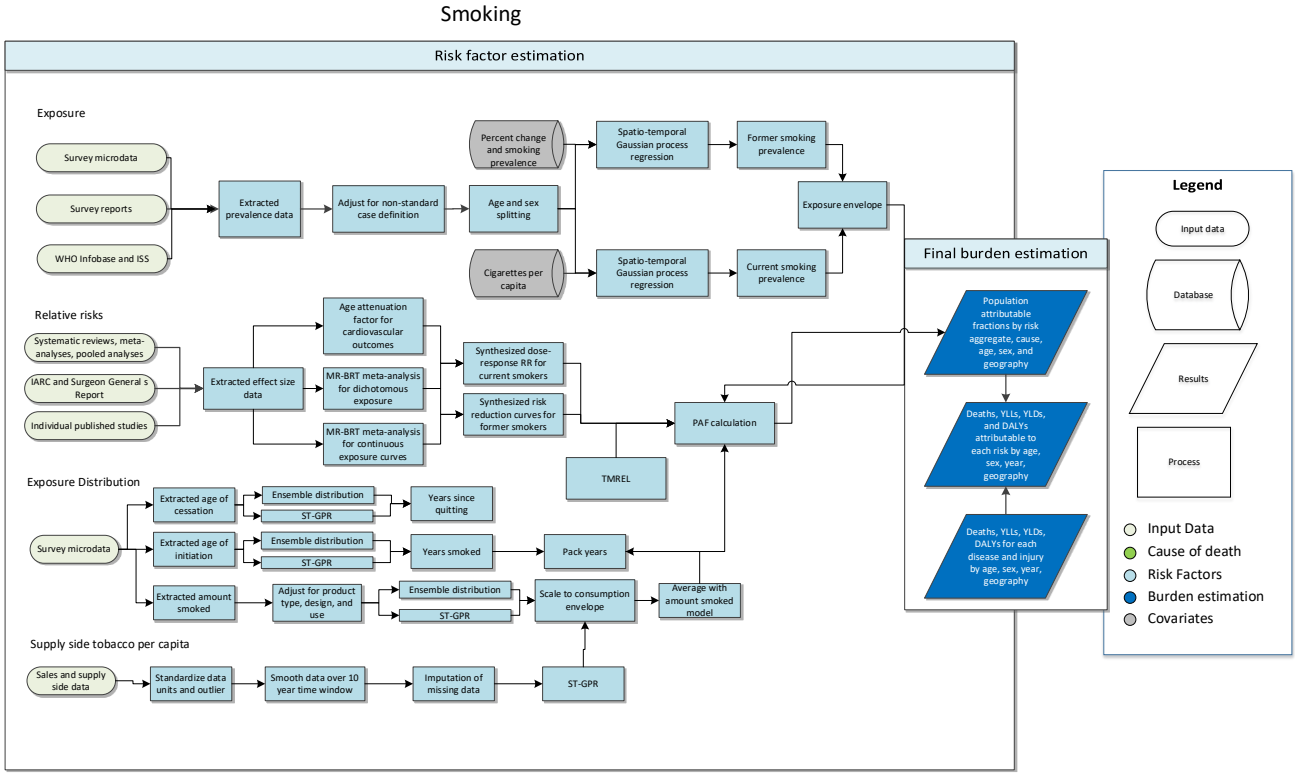
Smoking

Flowchart



Input data and methodological summary

Definition

Exposure

As in previous GBD cycles, we estimated the prevalence of current smoking and the prevalence of former smoking using data from cross-sectional nationally representative household surveys. We defined current smokers as individuals who currently use any smoked tobacco product on a daily or occasional basis. We defined former smokers as individuals who quit using all smoked tobacco products for at least six months, where possible, or according to the definition used by the given survey.

Input data

Exposure

Our survey data extraction method for smoking exposure has not changed from previous GBD cycles. A systematic review of literature was performed to extract data on our primary exposure indicators. We searched Global Health Data Exchange (GHDx), a comprehensive online catalog of health-related data created by IHME, for population survey data. We also included surveys that were recommended by our collaborators but were not in the GHDx. Regarding inclusion and exclusion criteria, we only included

surveys that are nationally or sub-nationally representative (on state/province level). Surveys conducted among specific populations (e.g., pregnant women, physicians) were excluded.

We extracted primary data from individual-level microdata and survey report tabulations. Specifically, we extracted data on current, former, and/or ever smoked tobacco use reported as any combination of frequency of use (daily, occasional, and unspecified, which includes both daily and occasional smokers) and type of smoked tobacco used (all smoked tobacco, cigarettes, hookah, and other smoked tobacco products such as cigars or pipes), resulting in 36 possible combinations. Other variants of tobacco products, for example hand-rolled cigarettes, were grouped into the four type categories listed above based on product similarities.

