BMJ Open

Statistical projection methods for lung cancer incidence and mortality: a systematic review

Xue Qin Yu,[1,2] Qingwei Luo,[1] Suzanne Hughes,[1] Stephen Wade,[1] Michael Caruana,[1] Karen Canfell,[1,2,3] Dianne L O'Connell[1,2,4]

To cite: Yu XQ, Luo Q, Hughes S, *et al*. Statistical projection methods for lung cancer incidence and mortality: a systematic review. *BMJ Open* 2019;**9**:e028497. doi:10.1136/bmjopen-2018-028497

► Prepublication history and additional material for this paper are available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2018-028497).

XQY and QL contributed equally.

Received 11 December 2018
Revised 05 June 2019
Accepted 30 July 2019

[1]Cancer Research Division, Cancer Council NSW, Sydney, New South Wales, Australia
[2]The University of Sydney School of Public Health, Faculty of Medicine and Health, The University of Sydney, Sydney, New South Wales, Australia
[3]Prince of Wales Clinical School, University of New South Wales, Sydney, New South Wales, Australia
[4]School of Medicine and Public Health, University of Newcastle, Newcastle, New South Wales, Australia

Correspondence to
Dr Qingwei Luo;
qingweil@nswcc.org.au

ABSTRACT

Objectives To identify and summarise all studies using statistical methods to project lung cancer incidence or mortality rates more than 5 years into the future.

Study type Systematic review.

Methods We performed a systematic literature search in multiple electronic databases to identify studies published from 1 January 1988 to 14 August 2018, which used statistical methods to project lung cancer incidence and/or mortality rates. Reference lists of relevant articles were checked for additional potentially relevant articles. We developed an organisational framework to classify methods into groups according to the type of data and the statistical models used. Included studies were critically appraised using prespecified criteria.

Results One hundred and one studies met the inclusion criteria; six studies used more than one statistical method. The number of studies reporting statistical projections for lung cancer increased substantially over time. Eighty-eight studies used projection methods, which did not incorporate data on smoking in the population, and 16 studies used a method which did incorporate data on smoking. Age–period–cohort models (44 studies) were the most commonly used methods, followed by other generalised linear models (35 studies). The majority of models were developed using observed rates for more than 10 years and used data that were considered to be good quality. A quarter of studies provided comparisons of fitted and observed rates. While validation by withholding the most recent observed data from the model and then comparing the projected and observed rates for the most recent period provides important information on the model's performance, only 12 studies reported doing this.

Conclusion This systematic review provides an up-to-date summary of the statistical methods used in published lung cancer incidence or mortality projections. The assessment of the strengths of existing methods will help researchers to better apply and develop statistical methods for projecting lung cancer rates. Some of the common methods described in this review can be applied to the projection of rates for other cancer types or other non-infectious diseases.

Strengths and limitations of this study

- ► This is the first systematic review summarising statistical methods used in projecting lung cancer incidence or mortality rates over the past three decades.
- ► The review was conducted according to the published guidelines.
- ► Using predefined assessment criteria and a standardised data extraction form resulted in a high level of agreement in the data extractions performed by two independent reviewers.
- ► The review provided theoretical and practical information, including a comprehensive summary of the methods and relevant software.
- ► Meta-analysis was not possible due to the wide variation in study populations and time periods used in the projections.

INTRODUCTION

Lung cancer has been the most commonly diagnosed cancer in the world for several decades and is the leading cause of cancer deaths worldwide, accounting for nearly 20% of all cancer deaths.[1] Reliable projections of future patterns of lung cancer incidence and mortality are, therefore, of importance for the planning of health service requirements and the management of healthcare resources.[2 3] Given the well-documented association between tobacco smoking and lung cancer risk,[4 5] projections of lung cancer incidence and mortality are also important for evaluating the effectiveness of existing tobacco control programme and the forward projection of the potential impact of new evidence-based tobacco control strategies.[2 6 7] There have been a variety of statistical methods developed and reported in the literature for projecting cancer incidence or mortality rates.[2] These methods range from assuming the current rate remains unchanged into the future, to a more complex class of statistical models of past trends such as age–period–cohort (APC) models, which may involve a range of assumptions, software and techniques.

Projecting future cancer incidence and mortality trends is always a complex exercise due to the changing risk factor profiles over

time, and the long latency period between risk factor exposure and development of some cancers.[8] For lung cancer in particular, projections can be inaccurate if any changes in past smoking behaviour are not accurately taken into account.[2 3] Unfortunately, data on smoking behaviour are not always available with the requisite level of detail (eg, sex-age-specific data), so choosing and implementing an appropriate projection method largely depends on data availability and the purpose for the projections.[8] Given the complexity involved in such projections, information on the available statistical methods, utilisation of these methods and further developments in this area are of particular interest to researchers working in this field. However, while some of these methods have been reviewed and evaluated,[8–11] to our knowledge, there are currently no published systematic reviews of all statistical methods available for projecting lung cancer incidence or mortality rates.

Therefore, we carried out a methodological systematic review to identify and summarise published population-based studies that used statistical methods to project lung cancer incidence or mortality rates over the long term (eg, more than 5 years). The aim was to provide up-to-date and comprehensive information on the statistical methods that are available for projecting lung cancer rates. In doing so, our intention was to provide readers with an understanding of these various statistical methods, the availability of statistical software to implement these methods, and the utilisation of these methods in different circumstances, and to highlight the differences and similarities between methods.

METHODS

This systematic review adhered to the checklist presented in the Preferred Reporting Items for Systematic Reviews and Meta-Analyses.[12] A protocol was developed for this review and is included as online supplementary resource 1.

Patient and public involvement

As this was a systematic review of statistical methods used to obtain lung cancer rate projections, no patients or public were involved.

Literature search

In August 2016, Embase, Medline and PreMEDLINE databases were searched using text terms and, where available, database-specific subject headings, for studies published since 1988, which used statistical methods to project lung cancer incidence and/or mortality. Searches for lung cancer-related terms were combined with searches for terms related to projection, forecasting and statistical models. Reference lists of relevant articles were checked for additional potentially relevant articles. In August 2018, Embase and Medline, including Epub Ahead of Print, In-Process and other Non-Indexed Citations databases, were searched for studies published from 2016 onwards using an updated search strategy, which aimed to capture all newly published articles. A complete list of the terms used is included in online supplementary resource 2.

Selection criteria

Full inclusion and exclusion criteria are listed in table 1. Studies were included if they used a statistical method to project lung cancer incidence and/or mortality over a period greater than 5 years using population-based data and were published in English from 1 January 1988 to 14 August 2018. ‘Statistical method’ was defined as a method that analyses the observed data using traditional regression, correlation or other statistical summaries.

Table 1 Inclusion and exclusion criteria employed

Domain	Inclusion criteria	Exclusion criteria
Study type	Population-based original research studies	Any of: Editorial comment, literature review, case studies, clinical trials, case–control studies.
Study population	General population in any country	Restricted to selected groups, that is, selected patients with cancer or high-risk populations.
Outcomes	Reports projections of lung cancer incidence and/or mortality rates	No relevant outcomes are reported, that is, no lung cancer-specific outcomes.
Statistical method	Uses a statistical method for the projection, including studies, which used simulation methods to estimate confidence intervals, that is, Bayesian technique	Uses mathematical models, which generate outcomes based on a proposed theoretical model of the disease’s natural history.
No of years projected	Reports long-term projections, that is, greater than 5 years	Does not report projections of lung cancer rates, that is, only explains past trends, or reports projections less than or equal to 5 years.
Publication type	Full-text published	Conference proceedings, abstracts, posters.
Time of publication	Published from 1 January 1988 to 14 August 2018	Published before 1988.
Language	English	Language other than English.

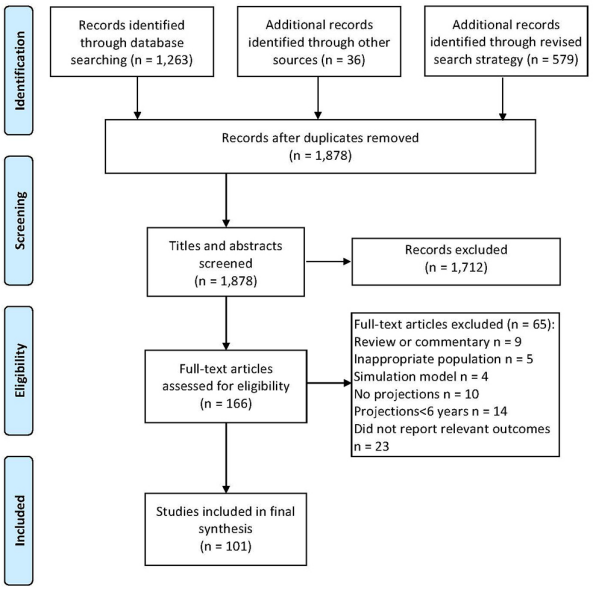

Figure 1 PRISMA flow chart of study selection process. PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analyses.

'Projection' was defined as the use of data including the whole or part of the observed data to forecast lung cancer incidence or mortality rates beyond the time period covered by the data included in the statistical models. Mathematical models, which generate outcomes based on a proposed theoretical model of the disease's natural history, were not included in this review.

