-intensity, low-duration smoking and low-intensity, high-duration smoking under the same total pack-years. Therefore, we applied a flexible excess relative risk (ERR) model to explore the association of exposure intensity/duration and cancer risk simultaneously. This model isolates the intensity effects while holding total pack-years constant, allowing for the comparison of ERRs for total exposure delivered at lower intensity (for longer duration) and at higher intensity (for shorter duration) across the 12 cancers. Additionally, the model adjusts for the ERR per pack-year after smoking cessation. The figure on the right presents the 12 cancers included in the analysis for which smoking is explicitly identified as a risk factor according to the WHO and US CDC.

**Abbreviations:** ERR, excess relative risk.

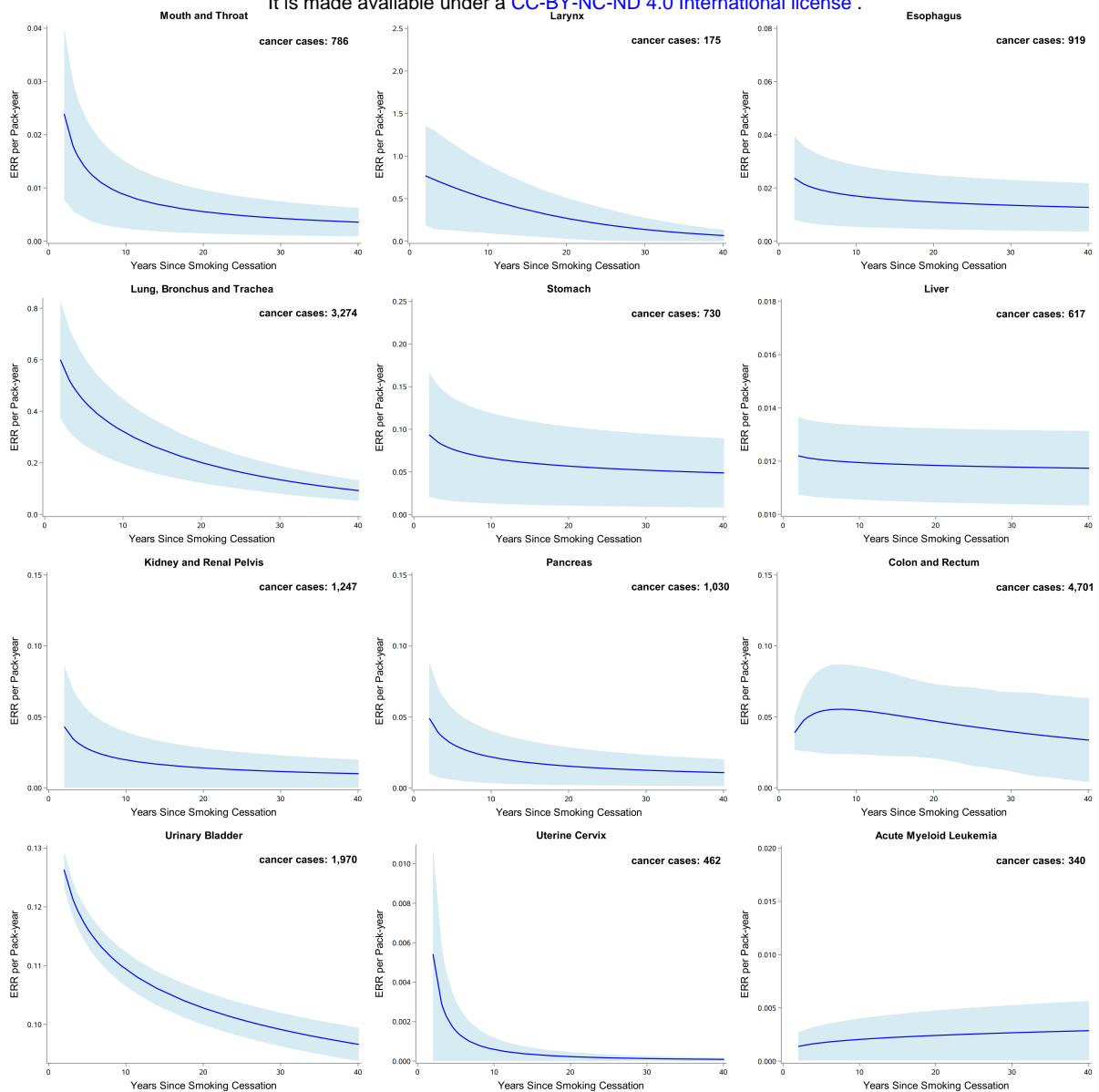


**Figure 2.** The excess relative risk (ERR) for pan-cancers per pack-year of smoking by smoking intensity.

Bootstrapped 95% confidence intervals are based on 1,000 replications. The blue solid line represents the prediction from the function  $g_1(n)$  in the model, with the area showing 95% confidence interval obtained from bootstrap analysis.

**Model adjustments:** age (<45, 45–50, 50–60, 60–65, ≥65, years), sex (men or women), ethnicity (White, Asian or Asian British, Black or Black British, Chinese, Mixed, other ethnic group, or unknown), BMI (<18.5, 18.5–25, 25–30, ≤30, or unknown, kg/m<sup>2</sup>), socioeconomic status (low, medium, high, or unknown), and alcohol consumption status (never, previous, current, or unknown).

