Tobacco smoking and risks of more than 470 diseases in China: a prospective cohort study

Ka Hung Chan, Neil Wright*, Dan Xiao, Yu Guo, Yiping Chen, Huaidong Du, Ling Yang, Iona Y Millwood, Pei Pei, Junzheng Wang, Iain Turnbull, Simon Gilbert, Daniel Avery, Christiana Kartsonaki, Canqing Yu, Junshi Chen, Jun Lv, Robert Clarke, Rory Collins, Richard Peto, Liming Li, Chen Wang, Zhengming Chen, on behalf of the China Kadoorie Biobank collaborative group†*

Summary

Background Tobacco smoking is estimated to account for more than 1 million annual deaths in China, and the epidemic continues to increase in men. Large nationwide prospective studies linked to different health records can help to periodically assess disease burden attributed to smoking. We aimed to examine associations of smoking with incidence of and mortality from an extensive range of diseases in China.

Methods We analysed data from the prospective China Kadoorie Biobank, which recruited 512 726 adults aged 30–79 years, of whom 210 201 were men and 302 525 were women. Participants who had no major disabilities were identified through local residential records in 100–150 administrative units, which were randomly selected by use of multistage cluster sampling, from each of the ten diverse study areas of China. They were invited and recruited between June 25, 2004, and July 15, 2008. Upon study entry, trained health workers administered a questionnaire assessing detailed smoking behaviours and other key characteristics (eg, sociodemographics, lifestyle, and medical history). Participants were followed up via electronic record linkages to death and disease registries and health insurance databases, from baseline to Jan 1, 2018. During a median 11-year follow-up (IQR 10–12), 285 542 (55·7%) participants were ever hospitalised, 48 869 (9·5%) died, and 5252 (1·0%) were lost to follow-up during the age-at-risk of 35–84 years. Cox regression yielded hazard ratios (HRs) associating smoking with disease incidence and mortality, adjusting for multiple testing.

Findings At baseline, 74·3% of men and 3·2% of women (overall 32·4%) ever smoked regularly. During follow-up, 1 137 603 International Classification of Diseases, 10th revision (ICD-10)-coded incident events occurred, involving 476 distinct conditions and 85 causes of death, each with at least 100 cases. Compared with never-regular smokers, ever-regular smokers had significantly higher risks for nine of 18 ICD-10 chapters examined at age-at-risk of 35–84 years. For individual conditions, smokers had significantly higher risks of 56 diseases (50 for men and 24 for women) and 22 causes of death (17 for men and nine for women). Among men, ever-regular smokers had an HR of 1·09 (95% CI 1·08–1·11) for any disease incidence when compared with never-regular smokers, and significantly more episodes and longer duration of hospitalisation, particularly those due to cancer and respiratory diseases. For overall mortality, the HRs were greater in men from urban areas than in men from rural areas (1·50 [1·42–1·58] *vs* 1·25 [1·20–1·30]). Among men from urban areas who began smoking at younger than 18 years, the HRs were 2·06 (1·89–2·24) for overall mortality and 1·32 (1·27–1·37) for any disease incidence. In this population, 19·6% of male (24·3% of men residing in urban settings and 16·2% of men residing in rural settings) and 2·8% of female deaths were attributed to ever-regular smoking.

Interpretation Among Chinese adults, smoking was associated with higher risks of morbidity and mortality from a wide range of diseases. Among men, the future smoking-attributed disease burden will increase further, highlighting a pressing need for reducing consumption through widespread cessation and uptake prevention.

Funding British Heart Foundation, Cancer Research UK, Chinese Ministry of Science and Technology, Kadoorie Charitable Foundation, UK Medical Research Council, National Natural Science Foundation of China, Wellcome Trust.

Introduction

Smoking accounted for about 100 million deaths worldwide in the 20th century and is projected to cause 1 billion deaths this century, mainly in low-income and middle-income countries (LMICs), including China.[1] About 40% of the world's tobacco is consumed by people in China, and almost exclusively by men.[2] In China, the main increase in consumption of manufactured cigarettes took place after 1980 and continued until the 2010s,[2] many decades behind that in high-income countries in Europe and North America.[1,3–5] Previous nationwide cohort studies that were established decades apart in China have reliably shown the increasing proportion of adult mortality that can be attributed to smoking in men,[6] which is estimated to cause more than 1 million annual deaths.[7] However, the morbidity burden that is attributed

Lancet Public Health 2022; 7: e1014–26

*Contributed equally

†Members listed in the appendix

Clinical Trial Service Unit and Epidemiological Studies Unit, Nuffield Department of Population Health (K H Chan DPhil, N Wright PhD, Y Chen DPhil, H Du PhD, L Yang PhD, I Y Millwood DPhil, I Turnbull MRCP, S Gilbert MSc, D Avery MSc, C Kartsonaki DPhil, Prof R Clarke MD, Prof R Collins FRS, Prof R Peto FRS, Prof Z Chen DPhil)**, MRC Population Health Research Unit, Nuffield Department of Population Health** (Y Chen, H Du L Yang, S Gilbert, C Kartsonaki, Prof Z Chen)**, and Oxford British Heart Foundation Centre of Research Excellence** (K H Chan)**, University of Oxford, Oxford, UK; WHO Collaborating Center for Tobacco Cessation and Respiratory Diseases Prevention, China-Japan Friendship Hospital, Beijing, China** (Prof D Xiao MD, Prof C Wang MD)**; Institute of Respiratory Medicine, Chinese Academy of Medical Sciences, Beijing, China** (Prof D Xiao, Prof C Wang)**; National Center for Cardiovascular Disease, Fuwai Hospital, Chinese Academy of Medical Sciences, Beijing, China** (Y Guo MSc)**; Licang Center of Disease Control and Prevention, Qingdao, China** (J Wang BSc)**; Department of Epidemiology and Biostatistics, School of Public Health** (C Yu PhD, Prof J Lv PhD, Prof L Li PhD) **and Peking University Center for Public Health and Epidemic Preparedness and Response** (C Yu, Prof J Lv, Prof L Li, P Pei BSc)**, Peking University, Beijing, China; China National Center for Food Safety Risk Assessment, Beijing, China** (Prof J Chen MD)

Correspondence to:
Prof Liming Li, Department of Epidemiology and Biostatistics, School of Public Health, Peking University, Beijing 100191, China
lmleeph@vip.163.com

or

Prof Chen Wang, Institute of Respiratory Medicine, Chinese Academy of Medical Sciences, Beijing 100730, China
wangchen@pumc.edu.cn

or

Prof Zhengming Chen, Clinical Trial Service Unit and Epidemiological Studies Unit, Nuffield Department of Population Health, University of Oxford, Oxford OX3 7LF, UK
zhengming.chen@ndph.ox.ac.uk

See Online for appendix

Research in context

Evidence before this study

We searched PubMed for relevant prospective cohort studies in China published in English from inception to Jan 1, 2022, using the terms "smoking OR tobacco" AND "China" AND "mortality OR death OR incidence OR risk". Most previous studies were conducted before 2000 and focused only on mortality from several major diseases, such as cancer, cardiovascular disease, and chronic respiratory diseases. A few large prospective cohort studies (each with >100 000 participants) in China, including our earlier reports based on the China Kadoorie Biobank (2004–08), also showed significant associations of smoking with incidence of several major diseases, including site-specific cancer and diabetes. Additionally, a pooled analysis of six prospective cohorts with more than 330 000 primarily Chinese adults reported significantly higher risk of nasopharyngeal cancer associated with smoking. However, there is no report in China assessing systematically the associations of smoking with a broader range of fatal and non-fatal diseases. Outside of China, only a few large prospective studies, such as the UK Million Women Study (1·3 million women) and pooled data from several US cohort studies (involving 0·5–1·0 million adults), have examined the associations of smoking with risks of mortality, but not incidence, from a wide range of disease outcomes simultaneously. Overall, the large UK study reported significant associations of smoking with 23 causes of death and the US studies reported significant associations of smoking with 35 causes of deaths.

Added value of this study

To our knowledge, this is the first large nationwide prospective cohort study in China to systematically examine the associations of smoking with diseases affecting all body systems. After accounting for multiple testing, smoking was significantly associated with increased disease incidence from 56 (50 for men and 24 for women) specific diseases (ie, ten cardiovascular, 14 respiratory, 14 cancer, five digestive, and 13 other diseases) and with higher risks of death from 22 (17 for men and nine for women) specific causes. This study provides important new evidence on the health effects of smoking beyond mortality from major diseases in China. It also provides additional evidence showing greater relative risk of morbidity and mortality among male smokers living in urban areas, especially among those who started young. Furthermore, for the first time, we estimated the excess total and cause-specific (ie, cardiovascular, respiratory, and cancer) hospitalisation episodes (ie, admission to hospital) and duration of hospitalisation associated with smoking in China. Apart from hazards of smoking, this study also showed that stopping smoking before the onset of major illness is beneficial.

Implications of all the available evidence

Decades of epidemiological research have produced strong evidence on the adverse health effects of smoking on an extensive range of major diseases. In China and many other low-income and middle-income countries, the future disease morbidity and mortality burden attributed to smoking will most likely increase. Large prospective cohort studies with reliable electronic linkage not only to mortality but also to hospital records in diverse populations could help to assess and monitor the evolution of the tobacco epidemic by providing population-specific estimates on the current and future disease burden attributed to smoking. Such evidence is crucial to inform appropriate policy action in tobacco control nationally and globally.

to smoking, from a much broader range of diseases than for all-cause and cause-specific mortality, has not been properly studied in China (and most other LMICs).

There is compelling evidence that smoking causes many diseases, chiefly cancer, cardiovascular disease, and chronic respiratory diseases.[5] Nevertheless, large prospective studies in high-income countries have also identified other diseases that are associated with smoking, including chronic kidney disease, influenza, and mental health disorders.[5,8–11] In most LMICs, the available prospective evidence on smoking hazards is mostly confined to mortality.[6,12–15] In China, although many studies have attempted to assess the effects of smoking on morbidity outcomes, they were constrained by the use of non-prospective study designs, involvement of few outcomes (ie, mainly site-specific cancer, chronic obstructive pulmonary disease, and cardiovascular disease), restriction to specific urban cities, and absence of objective validation of self-reported smoking status.[16–20] Reliable assessment of morbidity burden, in addition to mortality, attributed to smoking in different populations is needed to inform effective tobacco control nationally and globally.

