









OPEN

Health effects associated with smoking: a Burden of Proof study

Xiaochen Dai^{1,2}  , Gabriela F. Gil¹, Marissa B. Reitsma¹, Noah S. Ahmad¹, Jason A. Anderson¹, Catherine Bisignano¹, Sinclair Carr¹, Rachel Feldman¹, Simon I. Hay^{1,2} , Jiawei He^{1,2}, Vincent Iannucci¹, Hilary R. Lawlor¹, Matthew J. Malloy¹, Laurie B. Marczak¹, Susan A. McLaughlin¹, Larissa Morikawa¹ , Erin C. Mullany¹, Sneha I. Nicholson¹, Erin M. O'Connell¹, Chukwuma Okereke¹, Reed J. D. Sorensen¹, Joanna Whisnant¹, Aleksandr Y. Aravkin^{1,3}, Peng Zheng^{1,2}, Christopher J. L. Murray^{1,2}  and Emmanuela Gakidou^{1,2} 

As a leading behavioral risk factor for numerous health outcomes, smoking is a major ongoing public health challenge. Although evidence on the health effects of smoking has been widely reported, few attempts have evaluated the dose–response relationship between smoking and a diverse range of health outcomes systematically and comprehensively. In the present study, we re-estimated the dose–response relationships between current smoking and 36 health outcomes by conducting systematic reviews up to 31 May 2022, employing a meta-analytic method that incorporates between-study heterogeneity into estimates of uncertainty. Among the 36 selected outcomes, 8 had strong-to-very-strong evidence of an association with smoking, 21 had weak-to-moderate evidence of association and 7 had no evidence of association. By overcoming many of the limitations of traditional meta-analyses, our approach provides comprehensive, up-to-date and easy-to-use estimates of the evidence on the health effects of smoking. These estimates provide important information for tobacco control advocates, policy makers, researchers, physicians, smokers and the public.

Among both the public and the health experts, smoking is recognized as a major behavioral risk factor with a leading attributable health burden worldwide. The health risks of smoking were clearly outlined in a canonical study of disease rates (including lung cancer) and smoking habits in British doctors in 1950 and have been further elaborated in detail over the following seven decades^{1,2}. In 2005, evidence of the health consequences of smoking galvanized the adoption of the first World Health Organization (WHO) treaty, the Framework Convention on Tobacco Control, in an attempt to drive reductions in global tobacco use and second-hand smoke exposure³. However, as of 2020, an estimated 1.18 billion individuals globally were current smokers and 7 million deaths and 177 million disability-adjusted life-years were attributed to smoking, reflecting a persistent public health challenge⁴. Quantifying the relationship between smoking and various important health outcomes—in particular, highlighting any significant dose–response relationships—is crucial to understanding the attributable health risk experienced by these individuals and informing responsive public policy.

Existing literature on the relationship between smoking and specific health outcomes is prolific, including meta-analyses, cohort studies and case–control studies analyzing the risk of outcomes such as lung cancer^{5–7}, chronic obstructive pulmonary disease (COPD)^{8–10} and ischemic heart disease^{11–14} due to smoking. There are few if any attempts, however, to systematically and comprehensively evaluate the landscape of evidence on smoking risk across a diverse range of health outcomes, with most current research focusing on risk or attributable burden of smoking for a specific condition^{7,15}, thereby missing the opportunity to provide a comprehensive

picture of the health risk experienced by smokers. Furthermore, although evidence surrounding specific health outcomes, such as lung cancer, has generated widespread consensus, findings about the attributable risk of other outcomes are much more heterogeneous and inconclusive^{16–18}. These studies also vary in their risk definitions, with many comparing dichotomous exposure measures of ever smokers versus nonsmokers^{19,20}. Others examine the distinct risks of current smokers and former smokers compared with never smokers^{21–23}. Among the studies that do analyze dose–response relationships, there is large variation in the units and dose categories used in reporting their findings (for example, the use of pack-years or cigarettes per day)^{24,25}, which complicates the comparability and consolidation of evidence. This, in turn, can obscure data that could inform personal health choices, public health practices and policy measures. Guidance on the health risks of smoking, such as the *Surgeon General's Reports* on smoking^{26,27}, is often based on experts' evaluation of heterogeneous evidence, which, although extremely useful and well suited to carefully consider nuances in the evidence, is fundamentally subjective.

The present study, as part of the Global Burden of Diseases, Risk Factors, and Injuries Study (GBD) 2020, re-estimated the continuous dose–response relationships (the mean risk functions and associated uncertainty estimates) between current smoking and 36 health outcomes (Supplementary Table 1) by identifying input studies using a systematic review approach and employing a meta-analytic method²⁸. The 36 health outcomes that were selected based on existing evidence of a relationship included 16 cancers (lung cancer, esophageal cancer, stomach cancer, leukemia, liver cancer, laryngeal cancer, breast cancer, cervical cancer, colorectal

¹Institute for Health Metrics and Evaluation, University of Washington, Seattle, WA, USA. ²Department of Health Metrics Sciences, School of Medicine, University of Washington, Seattle, WA, USA. ³Department of Applied Mathematics, University of Washington, Seattle, WA, USA. ✉e-mail: xdai88@uw.edu

Table 1 | Policy summary

Background	There is widespread evidence that smoking is a leading behavioral risk factor for numerous health outcomes; despite this evidence, smoking remains a persistent public health challenge. Although there is substantial research on the subject, few attempts have been made to evaluate the dose–response relationships between smoking and its many potential health outcomes in a systematic and comprehensive way or to assess the strength of the evidence for these relationships. Due to limited assessments of this kind, current guidance on and regulation in response to the risks of smoking are often based on expert evaluation of heterogeneous evidence. Expert committees are valuable in large part because they are able to consider the nuances of the available evidence, but they are also inherently subjective. In the present study, we used a systematic review approach and employed a meta-analytic framework ²⁹ to assess all available evidence in a way that did not force log(linearity), more fully incorporated between-study heterogeneity into estimates of uncertainty and addressed many of the other limitations of existing meta-analytic approaches.
Main findings and limitations	<p>Even when between-study heterogeneity and other forms of uncertainty were incorporated into our assessment of risk—leading to a very conservative interpretation of the available data—smoking had a strong-to-very-strong (>50% increase in risk) association with 8 of the 36 outcomes selected for this analysis. A further 24 outcomes had weak-to-moderate evidence of an association (including 1 outcome for which smoking was protective), whereas 4 had no evidence of an association once between-study heterogeneity had been incorporated. On a 5-star scale with 1 suggesting no evidence of association (no increase in risk) and 5 very strong evidence of association (>85% increase in risk), the 8 highest pairs received 4–5 stars, the next 24 received 2–3 stars and the final 4 received just 1 star. These findings confirm that smoking is irrefutably highly harmful to human health.</p> <p>Limitations of the present study include the fact that the bias covariates we used were based on observable study characteristics and thus may not fully capture characteristics such as study quality or context; our approach to selecting which risk estimates from a given study should be included, if multiple estimates with different adjustment levels were reported, limited our ability to make full use of all available risk estimates in the literature, and we did not test for additional forms of bias such as whether studies are more consistent with each other than expected by chance.</p>
Policy implications	The available evidence demonstrates that smoking is a highly harmful risk factor for a wide array of serious health outcomes, most notably laryngeal cancer, aortic aneurysm, bladder cancer, lung cancer and other pharynx cancer (excluding nasopharynx cancer). Policy makers should pay particular attention to the 5- and 4-star smoking–outcome pairs, for which the evidence of an association is strongest, but should not ignore lower starred pairs, particularly those with a high prevalence or severity of outcome. Our findings further validate previous conclusions about the high risk of smoking and, using our meta-analytic methods, we offer the added value of minimizing the chance that risk has been overestimated. Policy makers can therefore be confident that smoking recommendations and regulations made based on the findings from the present study are not unnecessarily restrictive.

cancer, lip and oral cavity cancer, nasopharyngeal cancer, other pharynx cancer (excluding nasopharynx cancer), pancreatic cancer, bladder cancer, kidney cancer and prostate cancer), 5 cardiovascular diseases (CVDs: ischemic heart disease, stroke, atrial fibrillation and flutter, aortic aneurysm and peripheral artery disease) and 15 other diseases (COPD, lower respiratory tract infections, tuberculosis, asthma, type 2 diabetes, Alzheimer's disease and related dementias, Parkinson's disease, multiple sclerosis, cataracts, gallbladder diseases, low back pain, peptic ulcer disease, rheumatoid arthritis, macular degeneration and fractures). Definitions of the outcomes are described in Supplementary Table 1. We conducted a separate systematic review for each risk–outcome pair with the exception of cancers, which were done together in a single systematic review. This approach allowed us to systematically identify all relevant studies indexed in PubMed up to 31 May 2022, and we extracted relevant data on risk of smoking, including study characteristics, following a pre-specified template (Supplementary Table 2). The meta-analytic tool overcomes many of the limitations of traditional meta-analyses by incorporating between-study heterogeneity into the uncertainty of risk estimates, accounting for small numbers of studies, relaxing the assumption of log(linearity) applied to the risk functions, handling differences in exposure ranges between comparison groups, and systematically testing and adjusting for bias due to study designs and characteristics. We then estimated the burden-of-proof risk function (BPRF) for each risk–outcome pair, as proposed by Zheng et al.²⁹; the BPRF is a conservative risk function defined as the 5th quantile curve (for harmful risks) that reflects the smallest harmful effect at each level of exposure consistent with the available evidence. Given all available data for each outcome, the risk of smoking is at least as harmful as the BPRF indicates.

We used the BPRF for each risk–outcome pair to calculate risk–outcome scores (ROs) and categorize the strength of evidence for the association between smoking and each health outcome using a star rating from 1 to 5. The interpretation of the star ratings is

as follows: 1 star (*) indicates no evidence of association; 2 stars (**) correspond to a 0–15% increase in risk across average range of exposures for harmful risks; 3 stars (***) represent a 15–50% increase in risk; 4 stars (****) refer to >50–85% increase in risk; and 5 stars (***** equal >85% increase in risk. The thresholds for each star rating were developed in consultation with collaborators and other stakeholders.

The increasing disease burden attributable to current smoking, particularly in low- and middle-income countries⁴, demonstrates the relevance of the present study, which quantifies the strength of the evidence using an objective, quantitative, comprehensive and comparative framework. Findings from the present study can be used to support policy makers in making informed smoking recommendations and regulations focusing on the associations for which the evidence is strongest (that is, the 4- and 5-star associations). However, associations with a lower star rating cannot be ignored, especially when the outcome has high prevalence or severity. A summary of the main findings, limitations and policy implications of the study is presented in Table 1.

Results

Overview. We evaluated the mean risk functions and the BPRFs for 36 health outcomes that are associated with current smoking³⁰ (Table 2). Following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines³¹ for each of our systematic reviews, we identified studies reporting relative risk (RR) of incidence or mortality from each of the 36 selected outcomes for smokers compared with nonsmokers. We reviewed 21,108 records, which were identified to have been published between 1 May 2018 and 31 May 2022; this represents the most recent time period since the last systematic review of the available evidence for the GBD at the time of publication. The meta-analyses reported in the present study for each of the 36 health outcomes are based on evidence from a total of 793 studies published between 1970 and

Table 2 | Strength of the evidence for the relationship between current smoking and the 36 health outcomes analyzed

Risk-outcome pair	Risk unit	Mean risk at different exposure levels				85th percentile risk level	Mean risk at 85th percentile risk level	ROs	Average BPRF	Average increased risk (%)	Star rating	Pub. bias	No. of studies
		5	10	20	40								
Laryngeal cancer	Pack-years	2.30 (1.88, 2.84)	3.77 (2.73, 5.30)	7.25 (4.48, 12.05)	14.62 (7.62, 29.11)	50.50	17.73 (8.82, 37.07)	1.56	4.75	374.95	5	0	5
Aortic aneurism (ref. age: 55–59 years)	Cigarettes per day	2.52 (1.79, 3.60)	3.78 (2.31, 6.33)	5.39 (2.89, 10.36)	6.22 (3.17, 12.64)	30.00	6.08 (3.13, 12.27)	0.92	2.50	149.73	5	0	14
Peripheral artery disease (ref. age: 60–64 years)	Cigarettes per day	2.52 (1.67, 3.90)	3.80 (2.10, 7.14)	5.69 (2.62, 12.94)	7.82 (3.13, 20.68)	31.25	7.16 (2.98, 18.16)	0.86	2.37	136.53	5	0	6
Lung cancer	Pack-years	1.58 (1.19, 2.15)	2.48 (1.40, 4.53)	5.11 (1.84, 14.99)	11.62 (2.49, 58.73)	50.88	13.42 (2.63, 74.59)	0.73	2.07	106.66	5	1	78
Other pharynx cancer	Pack-years	1.65 (1.30, 2.13)	2.20 (1.51, 3.30)	3.02 (1.77, 5.30)	3.89 (2.02, 7.78)	63.75	4.72 (2.24, 10.45)	0.65	1.92	92.26	5	0	8
COPD	Pack-years	1.61 (1.21, 2.17)	2.16 (1.37, 3.51)	3.17 (1.60, 6.55)	5.05 (1.94, 13.97)	49.75	6.01 (2.08, 18.58)	0.54	1.72	72.11	4	1	13
Lower respiratory infection	Cigarettes per day	1.46 (1.23, 1.76)	1.83 (1.38, 2.46)	2.63 (1.68, 4.23)	2.97 (1.79, 5.06)	31.25	2.97 (1.79, 5.06)	0.43	1.54	54.45	4	0	7
Pancreatic cancer	Pack-years	1.26 (1.22, 1.31)	1.47 (1.38, 1.57)	1.72 (1.57, 1.88)	1.80 (1.63, 1.98)	51.25	1.87 (1.69, 2.08)	0.42	1.52	51.66	4	0	19
Bladder cancer	Pack-years	1.21 (1.08, 1.36)	1.43 (1.16, 1.77)	1.90 (1.31, 2.81)	2.92 (1.57, 5.63)	50.65	3.29 (1.65, 6.83)	0.34	1.40	40.18	3	1	30
Tuberculosis	Cigarettes per day	1.46 (1.13, 1.92)	1.88 (1.22, 2.97)	2.71 (1.38, 5.56)	3.47 (1.49, 8.50)	26.56	3.24 (1.46, 7.56)	0.27	1.31	31.04	3	0	19
Esophageal cancer	Pack-years	1.24 (1.06, 1.46)	1.48 (1.11, 2.00)	1.96 (1.20, 3.32)	3.12 (1.36, 7.57)	50.00	4.79 (1.53, 16.26)	0.26	1.29	29.36	3	1	14
Cervical cancer	Pack-years	1.60 (1.12, 2.35)	1.99 (1.18, 3.48)	2.37 (1.23, 4.79)	2.62 (1.26, 5.72)	25.50	2.53 (1.25, 5.36)	0.21	1.24	23.53	3	0	4
Multiple sclerosis	Cigarettes per day	1.14 (1.10, 1.17)	1.31 (1.23, 1.40)	1.78 (1.55, 2.06)	2.77 (2.17, 3.60)	20.00	1.78 (1.55, 2.06)	0.21	1.23	23.36	3	0	6
Rheumatoid arthritis	Cigarettes per day	1.20 (1.10, 1.31)	1.40 (1.20, 1.66)	1.67 (1.31, 2.14)	1.73 (1.34, 2.26)	26.25	1.72 (1.34, 2.25)	0.21	1.23	23.32	3	1	6
Lower back pain	Cigarettes per day	1.58 (1.11, 2.28)	1.96 (1.16, 3.39)	2.26 (1.20, 4.40)	2.29 (1.21, 4.50)	26.25	2.29 (1.21, 4.49)	0.20	1.22	21.84	3	0	6
Ischemic heart disease (ref. age: 55–59 years)	Cigarettes per day	1.66 (1.09, 2.57)	2.12 (1.14, 4.1)	2.43 (1.17, 5.27)	4.20 (1.28, 14.77)	31.25	3.13 (1.22, 8.50)	0.19	1.20	20.39	3	1	60
Peptic ulcer	Cigarettes per day	1.47 (1.11, 1.97)	1.79 (1.17, 2.80)	2.10 (1.22, 3.73)	2.16 (1.23, 3.92)	21.94	2.12 (1.22, 3.80)	0.18	1.20	19.84	3	0	7
Macular degeneration	Cigarettes per day	1.33 (1.07, 1.69)	1.67 (1.12, 2.54)	2.23 (1.20, 4.31)	2.42 (1.22, 5.01)	27.50	2.41 (1.22, 4.99)	0.18	1.19	19.44	3	0	2
Parkinson's disease (protective risk)	Cigarettes per day	0.79 (0.66, 0.93)	0.65 (0.47, 0.88)	0.54 (0.34, 0.84)	0.49 (0.29, 0.81)	26.25	0.51 (0.31, 0.83)	0.16	0.85	−14.88	3	1	14
Stomach cancer	Pack-years	1.10 (1.06, 1.15)	1.18 (1.10, 1.27)	1.29 (1.16, 1.44)	1.39 (1.22, 1.60)	51.13	1.61 (1.32, 1.97)	0.16	1.17	17.39	3	1	13
Stroke (ref. age: 55–59)	Cigarettes per day	1.40 (1.07, 1.88)	1.75 (1.11, 2.82)	2.23 (1.16, 4.45)	2.43 (1.18, 5.22)	29.50	2.42 (1.18, 5.19)	0.16	1.17	16.89	3	0	67
Type 2 diabetes	Cigarettes per day	1.23 (1.09, 1.39)	1.38 (1.15, 1.68)	1.49 (1.18, 1.90)	1.67 (1.24, 2.29)	26.25	1.54 (1.20, 2.01)	0.15	1.16	15.76	3	1	28
Cataracts	Cigarettes per day	1.10 (1.08, 1.12)	1.16 (1.13, 1.19)	1.29 (1.23, 1.35)	1.70 (1.54, 1.88)	25.00	1.43 (1.34, 1.53)	0.14	1.15	15.47	3	0	10
Nasopharyngeal cancer	Pack-years	1.16 (1.04, 1.31)	1.28 (1.06, 1.56)	1.39 (1.08, 1.83)	1.97 (1.17, 3.43)	50.00	2.67 (1.26, 5.94)	0.13	1.14	14.29	2	1	12
Alzheimer's and other dementia	Cigarettes per day	1.16 (1.04, 1.30)	1.29 (1.06, 1.58)	1.46 (1.10, 1.96)	1.74 (1.14, 2.70)	30.00	1.54 (1.11, 2.18)	0.09	1.10	9.70	2	1	8
Gallbladder diseases	Cigarettes per day	1.15 (1.03, 1.30)	1.24 (1.05, 1.49)	1.29 (1.05, 1.61)	1.51 (1.09, 2.14)	27.93	1.41 (1.07, 1.88)	0.06	1.06	6.34	2	0	4