For microdata, we extracted relevant demographic information, including age, sex, location, and year, as well as survey metadata, including survey weights, primary sampling units, and strata. This information allowed us to tabulate individual-level data in the standard GBD five-year age-sex groups and produce accurate estimates of uncertainty. For survey report tabulations, we extracted data at the most granular age-sex group provided. After data were extracted, we carefully vetted the extracted data, fixed any extraction error and cautiously outliered problematic data due to quality concerns based on expert opinion. We documented relevant survey variables from each data source as well as outliered data in spreadsheets. We extracted data using STATA 13.1 and R 3.3.

Table 1: Data inputs for exposure for smoked tobacco.

	Countries with data	New sources	Total sources
Exposure	201	157	3490

Relative risk

Since GBD 2016 we had performed systematic review and meta-analysis of all case-control and prospective cohort studies reporting a relative risk, hazard ratio, or odds ratio for any risk-outcome pair studied in GBD 2016. In GBD 2019, we had included 36 risk-outcome pairs for smoking. Studies were included if they reported a categorical or continuous dose for smoked tobacco consumption (pack-years or cigarettes per day) as well as uncertainty measures of the estimated risk, and the population under study was general population. Studies were excluded if they used cross-sectional or retrospective cohort design or if the study was conducted among specific populations (e.g., people with diabetes or drug users, etc.).

In GBD 2020, we undertook an effort to improve our relative risk curves by refining our search strings to capture a larger number of studies than was identified by previous search strings. Studies published between 01/01/1970 and 12/31/2019 were reviewed. Of those articles captured, prospective cohort and case-control studies were included if they reported the effect sizes (relative risk, hazard ratio, or odds ratio) of an association between a continuous or categorical dose for smoked tobacco consumption and a GBD outcome with uncertainty. Information on study design, confounders controlled for, sample representativeness, and measurement of exposure and outcomes was also extracted.

In GBD 2020, we also employed a new approach to produce age-specific relative risk (RR) curves for CVD outcomes, which involves estimating an age pattern of excess risk (i.e., RR-1) of smoking for CVD

outcomes. To estimate the age pattern, we performed a systematic review of literature on risk of smoking for five CVD outcomes, namely, ischemic heart disease, stroke, atrial fibrillation and flutter, aortic aneurysm, and peripheral arterial disease. We developed a search string to search for articles reporting any association of binary smoking status (i.e., current, former, and ever smokers) on the five CVD outcomes from 01/01/1970 to 12/31/2019 and only included studies reporting age-specific risk (RR, OR, HR) of smoking status, which is different from the estimation of dose-response risk of smoking for which we only included studies reporting dose-specific risk. Information on study design, confounders controlled for, sample representativeness, type of exposure (i.e., current, former, and ever smoker), measurement of exposure and outcomes was also extracted for bias adjustment. Table 2 summarizes the number of studies included for estimating the dose-response risk curve and the age pattern of risk for the CVD outcomes.

Table 2: Data inputs for relative risks for smoked tobacco use

	Countries with data	New sources	Total sources
Relative risks	55	218	730

Data processing

Crosswalk

Our GBD smoking case definitions were current smoking of any tobacco product and former smoking of any tobacco product. All other definitions were adjusted to be consistent with either of these definitions. Some sources contained information on more than one case definition and these sources were used to develop the adjustment coefficients to transform alternative case definitions to the GBD case definition. The adjustment coefficients were the beta values derived from linear regression models with one predictor and no intercept. We used the same crosswalk adjustment coefficients as in GBD 2019, and thus we have not included a methods explanation in this appendix, as it has been detailed previously.

Age and sex splitting

As in GBD 2019, we split data reported in broader age groups than the GBD 5-year age groups or as both sexes combined by adapting the method reported in Ng et al[1] to split using a sex- geography- time-specific reference age pattern. We separated the data into two sets: a training dataset, with data already falling into GBD sex-specific 5-year age groups, and a split dataset, which reported data in aggregated age or sex groups. We then used spatiotemporal Gaussian process regression (ST-GPR) to estimate sex-geography-time-specific age patterns using data in the training dataset. The estimated age patterns were used to split each source in the split dataset.

The ST-GPR model used to estimate the age patterns for age-sex splitting used an age weight parameter value that minimizes the effect of any age smoothing. This parameter choice allowed the estimated age pattern to be driven by data, rather than being enforced by any smoothing parameters of the model. These age-sex split data points were to be incorporated in the final ST-GPR exposure model; thus, we

did not want to doubly enforce a modelled age pattern for a given sex-location-year on a given aggregate data point.

Modelling strategy

Smoking prevalence modelling

We used ST-GPR to model current and former smoking prevalence. The model is identical to that in GBD 2019. Full details on the ST-GPR method are reported elsewhere in the appendix. Briefly, the mean function input to GPR is a complete time series of estimates generated from a mixed effects hierarchical linear model plus weighted residuals smoothed across time, space, and age. The linear model formula for current smoking, fit separately by sex using restricted maximum likelihood in R, is:

$$logit(p_{g,a,t}) = \beta_0 + \beta_1 CPC_{g,t} + \sum_{k=2}^{19} \beta_k I_{A[a]} + \alpha_s + \alpha_r + \alpha_g + \epsilon_{g,a,t}$$

Where $CPC_{g,t}$ is the tobacco consumption per capita covariate by geography g and time t, described above, $I_{A[a]}$ is a dummy variable indicating specific age group that the prevalence point $p_{g,a,t}$ captures, and and a_s, a_r, and a_g are super-region, region, and geography random intercepts, respectively. Random effects were used in model fitting but not in prediction.

