Online supplemental material 2. Statistical methods used for mortality rate projections

Previous studies have shown that historical patterns in tobacco smoking are a strong predictor of lung cancer mortality.[1-3] We therefore hypothesised that smoking patterns would also be an important factor in the projection of mortality rates for other cancers which have been shown to have a relationship with cigarette smoking.[4] In this study we fitted standard age-period-cohort (APC) models, and APC models and generalised linear models (GLMs) incorporating smoking data to each cancer type and the two large smoking-related cancer groups.[2 5 6] Possible lags of 25-35 years between cigarette tar exposure and cancer mortality were examined for each fitted model. The most appropriate statistical projection model was selected using the model fit statistics and then validated with the observed data.[2] Model validation provides important information on the performance and reliability of the projection model. This was achieved by withholding the observed data for 2000-2019 from the model fitting and then comparing the projected rates for those years with the actual observed values.[2 7] The final model for each cancer type is listed in Table S2.

To project the mortality rates for lung cancer, we used previously validated GLMs including age, cohort and cigarette smoking exposure.[2] A standard APC model was used to project the mortality rate for each remaining cancer type, as the term for cigarette smoking exposure was not statistically significant in the full APC models.[8] We do not provide individual projections for cervical cancer mortality as the projections based on historical data are likely to be inaccurate due to recent changes in the cervical cancer screening policy. Although models including smoking patterns could not be fitted for individual cancer types due to relatively small numbers, these projection models could be used for groups of smoking-related cancer types. For the group of cancer types with >30% of cases currently caused by smoking, GLMs

Luo Q, *et al. J Epidemiol Community Health* 2022;0:1–8. doi: 10.1136/jech-2021-218252

including age, cohort and cigarette tar exposure (smoking exposure lagged 27 years for men and 29 years for women) were used to project mortality rates, as the period effect was not significant in models including the cigarette smoking variable (Figure S1).[2] For the group of cancer types with 8-30% of cases currently caused by smoking, APC models incorporating cigarette smoking exposure (smoking exposure lagged 32 years for men and 33 years for women)[6] were used to project mortality rates (Figure S2). The final models for both cancer groups confirmed that historical levels of tobacco smoking are significant predictors of mortality rates for smoking-related cancers. The estimated numbers of deaths for these two cancer groups were then combined to estimate the overall projected mortality rate for all smoking-related cancers. In the validation, the 20-year projected mortality rates for both men and women were found to be close to the observed values, suggesting that the final models provide valid projections of mortality for smoking-related cancers in Australia (Figure S3).

Age-period-cohort models

To project cancer mortality for individual cancer types except lung cancer, we used APC models including age, period and cohort components within the framework of a GLM with Poisson distribution. The APC models were fitted by the apcspline command in Stata 17 with natural cubic splines for smoothing. [5] Briefly, the APC model with the log-link function can be expressed as:

$$lnD_{ij} = lnN_{ij} + \alpha Age_i + \beta Period_j + \gamma Cohort_k$$

where D_{ij} denotes the number of deaths from cancer for the i^{th} age group during the j^{th} calendar period; N_{ij} denotes the number at risk in the population for the i^{th} age group during the j^{th} calendar period; α is the coefficient of the age component for age group i; β is the non-linear coefficient of the period component for period j, and γ is the non-linear coefficient of the cohort component for birth cohort k. For each cancer type, we compared several APC

Luo Q, *et al. J Epidemiol Community Health* 2022;0:1–8. doi: 10.1136/jech-2021-218252

models with different numbers of knots for the age, period and cohort effects, to identify the one with the lowest Bayesian Information Criterion (BIC). The final models are listed in Table S2.

To project mortality rates beyond the observed period, future periods and cohorts were assumed to have the same effect as those for the most recent observed period and cohort. As these historical trends will not continue indefinitely, the default setting for the damping factor (equal to 0.92) was used, so that the drift was reduced by 8% for each year following the last observation.[5]

While a cervical cancer screening programme was first introduced in Australia in the early 1990s, a new protocol and screening technology were introduced in 2017 and so projections based on historical data are likely to be inaccurate.[9] Furthermore, only limited data on the coverage achieved by the human papilloma virus vaccination programme are available and there are no data for participation in the updated screening programme. Therefore, we do not provide individual projections for cervical cancer mortality and the smoking attributable deaths for cancer of the cervix were not estimated in this study.

Generalised linear model including age, cohort and smoking as a covariate

We previously developed and validated a statistical model for the projection of lung cancer mortality rates which included tobacco consumption as one of the covariates.[2] A detailed explanation of the method is provided elsewhere.[2] Here, we applied this model both for the projection of lung cancer mortality rates and for the projection of mortality rates for the group of smoking-related cancers with >30% of cases caused by smoking.