Continued

Table 2 | Strength of the evidence for the relationship between current smoking and the 36 health outcomes analyzed (continued)

Risk-outcome pair	Risk unit	Mean risk at different exposure levels				85th percentile risk level	Mean risk at 85th percentile risk level	ROs	Average BPRF	Average increased risk (%)	Star rating	Pub. bias	No. of studies
		5	10	20	40								
Atrial fibrillation and flutter (ref. age: 55–59 years)	Cigarettes per day	1.34 (1.03, 1.78)	1.40 (1.03, 1.93)	1.40 (1.03, 1.93)	1.40 (1.03, 1.93)	25.00	1.40 (1.03, 1.93)	0.06	1.06	5.67	2	0	5
Lip and oral cavity cancer	Pack-years	1.15 (1.00, 1.34)	1.37 (0.99, 1.94)	2.05 (0.98, 4.50)	3.50 (0.97, 13.93)	49.68	3.91 (0.96, 17.56)	0.05	1.05	4.81	2	0	10
Breast cancer	Pack-years	1.08 (1.02, 1.14)	1.13 (1.04, 1.24)	1.17 (1.04, 1.31)	1.17 (1.04, 1.31)	34.10	1.17 (1.04, 1.31)	0.04	1.04	4.46	2	0	23
Colon and rectum cancer	Pack-years	1.07 (0.99, 1.16)	1.12 (0.98, 1.29)	1.19 (0.97, 1.46)	1.20 (0.97, 1.50)	50.00	1.20 (0.97, 1.50)	−0.01	0.99	N/A	1	0	16
Kidney cancer	Pack-years	1.01 (1.00, 1.02)	1.04 (0.99, 1.09)	1.15 (0.98, 1.36)	1.52 (0.94, 2.53)	45.86	1.59 (0.93, 2.79)	−0.01	0.99	N/A	1	1	18
Leukemia	Pack-years	1.04 (0.98, 1.11)	1.07 (0.97, 1.19)	1.10 (0.95, 1.29)	1.62 (0.79, 3.47)	37.50	1.50 (0.82, 2.80)	−0.04	0.96	N/A	1	0	7
Fracture	Binary	N/A	N/A	N/A	N/A	1 (binary)	1.34 (0.84, 1.97)	−0.05	0.95	N/A	1	0	59
Prostate cancer	Cigarettes per day	1.03 (0.98, 1.10)	1.07 (0.95, 1.21)	1.16 (0.89, 1.53)	1.30 (0.82, 2.10)	29.73	1.25 (0.85, 1.89)	−0.06	0.94	N/A	1	1	22
Liver cancer	Pack-years	1.15 (0.84, 1.59)	1.26 (0.75, 2.19)	1.39 (0.65, 3.10)	1.42 (0.64, 3.30)	62.50	1.42 (0.64, 3.30)	−0.32	0.72	N/A	1	0	11
Asthma	Cigarettes per day	1.41 (0.54, 3.96)	1.59 (0.43, 6.39)	1.61 (0.42, 6.74)	1.64 (0.41, 7.23)	26.25	1.63 (0.41, 7.04)	−0.64	0.53	N/A	1	0	7

The ROS represents the signed value of the log(BPRF) averaged across the 15th–85th percentiles of exposure. The BPRF corresponds to the lower (if harmful) or higher (if protective) UI—inclusive of between-study heterogeneity—for each risk–outcome pair’s RR curve. ROSs are directly comparable across outcomes and each risk–outcome pair receives an ROS based on the final formulation of the risk curve. For Parkinson’s disease, the ROS reflects a protective effect of smoking, whereas for the other outcomes it reflects a harmful effect. Negative ROSs indicate that a conservative interpretation of the available evidence suggests that there may be no association between risk and outcome. For ease of interpretation, we have transformed the ROS and BPRF into a star rating (1–5), with a higher rating representing a larger effect and stronger evidence. Average BPRF, which is the exponential ROS for harmful effects (or exponential negative ROS for protective effects), is the conservative exposure-averaged RR consistent with all the available data. Average increased risk, which equates to (average BPRF – 1) × 100% for harmful effects or (1 – average BPRF) × 100% for protective effects, refers to the percentage increase in RR based on a conservative interpretation of the evidence. For harmful risks, this percentage is positive and, for protective risks, negative, indicating the percentage decrease in RR. The average increased risk is not applicable for pairs with negative ROSs. N/A, not available; Pub., Publication; ref., reference.

2022 (Extended Data Fig. 1–5 and Supplementary Information 1.5 show the PRISMA diagrams for each outcome). Only prospective cohort and case–control studies were included for estimating dose–response risk curves, but cross-sectional studies were also included for estimating the age pattern of smoking risk on cardiovascular and circulatory disease (CVD) outcomes. Details on each, including the study’s design, data sources, number of participants, length of follow-up, confounders adjusted for in the input data and bias covariates included in the dose–response risk model, can be found in Supplementary Information 2 and 3. The theoretical minimum risk exposure level used for current smoking was never smoking or zero³⁰.

Five-star associations. When the most conservative interpretation of the evidence, that is, the BPRF, suggests that the average exposure (15th–85th percentiles of exposure) of smoking increases the risk of a health outcome by >85% (that is, ROS > 0.62), smoking and that outcome are categorized as a 5-star pair. Among the 36 outcomes, there are 5 that have a 5-star association with current smoking: laryngeal cancer (375% increase in risk based on the BPRF, 1.56 ROS), aortic aneurysm (150%, 0.92), peripheral artery disease (137%, 0.86), lung cancer (107%, 0.73) and other pharynx cancer (excluding nasopharynx cancer) (92%, 0.65).

Results for all 5-star risk–outcome pairs are available in Table 2 and Supplementary Information 4.1. In the present study, we provide detailed results for one example 5-star association: current smoking and lung cancer. We extracted 371 observations from 25 prospective cohort studies and 53 case–control studies across 25 locations (Supplementary Table 3)^{5,6,32–107}. Exposure ranged from

1 pack-year to >112 pack-years, with the 85th percentile of exposure being 50.88 pack-years (Fig. 1a).

We found a very strong and significant harmful relationship between pack-years of current smoking and the RR of lung cancer (Fig. 1b). The mean RR of lung cancer at 20 pack-years of smoking was 5.11 (95% uncertainty interval (UI) inclusive of between-study heterogeneity = 1.84–14.99). At 50.88 pack-years (85th percentile of exposure), the mean RR of lung cancer was 13.42 (2.63–74.59). See Table 2 for mean RRs at other exposure levels. The BPRF, which represents the most conservative interpretation of the evidence (Fig. 1a), suggests that smoking in the 15th–85th percentiles of exposure increases the risk of lung cancer by an average of 107%, yielding an ROS of 0.73.

The relationship between pack-years of current smoking and RR of lung cancer is nonlinear, with diminishing impact of further pack-years of smoking, particularly for middle-to-high exposure levels (Fig. 1b). To reduce the effect of bias, we adjusted observations that did not account for more than five confounders, including age and sex, because they were the significant bias covariates identified by the bias covariate selection algorithm²⁹ (Supplementary Table 7). The reported RRs across studies were very heterogeneous. Our meta-analytic method, which accounts for the reported uncertainty in both the data and between-study heterogeneity, fit the data and covered the estimated residuals well (Fig. 1c). After trimming 10% of outliers, we still detected publication bias in the results for lung cancer. See Supplementary Tables 4 and 7 for study bias characteristics and selected bias covariates, Supplementary Fig. 5 for results without 10% trimming and Supplementary Table 8 for observed RR data and alternative exposures across studies for the remaining 5-star pairs.

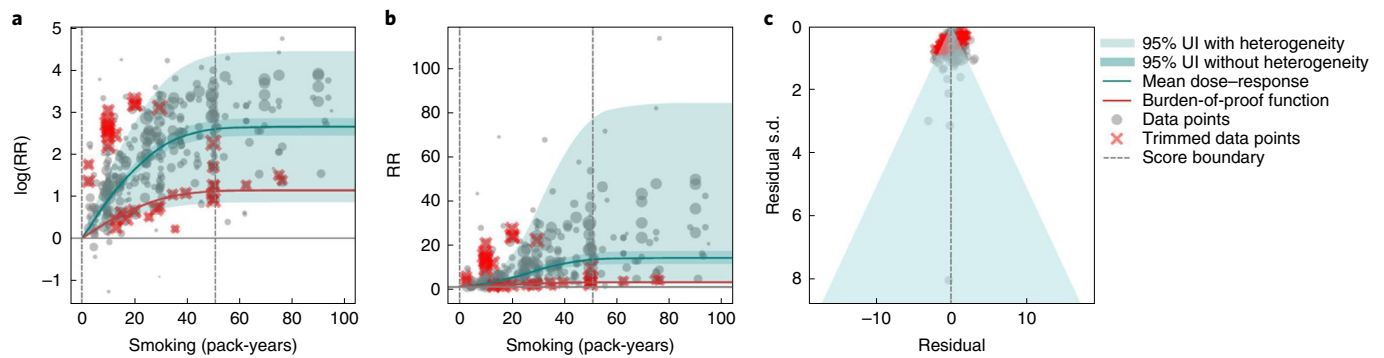


Fig. 1 | Smoking and lung cancer. **a**, The $\log(\text{RR})$ function. **b**, RR function. **c**, A modified funnel plot showing the residuals (relative to 0) on the x axis and the estimated s.d. that includes reported s.d. and between-study heterogeneity on the y axis.

Four-star associations. When the BPRF suggests that the average exposure of smoking increases the risk of a health outcome by 50–85% (that is, $\text{ROS} > 0.41\text{--}0.62$), smoking is categorized as having a 4-star association with that outcome. We identified three outcomes with a 4-star association with smoking: COPD (72% increase in risk based on the BPRF, 0.54 ROS), lower respiratory tract infection (54%, 0.43) and pancreatic cancer (52%, 0.42).

In the present study, we provide detailed results for one example 4-star association: current smoking and COPD. We extracted 51 observations from 11 prospective cohort studies and 4 case–control studies across 36 locations (Supplementary Table 3)^{6,8–10,78,108–116}. Exposure ranged from 1 pack-year to 100 pack-years, with the 85th percentile of exposure in the exposed group being 49.75 pack-years.

We found a strong and significant harmful relationship between pack-years of current smoking and RR of COPD (Fig. 2b). The mean RR of COPD at 20 pack-years was 3.17 (1.60–6.55; Table 2 reports RRs at other exposure levels). At the 85th percentile of exposure, the mean RR of COPD was 6.01 (2.08–18.58). The BPRF suggests that average smoking exposure raises the risk of COPD by an average of 72%, yielding an ROS of 0.54. The results for the other health outcomes that have an association with smoking rated as 4 stars are shown in Table 2 and Supplementary Information 4.2.

The relationship between smoking and COPD is nonlinear, with diminishing impact of further pack-years of current smoking on risk of COPD, particularly for middle-to-high exposure levels (Fig. 2a). To reduce the effect of bias, we adjusted observations that did not account for age and sex and/or were generated for individuals aged >65 years¹¹⁶, because they were the two significant bias covariates identified by the bias covariate selection algorithm (Supplementary Table 7). There was large heterogeneity in the reported RRs across studies, and our meta-analytic method fit the data and covered the estimated residuals well (Fig. 2b). Although we trimmed 10% of outliers, publication bias was still detected in the results for COPD. See Supplementary Tables 4 and 7 for study bias characteristics and selected bias covariates, Supplementary Fig. 5 for results without 10% trimming and Supplementary Table 8 for reported RR data and alternative exposures across studies for the remaining health outcomes that have a 4-star association with smoking.

Three-star associations. When the BPRF suggests that the average exposure of smoking increases the risk of a health outcome by 15–50% (or, when protective, decreases the risk of an outcome by 13–34%; that is, $\text{ROS} > 0.14\text{--}0.41$), the association between smoking and that outcome is categorized as having a 3-star rating. We identified 15 outcomes with a 3-star association: bladder cancer (40% increase in risk, 0.34 ROS); tuberculosis (31%, 0.27); esophageal cancer (29%, 0.26); cervical cancer, multiple sclerosis and rheumatoid arthritis (each 23–24%, 0.21); lower back pain (22%, 0.20);

ischemic heart disease (20%, 0.19); peptic ulcer and macular degeneration (each 19–20%, 0.18); Parkinson's disease (protective risk, 15% decrease in risk, 0.16); and stomach cancer, stroke, type 2 diabetes and cataracts (each 15–17%, 0.14–0.16).

We present the findings on smoking and type 2 diabetes as an example of a 3-star risk association. We extracted 102 observations from 24 prospective cohort studies and 4 case–control studies across 15 locations (Supplementary Table 3)^{117–144}. The exposure ranged from 1 cigarette to 60 cigarettes smoked per day, with the 85th percentile of exposure in the exposed group being 26.25 cigarettes smoked per day.

We found a moderate and significant harmful relationship between cigarettes smoked per day and the RR of type 2 diabetes (Fig. 3b). The mean RR of type 2 diabetes at 20 cigarettes smoked per day was 1.49 (1.18–1.90; see Table 2 for other exposure levels). At the 85th percentile of exposure, the mean RR of type 2 diabetes was 1.54 (1.20–2.01). The BPRF suggests that average smoking exposure raises the risk of type 2 diabetes by an average of 16%, yielding an ROS of 0.15. See Table 2 and Supplementary Information 4.3 for results for the additional health outcomes with an association with smoking rated as 3 stars.

The relationship between smoking and type 2 diabetes is nonlinear, particularly for high exposure levels where the mean risk curve becomes flat (Fig. 3a). We adjusted observations that were generated in subpopulations, because it was the only significant bias covariate identified by the bias covariate selection algorithm (Supplementary Table 7). There was moderate heterogeneity in the observed RR data across studies and our meta-analytic method fit the data and covered the estimated residuals extremely well (Fig. 3b,c). After trimming 10% of outliers, we still detected publication bias in the results for type 2 diabetes. See Supplementary Tables 4 and 7 for study bias characteristics and selected bias covariates, Supplementary Fig. 5 for results without 10% trimming and Supplementary Table 8 for observed RR data and alternative exposures across studies for the remaining 3-star pairs.

Two-star associations. When the BPRF suggests that the average exposure of smoking increases the risk of an outcome by 0–15% (that is, $\text{ROS} 0.0\text{--}0.14$), the association between smoking and that outcome is categorized as a 2-star rating. We identified six 2-star outcomes: nasopharyngeal cancer (14% increase in risk, 0.13 ROS); Alzheimer's and other dementia (10%, 0.09); gallbladder diseases and atrial fibrillation and flutter (each 6%, 0.06); lip and oral cavity cancer (5%, 0.05); and breast cancer (4%, 0.04).