The linear model formula for former smoking is:

$$logit(p_{g,a,t}) = \beta_0 + \beta_1 PctChange_{A[a],g,t} + \beta_3 CSP_{A[a],g,t} + \sum_{k=3}^{20} \beta_k I_{A[a]} + \alpha_s + \alpha_r + \alpha_g + \epsilon_{g,a,t}$$

Where $PctChange_{A[a],g,t}$ is the percentage change in current smoking prevalence from the previous year, and $CSP_{A[a],g,t}$ is the current smoking prevalence by specific age group A, geography g, and time t that point $p_{g,a,t}$ captures, both derived from the current smoking ST-GPR model defined above.

Supply-side estimation

The methods for modelling supply-side-level data were consistent with those used in GBD 2019. The raw data were domestic supply (USDA Global Surveillance Database and UN FAO) and retail supply (Euromonitor) of tobacco. Domestic supply was calculated as production + imports - exports. The data went through three rounds of outliering based on reasonable consumption thresholds of number of cigarettes per smoker per day, distance from the 10-year rolling mean tobacco per capita, and manual outliering for edge cases. Finally, data smoothing was performed by taking a three-year rolling mean over each location-year.

Next, to impute the missing years for each series and remove compositional bias from our final estimates, we modelled the log ratio of each pair of sources as a function of an intercept and nested

random effects on super-region, region, and location. The appropriate predicted ratio was multiplied by each source that we did have, and then the predictions were averaged to get the final imputed value. For some locations where there was limited overlap between series, the predicted ratio did not make sense, and a regional ratio was used.

Finally, variance was calculated both across series (within a location-year) as well as across years (within a location-source). Additionally, if a location-year had one imputed point, the variance was multiplied by 2. If a location-year had two imputed points, the variance was multiplied by 4. The average estimates in each location-year were the input to an ST-GPR model. For this, we used a simple mixed effects model, which was modelled in log space with nested location random effects. Subnational estimates were then further modelled by splitting the country-level estimates using current smoking prevalence.

Theoretical minimum-risk exposure level

The theoretical minimum-risk exposure level is 0.

Exposure among current and former smokers

Identical to GBD 2019, we estimated exposure among current smokers for two continuous indicators: cigarettes per smoker per day and pack-years. Pack-years incorporates aspects of both duration and amount. One pack-year represents the equivalent of smoking one pack of cigarettes (assuming a 20-cigarette pack) per day for one year. Since the pack-years indicator collapses duration and intensity into a single dimension, one pack-year of exposure can reflect smoking 40 cigarettes per day for six months or smoking 10 cigarettes per day for two years.

To produce these indicators, we simulated individual smoking histories based on distributions of age of initiation and amount smoked. We informed the simulation with cross-sectional survey data capturing these indicators, modelled at the mean level for all locations, years, ages, and sexes using ST-GPR. We rescaled estimates of cigarettes per smoker per day to an envelope of cigarette consumption based on supply-side data. We estimated pack-years of exposure by summing samples from age- and time-specific distributions of cigarettes per smoker for a birth cohort in order to capture both age trends and time trends and avoid the common assumption that the amount someone currently smokes is the amount they have smoked since they began smoking. All distributions were age-, sex-, and region- specific ensemble distributions, which were found to outperform any single distribution.

We estimated exposure among former smokers using years since cessation. We utilized ST-GPR to model mean age of cessation using cross-sectional survey data capturing age of cessation. Using these estimates, we generated ensemble distributions of years since cessation for every location, year, age group, and sex.

Relative risk

The same risk-outcome pairs from GBD 2019 were used for GBD 2020: tuberculosis, lower respiratory tract infections, oesophageal cancer, stomach cancer, bladder cancer, liver cancer, laryngeal cancer, lung cancer, breast cancer, cervical cancer, colorectal cancer, lip and oral cancer, nasopharyngeal cancer, other pharyngeal cancer, pancreatic cancer, kidney cancer, leukaemia, ischaemic heart disease,

ischaemic stroke, haemorrhagic stroke, subarachnoid haemorrhage, atrial fibrillation and flutter, aortic aneurysm, peripheral arterial disease, chronic obstructive pulmonary disease, other chronic respiratory diseases, asthma, peptic ulcer disease, gallbladder and biliary tract diseases, Alzheimer disease and other dementias, Parkinson disease (protective), multiple sclerosis, type-II diabetes, rheumatoid arthritis, low back pain, cataracts, macular degeneration, and fracture.

For GBD 2020, the risk of all risk outcome pairs is evaluated by continuous smoking exposure level (i.e., pack-year, cigarettes per smoker per day and years since cessation), expect for fracture, whose risk is evaluated by binary smoking exposure (i.e., smoker vs non-smoker/former smoker).