The final fitted model for each sex can be presented as a parsimonious equation:

$$lnD_{ij} = lnN_{ij} + \alpha Age_i + \gamma Cohort_k + \delta CTC_{ij-L}$$

where D_{ij} denotes the number of deaths from lung cancer or the group of smoking-related cancers with >30% of cases caused by smoking for the i^{th} age group during the j^{th} calendar period; N_{ij} denotes the number at risk in the population for the i^{th} age group during the j^{th} calendar period; α is the coefficient of the age component for age group i; γ is the non-linear coefficient of the cohort component for birth cohort k; δ is the coefficient of the CTC_{ij-} , which denotes the sex-age-period-specific cigarette tar exposure in the population for the i^{th} age group during the j-L^{th} calendar period, which is lagged by L years (for lung cancer mortality, lagged 26 years for men and 29 years for women; for the group of smoking-related cancers with >30% of cases caused by smoking, lagged 27 years for men and 29 years for women). To project mortality rates beyond the observed calendar period, these models assumed that the age effects remained constant over time.[1 3] We used cohort-specific cigarette tar exposure to predict the future cohort effects as in previous work.[2] Estimated coefficients for the final model for smoking-related cancers with >30% of cases caused by smoking are shown in Figure S1.

Age-period-cohort model with smoking as a covariate

For the group of smoking-related cancers with 8-30% of cases caused by smoking, we used a modified APC model incorporating cigarette tar exposure as a covariate to project the mortality rates. The period effect was included in these models to capture potential changes in exposure to major risk factors other than smoking.

The final fitted model can be presented as a parsimonious equation:

$$lnD_{ij} = lnN_{ij} + \alpha Age_i + \beta Period_j + \gamma Cohort_k + \delta CTC_{ij-L}$$

where D_{ij} denotes the number of deaths from smoking-related cancers with 8-30% of cases caused by smoking for the i^{th} age group during the j^{th} calendar period; N_{ij} denotes the number

at risk in the population for the i^{th} age group during the j^{th} calendar period; α is the coefficient of the age component for age group i; β is the non-linear coefficient of the period component for period j, and γ is the non-linear coefficient of the cohort component for birth cohort k; δ is the coefficient of the CTC_{ij-L}, which denotes the sex-age-period-specific cigarette tar exposure in the population for the i^{th} age group during the j-L^{th} calendar period, which is lagged by L years (32 years for men and 33 years for women).

The process of using the modified APC model involves two steps: first, the models were fitted by the updated apcspline command in Stata 17 with natural cubic splines for smoothing which can include age-period-specific cigarette tar exposure as a covariate in the APC model.[6] Second, the coefficients for birth cohorts were extracted from the best model selected based on the Bayesian Information Criterion (BIC), which was merged with sex-cohort specific cigarette tar exposure. The future cohort parameters were estimated by fitting a linear regression model for cohort coefficients and the cohort-specific cigarette tar exposure. To project lung cancer incidence rates beyond the observed calendar period these models assumed that the age effects remained constant over time.[1 3] The default setting for the damping factor (equal to 0.92) was used, so that the drift was reduced by 8% for each year following the last observation, and future periods were assumed to have the same effect as those for the most recent observed period.[5] The original programme has been modified, so that we used cohort-specific cigarette tar exposure to predict the future cohort effects.[2] Estimated coefficients for the final model for smoking-related cancers with 8-30% of cases caused by smoking are shown in Figure S2.

Table S2. Statistical methods used for mortality projections for individual cancer types, the two large smoking-related cancer groups and all smoking-related cancers combined included in this study

Cancer groups (ICD 10-code)	Mortality	
	Data used for projection model	**Projection model**
Bladder (C67)	1970-2019	APC model
Colon and rectum (C18-20, C26.0)	1970-2019	Age-stratified APC model for age <50 years and age 50+ years.
Gallbladder and bile duct (C23-C24)	1970-2019	APC model
Kidney (C64)	1970-2019	APC model
Larynx (c32)	1970-2019	APC model
Myeloid leukaemia (C92)	1970-2019	APC model for men, AC model for women.
Lip, oral cavity and pharynx (C00-C14, C30-C31)	1970-2019	APC model
Liver (C22)	1970-2019	AC model for men, APC model for women
Lung (C33-C34)	Cancer mortality data 1955-2019; data on smoking patterns 1945-2019	GLM: AC model with cigarette smoking exposure as a covariate
Oesophagus (C15)	1970-2019	APC model
Ovary (C56-C57)	1970-2019	APC model
Pancreas (C25)	1970-2019	APC model
Stomach (C16)	1970-2019	APC model
Uterus (C54-C55)	1970-2019	APC model
Cancer types with >30% of cases caused by smoking	Cancer mortality data 1955-2019; data on smoking patterns 1945-2019	GLM: AC model with cigarette smoking exposure as a covariate
Cancer types with 8-30% of cases caused by smoking	Cancer mortality data 1970-2019; data on smoking patterns 1945-2019	GLM: APC model with cigarette smoking exposure as a covariate
All smoking-related cancers combined		Number of deaths estimated by summing the number of deaths for cancer types with >30% of cases caused by smoking and cancer types with 8-30% of cases caused by smoking