For more on the **China Kadoorie Biobank** see https://www.ckbiobank.org

To help to fill the evidence gap, we conducted detailed analyses of the health risks associated with smoking in the prospective China Kadoorie Biobank (CKB) of more than 512 000 adults.[21,22] We aimed to examine the associations of ever-regular smoking with incidence of 476 distinct diseases and 85 causes of mortality, both overall and in rural and urban areas separately; assess the disease risks associated with specific smoking patterns (including smoking cessation); and evaluate the disease burden associated with smoking through hospitalisation (ie, admission to hospital) and death.

Methods

Study design and participants

Details of the study design and baseline characteristics of CKB have been described elsewhere.[21,22] Briefly, we established a prospective cohort study of adults aged 30–79 years, who were recruited to take part in the CKB via multistage cluster sampling between June 25, 2004,

and July 15, 2008, from four urban districts, five rural counties, and one semi-rural township across China. These areas were selected through China's nationally representative Disease Surveillance Point System[23] to cover a diverse range of geographical areas, socioeconomic development, risk exposures, and disease patterns. For each study site, about 100–150 administrative units (ie, rural villages or urban street committees) were selected, and all eligible residents (ie, people without known major disabilities) aged 35–74 years (n=1 801 167) were identified through official residential records and were invited to participate. 499 439 (28%) people participated, plus 13 287 individuals just outside the targeted age range, who were included to encourage community participation (n=9817 aged 30–34; n=3470 aged 75–79 years), resulting in 512 726 participants with a baseline age range of 30–79 years, of whom 210 201 were men and 302 525 were women. Participants were followed up continuously from recruitment via electronic linkages to established disease and death registries and national health insurance databases. In 2008 and 2013–14, two resurveys were done in approximately 5% of surviving participants, who were randomly selected, to obtain repeated assessment (including smoking) and additional data for enrichment. By Jan 1, 2018, 49 459 (9·6%) of 512 726 participants had died and 5302 (1·0%) participants were lost to follow-up. Ten participants died or were lost to follow-up before reaching 35 years of age and were therefore excluded from analysis, so a total of 512 716 participants were included in the main analysis.

All participants provided written informed consent. Ethical approvals were obtained by the Ethical Review Committee of the China National Center for Disease Control and Prevention and the Oxford Tropical Research Ethics Committee, University of Oxford, Oxford, UK, before commencement of the field work.

Procedures

In local study assessment clinics, trained health workers took physical measurements (eg, height, weight, and blood pressure) and administered a laptop-based questionnaire interview covering sociodemographic status, lifestyle, environmental factors, female reproductive factors, and medical history during the baseline survey conducted on study recruitment. Sex was self-reported and possible options were "man" or "woman". Logic and error checks were built in to the questionnaire to avoid missing data.

Questions on current and past smoking behaviours included age when individuals first began to smoke regularly; frequency, amount, and type of tobacco smoked; degree of inhalation; and, for ex-smokers, age when they last stopped smoking and main reasons for cessation (ie, due to illnesses or other reasons). To validate self-reported smoking status (along with exposure to household air pollution), exhaled carbon monoxide was also measured using MicroCO meters (Carefusion, San Diego, CA, USA).[24] In this study, participants who had never smoked or smoked less than 100 cigarettes in their lifetime were classified as never-regular smokers, and those who smoked one cigarette or more (or ≥1 g tobacco) daily for at least 6 consecutive months were classified as ever-regular smokers. Among participants who had stopped for 6 months or more at the time of the questionnaire administration, those who had stopped smoking due to illnesses were grouped with baseline current smokers as regular smokers as they would have notably elevated disease risk even after stopping, whereas those who had stopped voluntarily for other reasons (eg, finances) constituted a separate category to assess the effects of cessation (ie, before ill health).[6]

After the baseline survey, participants were followed up for death and any episodes of hospitalisation through electronic linkage via unique personal identification number to established mortality and morbidity (for cancer, stroke, ischaemic heart disease, and diabetes) registries and to national health insurance systems. All the reported disease events were coded following the International Classification of Diseases, 10th revision (ICD-10)[25] by trained medical professionals who were masked to baseline information. For this study, participants were censored on death, loss to follow-up, or Jan 1, 2018, whichever came first. During 5·5 million person-years (median 11 years, IQR 10–12) of follow-up, 48 869 (9·5%) of 512 716 participants died, 285 542 (55·7%) were hospitalised, and 5252 (1·0%) were lost to follow-up at age-at-risk of 35–84 years.

To enable a phenome-wide investigation, all disease events coded up to the first three characters of ICD-10 codes (ie, the disease-category component) were reviewed, and, when appropriate, combined (on the basis of knowledge about the disease characteristics), to produce a list of distinct diseases. Several ICD-10 chapters that were considered irrelevant to the study population (eg, perinatal-origin diseases [chapter XVI] and congenital conditions [XVII]) were excluded. Specific outcomes (or a small number of specific disease groups) with at least 100 incident events (for individuals with multiple hospitalisation events of the same nature, only the first event was considered; 80 incident events per sex for sex-specific analyses) or deaths (80 deaths per sex for sex-specific analyses) recorded during follow-up were analysed separately to capture a wide spectrum of conditions while ensuring reasonable statistical precision. Under each ICD-10 chapter, outcomes with less than 100 (<80 for sex-specific analyses) events were combined and considered as other disease of the individual chapter to enable exploratory analysis that might shed light on rare outcomes of the same disease system. Since the classification for the other disease endpoints was based on event number, the ICD-10 criteria used vary slightly in the combined and sex-specific analyses.

For overall mortality and major diseases constituting the top five causes of disability-adjusted life-years in

China,[26] namely ischaemic heart disease (ICD-10 I20-I25), intracerebral haemorrhage (I61), ischaemic stroke (I63), chronic obstructive pulmonary disease (J41-J44), and lung cancer (C34), further in-depth analyses by detailed smoking characteristics were also done. The total number of hospitalisation episodes (ie, including both first and subsequent events) and duration spent in hospitals (ie, bed-days) attributed to all causes, cardiovascular disease, respiratory disease, cancer, and other causes were also examined as indicators of use of health service.

Statistical analysis

We used Cox regression to estimate hazard ratios (HRs) comparing disease incidence or mortality risks at age-at-risk[27] of 35–84 years (which was chosen to capture the otherwise preventable long-term health effect of smoking in middle-aged to older-aged adults) in ever-regular smokers (ie, overall and by particular smoking characteristics) versus never-regular smokers, both overall and separately by sex (given the substantial sex difference in smoking prevalence[6]). We excluded participants who died or were lost to follow-up before reaching age 35 years.

All analyses were stratified by age-at-risk (5-year groups), study areas (ten sites), and, when appropriate, sex and adjusted for education (no formal school, primary school, middle or high school, or college or university) and alcohol drinking (never, occasional, or ever-regular drinking). Among men, analyses were also conducted separately in urban (four) and rural or semi-rural (six) areas. All-cause mortality is a competing risk for disease events and cause-specific mortality. Hence, to facilitate observational analyses assessing questions related to cause,[28] participants were censored at death from any cause to estimate cause-specific HRs, which compared event rates in participants who were alive and free of the event of interest.

For smoking exposures involving more than two categories, group-specific CIs of HRs were calculated by use of the variance of the log hazard in each category, including the reference group. This approach enables direct comparison of HRs across any two categories of exposure, instead of just between a fixed reference group and other exposure categories.[29] The proportional hazards assumption for ever-regular smoking was investigated by plotting and testing correlations between transformed follow-up time and the scaled Schoenfeld residuals[30] for incidence and mortality of the top-five major diseases, among men and women.

For disease outcomes showing significant associations with smoking, sensitivity analyses were conducted with additional adjustment for BMI, self-rated health, and occupation; for outcomes showing significant inverse associations, further exclusion of the first 5 years of follow-up was done to assess reverse causation bias. For the top-five major diseases, we further excluded participants with previous medical history of relevant conditions at baseline (eg, excluding those with previous cardiovascular disease in analyses of ischaemic heart disease or stroke). Moreover, we also conducted separate analyses among current smokers and ex-smokers who had stopped due to illnesses, who together formed the regular smoker category.

To assess the cumulative burden of smoking, the total number of hospitalisations and days in hospital were estimated for ever-regular versus never-regular smokers by use of the mean cumulative count. Mean cumulative count estimates the total number of events occurring in a population and does not assume independence between hospitalisations and all-cause mortality.[31–33] Overall survival of ever-regular versus never-regular smokers was also examined by use of Kaplan-Meier curves by sex. Adjusted incidence rates for individual diseases were calculated within age (ie, <65 years or ≥65 years) and sex strata as

$$HR_i \times \frac{\text{overall rate}}{\frac{\sum(n_i \times HR_i)}{\sum n_i}}$$

where HR_i and n_i are HRs and number of events for ever-regular smokers and never-regular smokers. Overall incidence rates were calculated as weighted means of the age-specific and sex-specific rates, and total incidence rates are the sum of disease-specific incidence rates. The fraction of all deaths that is attributed to smoking in the study population (ie, the population-attributed fraction) was estimated by $P(HR-1)/HR$, where P is the prevalence of ever-regular smoking among participants who died during follow-up and HR is the associated HR for all-cause mortality.

Significance (at the 5% level) was evaluated by use of both conventional and false discovery rate (FDR)-adjusted p values applied within ICD-10 chapters (and separately in the combined and sex-specific analyses).[34] To control for multiple testing, FDR-adjusted p values control the expected proportion of false positives among all significant associations.[35,36] Unless otherwise specified, we highlight only associations that were statistically significant after FDR adjustment. All analyses were conducted by use of R software, version 3.6.2.