Dose-response risk curves

Since GBD 2016, we had used the studies identified through the systematic review to estimate dose-response risk of smoking on related health outcomes, using DisMod ODE. We chose DisMod ODE rather than a conventional mixed effects meta-regression because of its ability to estimate nonparametric splines over doses (i.e., there is usually a non-linear relationship between smoking exposure level and outcome risk) and incorporate heterogeneous doses through dose-integration (i.e., most studies report smoking exposure level categorically in wide ranges and DisMod ODE can estimate risk of specific exposure level when categories overlap across studies, through an integration step).

For GBD 2020, we used the studies identified through the updated systematic review to estimate new dose-response curves using MR-BRT for all outcomes. Importantly, this new method takes into account the risk of biases in the RR estimation by selecting and including important covariates of the risk estimates in the model (e.g., measurement of exposure and outcomes, representativeness and adjustment level of the risk estimates) and incorporates unexplained between-study heterogeneity into the uncertainty of the RR estimates. The results of the meta-regression were used to estimate a non-parametric curve for all doses between zero and 100 pack-year or cigarettes per smoker per day and their corresponding relative risks. For all outcomes, we assumed the relative risk was the same for both sexes, expect for breast cancer, cervical cancer, and prostate cancer, which were assumed to apply only to female or to male.

For data-sparse risk-outcome pairs, we implemented the Fisher Scoring correction to the heterogeneity parameter. When data are sparse, the between-study heterogeneity parameter estimate may be 0, simply due to lack of data. The Fisher Scoring correction uses a quantile of gamma, which is sensitive to the number of studies, study design, and reported uncertainty.

We have also added methodology that can detect and flag publication bias. The approach is based on the classic Egger's Regression strategy, which is applied to the residuals in our model. In the current implementation, we do not correct for publication bias, but flag the risk-outcome pairs where the risk for publication bias is significant.

For risk of former smokers, we estimated risk curves of former smokers compared to never smokers taking into account the rate of risk reduction among former smokers seen in the cohort and case-control studies, and the cumulative exposure among former smokers within each age, sex, location, and year

group. For GBD 2020, we did not include new data or change the method of estimating the risk curves of former smokers.

In the table below, we list each risk outcome pair that is updated in GBD 2020 along with several of the key modelling parameters and results. The formulation for MR-BRT is described in detail later in this appendix.

Table 5: MR BRT model specifications by risk-outcome pair

Risk-outcome	Type of risk	Spline degree, # interior knots	Priors & constraints
Atrial fibrillation and flutter	Continuous, Harmful	Quadratic, 3 I knots	Monotonic increasing, right linear tail, Gaussian max derivative prior on the right tail (0, 0.001)
Alzheimer and other dementia	Continuous, Harmful	Quadratic, 3 I knots	Monotonic increasing, right linear tail, Gaussian max derivative prior on the right tail (0, 0.001)
Aortic aneurism	Continuous, Harmful	Quadratic, 3 I knots	Monotonic increasing, right linear tail, Gaussian max derivative prior on the right tail (0, 0.001)
Asthma	Continuous, Harmful	Quadratic, 3 I knots	Monotonic increasing, right linear tail, Gaussian max derivative prior on the right tail (0, 0.001)
Bladder cancer	Continuous, Harmful	Quadratic, 3 I knots	Monotonic increasing, right linear tail, Gaussian max derivative prior on the right tail (0, 0.001)
Breast cancer	Continuous, Harmful	Quadratic, 3 I knots	Monotonic increasing, right linear tail, Gaussian max derivative prior on the right tail (0, 0.001)
Cataracts	Continuous, Harmful	Quadratic, 3 I knots	Monotonic increasing, right linear tail, Gaussian max derivative prior on the right tail (0, 0.001)
Cervical cancer	Continuous, Harmful	Quadratic, 3 I knots	Monotonic increasing, right linear tail, Gaussian max derivative prior on the right tail (0, 0.001)
Colon and rectum cancer	Continuous, Harmful	Quadratic, 3 I knots	Monotonic increasing, right linear tail, Gaussian max derivative prior on the right tail (0, 0.001)
COPD	Continuous, Harmful	Quadratic, 3 I knots	Monotonic increasing, right linear tail, Gaussian max derivative prior on the right tail (0, 0.001)
Diabetes	Continuous, Harmful	Quadratic, 3 I knots	Monotonic increasing, right linear tail, Gaussian max derivative prior on the right tail (0, 0.001)
Esophageal cancer	Continuous, Harmful	Quadratic, 3 I knots	Monotonic increasing, right linear tail, Gaussian max derivative prior on the right tail (0, 0.001)