**Abbreviations:** ERR, excess relative risk; BMI, body mass index.

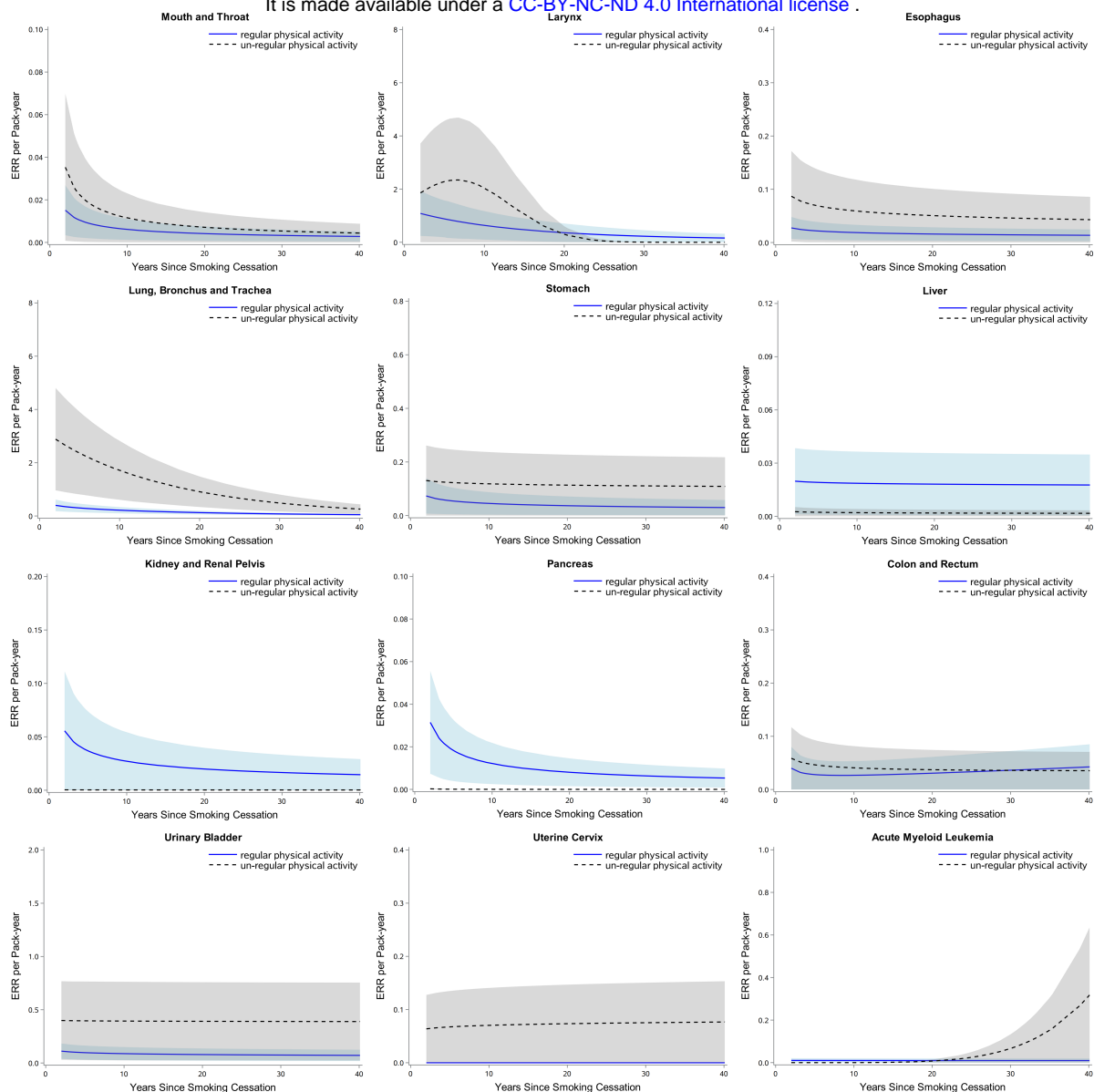


**Figure 3.** The excess relative risk (ERR) for pan-cancers per pack-year of smoking by time since smoking cessation.

Bootstrapped 95% confidence intervals are based on 1,000 replications. The blue solid line represents the prediction from the function  $g_2(t)$  in the model, with the area showing 95% confidence interval obtained from bootstrap analysis.

**Model adjustments:** age (<45, 45–50, 50–60, 60–65,  $\geq 65$ , years), sex (men or women), ethnicity (White, Asian or Asian British, Black or Black British, Chinese, Mixed, other ethnic group, or unknown), BMI (<18.5, 18.5–25, 25–30,  $\geq 30$ , or unknown,  $\text{kg/m}^2$ ), socioeconomic status (low, medium, high, or unknown), and alcohol consumption status (never, previous, current, or unknown).