Role of the funding source

The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

Results

Of the 512 716 participants included in this analysis, 210 201 (41·0%) were men and 302 515 (59·0%) were women, 339 794 (66·3%) resided in rural areas, and the mean age was 52·0 years (SD 10·7) at baseline. Overall, 166 081 (32·4%) ever smoked regularly, with a much higher

	Number of diseases		Number of cases		Positive associations without FDR adjustment		Inverse associations without FDR adjustment		Positive associations with FDR adjustment		Inverse associations with FDR adjustment	
	Incidence	Death	Incidence	Death	Incidence	Death	Incidence	Death	Incidence	Death	Incidence	Death
I: Infectious and parasitic	21	4	28572	624	3	1	0	0	1	0	0	0
II: Neoplasms	72	22	59716	15811	18	9	0	0	14	7	0	0
III: Blood and immune-related	7	1	6146	90	0	0	1	0	0	0	0	0
IV: Endocrine, nutritional, and metabolic	13	2	47809	1397	5	0	0	0	3	0	0	0
V: Mental and behavioural	12	4	8109	140	0	0	2	0	0	0	0	0
VI: Nerve-related	21	3	23065	375	1	1	1	1	0	0	1	0
VII: Eye and adnexa	24	NA*	43182	NA*	2	NA*	1	NA*	2	NA*	1	NA*
VIII: Ear and mastoid process	5	NA*	7317	NA*	0	NA*	0	NA*	0	NA*	0	NA*
IX: Circulatory	41	19	279306	19836	14	9	1	0	10	8	1	0
X: Respiratory	21	10	175637	4564	14	5	1	0	14	5	1	0
XI: Digestive	51	3	132817	1038	10	1	4	0	5	1	1	0
XII: Skin and subcutaneous tissue	11	1	12273	29	2	0	1	0	0	0	0	0
XIII: Musculoskeletal	34	1	99111	125	1	0	5	0	0	0	2	0
XIV: Genitourinary	54	2	78577	552	2	0	5	0	0	0	0	0
XV: Pregnancy-related	6	1	3745	1	1	0	0	0	0	0	0	0
XVIII: Other symptoms, signs, and abnormal findings	35	3	87254	831	10	2	0	0	7	1	0	0
XIX: Injury, poisoning, and other external causes	38	1	41073	79	4	0	0	0	0	0	0	0
XX: External causes	10	8	3894	3268	0	0	1	1	0	0	0	0
Total	476	85	1137603	48760	87	28	23	2	56	22	7	0

FDR=false discovery rate. ICD-10=International Classification of Diseases, 10th revision. NA=not applicable. *No eligible endpoint for analysis.

Table 1: **Number of morbidity and mortality events by ICD-10 chapter and their overall significant associations with ever-regular smoking, for men and women combined**

proportion in men (156284 [74·3%] of 210201) than in women (9797 [3·2%] of 302515). Both current smokers and ex-smokers who quit voluntarily tended to live in rural areas, to be less educated and more likely to consume alcohol, and to have slightly lower BMI and higher prevalence of previous chronic diseases than never-regular smokers (appendix p 4). Compared with female smokers, male smokers were more likely to start at a younger age, be heavy smokers, and habitually inhale tobacco smoke into the lungs. Among men, similar differences were generally seen between younger and older smokers, with younger smokers more likely to start at younger than 20 years and to habitually smoke manufactured cigarettes.

During follow-up, 1137603 ICD-10-coded incident events occurred, involving 476 distinct conditions and 85 causes of death, each with at least 100 cases. We performed the phenome-wide investigation across 18 ICD-10 chapters (table 1). Of the 18 ICD-10 disease chapters examined, ever-regular smoking was associated with significantly increased risks of disease incidence in nine chapters (and no significant positive or negative associations for the rest), including infectious and parasitic diseases (HR 1·07, 95% CI 1·03–1·11), neoplasms (1·34, 1·30–1·38), endocrine, nutritional, and metabolic diseases (1·05, 1·02–1·09), circulatory diseases (1·10, 1·08–1·12), respiratory diseases (1·18, 1·16–1·21), digestive diseases (1·03, 1·01–1·06), diseases of skin and subcutaneous tissue (1·14, 1·06–1·22), other symptoms, signs and abnormal findings (1·06, 1·04–1·09), and injury, poisoning, and other external causes (1·06, 1·02–1·10; appendix p 5). For the corresponding mortality analyses, the HRs generally appeared to be higher than for the disease incidence findings (appendix p 6). Although the HRs tended to be somewhat larger in men, the patterns of associations were broadly similar in men and women. Overall, male ever-regular smokers had an excess risk of morbidity from any disease (HR 1·09, 95% CI 1·08–1·11; appendix p 5), which was higher in urban than rural areas (1·17 [1·15–1·20] *vs* 1·05 [1·03–1·06]; appendix pp 7–8). Among female ever-regular smokers, the HR for any morbidity was 1·04 (95% CI 1·01–1·06; appendix p 5).

In men, 87 different diseases were significantly positively associated with smoking at the nominal level, and 50 diseases were still significantly associated after FDR adjustment (appendix p 9), whereas among women, there were 24 such diseases (appendix p 10). By contrast, seven conditions showed significant inverse associations with smoking after adjustment in sex-combined analyses eight significant inverse associations in men and none in women in sex-specific analyses). No other cause-specific associations were identified after adjustment.

Overall in the combined analyses, the FDR-adjusted HRs for positive associations ranged from 1·06 (95% CI 1·02–1·09) for diabetes (ICD-10 E10–E14) to 3·16 (1·98–5·05) for larynx cancer (C32; figure 1). The HRs were generally higher in men than in women, but exceptions included larynx cancer (C32), lung cancer (C34), and several major respiratory diseases, such as chronic bronchitis (J41), for which the HRs appeared similar or greater in women.

For the seven diseases showing an overall inverse association (ie, Parkinson's disease [G20], other disorders of conjunctiva [H11, H13], varicose veins [I83, I85, and I86], bronchiectasis [J47], inguinal hernia [K40], other arthrosis [M19], and gonarthrosis [M17]), the HRs ranged from 0·82 (95% CI 0·71–0·96) for bronchiectasis to 0·73 (0·59–0·89) for Parkinson's disease. Among women, there was no apparent association of ever-smoking with breast cancer (74 of 9797 ever-regular smokers and 2646 of 292 718 never-regular smokers had breast cancer; 0·94 [95% CI 0·74–1·20]) or cancer of the corpus uteri (mostly endometrial cancer; 15 ever-regular smokers and 562 never-regular smokers had cancer of the corpus uteri; 0·97 [0·56–1·68]). The HRs for all disease-specific morbidity under each ICD-10 chapter examined are shown in the appendix (pp 11–28). These associations were unaltered in sensitivity analyses with additional adjustment for BMI, self-rated health, and occupation (data not shown).

Of the 22 mortality endpoints that were significantly associated with ever-regular smoking after FDR adjustment (17 associations in men and nine associations in women), 16 showed significant, albeit generally more modest, positive associations with smoking in the morbidity analyses (figure 1; appendix pp 29–30). For the six other mortality endpoints that were not significant after FDR adjustment, all morbidity analyses showed directionally consistent, although non-significant, positive associations. For major causes of death, there were generally similar HRs between male and female smokers (appendix p 29). After FDR adjustment, no single cause of death had a lower risk among smokers than among non-smokers. Additional adjustments for potential confounders (ie, BMI, self-rated health, and occupation) did not alter the HRs for all the diseases showing either positive or inverse associations with smoking (appendix p 32).

For the five leading diseases, generally male smokers who had started smoking at younger ages or smoked more cigarettes (or equivalents) per day had consistently higher HRs for both morbidity and mortality than did those who started smoking at older ages or smoked fewer cigarettes, in a dose–response manner (appendix p 33), which was also true when men from urban settings and men from rural settings were analysed separately (table 2). Moreover, the HRs for ever-regular versus never-regular smokers were generally higher in men residing in urban areas than those in rural areas (table 2) and for mortality than morbidity (appendix p 34). Similar, albeit less extreme, findings were also shown in women (appendix p 33). Sensitivity analyses with exclusion of individuals with a previous history of specific diseases yielded similar results (appendix pp 35–36).

For overall mortality, the adjusted HR was 1·33 (95% CI 1·29–1·37) in male ever-regular smokers and the HR was higher in urban than in rural areas (1·50 [1·42–1·58] *vs* 1·25 [1·20–1·30]; appendix pp 7, 8, 37). Among women the corresponding HR was 1·44 (95% CI 1·36–1·52; appendix p 38), again higher in urban than rural areas (1·53 [1·41–1·66] *vs* 1·38 [1·29–1·48]). These HRs were not materially altered after exclusion of participants with previous history of cancer at baseline (appendix pp 7–8, 37–38). Moreover, generally regular smokers who started smoking regularly at a younger age had higher risks for major disease morbidity and mortality than did those who started smoking at an older age, particularly in urban areas (figure 2). Among men residing in urban areas who began smoking at younger than 18 years, the HRs were 2·06 (95% CI 1·89–2·24) for overall mortality and 1·32 (1·27–1·37) for any disease incidence (figure 2), whereas the HR was 2·35 (2·07–2·67) for overall mortality and 1·40 (1·30–1·50) for any disease incidence in people who began smoking at younger than 15 years.

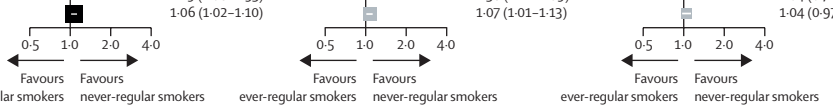
Compared with never-regular smokers, ever-regular smokers of either sex appeared to have worse survival from about age 55 years, and smokers reached median survival approximately 3·5 years earlier (appendix p 39). After accounting for excess mortality (ie, competing risk), ever-regular smokers had significantly higher total expected hospitalisations and (to a lesser extent) longer stay in hospital than did never-regular smokers, with more pronounced differences in men than in women (appendix p 40) and for hospitalisations attributed to cancer and respiratory disease than for hospitalisations attributed to other causes (appendix p 41). The differences broadly increased with age-at-risk, with apparent divergence beginning at around age 55–60 years.