Gallbladder diseases	Continuous, Harmful	Quadratic, 3 I knots	Monotonic increasing, right linear tail, Gaussian max derivative prior on the right tail (0, 0.001)
Fracture (hip and non-hip)	Dichotomous, Harmful	N/A	N/A
Ischemic health disease	Continuous, Harmful	Quadratic, 3 I knots	Monotonic increasing, right linear tail, Gaussian max derivative prior on the right tail (0, 0.001)
Kidney cancer	Continuous, Harmful	Quadratic, 3 I knots	Monotonic increasing, right linear tail, Gaussian max derivative prior on the right tail (0, 0.001)
Laryngeal cancer	Continuous, Harmful	Quadratic, 3 I knots	Monotonic increasing, right linear tail, Gaussian max derivative prior on the right tail (0, 0.001)
Lower back pain	Continuous, Harmful	Quadratic, 3 I knots	Monotonic increasing, right linear tail, Gaussian max derivative prior on the right tail (0, 0.001)
Leukemia	Continuous, Harmful	Quadratic, 3 I knots	Monotonic increasing, right linear tail, Gaussian max derivative prior on the right tail (0, 0.001)
Lip oral cavity cancer	Continuous, Harmful	Quadratic, 3 I knots	Monotonic increasing, right linear tail, Gaussian max derivative prior on the right tail (0, 0.001)
Liver cancer	Continuous, Harmful	Quadratic, 3 I knots	Monotonic increasing, right linear tail, Gaussian max derivative prior on the right tail (0, 0.001)
lower respiratory infections	Continuous, Harmful	Quadratic, 3 I knots	Monotonic increasing, right linear tail, Gaussian max derivative prior on the right tail (0, 0.001)
Lung cancer	Continuous, Harmful	Quadratic, 3 I knots	Monotonic increasing, right linear tail, Gaussian max derivative prior on the right tail (0, 0.001)
Macular degeneration	Continuous, Harmful	Quadratic, 3 I knots	Monotonic increasing, right linear tail, Gaussian max derivative prior on the right tail (0, 0.001)
Multiple sclerosis	Continuous, Harmful	Quadratic, 3 I knots	Monotonic increasing, right linear tail, Gaussian max derivative prior on the right tail (0, 0.001)
Nasopharyngeal cancer	Continuous, Harmful	Quadratic, 3 I knots	Monotonic increasing, right linear tail, Gaussian max derivative prior on the right tail (0, 0.001)
Other pharynx cancer	Continuous, Harmful	Quadratic, 3 I knots	Monotonic increasing, right linear tail, Gaussian max derivative prior on the right tail (0, 0.001)
Pancreatic cancer	Continuous, Harmful	Quadratic, 3 I knots	Monotonic increasing, right linear tail, Gaussian max derivative prior on the right tail (0, 0.001)

Parkinson	Continuous, Protective	Quadratic, 3 I knots	Monotonic decreasing, right linear tail, Gaussian max derivative prior on the right tail (0, 0.001)
Peptic ulcer	Continuous, Harmful	Quadratic, 3 I knots	Monotonic increasing, right linear tail, Gaussian max derivative prior on the right tail (0, 0.001)
Peripheral artery disease	Continuous, Harmful	Quadratic, 3 I knots	Monotonic increasing, right linear tail, Gaussian max derivative prior on the right tail (0, 0.001)
Prostate cancer	Continuous, Harmful	Quadratic, 3 I knots	Monotonic increasing, right linear tail, Gaussian max derivative prior on the right tail (0, 0.001)
Rheumatoid arthritis	Continuous, Harmful	Quadratic, 3 I knots	Monotonic increasing, right linear tail, Gaussian max derivative prior on the right tail (0, 0.001)
Stomach cancer	Continuous, Harmful	Quadratic, 3 I knots	Monotonic increasing, right linear tail, Gaussian max derivative prior on the right tail (0, 0.001)
Stroke (ischemic stroke, hemorrhagic stroke, and subarachnoid hemorrhage)	Continuous, Harmful	Quadratic, 3 I knots	Monotonic increasing, right linear tail, Gaussian max derivative prior on the right tail (0, 0.001)
Tuberculosis	Continuous, Harmful	Quadratic, 3 I knots	Monotonic increasing, right linear tail, Gaussian max derivative prior on the right tail (0, 0.001)

Table 6: MR BRT estimated parameters and bias covariates by risk-outcome pair

Risk-outcome	Unit of risk	Selected bias covariates	Mean gamma solution	publication bias
Atrial fibrillation and flutter	cigarettes per day	None	0.000	0
Alzheimer and other dementia	cigarettes per day	None	0.000	1
Aortic aneurism	cigarettes per day	None	0.000	0
Asthma	cigarettes per day	None	1.651	0
Bladder cancer	pack-year	None	0.000	0
Breast cancer	pack-year	None	0.000	0
Cataracts	cigarettes per day	None	0.000	0
Cervical cancer	pack-year	None	0.000	0
Colon and rectum cancer	pack-year	None	0.000	0
COPD	pack-year	cv_subpopulation, cv_adj_L1	0.000	1
Diabetes (type II)	cigarettes per day	cv_subpopulation	0.105	0
Esophageal cancer	pack-year	cv_exposure_selfreport	0.000	0
Gallbladder diseases	cigarettes per day	cv_adj_L0	0.000	0
Fracture (hip and non-hip)	Binary smoking status	cv_adj_L1	0.099	1