Application of selection criteria

The literature search and the review followed the stages described in figure 1. After removing duplicates, 1878 studies were retained for screening. One author (SH) screened the titles and abstracts against the inclusion criteria to exclude articles that were clearly irrelevant. The main reason for exclusion of papers at the screening stage was that the studies did not report on lung cancer incidence or mortality. Others were excluded because they used mathematical methods rather than being population-based studies. Further studies were excluded because they were an editorial commentary or literature review. After the screening process, a total of 166 studies were eligible for full-text review.

Full-text articles were independently reviewed and assessed for inclusion by two authors (XQY and QL) and a total of 101 studies were retained for final inclusion (92% agreement). Disagreements were discussed and if an agreement could not be reached the study was assessed for inclusion by a third reviewer (DLO). Excluded studies and the reasons for exclusion are listed in online supplementary resource 3. The main reasons for exclusion of studies at this stage were that they did not report lung cancer rates separately or the projections were for fewer than 6 years.

Critical appraisal

As the purpose of this methodological review was to provide an overview of statistical methods, and the projections of lung cancer rates were conducted in different populations and over different time periods, no meta-analysis was possible and specific projection results were not compared or analysed in this review. Therefore, the risk of bias evaluation of the included studies was not applicable.

The methodological quality of the studies was independently assessed by two reviewers using prespecified criteria (table 2): quality of the data source, length of period covered by the observed data, availability of software information, model fitting and validation. Validation provides information on the performance and reliability of the projection model and can be undertaken by withholding the most recent observed data from the model fitting and then comparing the projected rates for those years with the actual observed values.[7] As the use of scales for assessing study quality is discouraged in Cochrane reviews[13] and meta-analyses,[14] as the calculation of an overall score inevitably involves assigning (often arbitrary) weights to the quality criteria being assessed. It is difficult to justify the weights used and it has been shown that the overall quality score is not a reliable assessment of the study's validity.[13] Moreover, each method included in this review has its own merits and limitations, and depending on specific circumstances may be more or less reliable or relevant. Therefore, an overall score for the methodological quality of each study was not provided.

Data extraction

For each included study, two reviewers (XQY and QL) independently extracted details of the study including data sources, study population, year of publication, observed data period for the projections, statistical methods and software used, and whether the method incorporated information about smoking patterns, which is the main risk factor for lung cancer. The extracted data were collected using a standardised form (see online supplementary resource 4), which was pilot tested using 10 studies. Any differences between the two reviewers were discussed and when agreement could not be reached the studies were assessed by a third reviewer (DLO). The overall agreement between the two reviewers was 91.6%.

The selection of an appropriate statistical method for projecting cancer rates is largely restricted by the quality and availability of cancer data, which is generally better in more developed countries.[15] The Human Development Index (HDI), developed by the United Nations,[16] is a summary measure of life expectancy, education and gross domestic product per head of population. We, therefore, recorded HDI ranking for each of the study populations, so that we could describe the distribution of projections methods used according to the country's level of development.

Table 2 Prespecified criteria for assessing studies included in this review

Criterion	Yes	No or not clear
Strengths		
≥10 years observed data	Observed data period reported ≥10 years.	Observed data period reported <10 years, or there is insufficient information to make an assessment.
Good quality data source	Data source reported, and the majority of observed data used are included in IARC Cancer in Five Continents, or with high population coverage as stated in WHO database.	Data source reported but the majority of observed data used are not included in IARC Cancer in Five Continents, or with low population coverage as stated in WHO database, or there is insufficient information to make an assessment.
Provided fitted values of observed data	Reports both model estimates and observed data for the period used for model fitting.	Does not report both model estimates and observed data for the period used for model fitting.
Validated projections using observed data	The model was validated by excluding data for the most recent years from the model fitting, and then comparing the projected rates for those years with the observed data. Provides both model projections beyond the period included in model fitting and a comparison with the observed data for the same period.	Does not provide validation using observed data.
Advantage		
Provided software information	Software information was described or referenced.	Software information not provided.

IARC, International Agency for Research on Cancer; WHO, World Health Organization.

Classification of statistical methods

In order to summarise the differences and similarities between the methods reported, we developed an organisational framework to classify methods into groups according to both the type of observed data used and the statistical models reported (figure 2). As tobacco exposure is well known to be the most significant risk factor for lung cancer[4] and can be used as an important predictor for lung cancer incidence and mortality, we first divided the studies into two large categories according to whether or not they included data on smoking in the projection method. For each category, we then subdivided studies into groups according to the projection method used. Methods not incorporating data on smoking in the population were grouped as either: (1) APC models, a special form of generalised linear model (GLM), which includes age, period and cohort components, (2) other GLMs, where the number of cases (deaths) or the logarithm of this was modelled as a linear or non-linear function of the explanatory factors using the logarithm of the population size as an offset, with Poisson or negative binomial distribution and (3) present state method (eg, assumes that the age-specific rates in the future will be the same as the most recent observed rates, or assumes a constant annual rate of change as observed in a selected time period). Methods incorporating data on smoking were grouped into: (1) GLMs with a smoking variable as one of the covariates, (2) APC models that included an effect for smoking, (3) projections adjusted for the smoking attributable fraction (SAF) and (4) other methods (including all methods that do not use detailed historical cancer data or do not include detailed data on smoking). More detailed descriptions of each of these methods are provided in online supplementary resource 5.

RESULTS

A total of 101 eligible studies were included (table 3). All these studies are ecological studies that used single year or 5-year aggregated population incidence or mortality data, or are based on cancer rates reported in the literature. Table 4 shows the study characteristics grouped according to the method used for the projections. Eighty-eight studies used projection methods not incorporating data on smoking,[1 2 9 17–101] 16 studies used a method incorporating data on smoking,[3 7 33 41 42 102–112] and 6 studies used multiple methods.[18 33 36 41 42 62] Overall, APC models were the most commonly used method to project lung cancer rates (44 studies used this method),[2 9 17–58] and other GLMs were the next most commonly used (35 studies).[18 36 59–89 100 101] Only 12 studies used the present state method by assuming that the average cancer rates in the most recent years will remain constant into the future.[1 62 90–99] Of the 16 studies incorporating data on smoking, eight studies directly used GLMs with a variable reflecting detailed historical smoking-related behaviour as one of the covariates included.[3 7 33 103 106 108 111 112] These variables included number of cigarettes consumed and

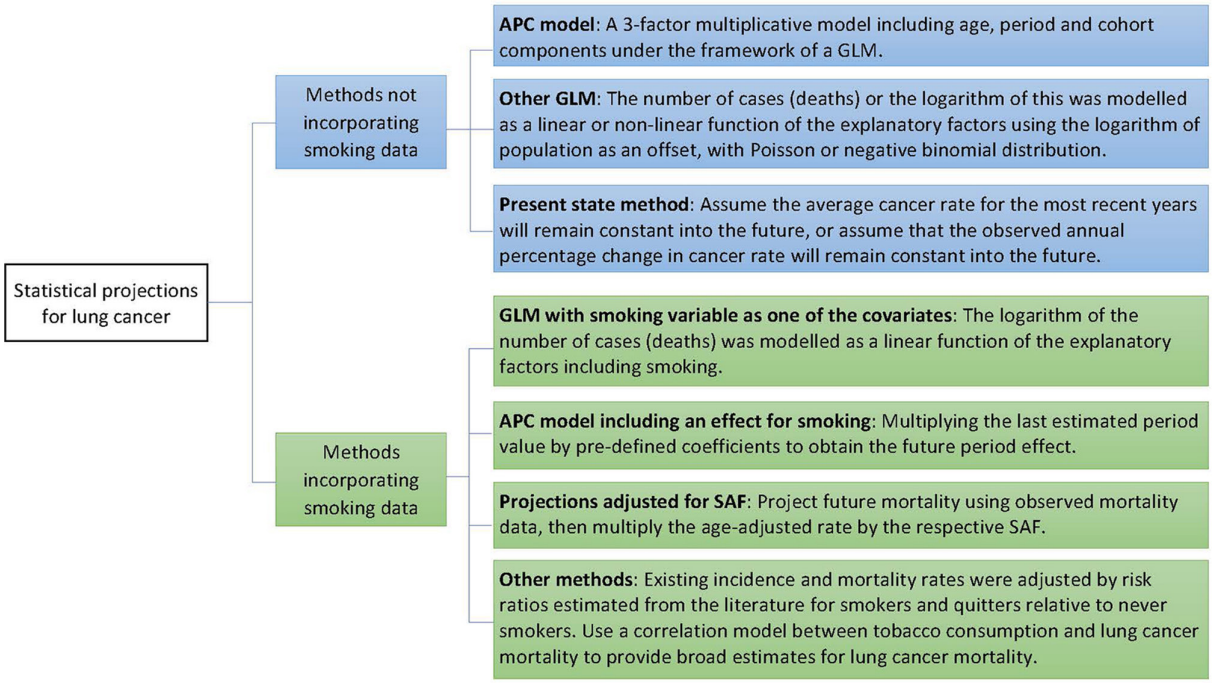

Figure 2 Organisational framework to categorise methods for lung cancer mortality projections. APC, age–period–cohort; GLM, generalised linear model; SAF, smoking attributable fraction.

average tar content,[3 7 33] smoking prevalence,[111] number of years of smoking[106 112] and smoking intensity.[103 108] Two studies used APC models and predefined coefficients based on recent trends in smoking prevalence and tar content to adjust the estimates for the period parameter.[41 42] Two studies made projections adjusted for the SAF, which required limited data on smoking behaviour,[102 107] and the remaining four studies used other methods, which required limited data on both cancer rates and smoking behaviour.[104 105 109 110]

The majority of models were developed using more than 10 years of observed data that was considered to be good quality, that is, incidence data included in the Cancer Incidence in Five Continents series,[15] or mortality data from a source considered by WHO to have a high population coverage.[113] Most studies provided projections for 10 years or more, and the proportion of studies providing projections for more than 19 years was higher for studies using methods incorporating data on smoking (50.0%) than for studies using methods which did not incorporate smoking patterns (18.2%). Only 25.7% of the studies provided comparisons of fitted and observed rates and 11.9% of the studies reported validation of the projection model using observed data.