Among male ex-smokers who had quit smoking due to illness, there were significant excess risks of overall

Figure 1: **Adjusted HRs for cause-specific disease incidence significantly associated with ever-regular smoking**

HRs were stratified by age-at-risk (5-year groups), sex, and study area and were adjusted for education and alcohol drinking. All analyses were restricted to age-at-risk range of 35–84 years. The solid boxes represent HRs, with the size inversely proportional to the variance of the logarithm of the HR, and the horizontal lines represent 95% CIs. The individual diseases listed included all that showed FDR-adjusted significant associations with smoking, in overall or sex-specific analyses. The black boxes represent FDR-adjusted p<0·05 and grey boxes indicate FDR-adjusted non-significant associations. FDR=false discovery rate. HR=hazard ratio. ICD-10=International Classification of Diseases, 10th revision. NA=not applicable. *Not included in the sex-specific phenome-wide investigation (<80 events in female participants). †Included diseases differ between combined and sex-specific phenome-wide investigations. ‡No cases were recorded in female participants.

	Number of events (ever-regular smokers/ never-regular smokers)	All	Men	Women
		HR (95% CI)	HR (95% CI)	HR (95% CI)
I Infectious and parasitic				
A16 Respiratory tuberculosis	968/951	1·35 (1·18–1·53)	1·36 (1·18–1·57)	1·25 (0·89–1·75)
A17–A19 Extrapulmonary tuberculosis	252/493	1·31 (1·01–1·68)	1·51 (1·12–2·03)	0·60 (0·28–1·30)
II Neoplasms				
C32 Larynx cancer	178/54	3·16 (1·98–5·05)	2·42 (1·52–3·85)	11·05 (4·37–27·92)*
D11 Benign neoplasm (salivary glands)	87/98	2·62 (1·56–4·40)	4·15 (1·89–9·11)	1·22 (0·41–3·60)
C34 Lung cancer	3759/2843	2·40 (2·22–2·59)	2·41 (2·18–2·65)	2·42 (2·11–2·77)
D02 Carcinoma in situ of middle ear and respiratory system	79/48	2·20 (1·24–3·90)	2·31 (1·18–4·52)	1·75 (0·53–5·73)*
D38 Uncertain neoplasm (middle ear, respiratory, or intrathoracic)	220/184	2·13 (1·57–2·91)	2·50 (1·72–3·63)	1·26 (0·61–2·60)
D41 Uncertain neoplasm (urinary organs)	77/81	2·08 (1·27–3·39)	2·03 (1·19–3·48)	2·31 (0·77–6·92)*
D36 Benign neoplasm (other)	102/154	1·92 (1·25–2·96)	1·81 (1·07–3·04)	2·11 (1·03–4·34)
C67 Bladder cancer	367/244	1·70 (1·35–2·14)	1·70 (1·32–2·20)	1·69 (0·96–2·95)
C78 Secondary cancer respiratory and digestive organs	360/380	1·61 (1·28–2·01)	1·83 (1·40–2·39)	1·00 (0·58–1·72)
C79 Secondary malignant neoplasm of other and unspecified sites	375/402	1·34 (1·07–1·67)	1·43 (1·12–1·84)	0·99 (0·57–1·74)
C15 Oesophagus cancer	1544/1157	1·33 (1·19–1·49)	1·34 (1·19–1·51)	1·14 (0·74–1·74)
C23–C24 Biliary cancer other than the liver	292/443	1·30 (1·03–1·64)	1·50 (1·12–2·00)	0·93 (0·59–1·48)
C16 Stomach cancer	2097/1701	1·30 (1·18–1·43)	1·31 (1·19–1·45)	1·18 (0·90–1·55)
C80 Other cancer	1829/3080	1·30 (1·18–1·43)	1·34 (1·20–1·50)	1·17 (0·97–1·41)
C22 Liver cancer	1754/1507	1·29 (1·16–1·42)	1·30 (1·17–1·45)	1·20 (0·92–1·56)
IV Endocrine, nutritional, and metabolic				
E88 Other metabolic disorders	241/226	1·58 (1·23–2·02)	1·29 (0·90–1·86)	1·88 (1·37–2·59)
E16 Other disorders of pancreatic internal secretion	154/296	1·36 (1·01–1·84)	1·11 (0·77–1·61)	1·87 (1·21–2·90)
E87 Other disorders of fluid, electrolyte, and acid–base balance	850/1267	1·23 (1·09–1·39)	1·19 (0·98–1·44)	1·26 (1·08–1·47)
E10–E14 Diabetes	10 211/22 734	1·06 (1·02–1·09)	1·04 (0·99–1·08)	1·12 (1·05–1·20)
VI Nerve-related				
G20 Parkinson's disease	267/571	0·73 (0·59–0·89)	0·72 (0·58–0·90)	0·76 (0·46–1·28)
VII Eye and adnexa				
H25, H26·9 Cataract	5096/11 244	1·11 (1·05–1·16)	1·12 (1·06–1·20)	1·07 (0·99–1·17)
H26 Other cataract	3550/7744	1·10 (1·04–1·17)	1·13 (1·05–1·22)	1·03 (0·93–1·15)
H11, H13 Other disorders of conjunctiva	850/2311	0·75 (0·66–0·85)	0·76 (0·66–0·87)	0·69 (0·49–0·96)
IX Circulatory				
I71 Aortic aneurysm and dissection	171/119	2·46 (1·71–3·54)	2·63 (1·71–4·03)	1·99 (0·94–4·21)
I74 Arterial embolism and thrombosis	152/120	1·99 (1·40–2·83)	1·99 (1·31–3·03)	1·98 (1·03–3·82)
I27 Other pulmonary heart diseases	3161/2445	1·78 (1·65–1·92)	1·78 (1·61–1·98)	1·77 (1·59–1·97)
I26 Pulmonary embolism	92/135	1·54 (1·03–2·30)	2·15 (1·27–3·64)	0·65 (0·25–1·67)
I21 Acute myocardial infarction	3810/4137	1·49 (1·39–1·59)	1·49 (1·38–1·60)	1·49 (1·30–1·71)
I46 Cardiac arrest	476/646	1·43 (1·20–1·70)	1·43 (1·18–1·74)	1·42 (0·95–2·12)
I70 Atherosclerosis	821/1055	1·32 (1·16–1·49)	1·40 (1·19–1·66)	1·22 (1·00–1·48)
I50 Heart failure	2399/3335	1·30 (1·21–1·40)	1·25 (1·13–1·37)	1·40 (1·24–1·59)
I51 Complications of heart disease	1444/2560	1·20 (1·08–1·33)	1·17 (1·04–1·33)	1·25 (1·02–1·53)
I25 Chronic ischaemic heart disease	17 159/33 311	1·18 (1·15–1·22)	1·22 (1·18–1·26)	1·10 (1·05–1·16)
I63 Cerebral infarction	18 388/32 726	1·12 (1·09–1·15)	1·12 (1·09–1·16)	1·10 (1·05–1·16)
I20 Angina pectoris	1949/3683	1·09 (1·01–1·19)	1·13 (1·02–1·24)	0·96 (0·80–1·15)
I69 Sequelae of cerebrovascular disease	2874/3938	1·06 (0·99–1·14)	1·02 (0·95–1·10)	1·28 (1·09–1·52)
I83, I85, I86 Varicose veins	987/1339	0·81 (0·72–0·92)	0·83 (0·73–0·94)	0·65 (0·40–1·07)
X Respiratory				
J93 Pneumothorax	300/105	2·86 (2·04–3·99)	2·97 (2·04–4·33)	2·47 (1·10–5·57)
J41 Simple and mucopurulent chronic bronchitis	107/83	2·36 (1·62–3·45)	2·15 (1·27–3·65)	2·50 (1·48–4·21)
Other respiratory ICD-10 codes†	192/131	1·96 (1·38–2·78)	1·70 (1·16–2·48)	2·39 (1·45–3·92)
J43 Emphysema	2177/1513	1·85 (1·68–2·03)	1·85 (1·63–2·09)	1·82 (1·58–2·10)
J85 Abscess of lung and mediastinum	92/54	1·84 (1·11–3·05)	2·08 (1·13–3·86)	1·25 (0·44–3·55)*
J96 Respiratory failure, not elsewhere classified	936/1211	1·74 (1·54–1·98)	1·66 (1·44–1·93)	1·97 (1·56–2·50)
J44 Other chronic obstructive pulmonary disease	8256/7294	1·72 (1·65–1·80)	1·74 (1·64–1·84)	1·70 (1·59–1·82)
J42 Unspecified chronic bronchitis	5214/5105	1·61 (1·52–1·71)	1·59 (1·47–1·71)	1·67 (1·51–1·85)
J45 Asthma	1126/2312	1·29 (1·14–1·45)	1·22 (1·07–1·40)	1·48 (1·19–1·84)
J90–J91 Pleural effusion	510/533	1·26 (1·06–1·49)	1·17 (0·95–1·43)	1·49 (1·10–2·01)
J12–J18 Pneumonia	12 002/19 743	1·22 (1·18–1·26)	1·25 (1·20–1·31)	1·14 (1·07–1·21)
J98 Other respiratory disorders	1268/2166	1·16 (1·04–1·30)	1·20 (1·05–1·37)	1·07 (0·87–1·32)
J40 Bronchitis, not specified as acute or chronic	4769/10 217	1·13 (1·06–1·20)	1·13 (1·06–1·21)	1·13 (0·99–1·30)
J00 Acute nasopharyngitis (common cold)	5608/10 982	1·10 (1·03–1·16)	1·09 (1·02–1·16)	1·22 (1·03–1·44)
J47 Bronchiectasis	505/1087	0·82 (0·71–0·96)	0·84 (0·71–1·01)	0·75 (0·54–1·05)
XI Digestive				
K70 Alcoholic liver disease	255/51	1·86 (1·27–2·72)	1·80 (1·22–2·65)	3·44 (0·83–14·30)*
K27 Peptic ulcer, site unspecified	440/484	1·58 (1·28–1·94)	1·68 (1·33–2·13)	1·24 (0·79–1·96)
K26 Duodenal ulcer	578/679	1·31 (1·10–1·55)	1·27 (1·06–1·52)	1·53 (0·98–2·37)
K25 Gastric ulcer	1364/1785	1·21 (1·08–1·35)	1·20 (1·05–1·36)	1·25 (1·01–1·56)
K92 Other diseases of digestive system	1976/2341	1·16 (1·06–1·27)	1·14 (1·03–1·26)	1·26 (1·02–1·55)
K74 Fibrosis and cirrhosis of liver	984/1211	1·08 (0·95–1·23)	0·99 (0·86–1·13)	1·64 (1·24–2·16)
K40 Inguinal hernia	698/435	0·75 (0·65–0·85)	0·75 (0·65–0·86)	0·65 (0·18–2·31)
XIII Musculoskeletal				
M19 Other arthrosis	4502/12 292	0·91 (0·86–0·96)	0·93 (0·87–0·99)	0·84 (0·74–0·94)
M17 Gonarthrosis (arthrosis of knee)	388/1557	0·74 (0·63–0·87)	0·73 (0·59–0·90)	0·76 (0·58–0·99)
M67 Other disorders of synovium and tendon	107/290	0·64 (0·45–0·91)	0·58 (0·41–0·84)	1·37 (0·57–3·26)
XIV Genitourinary				
N19 Unspecified kidney failure	801/1117	1·24 (1·08–1·42)	1·28 (1·09–1·50)	1·14 (0·89–1·47)
N40 Hyperplasia of prostate	2685/1163	0·90 (0·84–0·97)	0·90 (0·83–0·97)	NA
N35 Urethral stricture	78/71	0·56 (0·38–0·81)	0·55 (0·38–0·80)	NA‡
XVIII Other symptoms, signs, and abnormal findings				
R91 Abnormal findings on diagnostic imaging of lung	700/877	1·58 (1·32–1·88)	1·50 (1·23–1·81)	2·03 (1·35–3·06)
R99 Other ill defined and unspecified causes of mortality	242/303	1·49 (1·18–1·90)	1·63 (1·24–2·15)	1·04 (0·59–1·85)
R09 Other circulatory and respiratory symptoms	264/300	1·43 (1·10–1·85)	1·44 (1·07–1·94)	1·37 (0·78–2·40)
R22 Localised swelling	371/623	1·41 (1·13–1·77)	1·44 (1·11–1·85)	1·29 (0·76–2·19)
R04 Haemorrhage from respiratory passages	521/610	1·33 (1·10–1·61)	1·31 (1·06–1·61)	1·47 (0·96–2·26)
R50 Fever of other and unknown origin	475/916	1·29 (1·08–1·53)	1·38 (1·12–1·69)	1·04 (0·71–1·51)
R69 Unknown and unspecified causes of morbidity	7753/14 290	1·06 (1·02–1·10)	1·07 (1·01–1·13)	1·04 (0·97–1·11)