Ischemic health disease	cigarettes per day	cv_adj_L2, cv_subpopulation, cv_adj_L1, cv_adj_L0, cv_older	0.206	1
Kidney cancer	pack-year	None	0.000	1
Laryngeal cancer	pack-year	cv_adj_L0	0.000	0
Lower back pain	cigarettes per day	None	0.000	0
Leukemia	pack-year	None	0.000	0
Lip oral cavity cancer	pack-year	cv_adj_L0	0.105	1
Liver cancer	pack-year	None	0.214	1
lower respiratory infection	cigarettes per day	None	0.000	0
Lung cancer	pack-year	cv_adj_L1, cv_adj_L0	0.058	1
Macular degeneration	cigarettes per day	None	0.000	0
Multiple sclerosis	cigarettes per day	None	0.000	0
Nasopharyngeal cancer	pack-year	None	0.071	1
Other pharynx cancer	pack-year	cv_exposure_selfreport	0.000	0
Pancreatic cancer	pack-year	None	0.000	0
Parkinson	cigarettes per day	cv_adj_L2, cv_outcome_selfreport, cv_older	0.000	1
Peptic ulcer	cigarettes per day	cv_adj_L1, cv_subpopulation	0.000	0
Peripheral artery disease	cigarettes per day	cv_subpopulation	0.000	0
Prostate cancer	cigarettes per day	None	0.155	1
Rheumatoid arthritis	cigarettes per day	None	0.000	1
Stomach cancer	pack-year	None	0.000	0
Stroke (ischemic stroke, hemorrhagic stroke, and subarachnoid hemorrhage)	cigarettes per day	None	0.104	0
Tuberculosis	cigarettes per day	None	0.099	0

† definitions of bias covariates:

cv_subpopulation: 0 for risk estimates are likely generalizable to the general population because the sample was based on the general population with reasonable exclusions for pre-existing disease states; 1 for risk estimates of sub-groups such as high-risk groups

cv_adj_L0, cv_adj_L1, cv_adj_L2: cascading dummy variables for adjustment level of the risk estimates (i.e., how many confounders are adjusted for in the regression model for the risk estimate). There are four adjustment levels, namely, 1.no adjustment, 2.only adjusting for age and sex, 3.adjusting for age and sex and <= 3 other covariates, and 4.adjusting for age and sex and > 3 other covariates. If the adjustment level is 1, cv_adj_L0=1, cv_adj_L1=1, cv_adj_L2=1; if the adjustment level is 2, cv_adj_L0=1, cv_adj_L1=1, cv_adj_L2=0; if the adjustment level is 3, then cv_adj_L0=1, cv_adj_L1=0, cv_adj_L2=0; if the adjustment level is 4, then cv_adj_L0=0, cv_adj_L1=0, cv_adj_L2=0.

cv_exposure_selfreport: 0 for measurement of exposure based on assays, tests, or physician observation and 1 for self-report exposure.

cv_outcome_selfreport: 0 for measurement of outcome based on assays, tests, or physician observation and 1 for self-report outcome.

cv_older: 0 if the population contains both young and old people; 1 if the population only contains old people.

Age-specific dose-response risk curves for CVD outcomes

For all non-CVD outcomes, we assumed the risk curve to be the same for all ages. However, the risk of smoking on CVD outcomes (i.e., stroke, ischaemic heart disease, atrial fibrillation and flutter, aortic aneurysm, and peripheral arterial disease) is well known to attenuate with increasing age and thus we produced age-specific risk curves for all CVD outcomes. Previously, we used a linear relationship between age and log risk to adjust all RR data to a specific age group (e.g., 45-49). Then, we modeled the risk curve for each age group using the adjusted age-group specific data. This approach often produced curves with different shapes for different age groups and tended to underestimate the risk for older age groups since we set the log RR to be 0 for the terminal age group (e.g., 95+) in the linear function.

In GBD 2020, we adopted a new approach to produce the age-specific risk curves by producing an age pattern of smoking risk on CVD outcomes and adjusting the risk curve of the reference age group using the age pattern of risk to produce age-group specific risk curves. Briefly, we first estimated the reference dose-response risk of smoking for each CVD outcome using dose-specific RR data of each outcome regardless of the age group information. This step was the same with other non-CVD outcomes. Once we had the reference curve, we determined the age group of the reference curve by calculating the weighted mean age across all dose-specific RR data (weighted by 1/SE of each datum). For example, if the weighted mean age of all dose-specific RR data was 56.5, we determined the age group of the reference risk curve to be 55-59. For cohort studies, the mean age was calculated as mean age at baseline plus the mean/median years of follow-up (if only maximum years of follow-up is reported, we added half of the maximum years to the mean age at baseline). For the case-control studies, the mean age was just the reported mean age at baseline (in case the mean age is not reported, we used the midpoint of age range as the mean age instead). In the third step, we extracted age-group specific RR data and relevant bias covariates from literature identified in the systematic review mentioned above, and we used MR-BRT to model the age pattern of excess risk (i.e., RR-1) of smoking on CVD outcomes with age-group specific excess RR data of all CVD outcomes. In the final model, we included age as spline, random effects of study, and the bias covariates of exposure types (i.e., current, former, and ever smokers), which were selected by an algorithm described elsewhere.[2] 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. Figure S1 below shows the estimated age pattern of excess risk of smoking on CVD along with its 95% uncertainty intervals. In the fourth step, we calculated the age attenuation factors (AF) of excess risk compared with the reference age group for each CVD outcome as ratio of the estimated excess risk of each age group to that of the reference age group. We did the calculation at the draw level to obtain 1000 draws of the AF for each age group. Figure S2 below shows the AF for stroke along with its 95% uncertainty intervals. Once we had the AF, in the last step, we adjusted the risk curve of the reference age group from step 1 using equation (1) to produce the age-group specific risk curves for each CVD outcome.