We present the findings on smoking and breast cancer as an example of a 2-star association. We extracted 93 observations from 14 prospective cohort studies and 9 case–control studies across 14 locations (Supplementary Table 3)^{84,87,145–165}. The exposure ranged

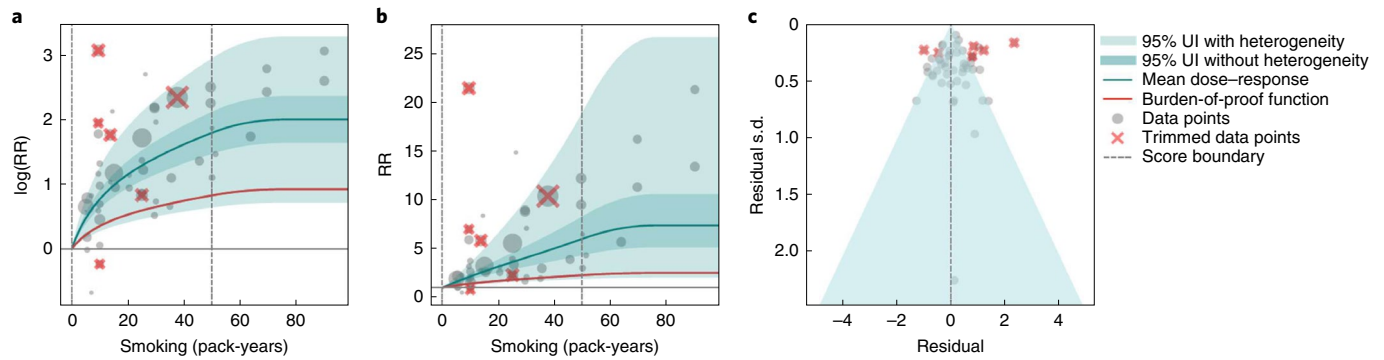


Fig. 2 | Smoking and COPD. **a**, The log(RR) function. **b**, RR function. **c**, A modified funnel plot showing the residuals (relative to 0) on the x axis and the estimated s.d. that includes the reported s.d. and between-study heterogeneity on the y axis.

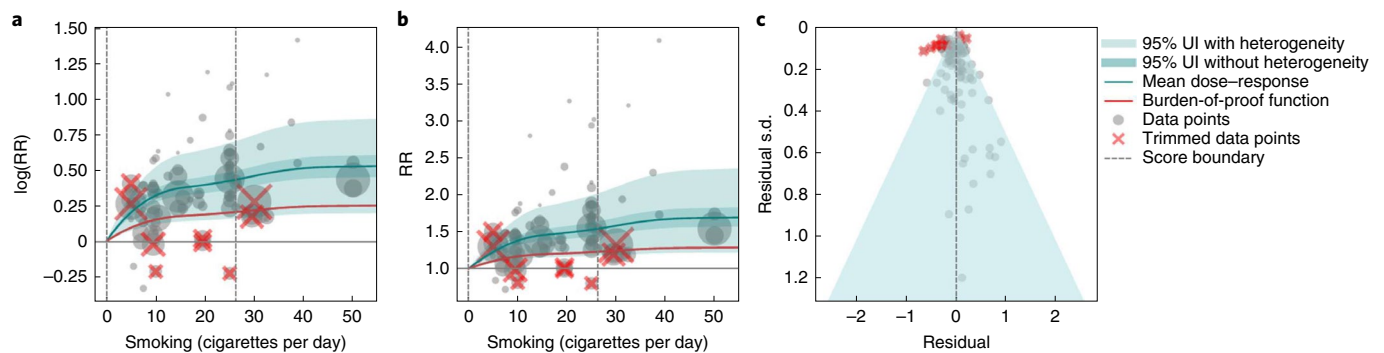


Fig. 3 | Smoking and type 2 diabetes. **a**, The log(RR) function. **b**, RR function. **c**, A modified funnel plot showing the residuals (relative to 0) on the x axis and the estimated s.d. that includes the reported s.d. and between-study heterogeneity on the y axis.

from 1 cigarette to >76 cigarettes smoked per day, with the 85th percentile of exposure in the exposed group being 34.10 cigarettes smoked per day.

We found a weak but significant relationship between pack-years of current smoking and RR of breast cancer (Extended Data Fig. 6). The mean RR of breast cancer at 20 pack-years was 1.17 (1.04–1.31; Table 2 reports other exposure levels). The BPRF suggests that average smoking exposure raises the risk of breast cancer by an average of 4%, yielding an ROS of 0.04. See Table 2 and Supplementary Information 4.4 for results on the additional health outcomes for which the association with smoking has been categorized as 2 stars.

The relationship between smoking and breast cancer is non-linear, particularly for high exposure levels where the mean risk curve becomes flat (Extended Data Fig. 6a). To reduce the effect of bias, we adjusted observations that were generated in subpopulations, because it was the only significant bias covariate identified by the bias covariate selection algorithm (Supplementary Table 7). There was heterogeneity in the reported RRs across studies, but our meta-analytic method fit the data and covered the estimated residuals (Extended Data Fig. 6b). After trimming 10% of outliers, we did not detect publication bias in the results for breast cancer. See Supplementary Tables 4 and 7 for study bias characteristics and selected bias covariates, Supplementary Fig. 5 for results without 10% trimming and Supplementary Table 8 for observed RR data and alternative exposures across studies for the remaining 2-star pairs.

One-star associations. When average exposure to smoking does not significantly increase (or decrease) the risk of an outcome, once between-study heterogeneity and other sources of uncertainty are accounted for (that is, $ROS < 0$), the association between smoking

and that outcome is categorized as 1 star, indicating that there is not sufficient evidence for the effect of smoking on the outcome to reject the null (that is, there may be no association). There were seven outcomes with an association with smoking that rated as 1 star: colorectal and kidney cancer (each -0.01 ROS); leukemia (-0.04); fractures (-0.05); prostate cancer (-0.06); liver cancer (-0.32); and asthma (-0.64).

We use smoking and prostate cancer as examples of a 1-star association. We extracted 78 observations from 21 prospective cohort studies and 1 nested case-control study across 15 locations (Supplementary Table 3)^{157,160,166–185}. The exposure among the exposed group ranged from 1 cigarette to 90 cigarettes smoked per day, with the 85th percentile of exposure in the exposed group being 29.73 cigarettes smoked per day.

Based on our conservative interpretation of the data, we did not find a significant relationship between cigarettes smoked per day and the RR of prostate cancer (Fig. 4B). The exposure-averaged BPRF for prostate cancer was 0.94, which was opposite null from the full range of mean RRs, such as 1.16 (0.89–1.53) at 20 cigarettes smoked per day. The corresponding ROS was -0.06 , which is consistent with no evidence of an association between smoking and increased risk of prostate cancer. See Table 2 and Supplementary Information 4.5 for results for the additional outcomes that have a 1-star association with smoking.

The relationship between smoking and prostate cancer is non-linear, particularly for middle-to-high exposure levels where the mean risk curve becomes flat (Fig. 4a). We did not adjust for any bias covariate because no significant bias covariates were selected by the algorithm (Supplementary Table 7). The RRs reported across studies were very heterogeneous, but our meta-analytic method fit

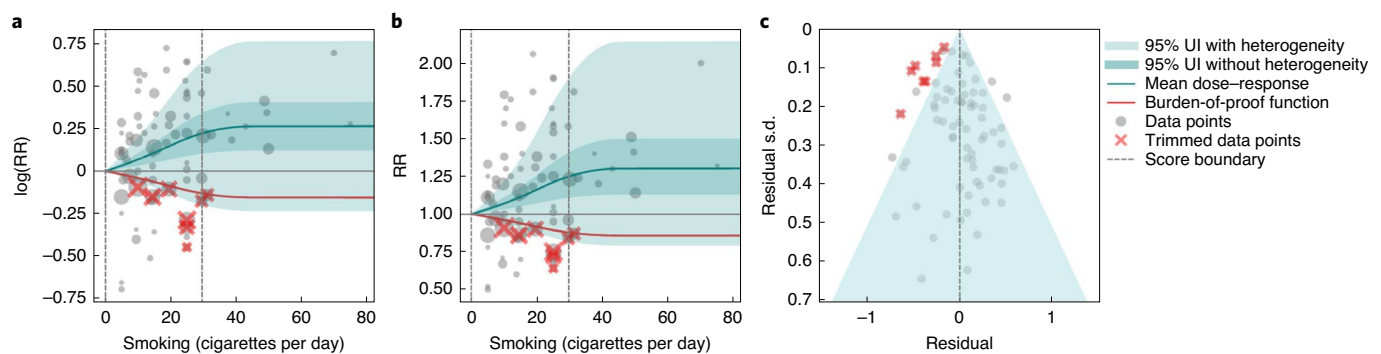


Fig. 4 | Smoking and prostate cancer. **a**, The $\log(\text{RR})$ function. **b**, RR function. **c**, A modified funnel plot showing the residuals (relative to 0) on the x axis and the estimated s.d. that includes reported s.d. and between-study heterogeneity on the y axis.

the data and covered the estimated residuals well (Fig. 4b,c). The ROS associated with the BPRF is -0.05 , suggesting that the most conservative interpretation of all evidence, after accounting for between-study heterogeneity, indicates an inconclusive relationship between smoking exposure and the risk of prostate cancer. After trimming 10% of outliers, we still detected publication bias in the results for prostate cancer, which warrants further studies using sample populations. See Supplementary Tables 4 and 7 for study bias characteristics and selected bias covariates, Supplementary Fig. 5 for results without 10% trimming and Supplementary Table 8 for observed RR data and alternative exposures across studies for the remaining 1-star pairs.

Age-specific dose–response risk for CVD outcomes. We produced age-specific dose–response risk curves for the five selected CVD outcomes (Methods). The ROS associated with each smoking–CVD pair was calculated based on the reference risk curve estimated using all risk data regardless of age information. Estimation of the BPRF, calculation of the associated ROS and star rating of the smoking–CVD pairs follow the same rules as the other non-CVD smoking–outcome pairs (Table 1 and Supplementary Figs. 2–4). Once we had estimated the reference dose–response risk curve for each CVD outcome, we determined the age group of the reference risk curve. The reference age group is 55–59 years for all CVD outcomes, except for peripheral artery disease, the reference age group for which is 60–64 years. We then estimated the age pattern of smoking on all CVD outcomes (Supplementary Fig. 2) and calculated age attenuation factors of the risk for each age group by comparing the risk of each age group with that of the reference age group, using the estimated age pattern (Supplementary Fig. 3). Last, we applied the draws of age attenuation factors of each age group to the dose–response risk curve for the reference age group to produce the age group-specific dose–response risk curves for each CVD outcome (Supplementary Fig. 4).

Discussion

Using our burden-of-proof meta-analytic methods, we re-estimated the dose–response risk of smoking on 36 health outcomes that had previously been demonstrated to be associated with smoking^{30,186}. Using these methods, which account for both the reported uncertainty of the data and the between-study heterogeneity, we found that 29 of the 36 smoking–outcome pairs are supported by evidence that suggests a significant dose–response relationship between smoking and the given outcome (28 with a harmful association and 1 with a protective association). Conversely, after accounting for between-study heterogeneity, the available evidence of smoking risk on seven outcomes (that is, colon and rectum cancer, kidney cancer, leukemia, prostate cancer, fractures, liver cancer and asthma) was insufficient to reject the null or draw definitive conclusions on

their relationship to smoking. Among the 29 outcomes that have evidence supporting a significant relationship to smoking, 8 had strong-to-very-strong evidence of a relationship, meaning that, given all the available data on smoking risk, we estimate that average exposure to smoking increases the risk of those outcomes by $>50\%$ (4- and 5-star outcomes). The currently available evidence for the remaining 21 outcomes with a significant association with current smoking was weak to moderate, indicating that smoking increases the risk of those outcomes by at least $>0–50\%$ (2- and 3-star associations).

Even under our conservative interpretation of the data, smoking is irrefutably harmful to human health, with the greatest increases in risk occurring for laryngeal cancer, aortic aneurysm, peripheral artery disease, lung cancer and other pharynx cancer (excluding nasopharynx cancer), which collectively represent large causes of death and ill-health. The magnitude of and evidence for the associations between smoking and its leading health outcomes are among the highest currently analyzed in the burden-of-proof framework²⁹. The star ratings assigned to each smoking–outcome pair offer policy makers a way of categorizing and comparing the evidence for a relationship between smoking and its potential health outcomes (<https://vizhub.healthdata.org/burden-of-proof>). We found that, for seven outcomes in our analysis, there was insufficient or inconsistent evidence to demonstrate a significant association with smoking. This is a key finding because it demonstrates the need for more high-quality data for these particular outcomes; availability of more data should improve the strength of evidence for whether or not there is an association between smoking and these health outcomes.

Our systematic review approach and meta-analytic methods have numerous benefits over existing systematic reviews and meta-analyses on the same topic that use traditional random effects models. First, our approach relaxes the $\log(\text{linear})$ assumption, using a spline ensemble to estimate the risk²⁹. Second, our approach allows variable reference groups and exposure ranges, allowing for more accurate estimates regardless of whether or not the underlying relative risk is $\log(\text{linear})$. Furthermore, it can detect outliers in the data automatically. Finally, it quantifies uncertainty due to between-study heterogeneity while accounting for small numbers of studies, minimizing the risk that conclusions will be drawn based on spurious findings.

We believe that the results for the association between smoking and each of the 36 health outcomes generated by the present study, including the mean risk function, BPRF, ROS, average excess risk and star rating, could be useful to a range of stakeholders. Policy makers can formulate their decisions on smoking control priorities and resource allocation based on the magnitude of the effect and the consistency of the evidence relating smoking to each of the 36 outcomes, as represented by the ROS and star rating for each smoking–outcome association¹⁸⁷. Physicians and public health practitioners can use the estimates of average increased risk and the

star rating to educate patients and the general public about the risk of smoking and to promote smoking cessation¹⁸⁸. Researchers can use the estimated mean risk function or BPRF to obtain the risk of an outcome at a given smoking exposure level, as well as uncertainty surrounding that estimate of risk. The results can also be used in the estimation of risk-attributable burden, that is, the deaths and disability-adjusted life-years due to each outcome that are attributable to smoking^{30,186}. For the general public, these results could help them to better understand the risk of smoking and manage their health¹⁸⁹.

Although our meta-analysis was comprehensive and carefully conducted, there are limitations to acknowledge. First, the bias covariates used, although carefully extracted and evaluated, were based on observable study characteristics and thus may not fully capture unobserved characteristics such as study quality or context, which might be major sources of bias. Second, if multiple risk estimates with different adjustment levels were reported in a given study, we included only the fully adjusted risk estimate and modeled the adjustment level according to the number of covariates adjusted for (rather than which covariates were adjusted for) and whether a standard adjustment for age and sex had been applied. This approach limited our ability to make full use of all available risk estimates in the literature. Third, although we evaluated the potential for publication bias in the data, we did not test for other forms of bias such as when studies are more consistent with each other than expected by chance²⁹. Fourth, our analysis assumes that the relationships between smoking and health outcomes are similar across geographical regions and over time. We do not have sufficient evidence to quantify how the relationships may have evolved over time because the composition of smoking products has also changed over time. Perhaps some of the heterogeneity of the effect sizes in published studies reflects this; however, this cannot be discerned with the currently available information.

In the future, we plan to include crude and partially adjusted risk estimates in our analyses to fully incorporate all available risk estimates, to model the adjusted covariates in a more comprehensive way by mapping the adjusted covariates across all studies comprehensively and systematically, and to develop methods to evaluate additional forms of potential bias. We plan to update our results on a regular basis to provide timely and up-to-date evidence to stakeholders.

To conclude, we have re-estimated the dose–response risk of smoking on 36 health outcomes while synthesizing all the available evidence up to 31 May 2022. We found that, even after factoring in the heterogeneity between studies and other sources of uncertainty, smoking has a strong-to-very-strong association with a range of health outcomes and confirmed that smoking is irrefutably highly harmful to human health. We found that, due to small numbers of studies, inconsistency in the data, small effect sizes or a combination of these reasons, seven outcomes for which some previous research had found an association with smoking did not—under our meta-analytic framework and conservative approach to interpreting the data—have evidence of an association. Our estimates of the evidence for risk of smoking on 36 selected health outcomes have the potential to inform the many stakeholders of smoking control, including policy makers, researchers, public health professionals, physicians, smokers and the general public.

Online content

Any methods, additional references, Nature Research reporting summaries, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41591-022-01978-x>.