GLM: Generalised linear model; APC: age-period-cohort; AC: age-cohort.

Figure S1. Model effects for mortality rates of the group of cancer types with >30% of cases currently caused by smoking

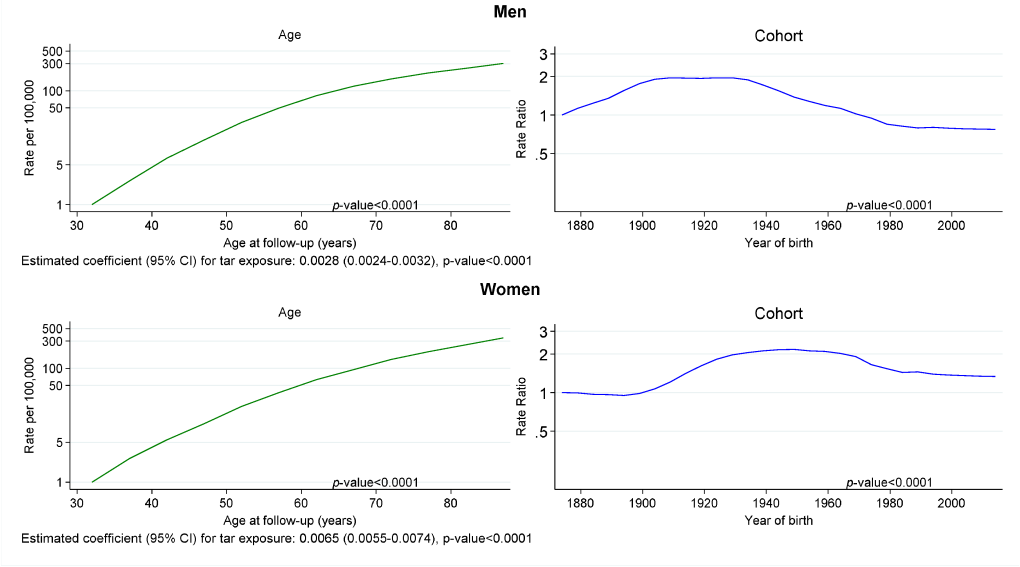


Figure S2. Model effects for mortality rates of the group of cancer types with 8-30% of cases currently caused by smoking