The numbers of studies by publication period and by the country's HDI rank are presented in figure 3. The number of publications increased substantially over time, especially the number of studies using APC models, which more than tripled in the most recent period (2008–2018) compared with 1998–2007. The majority of the articles included in this systematic review used data from countries with very high or high HDI including studies from the USA, Europe and Australia, 16 studies used data from countries in medium or low HDI groups including studies from China and India, and 22 studies used data from multiple countries.

The statistical software packages used by method and year of publication are shown in figure 4. Among the studies using APC models, the most commonly used software package was Nordpred (R package developed by Harald Fekjær and Bjørn Møller, Cancer Registry of Norway)[10 38] and most of these studies were published in recent years. GLIM (Oxford, UK)[114] was the second most commonly used software for APC modelling, but it was mainly used in the earlier years, with the latest study published in 2000.[45] Special software WinBUGS (Cambridge, UK),[115] INLA (R package developed by Rue and Martino, Department of Mathematical Sciences NTNU, Norway)[116] or BAMP (Institute of Biomedical Engineering, Imperial College, London, UK)[117] were used for studies employing Bayesian methods.[2 20 22 25 26 31–33 48 85] Among studies using other GLMs, Joinpoint (National Cancer Institute, USA)[118] and Stata[119] were the two most commonly used software packages. Most studies using the present state method did not mention which software was used. Nordpred, Stata, Joinpoint, SAS,[120] other R packages and WinBUGS were the software program most commonly used in the recent time period. An overview of these software packages is provided in table 5. Each of these packages has different features and some are freely available to researchers.

Table 3 Summary of included studies

First author and year	Lung cancer outcome(s)	Country	Observed data	No of years projected	Incorporated smoking data	Model	Software	Good data quality*	Provides fitted values†	Validation‡
AIHW 2012[59]	Incidence	Australia	1982–2007	10–19 years	No	Joinpoint analysis/GLM	Joinpoint	Yes	Yes	No
Alonso 2018[17]	Incidence	Uruguay	1990–2014	20+years	No	APC model	Stata, R, Nordpred	No	No	No
Arslanhan 2012[110]	Incidence/ mortality	Turkey	2002	20+years	Yes, smoking prevalence and smoking status	Relative risk	Not provided	No	No	No
Baade 2012[90]	Incidence	Australia	1982–2007	10–19 years	No	Assume same rate	Not provided	Yes	No	No
Bashir 2001[2]	Incidence/ ortality	Finland	1955–1974	20+years	No	Bayesian APC model	WinBUGs	Yes	Yes	Yes
Bosetti 2012[18]	Mortality	32 European countries	1970–2009	<10 years	No	Bayesian APC model / Joinpoint analysis	Joinpoint, GLIM	Yes	No	No
Bray 2012[60]	Incidence/ mortality	184 countries	1988–2002	20+ years	No	GLM/annual percentage change	Not provided	Yes	No	No
Brenner 1992[61]	Incidence	Germany	1968–1987	10–19 years	No	GLM	GLIM	Yes	No	No
Brown 1988[3]	Mortality	USA	1958–1982	20+ years	Yes, smoking prevalence, consumption and tar content	GLM with smoking as a covariate	Not provided	Yes	Yes	No
Byers 2006[62]	Mortality	USA	1990–2002	10–19 years	No	GLM/assume same rate	Not provided	Yes	No	No
Cancer Institute 2016[63]	Incidence/ mortality	Australia	1994–2008	10–19 years	No	GLM	SAS	Yes	No	No
Cancer Projections Network 2010[9]	Incidence/ mortality	Canada	1975–1994	10–19 years	No	Bayesian APC model/GAM	Nordpred, WinBUGs, GLIM	Yes	Yes	Yes
Carson 1993[64]	Mortality	USA	1979–1989	10–19 years	No	GLM/assume same rate	BMDP	Yes	No	No
Castro 2016[65]	Incidence	Portugal	1994–2009	10–19 years	No	Joinpoint analysis and GLM	Stata, Joinpoint	Yes	No	No
Cayuela 2011[19]	Mortality	Spain	1979–2008	20+ years	No	APC model	Nordpred	Yes	No	No
Chen 2011[20]	Incidence	China	1998–2007	10–19 years	No	Bayesian APC model	BAMP	No	No	No
Clements 2005[21]	Mortality	5 countries	1950–2001	10–19 years	No	Bayesian APC model	R, WinBUGs	Yes	Yes	Yes
Clèries 2016[22]	Mortality	Spain	1998–2012	10–19 years	No	Bayesian APC model	INLA	Yes	No	No
Clèries 2018[23]	Incidence/ mortality	Spain	1994–2013	10–19 years	No	Bayesian APC model	Not provided	Yes	No	No
Coupland 2010[24]	Incidence	UK	1985–2003	20+ years	No	APC model	Nordpred	Yes	No	No
Davis 2013[102]	Incidence/ mortality	USA	1990–2007	10–19 years	Yes, smoking prevalence	Annual percentage change and SAF	SAS	Yes	No	No
Didkowska 2009[66]	Mortality	Poland	1998–2006	10–19 years	No	GLM	Stata	No	No	No
D'Souza 2013[91]	Mortality	India	2001–2004	20+ years	No	Assume same rate	Not provided	No	No	No
D'Souza 2013b[92]	Incidence	India	2001–2004	20+ years	No	Assume same rate	Not provided	No	No	No
Dušek 2015[67]	Incidence	Czech Republic	1978–2011	<10 years	No	GLM	S	Yes	No	No
Dyba 1997[68]	Incidence	Sweden	1960–1984	20+ years	No	GLM	GLIM	Yes	No	No
Dyba 2000[69]	Incidence	Finland	1954–1978	10–19 years	No	GLM	GLIM	Yes	No	Yes

Continued

Table 3 Continued

First author and year	Lung cancer outcome(s)	Country	Observed data	No of years projected	Incorporated smoking data	Model	Software	Good data quality*	Provides fitted values†	Validation‡
Eilstein 2008[25]	Mortality	France	1978–2002	10–19 years	No	Bayesian APC model	WinBUGs	Yes	No	No
Eilstein 2012[26]	Mortality	France	1977–2006	10–19 years	No	Bayesian APC model/GAM	R, WinBUGs	Yes	Yes	No
Engeland 1995[70]	Mortality	Nordic countries	1958–1987	20+ years	No	GLM	Not provided	Yes	No	No
Ferlay 2010[27]	Incidence/ mortality	European countries	1978–2002	<10 years	No	APC model	Nordpred	Yes	No	No
Ferlay 2013[28]	Incidence/ mortality	European countries	1978–2006	<10 years	No	APC model	Nordpred	No	No	No
Ferlay 2013[1]	Incidence/ mortality	Worldwide	1989–2011	20+ years	No	Assume same rate	Not provided	Yes	No	No
French 2006[71]	Mortality	UK	1984–2004	10–19 years	No	Joinpoint analysis/GLM	Stata, Joinpoint	Yes	No	No
Fukuda 2002[72]	Mortality	Japan	1988–1997	10–19 years	No	GLM	Not provided	No	No	No
Galasso 2013[29]	Incidence/ mortality	Italy	1970–2002	10–19 years	No	APC model	MIAMOD	Yes	No	No
Godlewski 2012[73]	Incidence	Poland	1999–2008	10–19 years	No	GLM	Stata	No	No	No
Hakulinen 1994[74]	Incidence	Sweden	1960–1984	20+ years	No	GLM	GLIM	Yes	No	No
Heinävaara 2006[75]	Incidence/ mortality	Finland	1987–1997	10–19 years	No	GLM	Not provided	Yes	Yes	Yes
Hristova 1997[30]	Incidence	Bulgaria	1968–1992	20+ years	No	APC model	GLIM	No	No	No
Jee 1998[76]	Mortality	Korea (South)	1980–1994	10–19 years	No	GLM	Not provided	Yes	No	No
Jürgens 2014[31]	Mortality	Switzerland	1974–2008	10–19 years	No	Bayesian APC model	R, WinBUGs	Yes	Yes	Yes
Kaneko 2003[32]	Mortality	Japan	1952–2001	20+ years	No	Bayesian APC model	WinBUGs	Yes	Yes	No
Knorr-Held 2001[33]	Mortality	Germany	1952–1996	10–19 years	Yes, smoking prevalence and consumption	Bayesian APC model and GLM with smoking as a covariate	BAMP	Yes	No	No
Kubík 1998[34]	Mortality	4 European countries	1960–1989	20+ years	No	APC model	GLIM	Yes	No	No
Kuroishi 1992[77]	Mortality	Japan	1969–1989	20+ years	No	GLM	Not provided	Yes	No	No
Li 2017[35]	Mortality	China	1974–2014	10–19 years	No	APC model	Nordpred	No	No	Yes
Malvezzi 2013[36]	Mortality	33 European countries	1970–2009	<10 years	No	Joinpoint analysis/Bayesian APC model	R, Joinpoint, GLIM	Yes	No	No
Malvezzi 2015[78]	Mortality	28 European countries	1970–2009	<10 years	No	Joinpoint analysis/GLM	R, Joinpoint	Yes	No	No
Malvezzi 2018[100]	Mortality	6 countries	1970–2012	<10 years	No	Joinpoint analysis/GLM	Joinpoint	Yes	No	No
Martín-Sánchez 2016[79]	Mortality	Spain	2007–2013	<10 years	No	GLM	R, WinBUGs	Yes	No	No
Martín-Sánchez 2017[111]	Mortality	Spain	1980–2013	<10 years	Yes, smoking prevalence	GLM	Not provided	Yes	No	No
Martín-Sánchez 2018[101]	Mortality	52 countries	2008–2014	10–19 years	No	GLM	WinBUGs	No	No	No
Mistry 2011[37]	Incidence	UK	1975–2007	20+ years	No	APC model	Stata, Nordpred	Yes	Yes	No