$$rr_{age_i} = \left(rr_{ref} - 1\right) * AF_{age_i} + 1 \qquad (1)$$

We did the age adjustment on draw level so that the uncertainty of the AF can be naturally incorporated in the final adjusted age-specific RR curves. Figure S3 shows the age AF adjusted age-group specific RR curves for stroke outcome.

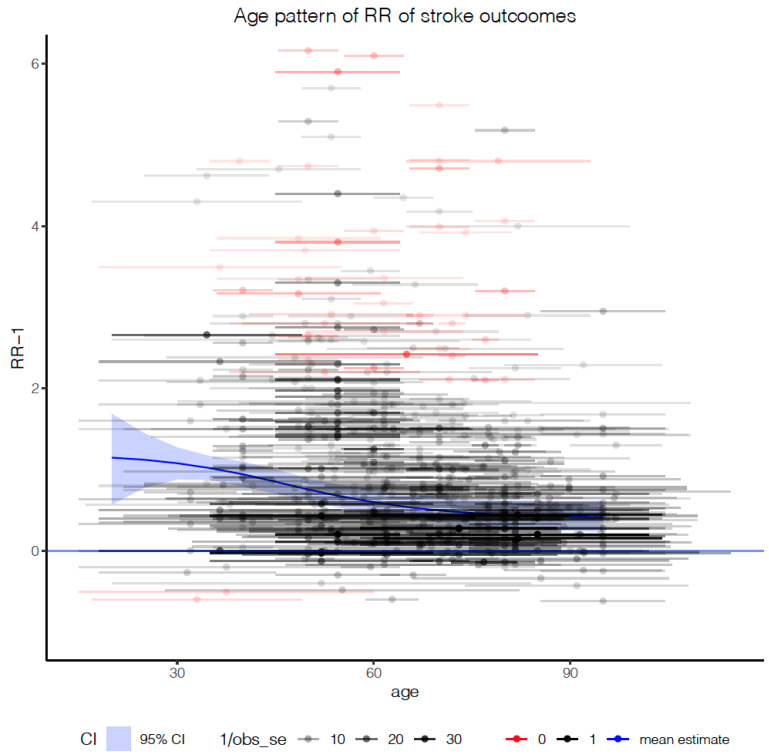


Figure S1: estimated age pattern of excess risk of smoking on CVD outcomes

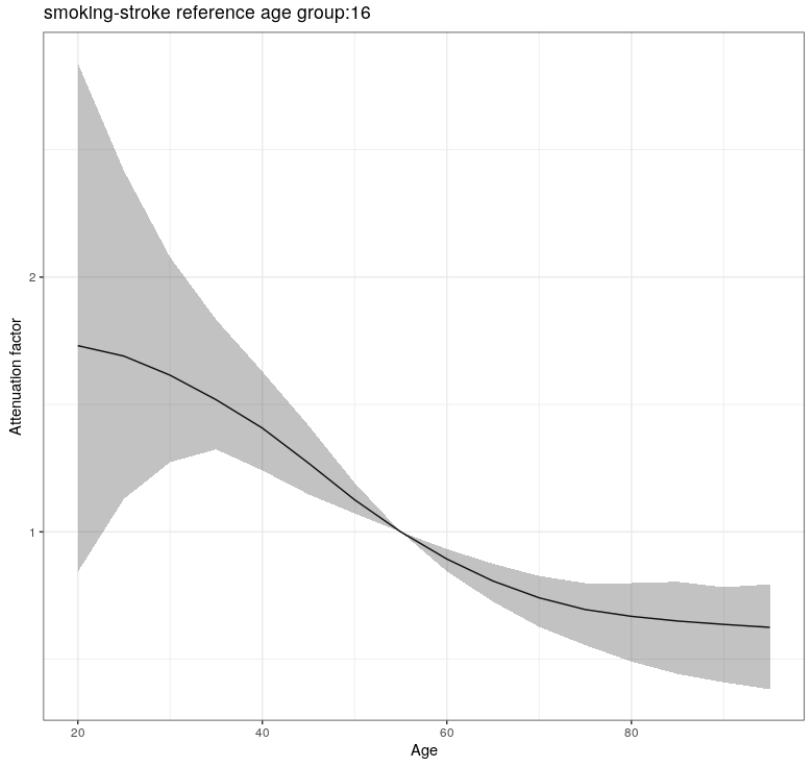


Figure S2: attenuation factors of excess risk of smoking on stroke campared with the reference age group