	Lung cancer		Ischaemic heart disease		Ischaemic stroke		Intracerebral haemorrhage		Chronic obstructive pulmonary disease	
	Number of events	HR (95% CI)	Number of events	HR (95% CI)	Number of events	HR (95% CI)	Number of events	HR (95% CI)	Number of events	HR (95% CI)
Men in urban areas										
Smoker category										
Never-regular smoker	242	1·00 (0·88–1·14)	3120	1·00 (0·96–1·04)	3593	1·00 (0·97–1·03)	473	1·00 (0·91–1·10)	463	1·00 (0·91–1·10)
Ever-regular smoker	1227	2·58 (2·24–2·98)	7220	1·27 (1·22–1·33)	7286	1·14 (1·09–1·19)	996	1·14 (1·02–1·28)	1576	2·26 (2·02–2·51)
Ex-smoker (by choice)*	134	1·48 (1·25–1·76)	1139	1·10 (1·03–1·16)	1201	1·01 (0·96–1·07)	152	1·01 (0·86–1·19)	205	1·40 (1·22–1·61)
Regular smoker	1093	2·87 (2·70–3·06)	6081	1·31 (1·28–1·35)	6085	1·17 (1·14–1·20)	844	1·17 (1·09–1·26)	1371	2·50 (2·37–2·64)
Ex-smoker (ill health)	190	2·35 (2·03–2·71)	1288	1·55 (1·46–1·63)	1162	1·16 (1·10–1·23)	180	1·37 (1·18–1·58)	415	3·13 (2·84–3·45)
Current smoker	903	3·04 (2·83–3·26)	4793	1·26 (1·22–1·30)	4923	1·17 (1·13–1·20)	664	1·12 (1·03–1·22)	956	2·28 (2·13–2·44)
Age participants began smoking†										
≥25 years	306	2·23 (1·99–2·49)	1954	1·23 (1·18–1·29)	2020	1·09 (1·04–1·14)	278	1·13 (1·01–1·27)	436	1·95 (1·77–2·14)
18–24 years	595	3·14 (2·89–3·41)	3161	1·33 (1·28–1·38)	3141	1·18 (1·14–1·23)	439	1·19 (1·08–1·31)	669	2·67 (2·47–2·88)
<18 years	192	3·64 (3·15–4·21)	966	1·55 (1·45–1·65)	924	1·33 (1·25–1·42)	127	1·19 (0·99–1·42)	266	3·73 (3·30–4·22)
Number of cigarettes smoked per day†										
<15	335	2·17 (1·94–2·42)	2280	1·20 (1·15–1·25)	2337	1·09 (1·05–1·14)	316	1·12 (1·00–1·25)	478	2·14 (1·95–2·35)
15–24	525	3·06 (2·81–3·34)	2722	1·33 (1·28– 1·38)	2747	1·19 (1·15–1·24)	387	1·18 (1·07–1·30)	630	2·56 (2·37–2·77)
≥25	233	4·03 (3·54–4·59)	1079	1·65 (1·56–1·76)	1001	1·29 (1·21–1·37)	141	1·26 (1·06–1·49)	263	3·11 (2·75–3·51)
Men in rural areas										
Smoker category										
Never-regular smoker	258	1·00 (0·88–1·13)	2425	1·00 (0·96–1·04)	2612	1·00 (0·96–1·04)	1037	1·00 (0·94–1·06)	1421	1·00 (0·95–1·05)
Ever-regular smoker	2215	2·23 (1·96–2·55)	9668	1·19 (1·14–1·25)	9429	1·11 (1·06–1·16)	3649	0·99 (0·92–1·06)	7799	1·52 (1·44–1·61)
Ex-smoker (by choice)*	137	1·56 (1·32–1·84)	723	1·11 (1·03–1·20)	738	0·99 (0·92–1·07)	271	0·95 (0·85–1·08)	470	1·14 (1·04–1·25)
Regular smoker	2078	2·30 (2·20–2·41)	8945	1·20 (1·18–1·23)	8691	1·12 (1·09–1·14)	3378	0·99 (0·96–1·03)	7329	1·56 (1·52–1·59)
Ex-smoker (ill health)	219	1·97 (1·72–2·25)	1187	1·51 (1·42–1·59)	1160	1·31 (1·24–1·39)	432	1·18 (1·08–1·30)	1218	2·25 (2·12–2·38)
Current smoker	1859	2·35 (2·24–2·47)	7758	1·16 (1·14–1·19)	7531	1·09 (1·07–1·12)	2946	0·97 (0·93–1·00)	6111	1·46 (1·43–1·50)
Age participant began smoking†										
≥25 years	526	1·71 (1·57–1·86)	2824	1·09 (1·05–1·14)	2802	1·06 (1·02–1·10)	1151	0·94 (0·89–1·00)	1926	1·22 (1·17–1·28)
18–24 years	1096	2·42 (2·28–2·57)	4459	1·22 (1·18–1·26)	4356	1·13 (1·09–1·16)	1612	0·99 (0·94–1·04)	3744	1·65 (1·60–1·70)
<18 years	456	3·13 (2·85–3·43)	1662	1·37 (1·31–1·44)	1533	1·21 (1·15–1·27)	615	1·09 (1·01–1·18)	1659	1·97 (1·87–2·07)
Number of cigarettes smoked per day†										
<15	524	1·70 (1·55–1·85)	3459	1·16 (1·12–1·20)	3531	1·08 (1·05–1·12)	1420	1·04 (0·99–1·10)	2234	1·43 (1·37–1·49)
15–24	994	2·40 (2·26–2·56)	3798	1·20 (1·16–1·24)	3614	1·12 (1·09–1·16)	1342	0·92 (0·88–0·98)	3313	1·61 (1·55–1·67)
≥25	560	3·27 (3·00–3·57)	1688	1·31 (1·25–1·38)	1546	1·21 (1·15–1·28)	616	1·01 (0·93–1·10)	1782	1·68 (1·60–1·76)

All analyses were stratified by study area and 5-year age-at-risk group and adjusted for education and alcohol drinking. HR=hazard ratio. *Includes only people who quit smoking voluntarily (ie, not due to ill health). †Includes current smokers and ex-smokers who had quit smoking due to ill health.

Table 2: **Adjusted HRs for incident risks of five major diseases associated with smoking, among men residing in urban and rural areas**

mortality (HR 1·61, 95% CI 1·55–1·66) and morbidity (1·25, 1·22–1·27; appendix p 37), which, although decreasing gradually, persisted beyond 15 years after quitting at baseline (appendix p 42). By contrast, people who had voluntarily stopped smoking (ie, before developing major diseases) had only small excess risks for overall mortality (1·06, 95% CI 1·01–1·11) and morbidity (1·05, 1·03–1·08), with the risks approaching those among never-smokers about 5–10 years after quitting (appendix p 42). Similarly, contrasting differences between the two groups of ex-smokers were also observed for the top-five disease incidence and mortality (table 2).

When considering all FDR-adjusted significant positive and inverse associations simultaneously, smoking was associated with an additional 13 incidence events (ie, death or hospitalisation; 20 for men and ten for women) for each one event prevented, corresponding to a net absolute excess of 1217 events (1473 for men and 1046 for women) per 100 000 person-years (figure 3). For mortality, smoking was associated with a net excess of 230 deaths (283 for men and 194 for women) per 100 000 person-years. Overall in this population, 19·6% of deaths in men (24·3% of men residing in urban settings and 16·2% of men residing in rural settings) and 2·8% of deaths in women at ages 35–84 years could be attributed to smoking if all the FDR-adjusted significant associations were causal.