Continued

Table 3 Continued

First author and year	Lung cancer outcome(s)	Country	Observed data	No of years projected	Incorporated smoking data	Model	Software	Good data quality*	Provides fitted values†	Validation‡
Møller 2002[38]	Incidence	Nordic countries	1958–1997	20+ years	No	APC model	Nordpred	Yes	No	No
Møller 2005[39]	Incidence	Nordic countries	1958–1987	10–19 years	No	APC model	R	Yes	No	Yes
Møller 2007[40]	Incidence	UK	1974–2003	20+ years	No	APC model	Nordpred	Yes	No	No
Murray 1997[103]	Mortality	47 countries	1950–1990	20+ years	Yes, smoking intensity	GLM	Not provided	Yes	No	No
Negri 1990[41]	Mortality	Italy	1955–1984	10–19 years	Yes, smoking prevalence	APC model involve smoking data	GLIM	Yes	Yes	No
Negri 1990[42]	Mortality	Switzerland	1950–1984	10–19 years	Yes, smoking prevalence	APC model involve smoking data	GLIM	Yes	Yes	No
Ng 2009[104]	Mortality	Indonesia, Vietnam, Ethiopia	2005–2006	10–19 years	Yes, smoking prevalence	GLM	SAS, Stata	No	No	No
Nowatzki 2011[43]	Incidence	Canada	1976–2005	20+ years	No	APC model	Nordpred	Yes	No	No
Oberaigner 2014[80]	Incidence	Austria	1990–2009	10–19 years	No	GLM	Stata	Yes	Yes	No
Olajide 2015[81]	Incidence	UK	2002–2011	<10 years	No	GLM	SAS, Stata	Yes	Yes	No
O'Lorcain 2004[82]	Mortality	Ireland	1954–2000	10–19 years	No	GLM	Stata	Yes	No	No
Olsen 2008[44]	Mortality	UK	1971–2005	20+ years	No	APC model	Nordpred	Yes	No	No
Parsons 2000[83]	Incidence	UK	1981–1995	20+ years	No	GLM	S-PLUS	Yes	Yes	No
Pearce 2016[93]	Mortality	Ireland	2007–2011	10–19 years	No	Assume same rate	SAS	Yes	No	No
Pierce 1992[105]	Mortality	8 countries	1975–1986	10–19 years	Yes, tobacco consumption	The simple tobacco consumption model	Not provided	No	No	No
Pisani 1993[94]	Mortality	24 geographical global areas	1985–1985	10–19 years	No	Assume same rate	Not provided	No	No	No
Pompe-Kirn 2000[45]	Incidence	Slovenia	1965–1994	10–19 years	No	APC model	GLIM	Yes	No	No
Preston 2014[106]	Mortality	USA	1940–2009	20+ years	Yes, smoking prevalence	GLM	Not provided	Yes	No	No
Quante 2016[95]	Incidence/ mortality	Germany	1998–2012	10–19 years	No	Joinpoint analysis/annual percentage change	SAS, Joinpoint	Yes	No	No
Rahib 2014[96]	Incidence/ mortality	USA	2006–2010	20+ years	No	Annual percentage change	Joinpoint	Yes	No	No
Rapiti 2014[46]	Incidence	Switzerland	1985–2009	10–19 years	No	APC model	Nordpred	Yes	No	No
Reissigova 1994[47]	Mortality	Czech Republic	1960–1989	10–19 years	No	APC model	GLIM	No	Yes	No
Ribes 2014[48]	Incidence/ mortality	Spain	1993–2007	10–19 years	No	Bayesian APC model	INLA	Yes	No	No
Riebler 2017[49]	Mortality	5 countries	1950–2011	10–19 years	No	Bayesian APC model	R, WinBUGs, INLA	Yes	Yes	Yes
Rutherford 2012[50]	Incidence	Finland	1957–1987	20+ years	No	APC model	Stata	Yes	Yes	Yes

Continued

Table 3 Continued

First author and year	Lung cancer outcome(s)	Country	Observed data	No of years projected	Incorporated smoking data	Model	Software	Good data quality*	Provides fitted values†	Validation‡
Sánchez 2010[51]	Incidence/ mortality	Spain	1981–2006	<10 years	No	APC model	MIAMOD	Yes	Yes	No
Shamseddine 2014[84]	Incidence	Lebanon	2003–2008	10–19 years	No	Joinpoint analysis/GLM	Joinpoint	No	No	No
Sharp 1996[52]	Incidence/ mortality	UK	1968–1992	<10 years	No	APC model	GLIM	Yes	No	No
Shibuya 2005[7]	Mortality	Four countries	1950–2000	20+ years	Yes, tobacco consumption and tar content	GLM with smoking as a covariate	Not provided	Yes	Yes	Yes
Smith 2009[97]	Incidence	USA	2003–2005	20+years	No	Assume same rate	SAS	Yes	No	No
Smittenaar 2016[53]	Incidence/ mortality	UK	1979–2014	20+years	No	APC model	Stata	Yes	Yes	No
Son 2016[54]	Mortality	Korea (South)	1983–2012	20+years	No	APC model	Nordpred	Yes	No	No
Stoeldraijer 2015[107]	Mortality	4 European countries	1950–2009	20+years	Yes, smoking prevalence	APC model / SAF	R	Yes	No	No
Stracci 2013[55]	Incidence/ mortality	Italy	1970–2002	10–19 years	No	APC model	MIAMOD	Yes	No	No
Strong 2008[108]	Mortality	107 countries	1950–2002	20+years	Yes, smoking intensity	GLM	Not provided	No	No	No
Swaminathan 2011[56]	Incidence	India	1982–2006	10–19 years	No	APC model	Nordpred	Yes	No	No
Torres-Avilés 2015[85]	Mortality	Chile	1990–2009	<10 years	No	GLM	WinBUGs	Yes	Yes	Yes
Tsoi 2017[86]	Incidence	China	1993–2007	20+years	No	GLM	R	Yes	Yes	No
Vardanjani 2017[98]	Incidence	Iran	2003–2009	<10 years	No	Joinpoint analysis/annual percentage change	Joinpoint	No	No	No
Virani 2017[57]	Incidence	Thailand	1989–2012	10–19 years	No	Joinpoint analysis/APC model	R, Joinpoint, Nordpred	No	Yes	No
Vogt 2017[112]	Mortality	German	1956–2013	20+years	Yes, years smoked	GLM	Not provided	Yes	No	No
Weir 2015[58]	Incidence	USA	1975–2009	10–19 years	No	APC model	Nordpred	Yes	No	No
Wiklund 1992[87]	Mortality	Sweden	1975–1984	20+years	No	GLM	CAN*TROL	Yes	No	No
Winkler 2015[109]	Mortality	South Africa	2010	10–19 years	Yes, smoking prevalence	GLM/relative risk for smokers	Not provided	No	No	No
Yabroff 2008[99]	Mortality	USA	1999–2003	10–19 years	No	Assume same rate	Not provided	Yes	No	No
Yang 2004[89]	Mortality	China	1990–1999	<10 years	No	GLM	Not provided	No	Yes	No
Yang 2005[88]	Incidence	China	1993–1997	<10 years	No	GLM	GLIM	No	No	No

*The majority of observed data used are included in the Cancer Incidence in Five Continents series published by the International Agency for Research on Cancer, or have high population coverage as stated in WHO mortality database.
†Provides fitted values of observed data to allow appraisal of the model fit to the observed data.
‡Validation using observed data: Paper compared the projected values with the observed data beyond the period included in model fitting. The model was validated by excluding data for the most recent years from the model fitting, and then compared the projected rates for those years with the observed data.
APC, age–period–cohort; GLM, generalised linear model; SAF, smoking attributable fraction.