Received: 11 April 2022; Accepted: 28 July 2022;
Published online: 10 October 2022

References

- Doll, R. & Hill, A. B. Smoking and carcinoma of the lung. *Br. Med. J.* **2**, 739–748 (1950).
- Di Cicco, M. E., Ragazzo, V. & Jacinto, T. Mortality in relation to smoking: the British Doctors Study. *Breathe* **12**, 275–276 (2016).
- World Health Organization. *WHO Framework Convention on Tobacco Control* 36 (WHO, 2003).
- Dai, X., Gakidou, E. & Lopez, A. D. Evolution of the global smoking epidemic over the past half century: strengthening the evidence base for policy action. *Tob. Control* **31**, 129–137 (2022).
- Dikshit, R. P. & Kanhere, S. Tobacco habits and risk of lung, oropharyngeal and oral cavity cancer: a population-based case-control study in Bhopal, India. *Int. J. Epidemiol.* **29**, 609–614 (2000).
- Liaw, K. M. & Chen, C. J. Mortality attributable to cigarette smoking in Taiwan: a 12-year follow-up study. *Tob. Control* **7**, 141–148 (1998).
- Gandini, S. et al. Tobacco smoking and cancer: a meta-analysis. *Int. J. Cancer* **122**, 155–164 (2008).
- Deng, X., Yuan, C. & Chang, D. Interactions between single nucleotide polymorphism of SERPINA1 gene and smoking in association with COPD: a case-control study. *Int. J. Chron. Obstruct. Pulmon. Dis.* **12**, 259–265 (2017).
- Leem, A. Y., Park, B., Kim, Y. S., Jung, J. Y. & Won, S. Incidence and risk of chronic obstructive pulmonary disease in a Korean community-based cohort. *Int. J. Chron. Obstruct. Pulmon. Dis.* **13**, 509–517 (2018).
- Forey, B. A., Thornton, A. J. & Lee, P. N. Systematic review with meta-analysis of the epidemiological evidence relating smoking to COPD, chronic bronchitis and emphysema. *BMC Pulmon. Med.* **11**, 36 (2011).
- Tan, J. et al. Smoking, blood pressure, and cardiovascular disease mortality in a large cohort of Chinese men with 15 years follow-up. *Int. J. Environ. Res. Public Health* **15**, E1026 (2018).
- Doll, R., Peto, R., Boreham, J. & Sutherland, I. Mortality in relation to smoking: 50 years' observations on male British doctors. *Br. Med. J.* **328**, 1519 (2004).
- Huxley, R. R. & Woodward, M. Cigarette smoking as a risk factor for coronary heart disease in women compared with men: a systematic review and meta-analysis of prospective cohort studies. *Lancet* **378**, 1297–1305 (2011).
- Hbejan, K. Smoking effect on ischemic heart disease in young patients. *Heart Views* **12**, 1–6 (2011).
- Chao, H. et al. A meta-analysis of active smoking and risk of meningioma. *Tob. Induc. Dis.* **19**, 34 (2021).
- Shi, H., Shao, X. & Hong, Y. Association between cigarette smoking and the susceptibility of acute myeloid leukemia: a systematic review and meta-analysis. *Eur. Rev. Med. Pharm. Sci.* **23**, 10049–10057 (2019).
- Macacu, A., Autier, P., Boniol, M. & Boyle, P. Active and passive smoking and risk of breast cancer: a meta-analysis. *Breast Cancer Res. Treat.* **154**, 213–224 (2015).
- Pujades-Rodriguez, M. et al. Heterogeneous associations between smoking and a wide range of initial presentations of cardiovascular disease in 1 937 360 people in England: lifetime risks and implications for risk prediction. *Int. J. Epidemiol.* **44**, 129–141 (2015).
- Kanazir, M. et al. Risk factors for hepatocellular carcinoma: a case-control study in Belgrade (Serbia). *Tumori* **96**, 911–917 (2010).
- Pytynia, K. B. et al. Matched-pair analysis of survival of never smokers and ever smokers with squamous cell carcinoma of the head and neck. *J. Clin. Oncol.* **22**, 3981–3988 (2004).
- Barengo, N. C., Antikainen, R., Harald, K. & Jousilahti, P. Smoking and cancer, cardiovascular and total mortality among older adults: the Finrisk Study. *Prev. Med. Rep.* **14**, 100875 (2019).
- Guo, Y. et al. Modifiable risk factors for cognitive impairment in Parkinson's disease: a systematic review and meta-analysis of prospective cohort studies. *Mov. Disord.* **34**, 876–883 (2019).
- Aune, D., Vatten, L. J. & Boffetta, P. Tobacco smoking and the risk of gallbladder disease. *Eur. J. Epidemiol.* **31**, 643–653 (2016).
- Qin, L., Deng, H.-Y., Chen, S.-J. & Wei, W. Relationship between cigarette smoking and risk of chronic myeloid leukaemia: a meta-analysis of epidemiological studies. *Hematology* **22**, 193–200 (2017).
- Petrick, J. L. et al. Tobacco, alcohol use and risk of hepatocellular carcinoma and intrahepatic cholangiocarcinoma: the Liver Cancer Pooling Project. *Br. J. Cancer* **118**, 1005–1012 (2018).
- United States Department of Health, Education and Welfare. *Smoking and Health. Report of the Advisory Committee on Smoking and Health to the Surgeon General of the United States Public Health Service* https://www.cdc.gov/tobacco/data_statistics/sgr/index.htm (US DHEW, 1964).
- United States Public Health Service Office of the Surgeon General & National Center for Chronic Disease Prevention and Health Promotion (US) Office on Smoking and Health. *Smoking Cessation: A Report of the Surgeon General*. (US Department of Health and Human Services, 2020).
- Zheng, P., Barber, R., Sorensen, R. J. D., Murray, C. J. L. & Aravkin, A. Y. Trimmed constrained mixed effects models: formulations and algorithms. *J. Comput. Graph Stat.* **30**, 544–556 (2021).

29. Zheng, P. et al. The Burden of Proof studies: assessing the evidence of risk. *Nat. Med.* in press (2022).
30. Reitsma, M. B. 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* **397**, 2337–2360 (2021).
31. Moher, D., Liberati, A., Tetzlaff, J. & Altman, D. G. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Br. Med. J.* **339**, b2535 (2009).
32. Liu, Z. Y., He, X. Z. & Chapman, R. S. Smoking and other risk factors for lung cancer in Xuanwei, China. *Int. J. Epidemiol.* **20**, 26–31 (1991).
33. Brownson, R. C., Reif, J. S., Keefe, T. J., Ferguson, S. W. & Pritzl, J. A. Risk factors for adenocarcinoma of the lung. *Am. J. Epidemiol.* **125**, 25–34 (1987).
34. Marugame, T. et al. Lung cancer death rates by smoking status: comparison of the Three-Prefecture Cohort study in Japan to the Cancer Prevention Study II in the USA. *Cancer Sci.* **96**, 120–126 (2005).
35. Dosemeci, M., Gokmen, I., Unsal, M., Hayes, R. B. & Blair, A. Tobacco, alcohol use, and risks of laryngeal and lung cancer by subsite and histologic type in Turkey. *Cancer Causes Control* **8**, 729–737 (1997).
36. Freedman, N. D. et al. Impact of changing US cigarette smoking patterns on incident cancer: risks of 20 smoking-related cancers among the women and men of the NIH-AARP cohort. *Int. J. Epidemiol.* **45**, 846–856 (2016).
37. Bae, J.-M. et al. Lung cancer incidence by smoking status in Korean men: 16 years of observations in the Seoul Male Cancer Cohort study. *J. Korean Med. Sci.* **28**, 636–637 (2013).
38. Everatt, R., Kuzmickienė, I., Virvičiūtė, D. & Tamošiūnas, A. Cigarette smoking, educational level and total and site-specific cancer: a cohort study in men in Lithuania. *Eur. J. Cancer Prev.* **23**, 579–586 (2014).
39. Nordlund, L. A., Carstensen, J. M. & Pershagen, G. Are male and female smokers at equal risk of smoking-related cancer: evidence from a Swedish prospective study. *Scand. J. Public Health* **27**, 56–62 (1999).
40. Siemiatycki, J., Krewski, D., Franco, E. & Kaiserman, M. Associations between cigarette smoking and each of 21 types of cancer: a multi-site case-control study. *Int. J. Epidemiol.* **24**, 504–514 (1995).
41. Chyou, P. H., Nomura, A. M. & Stemmermann, G. N. A prospective study of the attributable risk of cancer due to cigarette smoking. *Am. J. Public Health* **82**, 37–40 (1992).
42. Potter, J. D., Sellers, T. A., Folsom, A. R. & McGovern, P. G. Alcohol, beer, and lung cancer in postmenopausal women. The Iowa Women's Health Study. *Ann. Epidemiol.* **2**, 587–595 (1992).
43. Chyou, P. H., Nomura, A. M., Stemmermann, G. N. & Kato, I. Lung cancer: a prospective study of smoking, occupation, and nutrient intake. *Arch. Environ. Health* **48**, 69–72 (1993).
44. Pesch, B. et al. Cigarette smoking and lung cancer—relative risk estimates for the major histological types from a pooled analysis of case-control studies. *Int. J. Cancer* **131**, 1210–1219 (2012).
45. Jöckel, K. H. et al. Occupational and environmental hazards associated with lung cancer. *Int. J. Epidemiol.* **21**, 202–213 (1992).
46. Jöckel, K. H., Ahrens, W., Jahn, I., Pohlmann, H. & Bolm-Audorf, U. Occupational risk factors for lung cancer: a case-control study in West Germany. *Int. J. Epidemiol.* **27**, 549–560 (1998).
47. Lei, Y. X., Cai, W. C., Chen, Y. Z. & Du, Y. X. Some lifestyle factors in human lung cancer: a case-control study of 792 lung cancer cases. *Lung Cancer* **14**, S121–S136 (1996).
48. Pawlega, J., Rachtan, J. & Dyba, T. Evaluation of certain risk factors for lung cancer in Cracow (Poland)—a case-control study. *Acta Oncol.* **36**, 471–476 (1997).
49. Mao, Y. et al. Socioeconomic status and lung cancer risk in Canada. *Int. J. Epidemiol.* **30**, 809–817 (2001).
50. Barbone, F., Bovenzi, M., Cavallieri, F. & Stanta, G. Cigarette smoking and histologic type of lung cancer in men. *Chest* **112**, 1474–1479 (1997).
51. Matos, E., Vilensky, M., Boffetta, P. & Kogevinas, M. Lung cancer and smoking: a case-control study in Buenos Aires, Argentina. *Lung Cancer* **21**, 155–163 (1998).
52. Simonato, L. et al. Lung cancer and cigarette smoking in Europe: an update of risk estimates and an assessment of inter-country heterogeneity. *Int. J. Cancer* **91**, 876–887 (2001).
53. Risch, H. A. et al. Are female smokers at higher risk for lung cancer than male smokers? A case-control analysis by histologic type. *Am. J. Epidemiol.* **138**, 281–293 (1993).
54. Sankaranarayanan, R. et al. A case-control study of diet and lung cancer in Kerala, south India. *Int. J. Cancer* **58**, 644–649 (1994).
55. Band, P. R. et al. Identification of occupational cancer risks in British Columbia. Part I: methodology, descriptive results, and analysis of cancer risks, by cigarette smoking categories of 15,463 incident cancer cases. *J. Occup. Environ. Med.* **41**, 224–232 (1999).
56. Becher, H., Jöckel, K. H., Timm, J., Wichmann, H. E. & Drescher, K. Smoking cessation and nonsmoking intervals: effect of different smoking patterns on lung cancer risk. *Cancer Causes Control* **2**, 381–387 (1991).
57. Brockmüller, J., Kerb, R., Drakoulis, N., Nitz, M. & Roots, I. Genotype and phenotype of glutathione S-transferase class mu isoenzymes mu and psi in lung cancer patients and controls. *Cancer Res.* **53**, 1004–1011 (1993).
58. Vena, J. E., Byers, T. E., Cookfair, D. & Swanson, M. Occupation and lung cancer risk. An analysis by histologic subtypes. *Cancer* **56**, 910–917 (1985).
59. Cascorbi, I. et al. Homozygous rapid arylamine N-acetyltransferase (NAT2) genotype as a susceptibility factor for lung cancer. *Cancer Res.* **56**, 3961–3966 (1996).
60. Chiazzese, L., Watkins, D. K. & Fryar, C. A case-control study of malignant and non-malignant respiratory disease among employees of a fiberglass manufacturing facility. *Br. J. Ind. Med.* **49**, 326–331 (1992).
61. Ando, M. et al. Attributable and absolute risk of lung cancer death by smoking status: findings from the Japan Collaborative Cohort Study. *Int. J. Cancer* **105**, 249–254 (2003).
62. De Matteis, S. et al. Are women who smoke at higher risk for lung cancer than men who smoke? *Am. J. Epidemiol.* **177**, 601–612 (2013).
63. He, Y. et al. Changes in smoking behavior and subsequent mortality risk during a 35-year follow-up of a cohort in Xi'an, China. *Am. J. Epidemiol.* **179**, 1060–1070 (2014).
64. Nishino, Y. et al. Cancer incidence profiles in the Miyagi Cohort Study. *J. Epidemiol.* **14**, S7–S11 (2004).
65. Papadopoulos, A. et al. Cigarette smoking and lung cancer in women: results of the French ICARE case-control study. *Lung Cancer* **74**, 369–377 (2011).
66. Shimazu, T. et al. Alcohol and risk of lung cancer among Japanese men: data from a large-scale population-based cohort study, the JPHC study. *Cancer Causes Control* **19**, 1095–1102 (2008).
67. Tindle, H. A. et al. Lifetime smoking history and risk of lung cancer: results from the Framingham Heart Study. *J. Natl. Cancer Inst.* **110**, 1201–1207 (2018).
68. Yong, L. C. et al. Intake of vitamins E, C, and A and risk of lung cancer. The NHANES I epidemiologic followup study. First National Health and Nutrition Examination Survey. *Am. J. Epidemiol.* **146**, 231–243 (1997).
69. Hansen, M. S. et al. Sex differences in risk of smoking-associated lung cancer: results from a cohort of 600,000 Norwegians. *Am. J. Epidemiol.* **187**, 971–981 (2018).
70. Boffetta, P. et al. Tobacco smoking as a risk factor of bronchioloalveolar carcinoma of the lung: pooled analysis of seven case-control studies in the International Lung Cancer Consortium (ILCCO). *Cancer Causes Control* **22**, 73–79 (2011).
71. Yun, Y. D. et al. Hazard ratio of smoking on lung cancer in Korea according to histological type and gender. *Lung* **194**, 281–289 (2016).
72. Suzuki, I. et al. Risk factors for lung cancer in Rio de Janeiro, Brazil: a case-control study. *Lung Cancer* **11**, 179–190 (1994).
73. De Stefani, E., Deneo-Pellegrini, H., Carzoglio, J. C., Ronco, A. & Mendilaharsu, M. Dietary nitrosodimethylamine and the risk of lung cancer: a case-control study from Uruguay. *Cancer Epidemiol. Biomark. Prev.* **5**, 679–682 (1996).
74. Kreuzer, M. et al. Risk factors for lung cancer in young adults. *Am. J. Epidemiol.* **147**, 1028–1037 (1998).
75. Armadans-Gil, L., Vaqué-Rafart, J., Rosselló, J., Olona, M. & Alseda, M. Cigarette smoking and male lung cancer risk with special regard to type of tobacco. *Int. J. Epidemiol.* **28**, 614–619 (1999).
76. Kubik, A. K., Zatloukal, P., Tomásek, L. & Petruzelka, L. Lung cancer risk among Czech women: a case-control study. *Prev. Med.* **34**, 436–444 (2002).
77. Rachtan, J. Smoking, passive smoking and lung cancer cell types among women in Poland. *Lung Cancer* **35**, 129–136 (2002).
78. Thun, M. J. et al. 50-year trends in smoking-related mortality in the United States. *N. Engl. J. Med.* **368**, 351–364 (2013).
79. Zatloukal, P., Kubik, A., Pauk, N., Tomásek, L. & Petruzelka, L. Adenocarcinoma of the lung among women: risk associated with smoking, prior lung disease, diet and menstrual and pregnancy history. *Lung Cancer* **41**, 283–293 (2003).
80. Hansen, M. S., Licaj, I., Braaten, T., Lund, E. & Gram, I. T. The fraction of lung cancer attributable to smoking in the Norwegian Women and Cancer (NOWAC) Study. *Br. J. Cancer* **124**, 658–662 (2021).
81. Zhang, P. et al. Association of smoking and polygenic risk with the incidence of lung cancer: a prospective cohort study. *Br. J. Cancer* **126**, 1637–1646 (2022).
82. Weber, M. F. et al. Cancer incidence and cancer death in relation to tobacco smoking in a population-based Australian cohort study. *Int. J. Cancer* **149**, 1076–1088 (2021).
83. Guo, L.-W. et al. A risk prediction model for selecting high-risk population for computed tomography lung cancer screening in China. *Lung Cancer* **163**, 27–34 (2022).
84. Mezzouiso, A. G., Odono, A., Signorelli, C. & Russo, A. G. Association between smoking and cancers among women: results from the FRiCaM multisite cohort study. *J. Cancer* **12**, 3136–3144 (2021).
85. Hawrysz, I., Wadolowska, L., Slowinska, M. A., Czerwinska, A. & Golota, J. J. Adherence to prudent and mediterranean dietary patterns is inversely associated with lung cancer in moderate but not heavy male Polish smokers: a case-control study. *Nutrients* **12**, E3788 (2020).