Discussion

This study provided the first comprehensive assessment of the long-term health effects of tobacco smoking on a

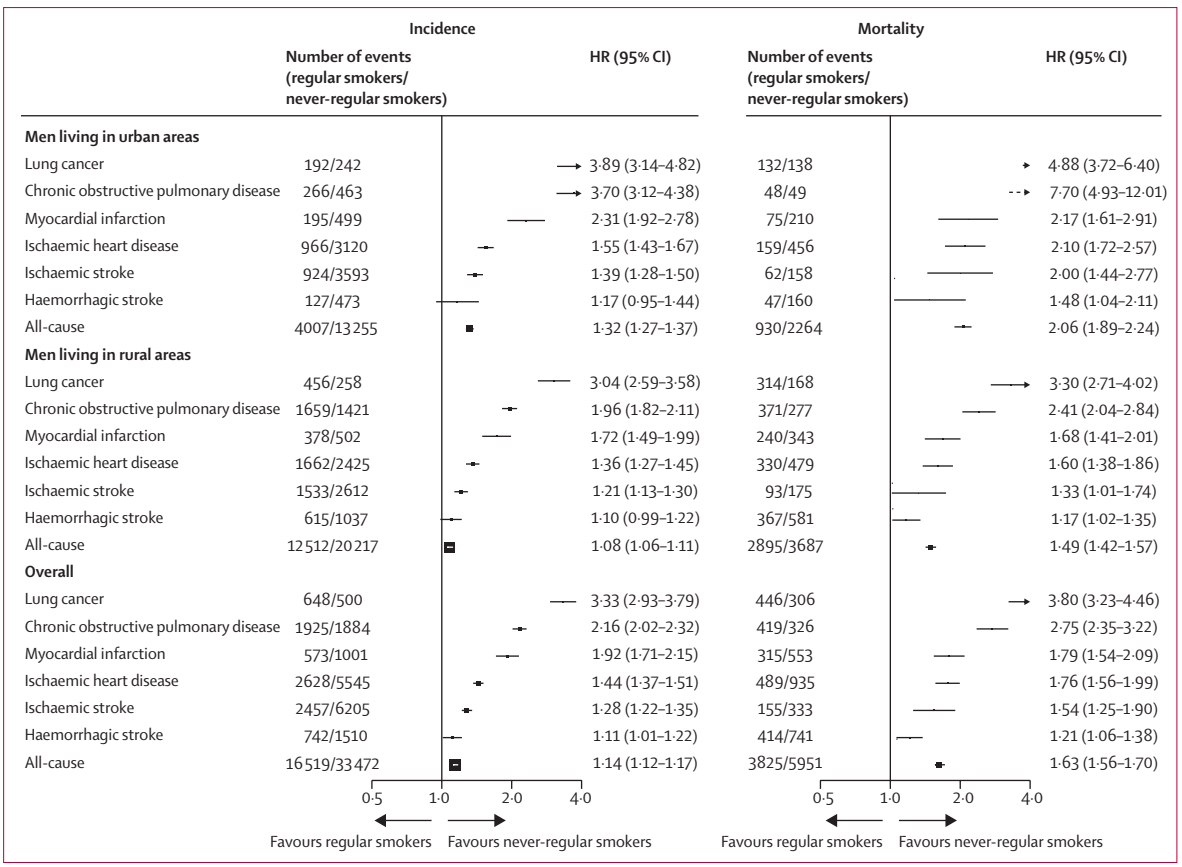


Figure 2: **Adjusted HRs for risks of selected major disease incidence and mortality in men who started smoking before age 18 years**
HRs were stratified by age-at-risk (5-year groups), sex, and study area and were adjusted for education and alcohol drinking. All analyses were restricted to age-at-risk range of 35–84 years. The solid boxes represent HRs, with the size inversely proportional to the variance of the logarithm of the HR, and the horizontal lines represent 95% CIs. The dashed line indicates that the lower 95% CI is beyond the upper limit of the x-axis. HR=hazard ratio.

wide range of diseases in adult men and women in China. Overall, ever-regular smoking was significantly associated with higher risks of 22 causes of death and 56 individual diseases across all major organ systems, as well as more episodes and longer durations of hospitalisation, than was never-regular smoking. The associations were stronger in urban than in rural areas, in people who started smoking at a younger age than in people who started at an older age, and in those who smoked a larger amount of tobacco than those who smoked less tobacco. Although the HRs associated with smoking were still modest for most diseases, among men who lived in urban areas who started smoking before age 18 years, the HR for overall mortality already approached those observed in high-income populations in Europe and North America where most smokers started during young adulthood. Furthermore, we also showed that stopping smoking before the onset of major illness is beneficial.

Decades of epidemiological studies in many high-income countries, such as the UK and the USA, have shown substantial hazards of tobacco smoking.[5,7] With a few exceptions,[37–39] most previous studies were unable to investigate simultaneously the associations with a broad range of diseases. By synthesising evidence from various sources, the 2014 US Surgeon General Report concluded that 26 diseases were likely to be causally associated with smoking.[5] The Global Burden of Diseases, Injuries, and Risk Factors Study (GBD) 2019 concluded that 39 diseases (and nine causes of injuries) were most likely associated with smoking (appendix p 43).[7] For many other diseases (eg, intracerebral haemorrhage, heart failure, influenza, or gastrointestinal conditions), the existing evidence is less conclusive, coming mainly from case-control or small cohort studies.[10,40,41] In China and other LMICs, most existing prospective evidence was on mortality from several major diseases.[6,12–14,42] In this large prospective study, we identified smoking to be associated with increased mortality from 22 causes and increased morbidity from 56 conditions, including many that are still largely understudied in LMICs (eg, intracerebral haemorrhage, heart failure, aortic aneurysm, and peptic ulcer).[5] Most of these associations overlap with those reported consistently in previous cohort studies of high-income populations.

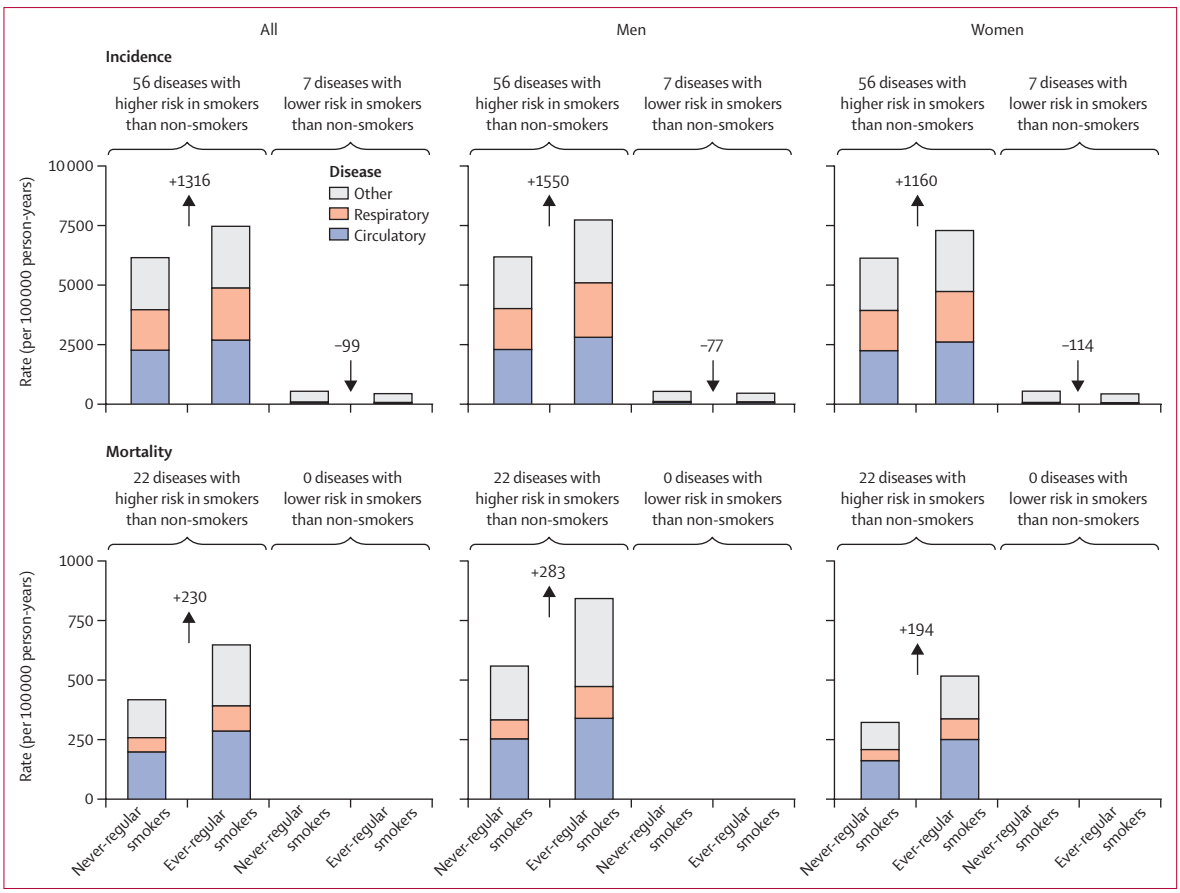


Figure 3: **Incidence and mortality rates from all diseases with an false-discovery-rate-adjusted significant association with ever-regular smoking**
The bar diagrams indicate the overall absolute morbidity and mortality rates per 100 000 person-years at age-at-risk 35–84 years in never-regular versus ever-regular smokers, overall and in men and women separately. The morbidity analyses included 56 diseases showing positive associations with smoking and seven showing inverse associations. The mortality analyses were based on the 22 causes of death showing significant positive associations with smoking. Upward arrows indicate the excess rates associated with ever-regular smoking. Downwards arrows indicate the decreased rates associated with ever-regular smoking.