**Abbreviations:** ERR, excess relative risk; BMI, body mass index.

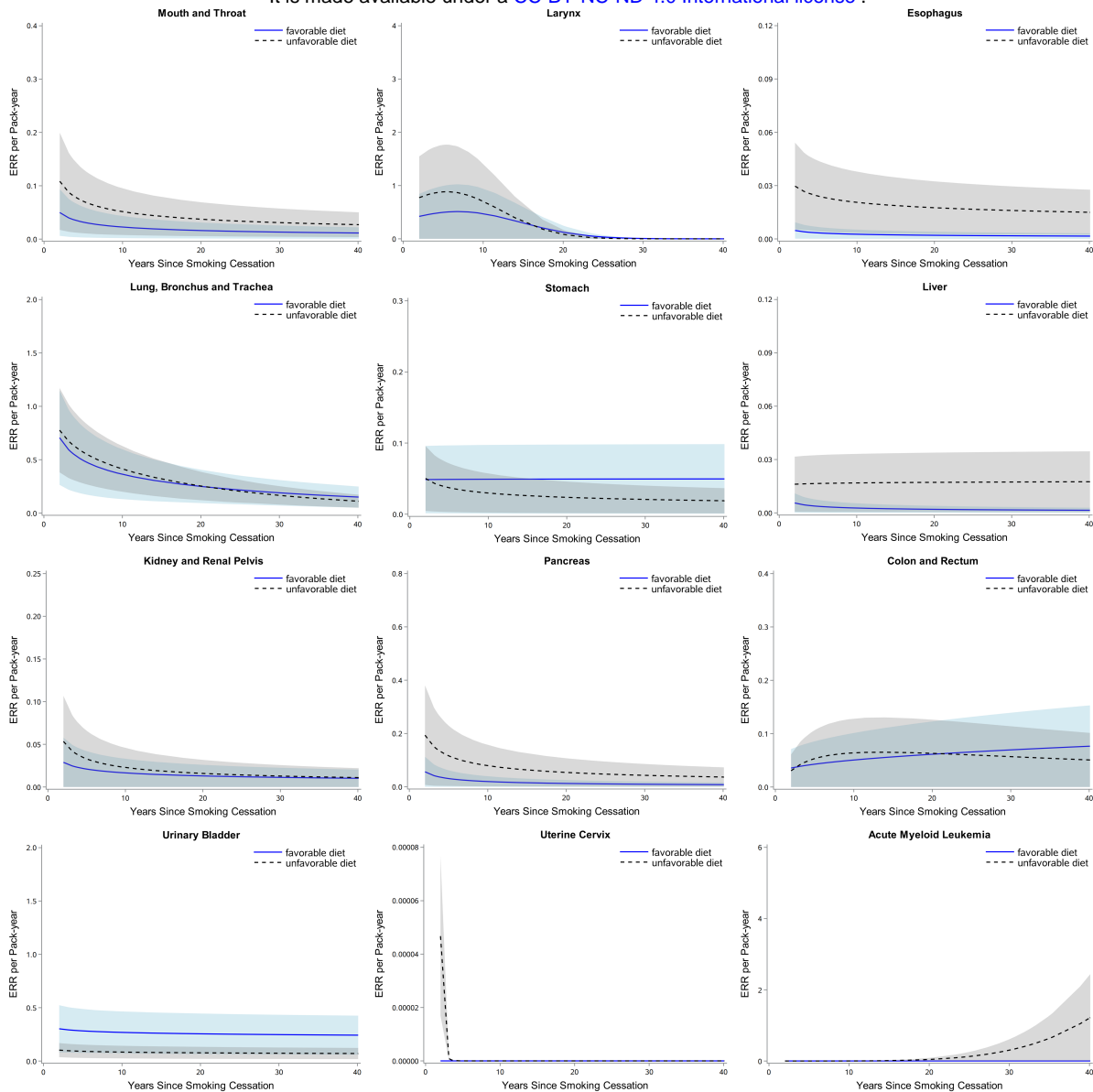


**Figure 4.** The excess relative risk (ERR) for pan-cancers per pack-year of smoking by time since smoking cessation within categories of physical activity.

Bootstrapped 95% confidence intervals are based on 1,000 replications. The blue solid line represents the prediction from the function  $g_2(t)$  in the model for individuals who engage in regular physical activity, with the area showing 95% confidence interval obtained from bootstrap analysis.

**Model adjustments:** age (<45, 45–50, 50–60, 60–65,  $\geq 65$ , years), sex (men or women), ethnicity (White, Asian or Asian British, Black or Black British, Chinese, Mixed, other ethnic group, or unknown), BMI (<18.5, 18.5–25, 25–30,  $\geq 30$ , or unknown,  $\text{kg/m}^2$ ), socioeconomic status (low, medium, high, or unknown), alcohol consumption status (never, previous, current, or unknown).

**Abbreviations:** ERR, excess relative risk; BMI, body mass index.



**Figure 5.** The excess relative risk (ERR) for pan-cancers per pack-year of smoking by time since smoking cessation within categories of a healthy diet.

Bootstrapped 95% confidence intervals are based on 1,000 replications. The blue solid line represents the prediction from the function  $g_2(t)$  in the model for individuals who engage in a healthy diet, while the dashed black line corresponds to those with an unhealthy diet.

**Model adjustments:** age (<45, 45–50, 50–60, 60–65,  $\geq 65$ , years), sex (men or women), ethnicity (White, Asian or Asian British, Black or Black British, Chinese, Mixed, other ethnic group, or unknown), BMI (<18.5, 18.5–25, 25–30,  $\geq 30$ , or unknown,  $\text{kg/m}^2$ ), socioeconomic status (low, medium, high, or unknown), alcohol consumption status (never, previous, current, or unknown).