86. Huang, C.-C., Lai, C.-Y., Tsai, C.-H., Wang, J.-Y. & Wong, R.-H. Combined effects of cigarette smoking, DNA methyltransferase 3B genetic polymorphism, and DNA damage on lung cancer. *BMC Cancer* **21**, 1066 (2021).
87. Viner, B., Barberio, A. M., Haig, T. R., Friedenreich, C. M. & Brenner, D. R. The individual and combined effects of alcohol consumption and cigarette smoking on site-specific cancer risk in a prospective cohort of 26,607 adults: results from Alberta's Tomorrow Project. *Cancer Causes Control* **30**, 1313–1326 (2019).
88. Park, E. Y., Lim, M. K., Park, E., Oh, J.-K. & Lee, D.-H. Relationship between urinary 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanol and lung cancer risk in the general population: a community-based prospective cohort study. *Front. Oncol.* **11**, 611674 (2021).
89. De Stefani, E., Deneo-Pellegrini, H., Mendilaharsu, M., Carzoglio, J. C. & Ronco, A. Dietary fat and lung cancer: a case-control study in Uruguay. *Cancer Causes Control* **8**, 913–921 (1997).
90. Wünsch-Filho, V., Moncau, J. E., Mirabelli, D. & Boffetta, P. Occupational risk factors of lung cancer in São Paulo, Brazil. *Scand. J. Work Environ. Health* **24**, 118–124 (1998).
91. Hu, J. et al. A case-control study of diet and lung cancer in northeast China. *Int. J. Cancer* **71**, 924–931 (1997).
92. Jia, G., Wen, W., Massion, P. P., Shu, X.-O. & Zheng, W. Incorporating both genetic and tobacco smoking data to identify high-risk smokers for lung cancer screening. *Carcinogenesis* **42**, 874–879 (2021).
93. Rusmaully, J. et al. Risk of lung cancer among women in relation to lifetime history of tobacco smoking: a population-based case-control study in France (the WELCA study). *BMC Cancer* **21**, 711 (2021).
94. Jin, K. et al. Tobacco smoking modifies the association between hormonal factors and lung cancer occurrence among post-menopausal Chinese women. *Transl. Oncol.* **12**, 819–827 (2019).
95. Tse, L. A., Wang, F., Wong, M. C.-S., Au, J. S.-K. & Yu, I. T.-S. Risk assessment and prediction for lung cancer among Hong Kong Chinese men. *BMC Cancer* **22**, 585 (2022).
96. Huang, C.-C. et al. Joint effects of cigarette smoking and green tea consumption with miR-29b and DNMT3b mRNA expression in the development of lung cancer. *Genes* **13**, 836 (2022).
97. Hosseini, M. et al. Environmental risk factors for lung cancer in Iran: a case-control study. *Int. J. Epidemiol.* **38**, 989–996 (2009).
98. Naghibzadeh-Tahami, A. et al. Is opium use associated with an increased risk of lung cancer? A case-control study. *BMC Cancer* **20**, 807 (2020).
99. Shimatani, K., Ito, H., Matsuo, K., Tajima, K. & Takezaki, T. Cumulative cigarette tar exposure and lung cancer risk among Japanese smokers. *Jpn J. Clin. Oncol.* **50**, 1009–1017 (2020).
100. Lai, C.-Y. et al. Genetic polymorphism of catechol-O-methyltransferase modulates the association of green tea consumption and lung cancer. *Eur. J. Cancer Prev.* **28**, 316–322 (2019).
101. Schwartz, A. G. et al. Hormone use, reproductive history, and risk of lung cancer: the Women's Health Initiative studies. *J. Thorac. Oncol.* **10**, 1004–1013 (2015).
102. Kreuzer, M., Gerken, M., Heinrich, J., Kreienbrock, L. & Wichmann, H.-E. Hormonal factors and risk of lung cancer among women? *Int. J. Epidemiol.* **32**, 263–271 (2003).
103. Sreeja, L. et al. Possible risk modification by CYP1A1, GSTM1 and GSTT1 gene polymorphisms in lung cancer susceptibility in a South Indian population. *J. Hum. Genet.* **50**, 618–627 (2005).
104. Siemiatycki, J. et al. Are the apparent effects of cigarette smoking on lung and bladder cancers due to uncontrolled confounding by occupational exposures? *Epidemiology* **5**, 57–65 (1994).
105. Chan-Yeung, M. et al. Risk factors associated with lung cancer in Hong Kong. *Lung Cancer* **40**, 131–140 (2003).
106. Lawania, S., Singh, N., Behera, D. & Sharma, S. Xeroderma pigmentosum complementation group D polymorphism toward lung cancer susceptibility survival and response in patients treated with platinum chemotherapy. *Future Oncol.* **13**, 2645–2665 (2017).
107. De Stefani, E. et al. Mate drinking and risk of lung cancer in males: a case-control study from Uruguay. *Cancer Epidemiol. Biomark. Prev.* **5**, 515–519 (1996).
108. Pérez-Padilla, R. et al. Exposure to biomass smoke and chronic airway disease in Mexican women. A case-control study. *Am. J. Respir. Crit. Care Med.* **154**, 701–706 (1996).
109. Zhang, X.-R. et al. Glucosamine use, smoking and risk of incident chronic obstructive pulmonary disease: a large prospective cohort study. *Br. J. Nutr.* <https://doi.org/10.1017/S000711452100372X> (2021).
110. Johannessen, A., Omenaas, E., Bakke, P. & Gulsvik, A. Incidence of GOLD-defined chronic obstructive pulmonary disease in a general adult population. *Int. J. Tuberc. Lung Dis.* **9**, 926–932 (2005).
111. Fox, J. Life-style and mortality: a large-scale census-based cohort study in Japan. *J. Epidemiol. Community Health* **45**, 173 (1991).
112. Thomson, B. et al. Low-intensity daily smoking and cause-specific mortality in Mexico: prospective study of 150 000 adults. *Int. J. Epidemiol.* **50**, 955–964 (2021).
113. van Durme, Y. M. T. A. et al. Prevalence, incidence, and lifetime risk for the development of COPD in the elderly: the Rotterdam study. *Chest* **135**, 368–377 (2009).
114. Li, L. et al. SERPINE2 rs16865421 polymorphism is associated with a lower risk of chronic obstructive pulmonary disease in the Uyghur population: a case-control study. *J. Gene Med.* **21**, e3106 (2019).
115. Ganbold, C. et al. The cumulative effect of gene-gene interactions between *GSTM1*, *CHRNA3*, *CHRNA5* and *SOD3* gene polymorphisms combined with smoking on COPD risk. *Int. J. Chron. Obstruct. Pulmon. Dis.* **16**, 2857–2868 (2021).
116. Omori, H. et al. Twelve-year cumulative incidence of airflow obstruction among Japanese males. *Intern. Med.* **50**, 1537–1544 (2011).
117. Manson, J. E., Ajani, U. A., Liu, S., Nathan, D. M. & Hennekens, C. H. A prospective study of cigarette smoking and the incidence of diabetes mellitus among US male physicians. *Am. J. Med.* **109**, 538–542 (2000).
118. Lv, J. et al. Adherence to a healthy lifestyle and the risk of type 2 diabetes in Chinese adults. *Int. J. Epidemiol.* **46**, 1410–1420 (2017).
119. Waki, K. et al. Alcohol consumption and other risk factors for self-reported diabetes among middle-aged Japanese: a population-based prospective study in the JPHC study cohort I. *Diabet. Med.* **22**, 323–331 (2005).
120. Meisinger, C., Döring, A., Thorand, B. & Löwel, H. Association of cigarette smoking and tar and nicotine intake with development of type 2 diabetes mellitus in men and women from the general population: the MONICA/KORA Augsburg Cohort Study. *Diabetologia* **49**, 1770–1776 (2006).
121. Huh, Y. et al. Association of smoking status with the risk of type 2 diabetes among young adults: a nationwide cohort study in South Korea. *Nicotine Tob. Res.* **24**, 1234–1240 (2022).
122. Sawada, S. S., Lee, I.-M., Muto, T., Matuszaki, K. & Blair, S. N. Cardiorespiratory fitness and the incidence of type 2 diabetes: prospective study of Japanese men. *Diabetes Care* **26**, 2918–2922 (2003).
123. Will, J. C., Galuska, D. A., Ford, E. S., Mokdad, A. & Calle, E. E. Cigarette smoking and diabetes mellitus: evidence of a positive association from a large prospective cohort study. *Int. J. Epidemiol.* **30**, 540–546 (2001).
124. Nakanishi, N., Nakamura, K., Matsuo, Y., Suzuki, K. & Tataru, K. Cigarette smoking and risk for impaired fasting glucose and type 2 diabetes in middle-aged Japanese men. *Ann. Intern. Med.* **133**, 183–191 (2000).
125. Sairenchi, T. et al. Cigarette smoking and risk of type 2 diabetes mellitus among middle-aged and elderly Japanese men and women. *Am. J. Epidemiol.* **160**, 158–162 (2004).
126. Hou, X. et al. Cigarette smoking is associated with a lower prevalence of newly diagnosed diabetes screened by OGTT than non-smoking in Chinese men with normal weight. *PLoS ONE* **11**, e0149234 (2016).
127. Hu, F. B. et al. Diet, lifestyle, and the risk of type 2 diabetes mellitus in women. *N. Engl. J. Med.* **345**, 790–797 (2001).
128. Teratani, T. et al. Dose-response relationship between tobacco or alcohol consumption and the development of diabetes mellitus in Japanese male workers. *Drug Alcohol Depend.* **125**, 276–282 (2012).
129. Kawakami, N., Takatsuka, N., Shimizu, H. & Ishibashi, H. Effects of smoking on the incidence of non-insulin-dependent diabetes mellitus. Replication and extension in a Japanese cohort of male employees. *Am. J. Epidemiol.* **145**, 103–109 (1997).
130. Patja, K. et al. Effects of smoking, obesity and physical activity on the risk of type 2 diabetes in middle-aged Finnish men and women. *J. Intern. Med.* **258**, 356–362 (2005).
131. White, W. B. et al. High-intensity cigarette smoking is associated with incident diabetes mellitus in Black adults: the Jackson Heart Study. *J. Am. Heart Assoc.* **7**, e007413 (2018).
132. Uchimoto, S. et al. Impact of cigarette smoking on the incidence of Type 2 diabetes mellitus in middle-aged Japanese men: the Osaka Health Survey. *Diabet. Med.* **16**, 951–955 (1999).
133. Rimm, E. B., Chan, J., Stampfer, M. J., Colditz, G. A. & Willett, W. C. Prospective study of cigarette smoking, alcohol use, and the risk of diabetes in men. *Br. Med. J.* **310**, 555–559 (1995).
134. Hilawe, E. H. et al. Smoking and diabetes: is the association mediated by adiponectin, leptin, or C-reactive protein? *J. Epidemiol.* **25**, 99–109 (2015).
135. InterAct Consortium et al. Smoking and long-term risk of type 2 diabetes: the EPIC-InterAct study in European populations. *Diabetes Care* **37**, 3164–3171 (2014).
136. Jee, S. H., Foong, A. W., Hur, N. W. & Samet, J. M. Smoking and risk for diabetes incidence and mortality in Korean men and women. *Diabetes Care* **33**, 2567–2572 (2010).
137. Rasouli, B. et al. Smoking and the risk of LADA: results from a Swedish population-based case-control study. *Diabetes Care* **39**, 794–800 (2016).
138. Wannamethee, S. G., Shaper, A. G. & Perry, I. J. British Regional Heart Study. Smoking as a modifiable risk factor for type 2 diabetes in middle-aged men. *Diabetes Care* **24**, 1590–1595 (2001).

139. Radzeviciene, L. & Ostrauskas, R. Smoking habits and type 2 diabetes mellitus in women. *Women Health* **58**, 884–897 (2018).
140. Carlsson, S., Midtjell, K. & Grill, V., Nord-Trøndelag Study. Smoking is associated with an increased risk of type 2 diabetes but a decreased risk of autoimmune diabetes in adults: an 11-year follow-up of incidence of diabetes in the Nord-Trøndelag study. *Diabetologia* **47**, 1953–1956 (2004).
141. Akter, S. et al. Smoking, smoking cessation, and the risk of type 2 diabetes among Japanese adults: Japan Epidemiology Collaboration on Occupational Health Study. *PLoS ONE* **10**, e0132166 (2015).
142. Pirie, K. et al. The 21st century hazards of smoking and benefits of stopping: a prospective study of one million women in the UK. *Lancet* **381**, 133–141 (2013).
143. Park, C.-H. et al. [The effect of smoking status upon occurrence of impaired fasting glucose or type 2 diabetes in Korean men]. *J. Prev. Med. Public Health* **41**, 249–254 (2008).
144. Doi, Y. et al. Two risk score models for predicting incident Type 2 diabetes in Japan. *Diabet. Med.* **29**, 107–114 (2012).
145. van den Brandt, P. A. A possible dual effect of cigarette smoking on the risk of postmenopausal breast cancer. *Eur. J. Epidemiol.* **32**, 683–690 (2017).
146. Dossus, L. et al. Active and passive cigarette smoking and breast cancer risk: results from the EPIC cohort. *Int. J. Cancer* **134**, 1871–1888 (2014).
147. Kawai, M., Malone, K. E., Tang, M.-T. C. & Li, C. I. Active smoking and the risk of estrogen receptor-positive and triple-negative breast cancer among women ages 20 to 44 years. *Cancer* **120**, 1026–1034 (2014).
148. Reynolds, P. et al. Active smoking, household passive smoking, and breast cancer: evidence from the California Teachers Study. *J. Natl Cancer Inst.* **96**, 29–37 (2004).
149. Ellingjord-Dale, M. et al. Alcohol, physical activity, smoking, and breast cancer subtypes in a large, nested case-control study from the Norwegian Breast Cancer Screening Program. *Cancer Epidemiol. Biomark. Prev.* **26**, 1736–1744 (2017).
150. Arthur, R. et al. Association between lifestyle, menstrual/reproductive history, and histological factors and risk of breast cancer in women biopsied for benign breast disease. *Breast Cancer Res. Treat.* **165**, 623–631 (2017).
151. Luo, J. et al. Association of active and passive smoking with risk of breast cancer among postmenopausal women: a prospective cohort study. *Br. Med. J.* **342**, d1016 (2011).
152. White, A. J., D'Aloisio, A. A., Nichols, H. B., DeRoo, L. A. & Sandler, D. P. Breast cancer and exposure to tobacco smoke during potential windows of susceptibility. *Cancer Causes Control* **28**, 667–675 (2017).
153. Gram, I. T. et al. Breast cancer risk among women who start smoking as teenagers. *Cancer Epidemiol. Biomark. Prev.* **14**, 61–66 (2005).
154. Gammon, M. D. et al. Cigarette smoking and breast cancer risk among young women (United States). *Cancer Causes Control* **9**, 583–590 (1998).
155. Magnusson, C., Wedrén, S. & Rosenberg, L. U. Cigarette smoking and breast cancer risk: a population-based study in Sweden. *Br. J. Cancer* **97**, 1287–1290 (2007).
156. Chu, S. Y. et al. Cigarette smoking and the risk of breast cancer. *Am. J. Epidemiol.* **131**, 244–253 (1990).
157. Lemogne, C. et al. Depression and the risk of cancer: a 15-year follow-up study of the GAZEL cohort. *Am. J. Epidemiol.* **178**, 1712–1720 (2013).
158. Morabia, A., Bernstein, M., Héritier, S. & Khatchatrian, N. Relation of breast cancer with passive and active exposure to tobacco smoke. *Am. J. Epidemiol.* **143**, 918–928 (1996).
159. Conlon, M. S. C., Johnson, K. C., Bewick, M. A., Lafrenie, R. M. & Donner, A. Smoking (active and passive), N-acetyltransferase 2, and risk of breast cancer. *Cancer Epidemiol.* **34**, 142–149 (2010).
160. Ozasa, K., Japan Collaborative Cohort Study for Evaluation of Cancer. Smoking and mortality in the Japan Collaborative Cohort Study for Evaluation of Cancer (JACC). *Asian Pac. J. Cancer Prev.* **8**, 89–96 (2007).
161. Jones, M. E., Schoemaker, M. J., Wright, L. B., Ashworth, A. & Swerdlow, A. J. Smoking and risk of breast cancer in the Generations Study cohort. *Breast Cancer Res.* **19**, 118 (2017).
162. Bjerkaas, E. et al. Smoking duration before first childbirth: an emerging risk factor for breast cancer? Results from 302,865 Norwegian women. *Cancer Causes Control* **24**, 1347–1356 (2013).
163. Gram, I. T., Little, M. A., Lund, E. & Braaten, T. The fraction of breast cancer attributable to smoking: the Norwegian women and cancer study 1991–2012. *Br. J. Cancer* **115**, 616–623 (2016).
164. Li, C. I., Malone, K. E. & Daling, J. R. The relationship between various measures of cigarette smoking and risk of breast cancer among older women 65–79 years of age (United States). *Cancer Causes Control* **16**, 975–985 (2005).
165. Xue, F., Willett, W. C., Rosner, B. A., Hankinson, S. E. & Michels, K. B. Cigarette smoking and the incidence of breast cancer. *Arch. Intern. Med.* **171**, 125–133 (2011).
166. Parker, A. S., Cerhan, J. R., Putnam, S. D., Cantor, K. P. & Lynch, C. F. A cohort study of farming and risk of prostate cancer in Iowa. *Epidemiology* **10**, 452–455 (1999).
167. Sawada, N. et al. Alcohol and smoking and subsequent risk of prostate cancer in Japanese men: the Japan Public Health Center-based prospective study. *Int. J. Cancer* **134**, 971–978 (2014).
168. Hiatt, R. A., Armstrong, M. A., Klatsky, A. L. & Sidney, S. Alcohol consumption, smoking, and other risk factors and prostate cancer in a large health plan cohort in California (United States). *Cancer Causes Control* **5**, 66–72 (1994).
169. Cerhan, J. R. et al. Association of smoking, body mass, and physical activity with risk of prostate cancer in the Iowa 65+ Rural Health Study (United States). *Cancer Causes Control* **8**, 229–238 (1997).
170. Watters, J. L., Park, Y., Hollenbeck, A., Schatzkin, A. & Albanes, D. Cigarette smoking and prostate cancer in a prospective US cohort study. *Cancer Epidemiol. Biomark. Prev.* **18**, 2427–2435 (2009).
171. Butler, L. M., Wang, R., Wong, A. S., Koh, W.-P. & Yu, M. C. Cigarette smoking and risk of prostate cancer among Singapore Chinese. *Cancer Causes Control* **20**, 1967–1974 (2009).
172. Lotufo, P. A., Lee, I. M., Ajani, U. A., Hennekens, C. H. & Manson, J. E. Cigarette smoking and risk of prostate cancer in the physicians' health study (United States). *Int. J. Cancer* **87**, 141–144 (2000).
173. Hsing, A. W. et al. Diet, tobacco use, and fatal prostate cancer: results from the Lutheran Brotherhood Cohort Study. *Cancer Res.* **50**, 6836–6840 (1990).
174. Veierød, M. B., Laake, P. & Thelle, D. S. Dietary fat intake and risk of prostate cancer: a prospective study of 25,708 Norwegian men. *Int. J. Cancer* **73**, 634–638 (1997).
175. Meyer, J., Rohrmann, S., Bopp, M. & Faeh, D. & Swiss National Cohort Study Group. Impact of smoking and excess body weight on overall and site-specific cancer mortality risk. *Cancer Epidemiol. Biomark. Prev.* **24**, 1516–1522 (2015).
176. Putnam, S. D. et al. Lifestyle and anthropometric risk factors for prostate cancer in a cohort of Iowa men. *Ann. Epidemiol.* **10**, 361–369 (2000).
177. Taghizadeh, N., Vonk, J. M. & Boezen, H. M. Lifetime smoking history and cause-specific mortality in a cohort study with 43 years of follow-up. *PLoS ONE* **11**, e0153310 (2016).
178. Park, S.-Y. et al. Racial/ethnic differences in lifestyle-related factors and prostate cancer risk: the Multiethnic Cohort Study. *Cancer Causes Control* **26**, 1507–1515 (2015).
179. Nomura, A. M., Lee, J., Stemmermann, G. N. & Combs, G. F. Serum selenium and subsequent risk of prostate cancer. *Cancer Epidemiol. Biomark. Prev.* **9**, 883–887 (2000).
180. Rodriguez, C., Tatham, L. M., Thun, M. J., Calle, E. E. & Heath, C. W. Smoking and fatal prostate cancer in a large cohort of adult men. *Am. J. Epidemiol.* **145**, 466–475 (1997).
181. Rohrmann, S. et al. Smoking and risk of fatal prostate cancer in a prospective U.S. study. *Urology* **69**, 721–725 (2007).
182. Giovannucci, E. et al. Smoking and risk of total and fatal prostate cancer in United States health professionals. *Cancer Epidemiol. Biomark. Prev.* **8**, 277–282 (1999).
183. Rohrmann, S. et al. Smoking and the risk of prostate cancer in the European Prospective Investigation into Cancer and Nutrition. *Br. J. Cancer* **108**, 708–714 (2013).
184. Lund Nilssen, T. I., Johnsen, R. & Vatten, L. J. Socio-economic and lifestyle factors associated with the risk of prostate cancer. *Br. J. Cancer* **82**, 1358–1363 (2000).
185. Hsing, A. W., McLaughlin, J. K., Hrubec, Z., Blot, W. J. & Fraumeni, J. F. Tobacco use and prostate cancer: 26-year follow-up of US veterans. *Am. J. Epidemiol.* **133**, 437–441 (1991).
186. Murray, C. J. L. 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* **396**, 1223–1249 (2020).
187. Bero, L. A. & Jadad, A. R. How consumers and policymakers can use systematic reviews for decision making. *Ann. Intern. Med.* **127**, 37–42 (1997).
188. Centers for Disease Control and Prevention (CDC). Cigarette smoking among adults and trends in smoking cessation—United States, 2008. *MMWR Morb. Mortal. Wkly Rep.* **58**, 1227–1232 (2009).
189. Prochaska, J. O. & Goldstein, M. G. Process of smoking cessation: implications for clinicians. *Clin. Chest Med.* **12**, 727–735 (1991).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022