Table 4 Summary of study characteristics grouped according to projection method used

Method	Total studies*	Incidence	Mortality	≥10 years observed data	Good data quality†	No of years projected			Provide fitted values‡	Validation§
						6–9	10–20	>20		
Methods without smoking factor, (%)	**88 (87.1)**	**50 (56.8)**	**55 (62.5)**	**75 (85.2)**	**71 (80.7)**	**15 (17.0)**	**57 (64.8)**	**16 (18.2)**	**23 (26.1)**	**11 (12.5)**
APC models, (%)[2 9 17–58]	44 (**43.6**)	26 (59.1)	29 (65.9)	44 (100.0)	37 (84.1)	6 (13.6)	31 (70.5)	7 (15.9)	15 (34.1)	8 (18.2)
Other GLMs, (%)[18 36 59–89 100 101]	35 (**34.7**)	17 (48.6)	21 (60.0)	30 (85.7)	29 (82.9)	10 (28.6)	20 (57.1)	5 (14.3)	8 (22.9)	3 (8.6)
Present state methods, (%)[1 62 90–99]	12 (**11.9**)	7 (58.3)	8 (66.7)	4 (33.3)	8 (66.7)	1 (8.3)	7 (58.3)	4 (33.3)	0 (0.0)	0 (0.0)
Methods incorporating smoking data, (%)	**16 (15.8)**	**2 (12.5)**	**16 (100.0)**	**13 (81.3)**	**11 (68.8)**	**1 (6.3)**	**7 (43.8)**	**8 (50.0)**	**5 (31.3)**	**1 (6.3)**
GLM with a smoking variable as one of the covariates, (%)[3 7 33 103 106 108 111 112]	8 (**7.9**)	0 (0.0)	8 (100.0)	8 (100.0)	7 (87.5)	1 (12.5)	1 (12.5)	6 (75.0)	3 (37.5)	1 (12.5)
APC model including an effect for smoking, (%)[41 42]	2 (**2.0**)	0 (0.0)	2 (100.0)	2 (100.0)	2 (100.0)	0 (0.0)	2 (100.0)	0 (0.0)	2 (100.0)	0 (0.0)
Projections adjusted for the SAF, (%)[102 107]	2 (**2.0**)	1 (50.0)	2 (100.0)	2 (100.0)	2 (100.0)	0 (0.0)	1 (50.0)	1 (50.0)	0 (0.0)	0 (0.0)
Other methods, (%)[104 105 109 110]	4 (**4.0**)	1 (25.0)	4 (100.0)	1 (25.0)	0 (0.0)	0 (0.0)	3 (75.0)	1 (25.0)	0 (0.0)	0 (0.0)
Total, (%)	**101 (100.0)**	**52 (51.5)**	**68 (67.3)**	**85 (84.2)**	**79 (78.2)**	**16 (15.8)**	**61 (60.4)**	**24 (23.8)**	**26 (25.7)**	**12 (11.9)**

*Numbers of studies are not mutually exclusive, with six studies using more than one method.
†The majority of observed data used are included in the Cancer Incidence in Five Continents series published by the International Agency for Research on Cancer, or have high population coverage as stated in WHO mortality database.
‡Provides fitted values of observed data to allow appraisal of the model fit to the observed data.
§Validation using observed data: Paper compared the projected values with the observed data beyond the period included in model fitting. The model was validated by excluding data for the most recent years from the model fitting, and then compared the projected rates for those years with the observed data.
APC, age–period–cohort; GLM, generalised linear model; SAF, smoking attributable fraction.

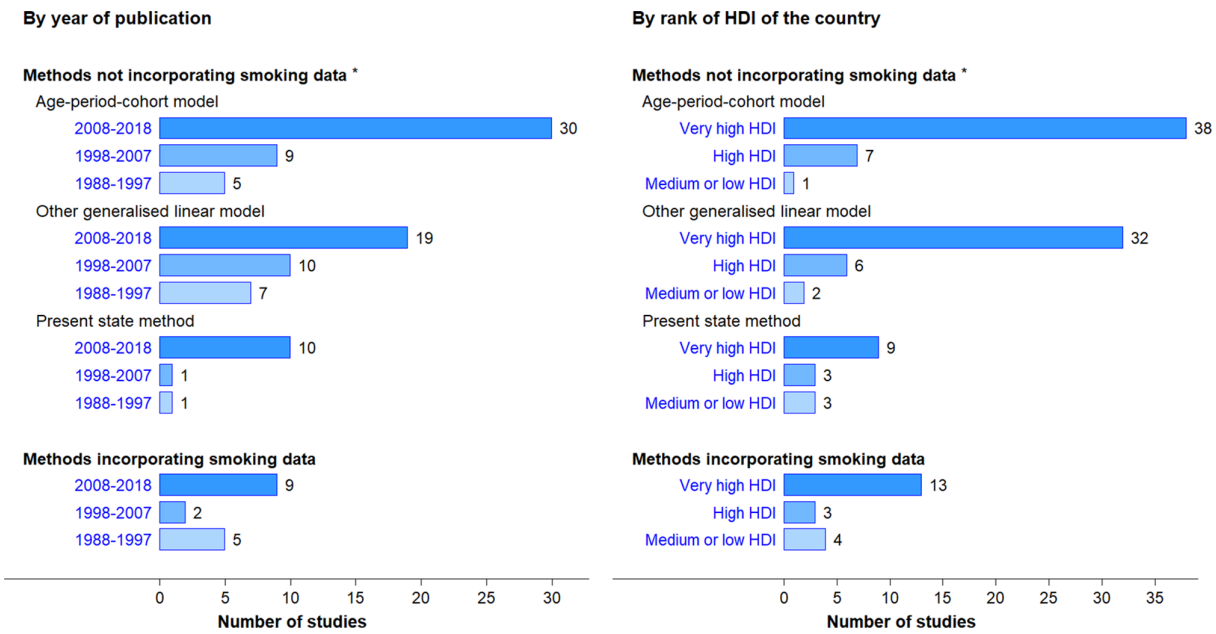

Figure 3 Studies included by year of publication, 1988–2018 and level of human development of the country providing the data, stratified by method. *Six studies used more than one method, and 22 studies used data from multiple countries. HDI, Human development index.

DISCUSSION

This review highlights the scope and diversity of the statistical methods used to project lung cancer rates for the longer term, and provides a summary of the main methods used in studies conducted over the last three decades. These methods range from using a basic assumption that the current rate will remain unchanged into the future, to more complex statistical models involving a range of different assumptions, statistical techniques and software packages. We found that both lung cancer incidence and mortality projections were commonly based solely on past cancer trends, and only a limited number of studies incorporated smoking data in the projection models, most likely due to the scarcity of data on past

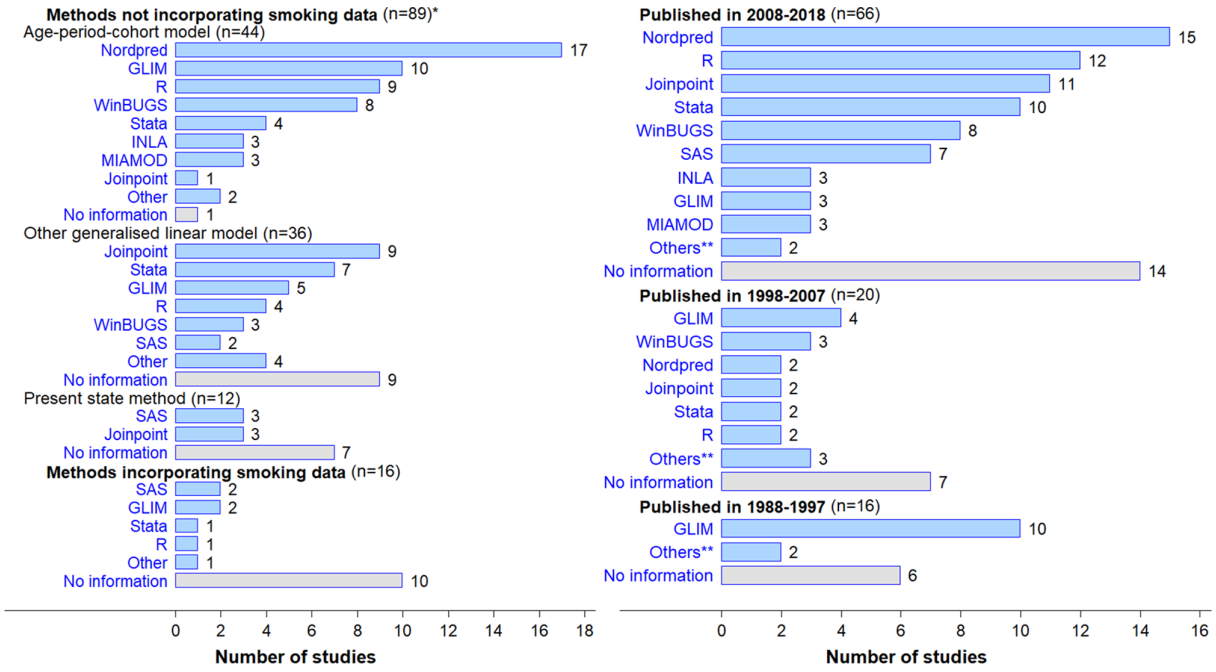

Figure 4 Statistical software packages used by method and year of publication. *Six studies used more than one method, 20 studies used more than one software package. **Others include BMDP, BAMP, S-Plus, S and Can*Trol.