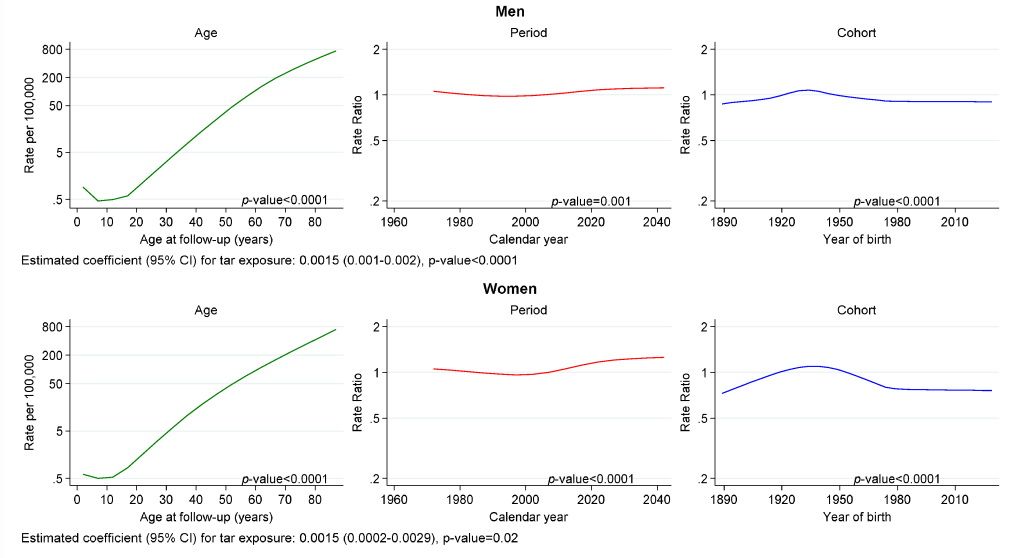
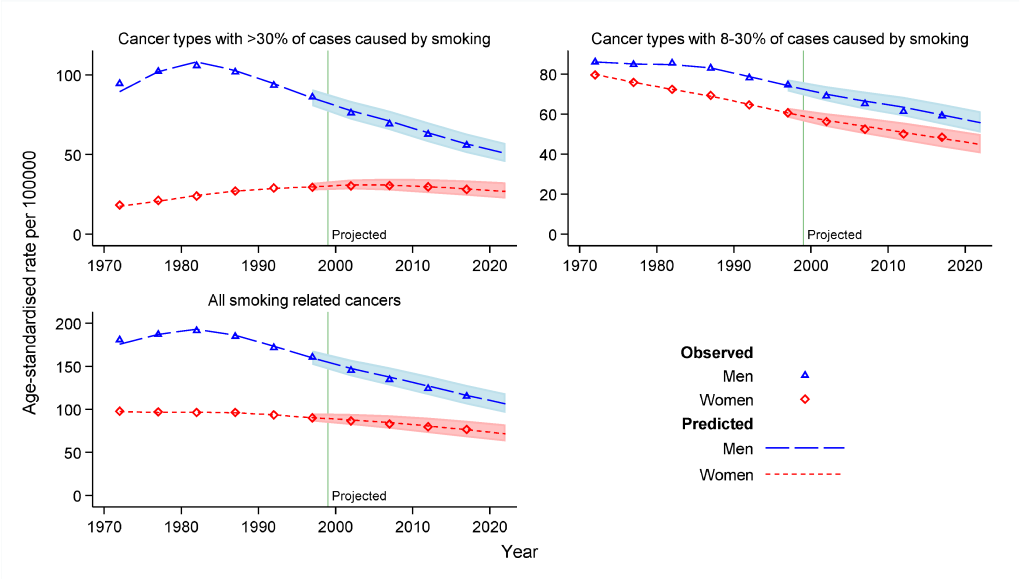


Figure S3. Validation of 20-year projections for mortality rates for smoking-related cancers in Australia using observed data from 1970 to 1999 projected to 2019, compared to observed data for 1970-2019



* Age-standardised using the 2001 Australian population. Shaded areas represent 95% confidence intervals.

Luo Q, *et al. J Epidemiol Community Health* 2022;0:1–8. doi: 10.1136/jech-2021-218252

References:

1. Brown CC, Kessler LG. Projections of lung cancer mortality in the United States: 1985-2025. *J Natl Cancer Inst* 1988;80(1):43-51.
2. Luo Q, Yu XQ, Wade S, et al. Lung cancer mortality in Australia: Projected outcomes to 2040. *Lung Cancer* 2018;125:68-76. doi: https://doi.org/10.1016/j.lungcan.2018.09.001
3. Shibuya K, Inoue M, Lopez AD. Statistical modeling and projections of lung cancer mortality in 4 industrialized countries. *IJC* 2005;117(3):476-85.
4. IARC Working Group on the Evaluation of Carcinogenic Risks to Humans. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans. Personal Habits and Indoor Combustions. Vol 100 – A Review of Human Carcinogens. Part E. Lyon (FRC): World Health Organisation, 2012.
5. Sasieni P. Age-period-cohort models in Stata. *The Stata Journal* 2012;12(1):15.
6. Sasieni P. Software Updates. *The Stata Journal* 2017;17(4):1.
7. Yu XQ, Luo Q, Hughes S, et al. Statistical projection methods for lung cancer incidence and mortality: a systematic review. *BMJ Open* 2019;9(8):e028497. doi: 10.1136/bmjopen-2018-028497 [published Online First: 2019/08/30]
8. Luo Q, O'Connell DL, Yu XQ, et al. Cancer incidence and mortality in Australia from 2020 to 2044 and an exploratory analysis of the potential effect of treatment delays during the COVID-19 pandemic: a statistical modelling study. *The Lancet Public Health* 2022;7(6):e537-e48. doi: https://doi.org/10.1016/S2468-2667(22)00090-1
9. Hall MT, Simms KT, Lew J-B, et al. The projected timeframe until cervical cancer elimination in Australia: a modelling study. *The Lancet Public Health* 2019;4(1):e19-e27. doi: 10.1016/S2468-2667(18)30183-X