Methods

Overview. For the present study, we used a meta-analytic tool, MR-BRT (metaregression—Bayesian, regularized, trimmed), to estimate the dose–response risk curves of the risk of a health outcome across the range of current smoking levels along with uncertainty estimates²⁸. Compared with traditional meta-analysis using linear mixed effect models, MR-BRT relaxes the assumption of a log(linear) relationship between exposure and risk, incorporates between-study heterogeneity into the uncertainty of risk estimates, handles estimates reported across different exposure categories, automatically identifies and trims outliers, and systematically tests and adjusts for bias due to study designs and characteristics. The meta-analytic methods employed by the present study followed the six main steps proposed by Zheng et al.^{28,29}, namely: (1) enacting a systematic review approach and data extraction following a pre-specified and standardized protocol; (2) estimating the shape of the relationship between exposure and RR; (3) evaluating and adjusting for systematic bias as a function of study characteristics and risk estimation; (4) quantifying between-study heterogeneity while adjusting for within-study correlation and the number of studies; (5) evaluating potential publication or reporting biases; and (6) estimating the mean risk function and the BPRF, calculating the ROS and categorizing smoking–outcome pairs using a star-rating scheme from 1 to 5.

The estimates for our primary indicators of this work—mean RRs across a range of exposures, BRPFs, ROSs and star ratings for each risk–outcome pair—are not specific to or disaggregated by specific populations. We did not estimate RRs separately for different locations, sexes (although the RR of prostate cancer was estimated only for males and of cervical and breast cancer only for females) or age groups (although this analysis was applied to disease endpoints in adults aged ≥ 30 years only and, as detailed below, age-specific estimates were produced for the five CVD outcomes).

The present study complies with the PRISMA guidelines¹⁹⁰ (Supplementary Tables 9 and 10 and Supplementary Information 1.5) and Guidelines for Accurate and Transparent Health Estimates Reporting (GATHER) recommendations¹⁹¹ (Supplementary Table 11). The study was approved by the University of Washington Institutional Review Board (study no. 9060). The systematic review approach was not registered.

Selecting health outcomes. In the present study, current smoking is defined as the current use of any smoked tobacco product on a daily or occasional basis. Health outcomes were initially selected using the World Cancer Research Fund criteria for convincing or probable evidence as described in Murray et al.¹⁸⁶. The 36 health outcomes that were selected based on existing evidence of a relationship included 16 cancers (lung cancer, esophageal cancer, stomach cancer, leukemia, liver cancer, laryngeal cancer, breast cancer, cervical cancer, colorectal cancer, lip and oral cavity cancer, nasopharyngeal cancer, other pharynx cancer (excluding nasopharynx cancer), pancreatic cancer, bladder cancer, kidney cancer and prostate cancer), 5 CVDs (ischemic heart disease, stroke, atrial fibrillation and flutter, aortic aneurysm and peripheral artery disease) and 15 other diseases (COPD, lower respiratory tract infections, tuberculosis, asthma, type 2 diabetes, Alzheimer's disease and related dementias, Parkinson's disease, multiple sclerosis, cataracts, gallbladder diseases, low back pain, peptic ulcer disease, rheumatoid arthritis, macular degeneration and fracture). Definitions of the outcomes are described in Supplementary Table 1.

Step 1: systematic review approach to literature search and data extraction.

Informed by the systematic review approach we took for the GBD 2019 (ref. 30), for the present study we identified input studies in the literature using a systematic review approach for all 36 smoking–outcome pairs using updated search strings to identify all relevant studies indexed in PubMed up to 31 May 2022 and extracted data on smoking risk estimates. Briefly, the studies that were extracted represented several types of study design (for example, cohort and case–control studies), measured exposure in several different ways and varied in their choice of reference categories (where some compared current smokers with never smokers, whereas others compared current smokers with nonsmokers or former smokers). All these study characteristics were catalogued systematically and taken into consideration during the modeling part of the analysis.

In addition, for CVD outcomes, we also estimated the age pattern of risk associated with smoking. We applied a systematic review of literature approach for smoking risk for the five CVD outcomes. We developed a search string to search for studies reporting any association between binary smoking status (that is, current, former and ever smokers) and the five CVD outcomes from 1 January 1970 to 31 May 2022, and included only studies reporting age-specific risk (RR, odds ratio (OR), hazard ratio (HR)) of smoking status. The inclusion criteria and results of the systematic review approach are reported in accordance with PRISMA guidelines³¹. Details for each outcome on the search string used in the systematic review approach, refined inclusion and exclusion criteria, data extraction template and PRISMA diagram are given in Supplementary Information 1. Title and/or abstract screening, full text screening and data extraction were conducted by 14 members of the research team and extracted data underwent manual quality assurance by the research team to verify accuracy.

Selecting exposure categories. Cumulative exposure in pack-years was the measure of exposure used for COPD and all cancer outcomes except for prostate cancer, to reflect the risk of both duration and intensity of current smoking on these outcomes. For prostate cancer, CVDs and all the other outcomes except for fractures, we used cigarette-equivalents smoked per day as the exposure for current smoking, because smoking intensity is generally thought to be more important than duration for these outcomes. For fractures, we used binary exposure, because there were few studies examining intensity or duration of smoking on fractures. The smoking–outcome pairs and the corresponding exposures are summarized in Supplementary Table 4 and are congruent with the GBD 2019 (refs. 30,186).

Steps 2–5: modeling dose–response RR of smoking on the selected health outcomes. Of the six steps proposed by Zheng et al.²⁹, steps 2–5 cover the process of modeling dose–response risk curves. In step 2, we estimated the shape (or the ‘signal’) of the dose–response risk curves, integrating over different exposure ranges. To relax the log(linear) assumption usually applied to continuous dose–response risk and make the estimates robust to the placement of spline knots, we used an ensemble spline approach to fit the functional form of the dose–response relationship. The final ensemble model was a weighted combination of 50 models with random knot placement, with the weight of each model proportional to measures of model fit and total variation. To avoid the influence of extreme data and reduce publication bias, we trimmed 10% of data for each outcome as outliers. We also applied a monotonicity constraint to ensure that the mean risk curves were nondecreasing (or nonincreasing in the case of Parkinson's disease).

In step 3, following the GRADE approach^{192,193}, we quantified risk of bias across six domains, namely, representativeness of the study population, exposure, outcome, reverse causation, control for confounding and selection bias. Details about the bias covariates are provided in Supplementary Table 4. We systematically tested for the effect of bias covariates using metaregression, selected significant bias covariates using the Lasso approach^{194,195} and adjusted for the selected bias covariates in the final risk curve.

In step 4, we quantified between-study heterogeneity accounting for within-study correlation, uncertainty of the heterogeneity, as well as small number of studies. Specifically, we used a random intercept in the mixed-effects model to account for the within-study correlation and used a study-specific random slope with respect to the ‘signal’ to capture between-study heterogeneity. As between-study heterogeneity can be underestimated or even zero when the number of studies is small^{196,197}, we used Fisher's information matrix to estimate the uncertainty of the heterogeneity¹⁹⁸ and incorporated that uncertainty into the final results.

In step 5, in addition to generating funnel plots and visually inspecting for asymmetry (Figs. 1c, 2c, 3c and 4c and Extended Data Fig. 6c) to identify potential publication bias, we also statistically tested for potential publication or reporting bias using Egger's regression¹⁹⁹. We flagged potential publication bias in the data but did not correct for it, which is in line with the general literature^{10,200,201}. Full details about the modeling process have been published elsewhere²⁹ and model specifications for each outcome are in Supplementary Table 6.

Step 6: estimating the mean risk function and the BPRF. In the final step, step 6, the metaregression model inclusive of the selected bias covariates from step 3 (for example, the highest adjustment level) was used to predict the mean risk function and its 95% UI, which incorporated the uncertainty of the mean effect, between-study heterogeneity and the uncertainty in the heterogeneity estimate accounting for small numbers of studies. Specifically, 1,000 draws were created for each 0.1 level of doses from 0 pack-years to 100 pack-years or cigarette-equivalents smoked per day using the Bayesian metaregression model. The mean of the 1,000 draws was used to estimate the mean risk at each exposure level, and the 25th and 95th draws were used to estimate the 95% UIs for the mean risk at each exposure level.

The BPRF²⁹ is a conservative estimate of risk function consistent with the available evidence, correcting for both between-study heterogeneity and systemic biases related to study characteristics. The BPRF is defined as either the 5th (if harmful) or 95th (if protective) quantile curve closest to the line of $\log(\text{RR})$ of 0, which defines the null (Figs. 1a, 2b, 3a and 4a). The BPRF represents the smallest harmful (or protective) effect of smoking on the corresponding outcome at each level of exposure that is consistent with the available evidence. A BPRF opposite null from the mean risk function indicates that insufficient evidence is available to reject null, that is, that there may not be an association between risk and outcome. Likewise, the further the BPRF is from null on the same side of null as the mean risk function, the higher the magnitude and evidence for the relationship. The BPRF can be interpreted as indicating that, even accounting for between-study heterogeneity and its uncertainty, the $\log(\text{RR})$ across the studied smoking range is at least as high as the BPRF (or at least as low as the BPRF for a protective risk).

To quantify the strength of the evidence, we calculated the ROS for each smoking–outcome association as the signed value of the $\log(\text{BPRF})$ averaged between the 15th and 85th percentiles of observed exposure levels for each outcome. The ROS is a single summary of the effect of smoking on the outcome, with higher positive ROSs corresponding to stronger and more consistent evidence and a higher average effect size of smoking and a negative ROS, suggesting that, based on the available evidence, there is no significant effect of smoking on the outcome after accounting for between-study heterogeneity.

For ease of communication, we further classified each smoking–outcome association into a star rating from 1 to 5. Briefly, 1-star associations have an ROS <0, indicating that there is insufficient evidence to find a significant association between smoking and the selected outcome. We divided the positive ROSs into ranges 0.0–0.14 (2-star), >0.14–0.41 (3-star), >0.41–0.62 (4-star) and >0.62 (5-star). These categories correspond to excess risk ranges for harmful risks of 0–15%, >15–50%, >50–85% and >85%. For protective risks, the ranges of exposure-averaged decreases in risk by star rating are 0–13% (2 stars), >13–34% (3 stars), >34–46% (4 stars) and >46% (5 stars).

Among the 36 smoking–outcome pairs analyzed, smoking fracture was the only binary risk–outcome pair, which was due to limited data on the dose–response risk of smoking on fracture²⁰². The estimation of binary risk was simplified because the RR was merely a comparison between current smokers and nonsmokers or never smokers. The concept of ROS for continuous risk can naturally extend to binary risk because the BPRF is still defined as the 5th percentile of the effect size accounting for data uncertainty and between-study heterogeneity. However, binary ROSs must be divided by 2 to make them comparable with continuous ROSs, which were calculated by averaging the risk over the range between the 15th and the 85th percentiles of observed exposure levels. Full details about estimating mean risk functions, BPRFs and ROSs for both continuous and binary risk–outcome pairs can be found elsewhere²⁹.

Estimating the age-specific risk function for CVD outcomes. For non-CVD outcomes, we assumed that the risk function was the same for all ages and all sexes, except for breast, cervical and prostate cancer, which were assumed to apply only to females or males, respectively. As the risk of smoking on CVD outcomes is known to attenuate with increasing age^{203–206}, we adopted a four-step approach for GBD 2020 to produce age-specific dose–response risk curves for CVD outcomes.

First, we estimated the reference dose–response risk of smoking for each CVD outcome using dose-specific RR data for each outcome regardless of the age group information. This step was identical to that implemented for the other non-CVD outcomes. Once we had generated the reference curve, we determined the age group associated with it by calculating the weighted mean age across all dose-specific RR data (weighted by the reciprocal of the s.e.m. of each datum). For example, if the weighted mean age of all dose-specific RR data was 56.5, we estimated the age group associated with the reference risk curve to be aged 55–59 years. For cohort studies, the age range associated with the RR estimate was calculated as a mean age at baseline plus the mean/median years of follow-up (if only the maximum years of follow-up were reported, we would halve this value and add it to the mean age at baseline). For case–control studies, the age range associated with the OR estimate was simply the reported mean age at baseline (if mean age was not reported, we used the midpoint of the age range instead).

In the third step, we extracted age group-specific RR data and relevant bias covariates from the studies identified in our systematic review approach of age-specific smoking risk on CVD outcomes, and used MR-BRT to model the age pattern of excess risk (that is, RR-1) of smoking on CVD outcomes with age group-specific excess RR data for all CVD outcomes. We modeled the age pattern of smoking risk on CVDs following the same steps we implemented for modeling dose–response risk curves. In the final model, we included a spline on age, random slope on age by study and the bias covariate encoding exposure definition (that is, current, former and ever smokers), which was picked by the variable selection algorithm^{28,29}. When predicting the age pattern of the excess risk of smoking on CVD outcomes using the fitted model, we did not include between-study heterogeneity to reduce uncertainty in the prediction.

In the fourth step, we calculated the age attenuation factors of excess risk compared with the reference age group for each CVD outcome as the ratio of the estimated excess risk for each age group to the excess risk for the reference age group. We performed the calculation at the draw level to obtain 1,000 draws of the age attenuation factors for each age group. Once we had estimated the age attenuation factors, we carried out the last step, which consisted of adjusting the risk curve for the reference age group from step 1 using equation (1) to produce the age group-specific risk curves for each CVD outcome:

$$\begin{aligned} RR_{age_i} &= (RR_{ref} - 1) \times AF_{age_i} + 1 \\ RR_{age_i} &= \text{Relative risk at age group } i \\ RR_{ref} &= \text{Risk curve at the reference age group} \\ AF_{age_i} &= \text{Age attenuation factor for age group } i \end{aligned} \quad (1)$$

We implemented the age adjustment at the draw level so that the uncertainty of the age attenuation factors could be naturally incorporated into the final adjusted age-specific RR curves. A PRISMA diagram detailing the systematic review approach, a description of the studies included and the full details about the methods are in Supplementary Information 1.5 and 5.2.