Table 5 Summary of software packages commonly used in 2008–2018

Methods group	Software/ package	Free software	References	Descriptions	Programming requirement
APC model	Nordpred[10]	Yes	9 17 27 28 35 37 43 44 46 56–58	Nordpred is an R package for projection up to 25 years, based on log-link or the power 5 model, and provide significance test for use of recent slope or average slope for the whole period. Requires specific data format by 5-year age group and 5-year period and cannot incorporate other covariates.	Requires a specific data format and basic R programming. Assumes that the last non-linear period component applies to all future periods, and the non-linear cohort component was projected for estimated cohorts.
	Stata[122 124]	No	37 50 53	User-written command, published packages include 'apcfit'[122] and 'apcspline'[124] using restricted cubic splines, and the latter command has not been used for lung cancer projections. Can be used with single year data or 5-year grouped data.	Apcfit requires some programming when projecting beyond the observed data. User defines the number of knots for age, period and cohort, therefore involves model selection and comparison.
	R-other[9 123]	Yes	26 31 36 49 57	Other packages include 'Epi'[9] and 'apcfit'[123] which incorporate a smoothing method and the lexis diagram method. Can be used with single year data or 5-year grouped data. Allows user to adjust the boundary knot for period and cohort projections.	Requires R programming when projecting beyond the observed data. User defines the number of knots for age, period and cohort, and allows user to specify the centering of period and cohort.
	WinBUGS[115]	Yes	9 25 26 31 49	Commonly used for Bayesian models, with Markov Chain Monte Carlo (MCMC) techniques. Trends for age, period and cohort effects are smoothed. MCMC is inherently less robust than analytic statistical methods. There is no in-built protection against misuse.	Requires knowledge of Bayesian methods including recognition of the importance of prior distributions.
GLM	Stata[119]	No	65 66 73 80 81 104	Stata's glm fits GLMs by using either maximum quasi likelihood or Newton-Raphson (maximum likelihood) optimisation, which is the default.	Requires basic programming and user can define link functions, distributions, or perform analyses via a menu.
	SAS[120]	No	63 81 104	SAS's genmod procedure fits a GLM to the data by maximum likelihood estimation of the parameter vector beta.	Requires some SAS programming experience.
	Joinpoint[118]	Yes	18 59 78 84 95 100	Analyse Joinpoint models based on linear or log-linear regression, the tests of significance for change in trend use a Monte Carlo Permutation method.	No programming required. Can be easily learnt using the sample analyses provided on their website.

APC, age-period-cohort; GLM, generalised linear model.

smoking behaviour in the population.[21] Methods, which do not incorporate smoking data, are also generalisable to projections for other cancer types. We found that the number of studies reporting statistical projections for lung cancer increased substantially over time, and that the majority of these studies used good quality data from countries with a very high or high level of HDI.

The three-factor APC model was the most commonly used method for projecting lung cancer rates. This method does not require knowledge of aetiological factors,[25] as the period and cohort effects are considered to be surrogates for exposure to a range of risk factors.[11] For example, period effects can reflect diagnostic and treatment factors, which lead to changes in disease incidence and survival across all age groups.[11] On the other hand, the cohort effect may represent risk factors such as smoking behaviour that change from generation to generation.[7 11 106] This method is considered to be appropriate for long-term projections.[10] However, due to the non-identifiability of the linear components of the age, period and cohort parameter estimates, there is no way to distinguish the period effect and the cohort effect. This non-identifiability issue for APC models can be addressed by introducing constraints to the time effects, however, the parameter estimates can be sensitive to the choice of constraint on period and cohort factors.[3 121] In addition, the APC model used in this context generally assumes that current and past trends continue into the future, and such an assumption would be questionable if any interventions have significant impacts on the cancer rates. Given the latency period between exposure to a cancer agent and development of some cancers, projections that are based on past trends may be inaccurate.[8] Nonetheless, with the development of strategies to deal with the inherent non-identifiability problem in such models, the APC model has been implemented in various statistical software packages in recent years.[10 122–124]

In contrast to the APC model, other methods using GLMs do not include all three time components in the same model, making them less complicated to use. GLMs are more flexible and can be easily implemented using commonly available software including Stata,[119] SAS,[120] R and Joinpoint. The interpretation of the results from the standard GLM seems to be straightforward, and it can be extended to incorporate other factors.[125] This method has been evaluated using Finnish Cancer Registry data and it was concluded that the GLM performed reasonably well for short-term (eg, 5 years) projections.[125] However, GLMs may not be appropriate for long-term projections (>10 years) as the model does not consider period and cohort effects at the same time. For example, a GLM without a cohort component may not be appropriate for cancer types where significant changes in risk factors have occurred, due to the lack of cohort-specific effects in the projections.[125] On the other hand, a GLM without a period component will not be able to capture the changes in period effects for cancer types with screening programme or improvements in treatments over time.[33] It is recommended that the potential significance of period and cohort effects should be examined and determined prior to implementing any projections using GLMs.[9]

The present state method is the simplest projection method, which projects future numbers of lung cancer cases or deaths by applying the average of the age-specific incidence/mortality rates observed in the most recent years to the projected future age-specific population estimates. The projection is based on a very strong assumption that the rates will remain constant over the projection period, which could be 20 or 30 years long. This method does not need special software, and it is a practical method to use when long-term historical data are not available. Although the validity of this assumption may not be realistic, especially for long-term projections, the results of present state projections can provide base assumptions from which to examine the impact of population growth and ageing on the cancer burden, and can provide a benchmark which is useful for evaluating the effect of cancer prevention or intervention activities.

Due to the association between tobacco smoking and lung cancer risk,[4 5] past smoking behaviour is considered to be an important predictor for lung cancer rates.[3 7] The accuracy of lung cancer projections can, therefore, be improved if historical data on smoking exposure in the population are incorporated into the models. This is likely to be particularly important if smoking trends peak and then reverse over time, as has occurred in a number of high-income countries,[126] since the simple projection of lung cancer trends based only on data reflecting the burgeoning epidemic will not reflect the impact of a turnaround in smoking prevalence.[3] However, our review found that only a very limited number of published studies incorporated smoking data in the projection models, with only eight studies including detailed historical data on smoking exposure along with lung cancer data in their projection models.[3 7 33 103 106 108 111 112] Another eight studies used less detailed information or a limited amount of smoking data, which was not directly included in the projection models.[41 42 102 104 105 107 109 110] Negri *et al* developed a method to incorporate smoking patterns into an APC model, multiplying the estimated period parameters by predefined coefficients based on recent trends in smoking prevalence and the tar yield of cigarettes.[41 42] Two studies reported projections adjusted for the SAF, which involved modelling projections based on observed cancer data and then modifying the projected rates by multiplying by the SAF, which was estimated from a previous population-based study.[102 107] This method can be used for data from any country where lung cancer is primarily caused by smoking,[107] but is more suitable for countries where lung cancer mortality for males had reached its peak some time ago and recent smoking prevalence is similar for males and females.[107] In addition, it should be noted that the SAF based on the relative risk of death for current smokers estimated by the American Cancer Society's Cancer Prevention Study II (ACS CPS-II) in the USA may not be applicable to other countries.[107]

A few other studies used methods which were based on cancer rates reported in the literature or on less detailed data,[104 105 109] these methods are useful for countries where it is not realistic to use more sophisticated models due to the lack of detail in the available cancer and smoking data. However, for projections in populations at an earlier stage in the smoking epidemic more detailed information on tobacco exposure would be necessary so that the complex changes over time in the smoking behaviour of the population are captured.[127]

As previously discussed, GLMs are flexible and can be extended to incorporate other covariates, including smoking exposure, at the requisite level of detail. Log-linear models assuming a Poisson distribution based on age, cohort and cigarette tar exposure were reported by Brown and Kessler[3] using data from the USA, and by Shibuya *et al*[7] using data for four countries—the USA, UK, Canada and Australia. Both studies were based on sex-specific tobacco consumption over time for two large age groups (30–49 years and ≥50 years).[3 7] These studies take into account the effects of changes in tobacco consumption and differences in exposure among birth cohorts, and both studies demonstrated improvements in projections by incorporating tar exposure measurements into the projection models. This approach was also reported by Knorr-Held and Rainer[33] using data from Germany, but they concluded that the available smoking data in Germany were not able to improve their projections, because there was no available information on sex-specific cigarette consumption, nor on the average tar content per cigarette. This confirmed that accurate projections and the selection of appropriate projection methods depend on the quality and availability of data at the requisite level of detail. Some other smoking-related variables have also been used, including smoking intensity[103 108] and the number of years of smoking prior to age 40.[106] All the studies using GLMs did not include constraints on the period and cohort components. This method has the advantage of flexibility and is able to piecewise examine the performance of various models based on different covariates, which is particularly relevant when detailed data on risk factors are available. However, the application of this method for a specific cancer type requires reasonable justification and validation, to ensure that the covariates included in the projection model are sufficient to reflect the factors that impact cancer rates in the population. In addition, the potential risk of ecological bias should be considered.