Long-term prospective studies in high-income countries, especially those in Europe and North America, have also clearly shown the long delay between widespread uptake of cigarette smoking in the young adult population, first in men then in women, and subsequent rising morbidity and mortality risks.[43] In these high-income populations where the smoking epidemic has matured, most adult smokers started at a young age, smoked a large amount, and persistently consumed manufactured cigarettes (as opposed to traditional tobacco products) for decades, which are the hallmarks of high-risk smoking patterns for disease. Therefore, contemporary studies in those populations could reliably capture the full effects of long-term cigarette smoking.[37,44] By contrast, as the widespread uptake and peak of cigarette smoking in China was much more recent, adult smokers in CKB who were born mainly before the 1970s had a later starting age, consumed fewer cigarettes daily, and used more traditional tobacco products (which are less harmful than cigarettes) than did smokers in high-income countries[6,12] or younger smokers (ie, people born after the 1970s) in China. These differences in smoking patterns most likely explain the generally weaker associations in CKB than those reported in studies of high-income countries in Europe and North America.[37,38] Our findings are, however, highly consistent with previous studies in China[6,12,13] and other LMICs where the widespread use of cigarettes is relatively recent.[12,13,15]

Similarly, the greater excess risks for most diseases in male smokers in urban areas compared with those in rural areas most likely reflect a more advanced smoking epidemic in urban than in rural China, as cigarettes had been less available and affordable in rural areas until recent decades.[45] The less mature smoking epidemic in rural areas of China can also partly explain the absence of apparent associations in this study for some conditions that are known to be linked, albeit modestly, to smoking (eg, prostate cancer, colorectal cancer, and dementia). Likewise, the elevated background disease risk for many diseases, especially in rural areas, due to greater exposure to traditional risk factors (eg, household air pollution and

chronic infection) might also have a role. For some diseases (eg, intracerebral haemorrhage) that are much less common in high-income populations than in low-income populations, previous studies have reported mixed findings.[40] With a much larger number of well characterised intracerebral haemorrhage cases (>10 000; >80% confirmed by imaging), we identified significantly elevated risks of mortality and incidence, particularly among men living in urban areas.

Among women, despite the low smoking prevalence (<3%) and intensity among smokers, we identified significant excess risks of 20 distinct conditions. With a few exceptions (eg, lung cancer and chronic obstructive pulmonary disease), the HRs for most disease-specific morbidities, but not necessarily mortalities, were more modest in female than in male smokers. The absence of apparent sex difference in HRs for lung cancer and chronic obstructive pulmonary disease corroborates previous findings,[38,46] suggesting that women's respiratory systems are more susceptible to the harm of smoking than men's. Women's susceptibility might partly reflect the potential protective role of oestrogen in these diseases and the anti-oestrogenic effects of tobacco smoke,[47] as well as genetic predisposition of early-onset chronic obstructive pulmonary disease and lung cancer in women.[46] For breast cancer, we did not find any significant associations, consistent with previous Chinese studies[48] but not with those in high-income populations, which generally showed positive association.[5,37] However, the recorded case numbers among smokers in this and other Chinese studies were small.

Previous high-income studies have also reported lower risks of some diseases among smokers than among non-smokers, particularly Parkinson's disease and endometrial cancer.[5] For Parkinson's disease, we also identified a significantly lower risk among smokers. For the other conditions that have previously been reported as inversely associated with smoking, we identified no clear associations with smoking, but the case numbers were small (eg, only 15 cases of corpus uteri cancer in female smokers). For six other conditions, however, we identified lower risks among smokers, but previous evidence was scarce. Among these, varicose veins, inguinal hernia, gonarthrosis, and other arthrosis are all strongly linked to family history that cannot be controlled for in CKB,[49–53] and all except inguinal hernia have been associated with higher BMI,[49–53] whereas smoking is known to have a weight reduction effect.[54] Further adjustment for BMI, occupation, and self-rated health in sensitivity analyses, however, did not substantially change the associations, suggesting no strong residual confounding by these factors. Bronchiectasis is associated with severe lung infections, asthma, and cystic fibrosis but is not known to be linked to smoking.[49] Reverse causation bias can arise from the inclusion of individuals with pre-existing disease who were less likely to start smoking or more likely to quit early in life than those without disease. Similar bias can also apply to conjunctiva disorders, which tend to develop during early adulthood with recurrent irritating symptoms or even vision impairment. Sensitivity analyses to exclude the first 5 years of follow-up, however, did not substantially alter the results (appendix p 44). Nonetheless, the inverse associations could also represent chance findings, given the large number of tests conducted. With the scarcity of previous evidence, our findings could be considered only as hypothesis generating and require further verifications in other studies.

Our study has several strengths, including a large sample size, reliable assessment of smoking exposure, completeness of follow-up, and the broad range of diseases included. However, it also has limitations. First, for rare conditions (eg, leukaemia and peripheral artery disease) the case numbers were small, especially for women. Second, we assessed previous medical history of only approximately 20 major diseases, so we were unable to account fully for reverse causality for many other outcomes. Third, the use of FDR adjustment might have obscured any true but modest associations for some conditions (eg, hyperthyroidism) that have been associated with smoking. Fourth, although the baseline smoking habits had been objectively validated,[24] subsequent changes in smoking patterns (eg, quitting) could yield underestimated HRs. In the 2013–14 resurvey of a random subset of approximately 25 000 surviving participants, we identified that about a fifth of smokers stopped after the baseline survey. Despite the knowledge of changes over time, we could not directly assess the probable resulting underestimation of risks associated with changes in smoking patterns. Fifth, the record linkage might miss milder (eg, influenza) or under-diagnosed conditions (eg, chronic obstructive pulmonary disease and dementia), resulting in underestimation of the disease burden associated with smoking. Sixth, CKB was not nationally representative, and the study participation was voluntary, so healthy volunteer bias is inevitable. However, the large sample size, diverse areas covered, heterogeneity of exposure, and highly consistent smoking patterns with those reported in other representative surveys in China[6,19] mean that our HRs for smoking could still be largely generalisable to the Chinese population.[55,56] Finally, unlike for mortality, we were not able to estimate the total disease morbidity burden attributed to smoking in China, due to the absence of nationally representative disease incidence data and regional variations in accessing health services.

Various estimates have been made about the smoking-attributed mortality burden in China. We showed that smoking now accounts for 19·6% of male deaths and 2·8% of female deaths at age 35–84 years. These estimations would translate into more than 1 million total deaths per year in China, in line with our previous

projection[6] but much lower than that estimated in GBD 2019 (ie, 2·1 million men and 0·3 million women).[7] These estimates would suggest that about half of all adult male deaths in China (approximately 4·5 million in 2019) could be attributed to smoking. The GBD report used relative risk estimates for 36 causes of death (as opposed to all-cause mortality in this study) derived from pooled analyses of mainly high-income cohort and case-control studies.[7] Although various adjustments were made,[7] they did not appear to fully account for the delayed effects and large urban–rural differences of cigarette uptake in China. Nevertheless, our study provided reliable evidence that if the current trends in smoking persist, then the future smoking-attributed disease burden is likely to increase markedly in Chinese men, whereas in women the burden will stay low and continue to decline.[6]

In summary, this study showed substantial hazards of smoking from a wide range of conditions among Chinese men and women. Given the delayed effects, the future risk per individual smoker and overall disease burden attributed to smoking among adult Chinese men will be much greater than what is currently estimated, especially in those who were born in the 1970s–80s, who reached adulthood when nationwide cigarette consumption was high. As shown in this study and many previous studies,[6,37] stopping smoking before the onset of major illness is beneficial, and widespread smoking cessation, facilitated through increased tobacco tax, effective package warnings, and cessation clinics and helplines,[57] offers China one of the most effective strategies to control the rising burden of chronic diseases over the next few decades. Future research should continue to monitor the growing health effects of smoking in China, especially among men born after 1970. Our work also emphasised the importance of seeking evidence independently in other LMICs, many of which have a similarly increasing tobacco epidemic as in China.

Contributors

ZC, RCl, RCo, RP, CW, JC, and LL conceived and designed the study. YG, YC, HD, LY, IYM, PP, JW, and CY acquired the data. NW analysed the data. KHC and ZC drafted the initial manuscript. ZC, RP, CW, and LL supervised the study. DX, IT, DA, CY, SG, JL, CK, and RCl provided administrative, technical, clinical, or material support. ZC, KHC, NW, and LL accessed and verified the data. All authors reviewed and revised the manuscript. All authors had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Declaration of interests

We declare no competing interests.

Data sharing

The CKB is a global resource for the investigation of lifestyle, environmental, blood biochemical, and genetic factors as determinants of common diseases. The CKB collaborator group is committed to making the cohort data available to the scientific community worldwide to advance knowledge about the causes, prevention, and treatment of disease. For detailed information on what data are currently available to open access users and how to apply for data, visit http://www.ckbiobank.org/site/Data+Access. Researchers who are interested in obtaining the raw data from the CKB study that underlines this paper should contact ckbaccess@ndph.ox.ac.uk. A research proposal will be requested to ensure that any analysis is performed by bona fide researchers and—where data are not currently available to open access researchers—is restricted to the topic covered in this Article.

Acknowledgments

The CKB baseline survey and the first re-survey were supported by the Kadoorie Charitable Foundation in Hong Kong. The long-term continuation of the study and research activities has been supported by grants from the National Natural Science Foundation of China (82192900, 82192901, 82192904, and 91846303), the National Key Research and Development Program of China (2016YFC0900500, 2016YFC0900501, 2016YFC0900504, and 2016YFC1303904), and Wellcome to Oxford University (212946/Z/18/Z, 202922/Z/16/Z, 104085/Z/14/Z, and 088158/Z/09/Z). The UK Medical Research Council (MC_UU_00017/1, MC_UU_12026/2, and MC_U137686851), Cancer Research UK (C16077/A29186 and C500/A16896), and the British Heart Foundation (CH/1996001/9454) provide core funding to the Clinical Trial Service Unit and Epidemiological Studies Unit at the University of Oxford for the project. KHC acknowledges support from the British Heart Foundation Centre of Research Excellence, University of Oxford (RE/18/3/34214).