Estimating the theoretical minimum risk exposure level. The theoretical minimum risk exposure level for smoking was 0, that is, no individuals in the population are current or former smokers.

Model validation. The validity of the meta-analytic tool has been extensively evaluated by Zheng and colleagues using simulation experiments^{28,29}. For the present study, we conducted two additional sensitivity analyses to examine how the shape of the risk curves was impacted by applying a monotonicity constraint and trimming 10% of data. We present the results of these sensitivity analyses in Supplementary Information 6. In addition to the sensitivity analyses, the dose–response risk estimates were also validated by plotting the mean risk function along with its 95% UI against both the extracted dose-specific RR data from the studies included and our previous dose–response risk estimates from the GBD 2019 (ref. ³⁰). The mean risk functions along with the 95% UIs were validated based on data fit and the level, shape and plausibility of the dose–response risk curves. All curves were validated by all authors and reviewed by an external expert panel, comprising professors with relevant experience from universities including Johns Hopkins University, Karolinska Institute and University of Barcelona; senior scientists working in relevant departments at the WHO and the Center for Disease Control and Prevention (CDC) and directors of nongovernmental organizations such as the Campaign for Tobacco-Free Kids.

Statistical analysis. Analyses were carried out using R v.3.6.3, Python v.3.8 and Stata v.16.

Statistics and reproducibility. The study was a secondary analysis of existing data involving systematic reviews and meta-analyses. No statistical method was used to predetermine sample size. As the study did not involve primary data collection, randomization and blinding, data exclusions were not relevant to the present study, and, as such, no data were excluded and we performed no randomization or blinding. We have made our data and code available to foster reproducibility.

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

Data availability

The findings from the present study are supported by data available in the published literature. Data sources and citations for each risk–outcome pair can be downloaded using the ‘download’ button on each risk curve page currently available at <https://vizhub.healthdata.org/burden-of-proof>. Study characteristics and citations for all input data used in the analyses are also provided in Supplementary Table 3, and Supplementary Table 2 provides a template of the data collection form.

Code availability

All code used for these analyses is publicly available online (<https://github.com/ihmeuw-msca/burden-of-proof>).

References

- Page, M. J. et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Br. Med. J.* **372**, n71 (2021).
- Stevens, G. A. et al. Guidelines for Accurate and Transparent Health Estimates Reporting: the GATHER statement. *Lancet* **388**, e19–e23 (2016).
- BMJ Best Practice. *What is GRADE?* <https://bestpractice.bmj.com/info/us/toolkit/learn-ebm/what-is-grade> (BMJ, 2021).
- The GRADE Working Group. *GRADE handbook*. <https://gdt.grade.pro/org/app/handbook/handbook.html> (The GRADE Working Group, 2013).
- Efron, B., Hastie, T., Johnstone, I. & Tibshirani, R. Least angle regression. *Ann. Stat.* **32**, 407–499 (2004).
- Tibshirani, R. Regression shrinkage and selection via the lasso. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **58**, 267–288 (1996).
- von Hippel, P. T. The heterogeneity statistic I2 can be biased in small meta-analyses. *BMC Med. Res. Methodol.* **15**, 35 (2015).
- Kontopantelis, E., Springate, D. A. & Reeves, D. A re-analysis of the Cochrane Library data: the dangers of unobserved heterogeneity in meta-analyses. *PLoS ONE* **8**, e69930 (2013).
- Biggerstaff, B. J. & Tweedie, R. L. Incorporating variability in estimates of heterogeneity in the random effects model in meta-analysis. *Stat. Med.* **16**, 753–768 (1997).
- Egger, M., Smith, G. D., Schneider, M. & Minder, C. Bias in meta-analysis detected by a simple, graphical test. *Br. Med. J.* **315**, 629–634 (1997).
- Lee, P. N., Forey, B. A. & Coombs, K. J. Systematic review with meta-analysis of the epidemiological evidence in the 1900s relating smoking to lung cancer. *BMC Cancer* **12**, 385 (2012).
- Rücker, G., Carpenter, J. R. & Schwarzer, G. Detecting and adjusting for small-study effects in meta-analysis. *Biometr.* **53**, 351–368 (2011).
- Wu, Z.-J., Zhao, P., Liu, B. & Yuan, Z.-C. Effect of cigarette smoking on risk of hip fracture in men: a meta-analysis of 14 prospective cohort studies. *PLoS ONE* **11**, e0168990 (2016).
- Thun, M. J. et al. in *Cigarette Smoking Behaviour in the United States: changes in cigarette-related disease risks and their implication for prevention*

- and control (eds Burns, D.M. et al.) Tobacco Control Monograph No. 8 Ch. 4 (National Cancer Institute, 1997).
204. Tolstrup, J. S. et al. Smoking and risk of coronary heart disease in younger, middle-aged, and older adults. *Am. J. Public Health* **104**, 96–102 (2014).
 205. Jonas, M. A., Oates, J. A., Ockene, J. K. & Hennekens, C. H. Statement on smoking and cardiovascular disease for health care professionals. American Heart Association. *Circulation* **86**, 1664–1669 (1992).
 206. Khan, S. S. et al. Cigarette smoking and competing risks for fatal and nonfatal cardiovascular disease subtypes across the life course. *J. Am. Heart Assoc.* **10**, e021751 (2021).

Acknowledgements

Research reported in this publication was supported by the Bill & Melinda Gates Foundation and Bloomberg Philanthropies. The content is solely the responsibility of the authors and does not necessarily represent the official views of the funders. The study funders had no role in study design, data collection, data analysis, data interpretation, writing of the final report or the decision to publish.

We thank the Tobacco Metrics Team Advisory Group for their valuable input and review of the work. The members of the Advisory Group are: P. Allebeck, R. Chandora, J. Drope, M. Eriksen, E. Fernández, H. Gouda, R. Kennedy, D. McGoldrick, L. Pan, K. Schotte, E. Sebrie, J. Soriano, M. Tynan and K. Welding.

Author contributions

X.D., S.I.H., S.A.M., E.C.M., E.M.O., C.J.L.M. and E.G. managed the estimation or publications process. X.D. and G.F.G. wrote the first draft of the manuscript. X.D. and

P.Z. had primary responsibility for applying analytical methods to produce estimates. X.D., G.F.G., N.S.A., J.A.A., S.C., R.F., V.I., M.J.M., L.M., S.I.N., C.O., M.B.R. and J.W. had primary responsibility for seeking, cataloguing, extracting or cleaning data, and for designing or coding figures and tables. X.D., G.F.G., M.B.R., N.S.A., H.R.L., C.O. and J.W. provided data or critical feedback on data sources. X.D., J.H., R.J.D.S., A.Y.A., P.Z., C.J.L.M. and E.G. developed methods or computational machinery. X.D., G.F.G., M.B.R., S.I.H., J.H., R.J.D.S., A.Y.A., P.Z., C.J.L.M. and E.G. provided critical feedback on methods or results. X.D., G.F.G., M.B.R., C.B., S.I.H., L.B.M., S.A.M., A.Y.A. and E.G. drafted the work or revised it critically for important intellectual content. X.D., S.I.H., L.B.M., E.C.M., E.M.O. and E.G. managed the overall research enterprise.

Competing interests

The authors declare no competing interests.

Additional information

Extended data are available for this paper at <https://doi.org/10.1038/s41591-022-01978-x>.

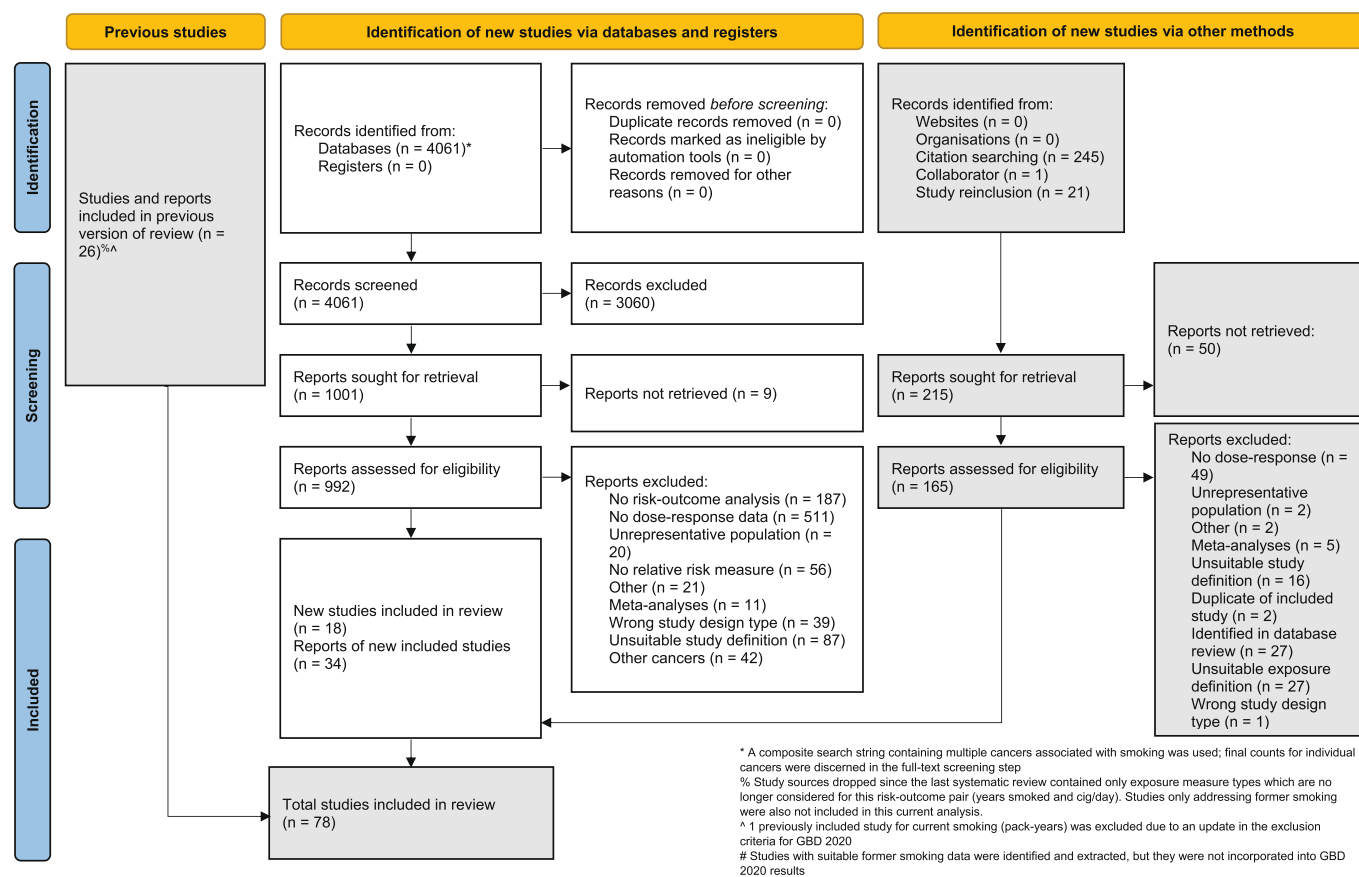
Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41591-022-01978-x>.

Correspondence and requests for materials should be addressed to Xiaochen Dai.

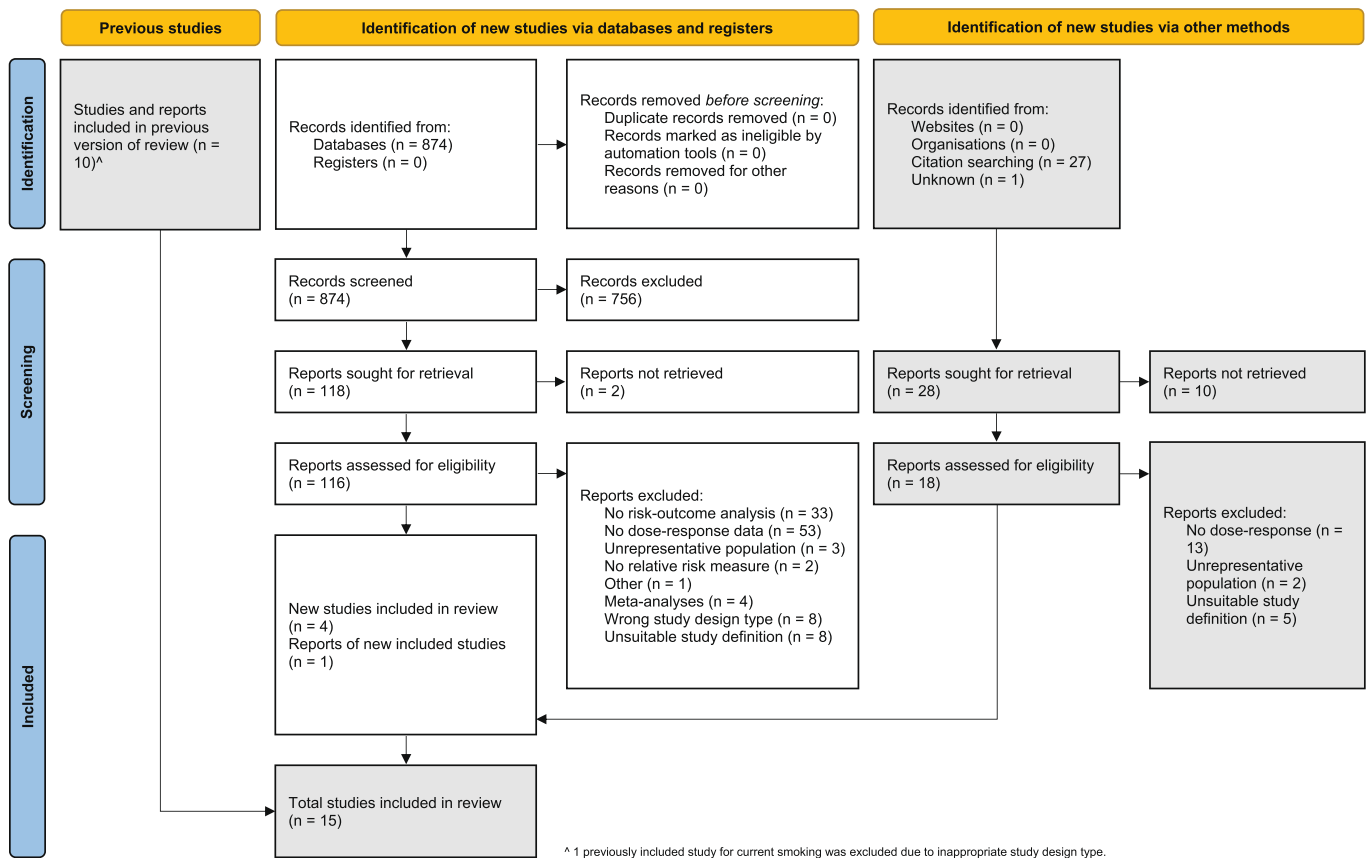
Peer review information *Nature Medicine* thanks Frederic Sitas and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Primary Handling Editors: Jennifer Sargent and Ming Yang, in collaboration with the *Nature Medicine* team.