The availability of suitable software is paramount when dealing with complex models and inferences, such as when using APC models. The increasing number of studies using APC models is likely to be due to recent developments in statistical software packages including R and Stata. Norpred is a free-software package in R and S-PLUS for APC modelling which was developed by Møller *et al* at the Cancer Registry of Norway.[10] It incorporates a smoothing technique and has become the most commonly used software for fitting APC models in recent years. However, Norpred only provides projections for a maximum of 25 years beyond the observed data, and no other covariates can be incorporated into the model. Other R packages, including 'Epi',[9] 'apc'[123] and 'INLA',[48] can also be used for cancer incidence or mortality projections. Two packages in Stata were developed for APC models in the early 2010s and have the advantage of more flexible modelling implementation,[122 124] although one package requires additional programming when projecting beyond the observed data.[122] Joinpoint[118] is another popular package that has been increasingly used to project cancer rates into the future by extrapolating the most recent trend.[128] However, Joinpoint is only considered to be suitable for short-term projections.[118]

We acknowledge that each method included in this review has its own merits and limitations depending on the length of projections, data quality and availability, and the timing of analysis in relation to different stages of the smoking epidemic in a country (particularly, whether smoking prevalence is assumed to peak over the time frame of the analysis). It is important to note that all projections of cancer incidence and mortality based on historical trends may be inaccurate, regardless of the method used, if the underlying trends in risk or interventions change.[9] This is particularly relevant to lung cancer due to its strong relationship with tobacco exposure.[8] There is no way to identify the 'best model' for all situations or to conclude that one method is superior to another. Furthermore, even projections using the same method can be sensitive to the model setting and the length of the projection base.[10] Therefore, wherever possible, appropriate validation of the selected projection method should be performed, as such information is useful for checking the specifications of the model and helps researchers understand the potential limitations of the projection model. Performing a validation of the model being used for a projection by withholding the most recent observed data from the model fitting and then comparing the projected with the observed rates for the most recent period, can provide important information on the performance of the projection model.[7] Surprisingly, however, fewer than 12% of the studies reported on this, although as high-quality data on lung cancer rates is now available for several decades or longer for many countries it is likely that this type of validation will become more feasible and more frequently performed. In addition, as more data become available over time, prior statistical projections can be compared against the emergent data, which will allow for even greater understanding of the general strengths and pitfalls of the various methods—this exercise is underway and will yield further insights.

Strengths and weaknesses

Although we searched multiple electronic databases (Medline, Embase and PreMEDLINE databases), this review is limited to studies published in English. Thus, this review may not be complete if there were relevant studies published in other languages. It is also possible

that we may have missed articles in the initial search, as we were unable to search the grey literature completely for eligible studies. It should also be noted that this review is limited to lung cancer only (International Classification of Diseases 10th Revision, ICD 10 C33–C34), which means it will not capture the literature on every possible type of cancer related to the lungs (eg, mesothelioma). In addition, the wide variability in study populations and time periods made meta-analyses infeasible. Despite these potential limitations, we believe this review is still a valuable resource and has many strengths. By searching the reference lists of all included articles, we should have ensured a thorough and extensive coverage of the literature, and developing prespecified assessment criteria to provide clear definitions for the different assessment areas allowed for objective assessment of the studies. Also, a pretested and revised standardised form was used for data extraction, which should have minimised differences between the data extraction by different reviewers, as confirmed by the high agreement for the data extracted by the two reviewers (91.6%). Also, we developed an organisational framework to categorise and summarise the projection methods used in the literature, which provides the comprehensive information and highlights the similarities and differences across methods. To our knowledge, this systematic review is the first to provide comprehensive, up-to-date coverage of the literature on statistical methods for projecting lung cancer rates.

Implications for research

This systematic review provides a comprehensive summary of the statistical methods over the past three decades used in published lung cancer incidence or mortality projections. The assessment of the strengths and advantages of existing methods will help researchers to better understand the currently used statistical methods for projecting lung cancer rates. In this review, we summarised both theoretical and practical aspects, including software information and generalisability of the methods, and some of the common methods described in this review can be applied to other cancer types, so it is hoped that this review will serve as a resource for researchers who are interested in using or developing one or more of these methods for projecting cancer rates. In particular, the methods incorporating a covariate such as smoking may be also applicable to projection of rates for other cancers with data on risk factors or diagnostic factors at the requisite level of detail, such as prostate specific antigen (PSA) testing rates for prostate cancer.

Acknowledgements We would like to thank Clare Kahn for editorial assistance, Harriet Hui for assistance in collecting full-text articles, and Victoria Freeman for assistance in updating the final search for this review.

Contributors All authors contributed substantially to the conception and design of the study. KC and DLO conceived the study. SH, QL and XQY drafted the study protocol and designed the data extraction form with input from DLO. SH did the initial scan of the literature search results to exclude articles that were clearly irrelevant. XQY, QL and DLO acted as reviewers. XQY and QL conducted the data extraction, data analysis and drafted the initial manuscript. DLO contributed to the interpretation of results and drafting of the manuscript. SW and MC contributed to the interpretation of results. All authors critically reviewed the manuscript and approved the final version.

Funding This project has not received any funding and the authors are employed by Caner Council NSW, Australia.

Competing interests None declared.

Patient consent for publication Not required.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement All data relevant to the study are included in the article or uploaded as online supplementary information.

REFERENCES

1. Ferlay J, Soerjomataram I, Ervik M, *et al*. GLOBOCAN 2012 v1.0, Cancer incidence and mortality worldwide: IARC CancerBase No. 11 Lyon, France: International Agency for Research on Cancer, 2013. Available: http://globocan.iarc.fr
2. Bashir SA, Estève J. Projecting cancer incidence and mortality using Bayesian age-period-cohort models. *J Epidemiol Biostat* 2001;6:287–96.
3. Brown CC, Kessler LG. Projections of lung cancer mortality in the United States: 1985-2025. *J Natl Cancer Inst* 1988;80:43–51.
4. Doll R, Hill AB. Smoking and carcinoma of the lung. *BMJ* 1950;2:739–48.
5. Doll R, Peto R. Cigarette smoking and bronchial carcinoma: dose and time relationships among regular smokers and lifelong non-smokers. *J Epidemiol Community Health* 1978;32:303–13.
6. Hakulinen T, Hakama M. Predictions of epidemiology and the evaluation of cancer control measures and the setting of policy priorities. *Soc Sci Med* 1991;33:1379–83.
7. Shibuya K, Inoue M, Lopez AD. Statistical modeling and projections of lung cancer mortality in 4 industrialized countries. *Int J Cancer* 2005;117:476–85.
8. Bray F, Møller B. Predicting the future burden of cancer. *Nat Rev Cancer* 2006;6:63–74.
9. Cancer Projections Network. *Long-Term projection methods: comparison of age-period-cohort model-based approaches*. Alberta Health Services, 2010.
10. Møller B, Fekjaer H, Hakulinen T, *et al*. Prediction of cancer incidence in the Nordic countries: empirical comparison of different approaches. *Stat Med* 2003;22:2751–66.
11. Smith TR, Wakefield J. A review and comparison of age–period–cohort models for cancer incidence. *Statist Sci* 2016;31:591–610.
12. Moher D, Liberati A, Tetzlaff J, *et al*. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med* 2009;6:e1000097.
13. Higgins JPT, Green S, eds. Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 [updated March 2011]. The Cochrane Collaboration, 2011. Available: www.handbook.cochrane.org
14. Deeks J, Dinnes J, D'Amico R, *et al*. Evaluating non-randomised intervention studies. *Health Technol Assess* 2003;7:1–173.
15. Cancer Incidence in Five Continents. In: Forman D, Bray F, Brewster DH, *et al*, eds. *IARC Scientifc Publication No 164*. Lyon: International Agency for Research on Cancer, 2014.
16. United Nations Development Programme (UNDP). *Human development report 2016, human development for everyone*. Washington DC, USA: Communications Development Incorporated, 2016.
17. Alonso R, Piñeros M, Laversanne M, *et al*. Lung cancer incidence trends in Uruguay 1990–2014: an age-period-cohort analysis. *Cancer Epidemiol* 2018;55:17–22.
18. Bosetti C, Malvezzi M, Rosso T, *et al*. Lung cancer mortality in European women: trends and predictions. *Lung Cancer* 2012;78:171–8.
19. Cayuela A, Rodríguez-Dominguez S, López-Campos JL, *et al*. Lung cancer mortality in Spain: estimating the future burden to the year 2028. *int j tuberc lung dis* 2011;15:1117–21.