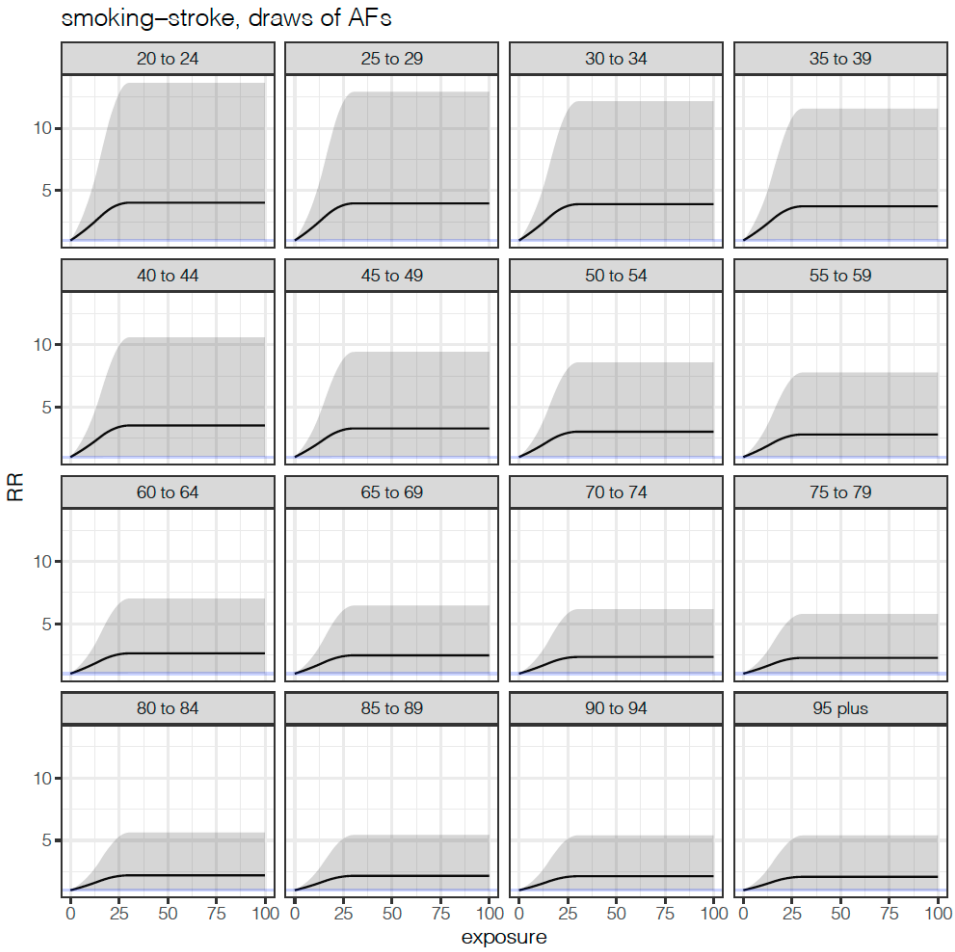


Figure S3: AF adjusted age-group specific RR curves for stroke, reference age group 55-59. The exposure is in units of cigarette-equivalents per smoker per day.

Population attributable fraction (PAF)

As in GBD 2019, we estimated PAFs based on the following equation:

$$PAF = \frac{p(n) + p(f)\int \exp(x) * rr(x) + p(c)\int \exp(y) * rr(y) - 1}{p(n) + p(f)\int \exp(x) * rr(x) + p(c)\int \exp(y) * rr(y)}$$

where $p(n)$ is the prevalence of never smokers, $p(f)$ is the prevalence of former smokers, $p(c)$ is the prevalence of current smokers, $exp(x)$ is a distribution of years since quitting among former smokers, $rr(x)$ is the relative risk for years since quitting, $exp(y)$ is a distribution of cigarettes per smoker per day or pack-years, and $rr(y)$ is the relative risk for cigarettes per smoker per day or pack-years.

We used pack-years as the exposure definition for cancers and chronic respiratory diseases, and cigarettes per smoker per day for cardiovascular diseases and all other health outcomes.

Citations

1. Ng M, Freeman MK, Fleming TD, Robinson M, Dwyer-Lindgren L, Thomson B, et al. Smoking prevalence and cigarette consumption in 187 countries, 1980-2012. JAMA. 2014 Jan 8;311(2):183–92.

2. Zheng P, Aravkin A, Barber R, Sorensen R, Murray C. Trimmed Constrained Mixed Effects Models: Formulations and Algorithms. bioRxiv. 2020 Jan 29;2020.01.28.923599.