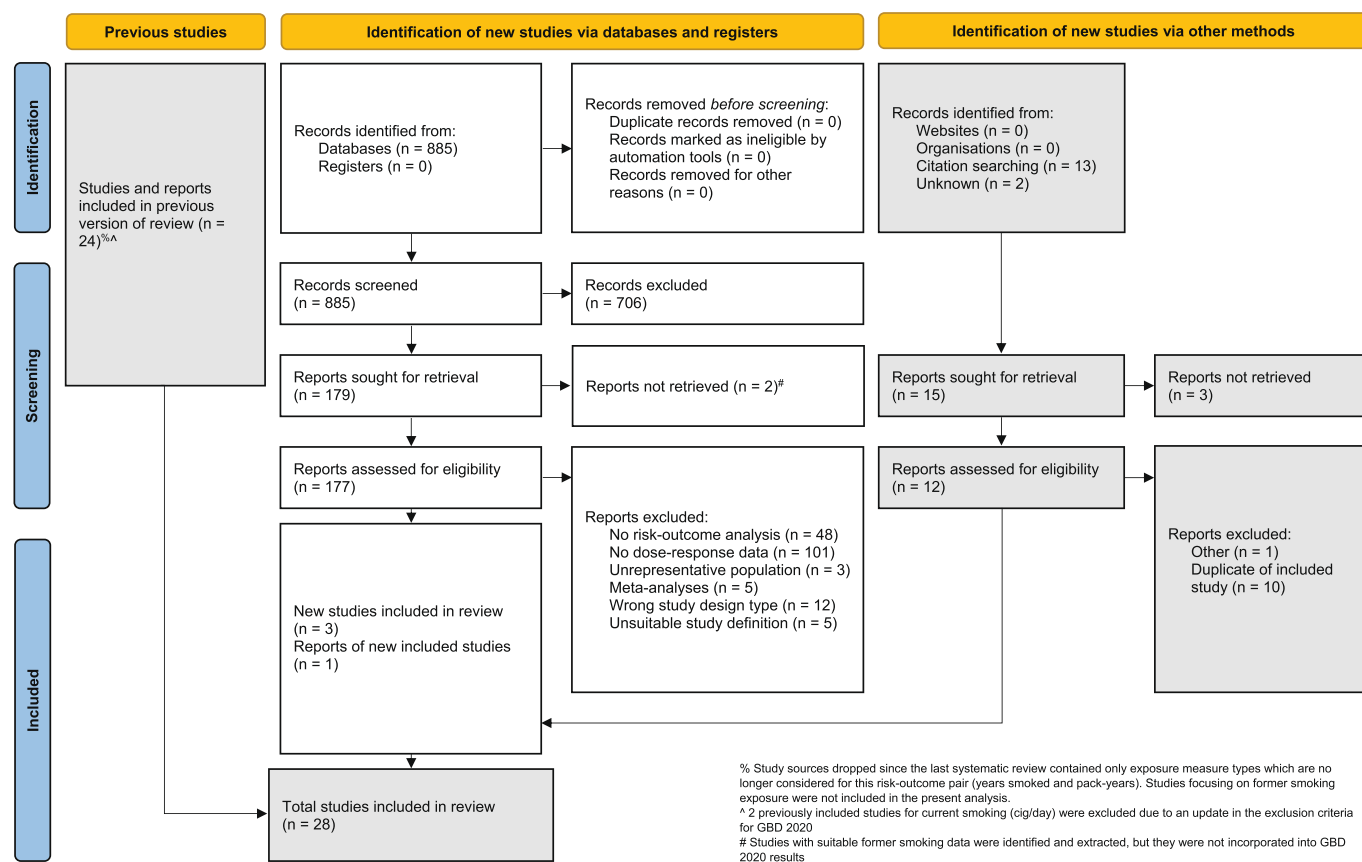
Reprints and permissions information is available at www.nature.com/reprints.



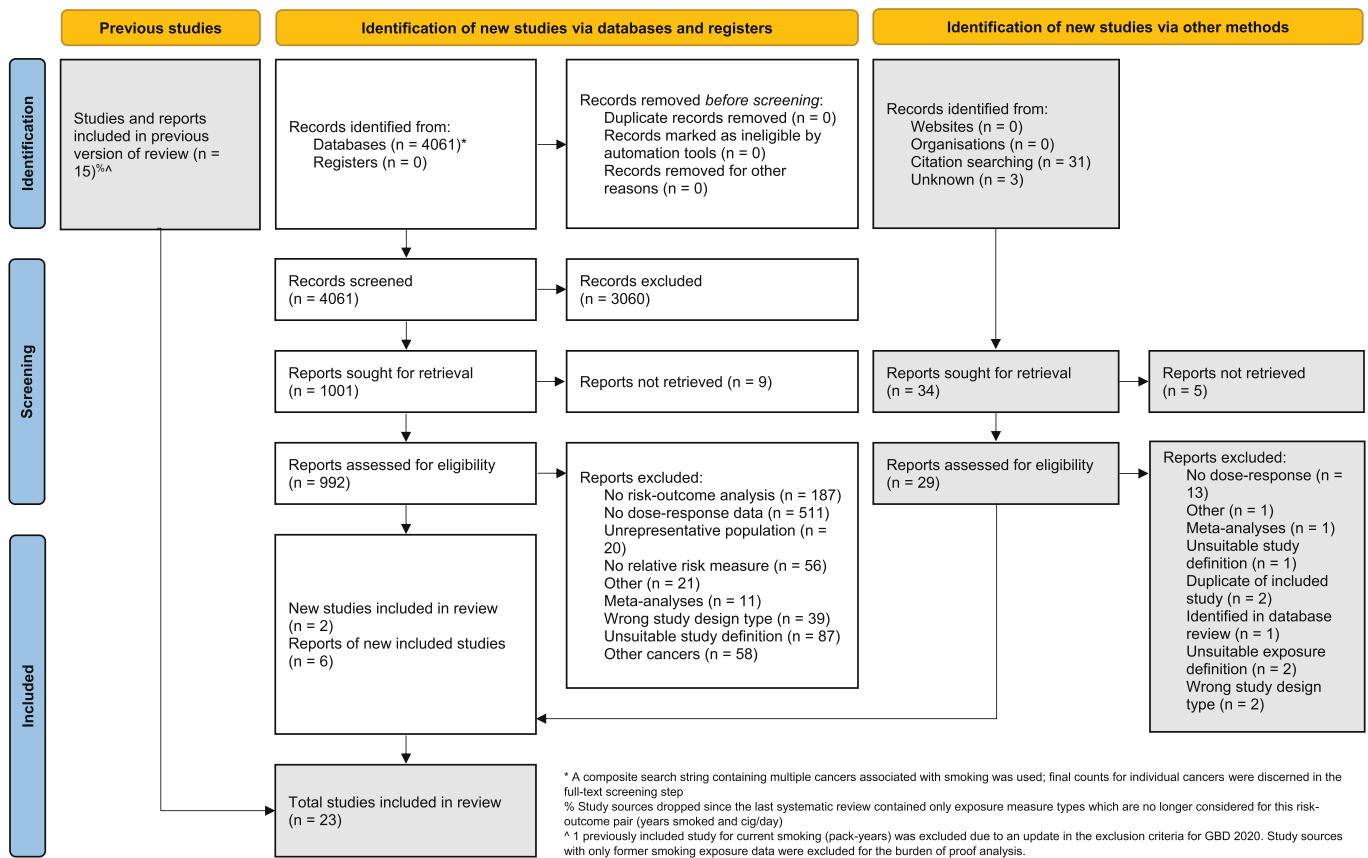
Extended Data Fig. 1 | PRISMA 2020 flow diagram for an updated systematic review of the Smoking and Tracheal, bronchus, and lung cancer risk-outcome pair. The PRISMA flow diagram of an updated systematic review on the relationship between smoking and lung cancer conducted on PubMed to update historical review from previous cycles of the Global Burden of Disease Study. Template is from: Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 2021;372:n71. doi: 10.1136/bmj.n71. For more information, visit: <http://www.prisma-statement.org/>.



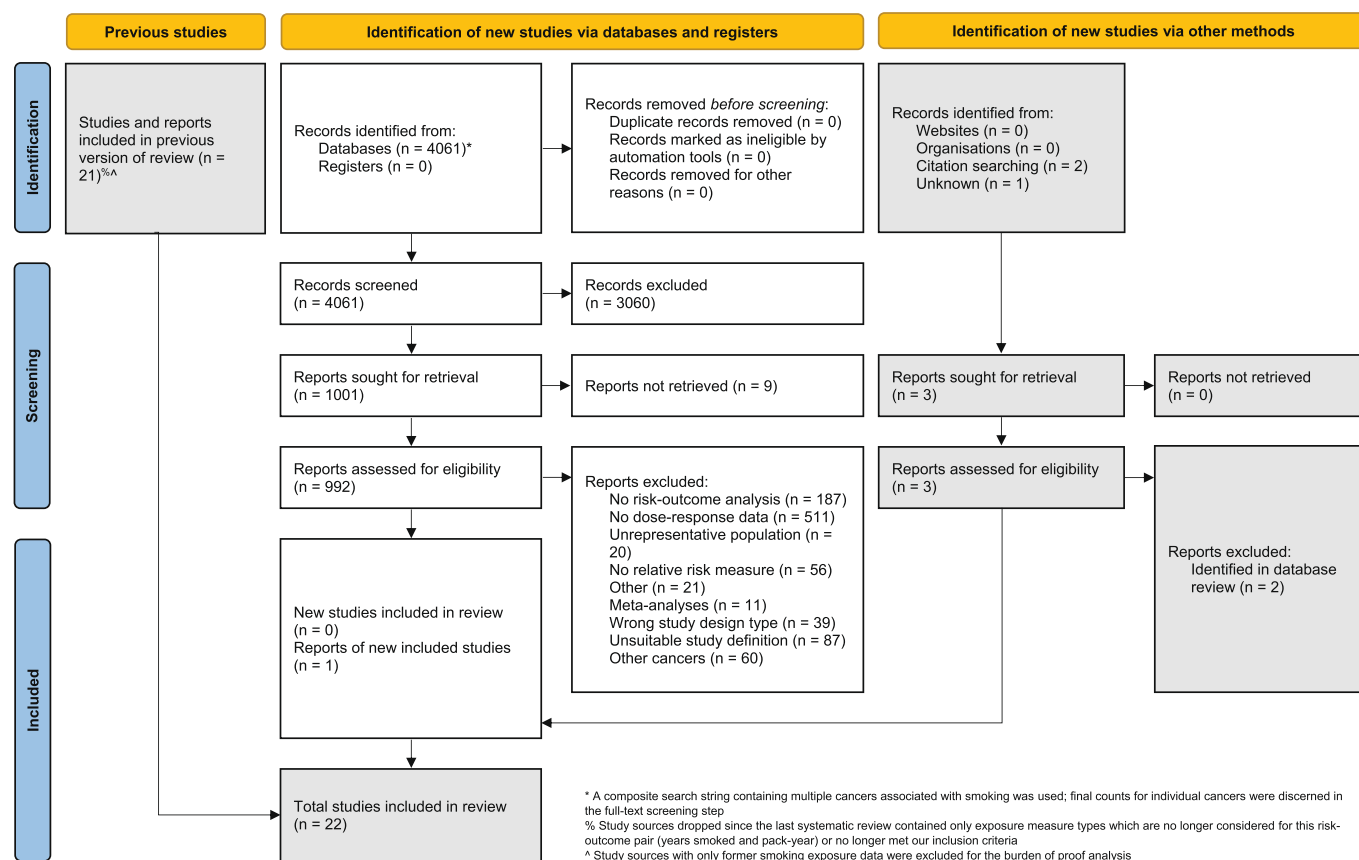
Extended Data Fig. 2 | PRISMA 2020 flow diagram for an updated systematic review of the Smoking and Chronic obstructive pulmonary disease risk-outcome pair. The PRISMA flow diagram of an updated systematic review on the relationship between smoking and chronic obstructive pulmonary disease conducted on PubMed to update historical review from previous cycles of the Global Burden of Disease Study. Template is from: Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 2021;372:n71. doi: 10.1136/bmj.n71. For more information, visit: <http://www.prisma-statement.org/>.



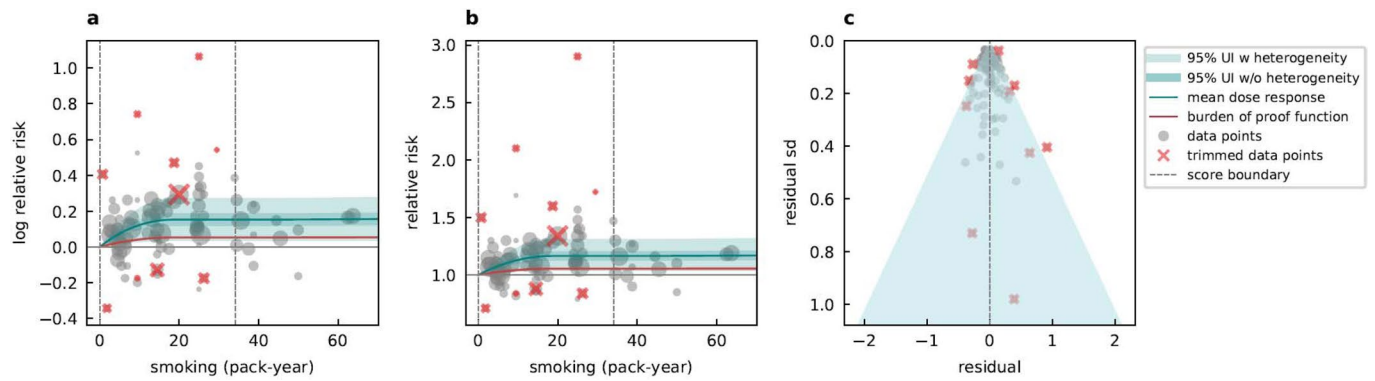
Extended Data Fig. 3 | PRISMA 2020 flow diagram for an updated systematic review of the Smoking and Diabetes mellitus type 2 risk- outcome pair. The PRISMA flow diagram of an updated systematic review on the relationship between smoking and type 2 diabetes conducted on PubMed to update historical review from previous cycles of the Global Burden of Disease Study. Template is from: Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 2021;372:n71. doi: 10.1136/bmj.n71. For more information, visit: <http://www.prisma-statement.org/>.



Extended Data Fig. 4 | PRISMA 2020 flow diagram for an updated systematic review of the Smoking and Breast cancer risk-outcome pair. The PRISMA flow diagram of an updated systematic review on the relationship between smoking and breast cancer conducted on PubMed to update historical review from previous cycles of the Global Burden of Disease Study. Template is from: Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 2021;372:n71. doi: 10.1136/bmj.n71. For more information, visit: <http://www.prisma-statement.org/>.



Extended Data Fig. 5 | PRISMA 2020 flow diagram for an updated systematic review of the Smoking and Prostate cancer risk-outcome pair. The PRISMA flow diagram of an updated systematic review on the relationship between smoking and prostate cancer conducted on PubMed to update historical review from previous cycles of the Global Burden of Disease Study. Template is from: Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 2021;372:n71. doi: 10.1136/bmj.n71. For more information, visit: <http://www.prisma-statement.org/>.



Extended Data Fig. 6 | Smoking and Breast Cancer. **a**, log-relative risk function. **b**, relative risk function. **c**, A modified funnel plot showing the residuals (relative to 0) on the x-axis and the estimated standard deviation (SD) that includes reported SD and between-study heterogeneity on the y-axis.

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection No primary data collection was carried out for this analysis.

Data analysis All codes used for these analyses are publicly available online (<https://github.com/ihmeuw-msca/burden-of-proof>). This includes codes for the meta-regression engine, the model specification interface, the data processing, and risk-specific custom code, as appropriate. Analyses were carried out using R version 3.6.1 and 4.0.5, Python version 3.8, and Stata version 17.

To validate key aspects of the meta-regression model used in this analysis, the following packages were used, as described in Zheng et al: metafor (R package available for download at <https://www.jstatsoft.org/article/view/v036i03>) and dosmesreta (R package available for download at <https://www.jstatsoft.org/article/view/v072c01>).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

The findings from this study are supported by data available in the published literature from PubMed database. Study sources and citations for each risk-outcome pair can be downloaded using the “download” button on each risk curve page at <https://vizhub.healthdata.org/burden-of-proof/>. Study characteristics and citations for all input data used in the analyses are also provided in the Supplementary Table 3. Also see Supplementary Table 2 for a template of the data collection form.

Human research participants

Policy information about [studies involving human research participants and Sex and Gender in Research](#).

Reporting on sex and gender

No primary data collection was carried out for this analysis, so the study does not involve human research participants. As stated in the methods overview, our estimates are not specific to or disaggregated by specific populations, including by sex. That said, relative risk of prostate cancer was only estimated for males and cervical and breast cancer were only estimated for females because of the sex-specific nature of these outcomes. Because estimates were not disaggregated by sex, we included all available data regardless of how or if the input study collected and reported data by sex or gender. We did not perform sex- or gender-based analyses due to limitations in and scarcity of the underlying data. There is no evidence to suggest that relative risk of an outcome due to smoking varies between males and females.

Population characteristics

No primary data collection was carried out for this analysis, so the study does not involve human research participants. The analysis evaluated the evidence on the association between smoking and 36 disease endpoints.

Recruitment

No primary data collection was carried out for this analysis, so we did not recruit participants.

Ethics oversight

This study was approved by the University of Washington IRB Committee (study #9060).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size

The study was a secondary analysis of existing data involving systematic reviews and meta-analyses. No sample size calculation was performed for this meta-analysis; all available data meeting inclusion criteria were included. As reported in the results section, we reviewed 21,108 studies published in PubMed; and 793 studies met our inclusion criteria and thus were included in the analysis.

Data exclusions

Studies were excluded if they did not meet our inclusion criteria. As described in Supplementary Information file, studies were excluded based on the following exclusion criteria: were an aggregate study: meta-analysis or pooled cohort; had the wrong study type: not a cohort study or a case-control study; were a duplicate study: cohort reported in paper was also reported elsewhere; had unmeasurable exposure: reported smoking consumption without details; had no measure of interest: reported RR for change in smoking or doesn't report RR; did not have outcome of interest: reported on all-cause-mortality or an outcome outside of the 36 studied in this paper; were not in English.

Replication

This is a meta-analysis of existing studies with many years of cohort and other data. The data and code have all been made available. It could theoretically be replicated.

Randomization

This analysis is a meta-analysis of existing studies and thus, there were no experimental groups.

Blinding

N/A. Blinding was not relevant to this study, as it was a meta-analysis of existing data and we did not collect primary data.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

- | n/a | Included in the study |
|-------------------------------------|--|
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Antibodies |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Eukaryotic cell lines |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Palaeontology and archaeology |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Animals and other organisms |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Clinical data |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Dual use research of concern |

Methods

- | n/a | Included in the study |
|-------------------------------------|---|
| <input checked="" type="checkbox"/> | <input type="checkbox"/> ChIP-seq |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Flow cytometry |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> MRI-based neuroimaging |