References

1 Eriksen M, Mackay J, Schluger N, Gomeshtapeh FI, Drope J. Smoking among women. The tobacco atlas, 5th edn. Atlanta, GA: American Cancer Society, 2015.
2 Chinese Center for Disease Control and Prevention. China adult tobacco survey report 2018. Nov 26, 2020. https://extranet.who.int/ncdsmicrodata/index.php/catalog/803 (accessed Jan 12, 2022).
3 Doll R, Peto R. The causes of cancer: quantitative estimates of avoidable risks of cancer in the United States today. *J Natl Cancer Inst* 1981; **66:** 1191–308.
4 Wald N, Nicolaides-Bouman A. UK smoking statistics. London: Wolfson Institute of Preventive Medicine, 1991.
5 US Department of Health and Human Services. The health consequences of smoking—50 years of progress: a report of the Surgeon General. Atlanta, GA: US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health, 2014.
6 Chen Z, Peto R, Zhou M, et al. Contrasting male and female trends in tobacco-attributed mortality in China: evidence from successive nationwide prospective cohort studies. *Lancet* 2015; **386:** 1447–56.
7 Reitsma MB, Kendrick PJ, Ababneh E, et al. Spatial, temporal, and demographic patterns in prevalence of smoking tobacco use and attributable disease burden in 204 countries and territories, 1990–2019: a systematic analysis from the Global Burden of Disease Study 2019. *Lancet* 2021; **397:** 2337–60.
8 Scott JG, Matuschka L, Niemelä S, Miettunen J, Emmerson B, Mustonen A. Evidence of a causal relationship between smoking tobacco and schizophrenia spectrum disorders. *Front Psychiatry* 2018; **9:** 607.
9 Xia J, Wang L, Ma Z, et al. Cigarette smoking and chronic kidney disease in the general population: a systematic review and meta-analysis of prospective cohort studies. *Nephrol Dial Transplant* 2017; **32:** 475–87.
10 Lawrence H, Hunter A, Murray R, Lim WS, McKeever T. Cigarette smoking and the occurrence of influenza—systematic review. *J Infect* 2019; **79:** 401–06.
11 Barbhaiya M, Tedeschi SK, Lu B, et al. Cigarette smoking and the risk of systemic lupus erythematosus, overall and by anti-double stranded DNA antibody subtype, in the Nurses' Health Study cohorts. *Ann Rheum Dis* 2018; **77:** 196–202.
12 Yang JJ, Yu D, Wen W, et al. Tobacco smoking and mortality in Asia: a pooled meta-analysis. *JAMA Netw Open* 2019; **2:** e191474.
13 Yang JJ, Yu D, Shu XO, et al. Quantifying the association of low-intensity and late initiation of tobacco smoking with total and cause-specific mortality in Asia. *Tob Control* 2021; **30:** 328–35.
14 Thomson B, Rojas NA, Lacey B, et al. Association of childhood smoking and adult mortality: prospective study of 120 000 Cuban adults. *Lancet Glob Health* 2020; **8:** e850–57.
15 Sathish T, Teo KK, Britz-McKibbin P, et al. Variations in risks from smoking between high-income, middle-income, and low-income countries: an analysis of data from 179 000 participants from 63 countries. *Lancet Glob Health* 2022; **10:** e216–26.

16 Wang B, Xiao D, Wang C. Smoking and chronic obstructive pulmonary disease in Chinese population: a meta-analysis. *Clin Respir J* 2015; **9:** 165–75.

17 Yu K, Zhu S, He M, et al. Epidemiological characteristics of 561 cases of intracerebral hemorrhage in Chengdu, China. *Medicine (Baltimore)* 2021; **100:** e24952.

18 Zhang R, Li H, Li N, et al. Risk factors for gastric cancer: a large-scale, population-based case-control study. *Chin Med J (Engl)* 2021; **134:** 1952–58.

19 Wang M, Luo X, Xu S, et al. Trends in smoking prevalence and implication for chronic diseases in China: serial national cross-sectional surveys from 2003 to 2013. *Lancet Respir Med* 2019; **7:** 35–45.

20 Pan A, Wang Y, Talaei M, Hu FB, Wu T. Relation of active, passive, and quitting smoking with incident type 2 diabetes: a systematic review and meta-analysis. *Lancet Diabetes Endocrinol* 2015; **3:** 958–67.

21 Chen Z, Lee L, Chen J, et al. Cohort profile: the Kadoorie Study of Chronic Disease in China (KSCDC). *Int J Epidemiol* 2005; **34:** 1243–49.

22 Chen Z, Chen J, Collins R, et al. China Kadoorie Biobank of 0.5 million people: survey methods, baseline characteristics and long-term follow-up. *Int J Epidemiol* 2011; **40:** 1652–66.

23 Yang GH, Stroup DF, Thacker SB. National public health surveillance in China: implications for public health in China and the United States. *Biomed Environ Sci* 1997; **10:** 1–13.

24 Zhang Q, Li L, Smith M, et al. Exhaled carbon monoxide and its associations with smoking, indoor household air pollution and chronic respiratory diseases among 512,000 Chinese adults. *Int J Epidemiol* 2013; **42:** 1464–75.

25 WHO. ICD-10: international statistical classification of diseases and related health problems: tenth revision, 2nd ed. https://apps.who.int/iris/handle/10665/42980 (accessed Oct 18, 2022).

26 Murray CJL, Aravkin AY, Zheng P, et al. Global burden of 87 risk factors in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet* 2020; **396:** 1223–49.

27 San Hong L, Lewington S. Lexis expansion—age-at-risk adjustment for survival analysis. https://www.lexjansen.com/phuse/2013/sp/SP09.pdf (accessed Oct 18, 2022).

28 Lau B, Cole SR, Gange SJ. Competing risk regression models for epidemiologic data. *Am J Epidemiol* 2009; **170:** 244–56.

29 Plummer M. Improved estimates of floating absolute risk. *Stat Med* 2004; **23:** 93–104.

30 Grambsch PM, Therneau TM. Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika* 1994; **81:** 515–26.

31 Dong H, Robison LL, Leisenring WM, Martin LJ, Armstrong GT, Yasui Y. Estimating the burden of recurrent events in the presence of competing risks: the method of mean cumulative count. *Am J Epidemiol* 2015; **181:** 532–40.

32 Ghosh D, Lin DY. Nonparametric analysis of recurrent events and death. *Biometrics* 2000; **56:** 554–62.

33 Cook RJ, Lawless JF. Marginal analysis of recurrent events and a terminating event. *Stat Med* 1997; **16:** 911–24.

34 Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Series B Stat Methodol* 1995; **57:** 289–300.

35 Braun JM, Kalloo G, Kingsley SL, Li N. Using phenome-wide association studies to examine the effect of environmental exposures on human health. *Environ Int* 2019; **130:** 104877.

36 Glickman ME, Rao SR, Schultz MR. False discovery rate control is a recommended alternative to Bonferroni-type adjustments in health studies. *J Clin Epidemiol* 2014; **67:** 850–57.

37 Pirie K, Peto R, Reeves GK, Green J, Beral V. The 21st century hazards of smoking and benefits of stopping: a prospective study of one million women in the UK. *Lancet* 2013; **381:** 133–41.

38 Carter BD, Abnet CC, Feskanich D, et al. Smoking and mortality—beyond established causes. *N Engl J Med* 2015; **372:** 631–40.

39 Lariscy JT, Hummer RA, Rogers RG. Cigarette smoking and all-cause and cause-specific adult mortality in the United States. *Demography* 2018; **55:** 1855–85.

40 Cho S, Rehni AK, Dave KR. Tobacco use: a major risk factor of intracerebral hemorrhage. *J Stroke* 2021; **23:** 37–50.

41 Lee H, Son YJ. Influence of smoking status on risk of incident heart failure: a systematic review and meta-analysis of prospective cohort studies. *Int J Environ Res Public Health* 2019; **16:** E2697.

42 Gajalakshmi V, Peto R, Kanimozhi VC, Whitlock G, Veeramani D. Cohort profile: the Chennai prospective study of mortality among 500,000 adults in Tamil Nadu, south India. *Int J Epidemiol* 2007; **36:** 1190–95.

43 Thun MJ, Carter BD, Feskanich D, et al. 50-year trends in smoking-related mortality in the United States. *N Engl J Med* 2013; **368:** 351–64.

44 Di Cicco ME, Ragazzo V, Jacinto T. Mortality in relation to smoking: the British Doctors Study. *Breathe (Sheff)* 2016; **12:** 275–76.

45 Huang X. The China tobacco history. Beijing: China Press, 2006.

46 Gut-Gobert C, Cavaillès A, Dixmier A, et al. Women and COPD: do we need more evidence? *Eur Respir Rev* 2019; **28:** 180055.

47 Tankó LB, Christiansen C. An update on the antiestrogenic effect of smoking: a literature review with implications for researchers and practitioners. *Menopause* 2004; **11:** 104–09.

48 Chen C, Huang YB, Liu XO, et al. Active and passive smoking with breast cancer risk for Chinese females: a systematic review and meta-analysis. *Chin J Cancer* 2014; **33:** 306–16.

49 Chalmers JD, Chang AB, Chotirmall SH, Dhar R, McShane PJ. Bronchiectasis. *Nat Rev Dis Primers* 2018; **4:** 45.

50 Fukaya E, Flores AM, Lindholm D, et al. Clinical and genetic determinants of varicose veins. *Circulation* 2018; **138:** 2869–80.

51 Sorensen LT, Friis E, Jorgensen T, et al. Smoking is a risk factor for recurrence of groin hernia. *World J Surg* 2002; **26:** 397–400.

52 Hemberg A, Holmberg H, Norberg M, Nordin P. Tobacco use is not associated with groin hernia repair, a population-based study. *Hernia* 2017; **21:** 517–23.

53 Veronese N, Maggi S. Remaining debates on osteoarthrosis prevention. In: Michel J-P, ed. Prevention of chronic diseases and age-related disability. Cham: Springer International Publishing, 2019: 111–15.

54 Chiolero A, Faeh D, Paccaud F, Cornuz J. Consequences of smoking for body weight, body fat distribution, and insulin resistance. *Am J Clin Nutr* 2008; **87:** 801–09.

55 Manolio TA, Collins R. Enhancing the feasibility of large cohort studies. *JAMA* 2010; **304:** 2290–91.

56 Rothman KJ, Gallacher JE, Hatch EE. Why representativeness should be avoided. *Int J Epidemiol* 2013; **42:** 1012–14.

57 US Department of Health and Human Services. Smoking cessation: a report of the Surgeon General. Rockville, MD: US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health, 2020.