**Abbreviations:** ERR, excess relative risk; BMI, body mass index.

**Table 1.** Summary of smoking data\* for 12 cancers in the UK Biobank by sex

	%Ever Smoked	%Current Smokers	Among Smokers		
			Duration (Mean No. Years)	Intensity (Mean No. Cigarettes/d)	Mean Pack-Years
<b>Mouth and Throat</b>					
Men					
Cases (n=506)	58.7	29.1	19.7	13.7	23.1
Non-cases (n=220,057)	35.8	8.6	9.3	7.2	9.4
Women					
Cases (n=280)	45.4	17.9	14.0	8.2	12.1
Non-cases (n=269,630)	26.8	6.8	6.9	4.3	5.5
<b>Larynx</b>					
Men					
Cases (n=154)	79.2	34.4	29.3	17.5	32.4
Non-cases (n=220,462)	35.8	8.7	9.3	7.2	9.4
Women					
Cases (n=21)	81.0	47.6	29.5	13.3	23.6
Non-cases (n=270,059)	26.9	6.8	6.9	4.3	5.5
<b>Esophagus</b>					
Men					
Cases (n=677)	58.8	15.2	18.2	14.0	21.6
Non-cases (n=220,083)	35.8	8.7	9.3	7.2	9.4
Women					
Cases (n=242)	41.3	16.5	13.8	8.0	13.3
Non-cases (n=269,854)	26.8	6.8	6.9	4.3	5.5
<b>Lung, Bronchus and Trachea</b>					
Men					
Cases (n=1,681)	81.3	37.6	31.3	18.6	35.3
Non-cases (n=219,004)	35.5	8.5	9.2	7.2	9.2
Women					
Cases (n=1,593)	70.1	32.6	26.3	13.2	24.5
Non-cases (n=268,359)	26.6	6.6	6.7	4.2	5.3
<b>Stomach</b>					
Men					
Cases (n=513)	54.4	15.0	17.2	12.2	19.3
Non-cases (n=220,232)	35.8	8.7	9.3	7.2	9.4
Women					
Cases (n=217)	37.8	11.5	11.5	6.5	9.5
Non-cases (n=269,828)	26.8	6.8	6.9	4.3	5.5
<b>Liver</b>					
Men					
Cases (n=351)	49.9	13.1	15.4	11.5	17.5
Non-cases (n=220,444)	35.9	8.7	9.3	7.2	9.4
Women					
Cases (n=266)	36.5	7.5	10.2	6.9	9.4
Non-cases (n=269,802)	26.9	6.8	6.9	4.3	5.5
<b>Kidney and Renal Pelvis</b>					
Men					
Cases (n=787)	46.9	12.2	13.9	10.0	14.9
Non-cases (n=219,757)	35.8	8.7	9.3	7.2	9.4
Women					
Cases (n=460)	36.3	13.3	11.1	6.3	9.4
Non-cases (n=269,443)	26.8	6.8	6.9	4.3	5.5
<b>Pancreas</b>					
Men					
Cases (n=540)	48.5	13.7	14.5	10.3	15.2

Non-cases (n=220,324)	35.9	8.7	9.3	7.2	9.4
<b>Women</b>					
Cases (n=490)	34.9	11.6	10.7	6.2	9.4
Non-cases (n=269,613)	26.8	6.8	6.9	4.3	5.5
<b>Colon and Rectum</b>					
<b>Men</b>					
Cases (n=2,651)	46.0	8.3	13.0	10.2	14.5
Non-cases (n=216,802)	35.7	8.7	9.3	7.2	9.4
<b>Women</b>					
Cases (n=2,050)	31.4	6.2	8.5	5.1	6.9
Non-cases (n=266,852)	26.8	6.8	6.8	4.3	5.5
<b>Urinary Bladder</b>					
<b>Men</b>					
Cases (n=1,453)	56.2	14.5	16.9	12.5	18.8
Non-cases (n=218,494)	35.7	8.6	9.3	7.2	9.4
<b>Women</b>					
Cases (n=517)	41.6	12.4	12.7	7.0	10.4
Non-cases (n=269,278)	26.8	6.7	6.8	4.3	5.4
<b>Uterine Cervix</b>					
<b>Women</b>					
Cases (n=462)	30.5	12.6	7.5	4.8	5.9
Non-cases (n=262,260)	26.4	6.5	6.7	4.2	5.3
<b>Acute Myeloid Leukemia</b>					
<b>Men</b>					
Cases (n=197)	41.6	8.1	11.9	9.2	13.4
Non-cases (n=220,655)	35.9	8.7	9.3	7.2	9.4
<b>Women</b>					
Cases (n=143)	37.1	4.9	9.5	7.1	8.5
Non-cases (n=269,922)	26.9	6.8	6.9	4.3	5.5

\*Includes never and cigarette-only smokers, “Ever smoked” includes anyone who has smoked at any point, whether currently or formerly. The “never smokers” proportion is calculated as 1 minus the “ever smoked” percentage.

**Abbreviations:** No. Years, numbers of years; No. Cigarettes/day, numbers of cigarettes per day.