20. Chen W-Q, Zheng R-S, Zeng H-M. Bayesian age-period-cohort prediction of lung cancer incidence in China. *Thoracic Cancer* 2011;2:149–55.
21. Clements MS, Armstrong BK, Moolgavkar SH. Lung cancer rate predictions using generalized additive models. *Biostatistics* 2005;6:576–89.
22. Clèries R, Buxó M, Martínez JM, *et al*. Contribution of changes in demography and in the risk factors to the predicted pattern of cancer mortality among Spanish women by 2022. *Cancer Epidemiol* 2016;40:113–8.
23. Clèries R, Ameijide A, Marcos-Gragera R, *et al*. Predicting the cancer burden in Catalonia between 2015 and 2025: the challenge of cancer management in the elderly. *Clin Transl Oncol* 2018;20:647–57.
24. Coupland VH, Okello C, Davies EA, *et al*. The future burden of cancer in London compared with England. *J Public Health* 2010;32:83–9.
25. Eilstein D, Uhry Z, Lim TA, *et al*. Lung cancer mortality in France. trend analysis and projection between 1975 and 2012, using a Bayesian age-period-cohort model. *Lung Cancer* 2008;59:282–90.
26. Eilstein D, Eshai K. Lung and breast cancer mortality among women in France: future trends. *Cancer Epidemiol* 2012;36:e341–8.
27. Ferlay J, Parkin DM, Steliarova-Foucher E. Estimates of cancer incidence and mortality in Europe in 2008. *Eur J Cancer* 2010;46:765–81.
28. Ferlay J, Steliarova-Foucher E, Lortet-Tieulent J, *et al*. Cancer incidence and mortality patterns in Europe: estimates for 40 countries in 2012. *Eur J Cancer* 2013;49:1374–403.
29. Galasso R, Capocaccia R, Del Riccio L, *et al*. Estimates of cancer burden in Basilicata and Calabria. *Tumori* 2013;99:390–8.
30. Hristova L, Dimova I, Iltcheva M. Projected cancer incidence rates in Bulgaria, 1968-2017. *Int J Epidemiol* 1997;26:469–75.
31. Jürgens V, Ess S, Cerny T, *et al*. A Bayesian generalized age-period-cohort power model for cancer projections. *Stat Med* 2014;33:4627–36.
32. Kaneko S, Ishikawa KB, Yoshimi I, *et al*. Projection of lung cancer mortality in Japan. *Cancer Sci* 2003;94:919–23.
33. Knorr-Held L, Rainer E. Projections of lung cancer mortality in West Germany: a case study in Bayesian prediction. *Biostatistics* 2001;2:109–29.
34. Kubík A, Plesko I, Reissigová J. Prediction of lung cancer mortality in four central European countries, 1990-2009. *Neoplasma* 1998;45:60–7.
35. Li M, Wang S, Han X, *et al*. Cancer mortality trends in an industrial district of Shanghai, China, from 1974 to 2014, and projections to 2029. *Oncotarget* 2017;8:92470–82.
36. Malvezzi M, Bosetti C, Rosso T, *et al*. Lung cancer mortality in European men: trends and predictions. *Lung Cancer* 2013;80:138–45.
37. Mistry M, Parkin DM, Ahmad AS, *et al*. Cancer incidence in the United Kingdom: projections to the year 2030. *Br J Cancer* 2011;105:1795–803.
38. Møller B, Fekjaer H, Hakulinen T, *et al*. Prediction of cancer incidence in the Nordic countries up to the year 2020. *Eur J Cancer Prev* 2002;11(Suppl 1):S1–96.
39. Møller B, Weedon-Fekjær H, Haldorsen T. Empirical evaluation of prediction intervals for cancer incidence. *BMC Med Res Methodol* 2005;5:21.
40. Møller H, Fairley L, Coupland V, *et al*. The future burden of cancer in England: incidence and numbers of new patients in 2020. *Br J Cancer* 2007;96:1484–8.
41. Negri E, La Vecchia C, Decarli A, *et al*. Projections to the end of the century of mortality from major cancer sites in Italy. *Tumori* 1990;76:420–8.
42. Negri E, La Vecchia C, Levi F, *et al*. The application of age, period and cohort models to predict Swiss cancer mortality. *J Cancer Res Clin Oncol* 1990;116:207–14.
43. Nowatzki J, Moller B, Demers A. Projection of future cancer incidence and new cancer cases in Manitoba, 2006-2025. *Chronic Dis Can* 2011;31:71–8.
44. Olsen AH, Parkin DM, Sasieni P. Cancer mortality in the United Kingdom: projections to the year 2025. *Br J Cancer* 2008;99:1549–54.
45. Pompe-Kirn V, Japelj B, Primic-Žakelj M. Future trends in breast, cervical, lung, mouth and pharyngeal cancer incidence in Slovenia. *Cancer Causes Control* 2000;11:309–18.
46. Rapiti E, Guarnori S, Pastoors B, *et al*. Planning for the future: cancer incidence projections in Switzerland up to 2019. *BMC Public Health* 2014;14:102.
47. Reissigova J, Luostarinen T, Hakulinen T, *et al*. Statistical modelling and prediction of lung cancer mortality in the Czech and Slovak Republics, 1960–1999. *Int J Epidemiol* 1994;23:665–72.
48. Ribes J, Esteban L, Clèries R, *et al*. Cancer incidence and mortality projections up to 2020 in Catalonia by means of Bayesian models. *Clin Transl Oncol* 2014;16:714–24.
49. Riebler A, Held L. Projecting the future burden of cancer: Bayesian age-period-cohort analysis with integrated nested Laplace approximations. *Biom J* 2017;59:531–49.
50. Rutherford MJ, Thompson JR, Lambert PC. Projecting cancer incidence using age-period-cohort models incorporating restricted cubic splines. *Int J Biostat* 2012;8:1–25.
51. Sánchez MJ, Payer T, De Angelis R, *et al*. Cancer incidence and mortality in Spain: estimates and projections for the period 1981-2012. *Ann Oncol* 2010;21(Suppl 3):iii30–6.
52. Sharp L, Black RJ, Muir CS, *et al*. Will the Scottish cancer target for the year 2000 be met? The use of cancer registration and death records to predict future cancer incidence and mortality in Scotland. *Br J Cancer* 1996;73:1115–21.
53. Smittenaar CR, Petersen KA, Stewart K, *et al*. Cancer incidence and mortality projections in the UK until 2035. *Br J Cancer* 2016;115:1147–55.
54. Son M, Yun J-W. Cancer mortality projections in Korea up to 2032. *J Korean Med Sci* 2016;31:892–901.
55. Stracci F, Petrucci MS, Ciampichini R, *et al*. Estimates of cancer burden in Umbria. *Tumori* 2013;99:342–50.
56. Swaminathan R, Shanta V, Ferlay J, *et al*. Trends in cancer incidence in Chennai city (1982-2006) and statewide predictions of future burden in Tamil Nadu (2007-16). *Natl Med J India* 2011;24:72–7.
57. Virani S, Bilheem S, Chansaard W, *et al*. National and subnational population-based incidence of cancer in Thailand: assessing cancers with the highest burdens. *Cancers* 2017;9:250.
58. Weir HK, Thompson TD, Soman A, *et al*. The past, present, and future of cancer incidence in the United States: 1975 through 2020. *Cancer* 2015;121:1827–37.
59. Australian Institute of Health Welfare. *Cancer incidence projections Australia, 2011 to 2020*. Canberra: AIHW, 2012.
60. Bray F, Jemal A, Grey N, *et al*. Global cancer transitions according to the human development index (2008–2030): a population-based study. *Lancet Oncol* 2012;13:790–801.
61. Brenner H, Ziegler H. Monitoring and projecting cancer incidence in Saarland, Germany, based on age-cohort analyses. *J Epidemiol Community Health* 1992;46:15–20.
62. Byers T, Barrera E, Fontham ETH, *et al*. A midpoint assessment of the American cancer Society challenge goal to halve the U.S. cancer mortality rates between the years 1990 and 2015. *Cancer* 2006;107:396–405.
63. Cancer Institute NSW. *Cancer incidence and projections 2011-2021*, 2016.
64. Carson CA, Zucconi SL. Health status indicators for the year 2000: projections for Allegheny County, Pennsylvania. *Public Health Rep* 1993;108:711–5.
65. Castro C, Antunes L, Lunet N, *et al*. Cancer incidence predictions in the North of Portugal: keeping population-based cancer registration up to date. *Eur J Cancer Prev* 2016;25:472–80.
66. Didkowska J, Wojiechowska U, Zatoriski W. *Prediction of cancer incidence and mortality in Poland up to the year 2025*. Warsaw: National Programme of Cancer Prevention, 2009.
67. Dušek L, Pavlík T, Májek O, *et al*. Estimating cancer incidence, prevalence, and the number of cancer patients treated with antitumor therapy in 2015 and 2020 – analysis of the Czech National Cancer Registry. *Klin Onkol* 2015;28:30–43.
68. Dyba T, Hakulinen T, Päivärinta L. A simple non-linear model in incidence prediction. *Stat Med* 1997;16:2297–309.
69. Dyba T, Hakulinen T. Comparison of different approaches to incidence prediction based on simple interpolation techniques. *Stat Med* 2000;19:1741–52.
70. Engeland A, Haldorsen T, Tretli S, *et al*. Prediction of cancer mortality in the Nordic countries up to the years 2000 and 2010, on the basis of relative survival analysis. A collaborative study of the five Nordic cancer registries. *APMIS Suppl* 1995;49:1–161.
71. French D, Catney D, Gavin AT. Modelling predictions of cancer deaths in Northern Ireland. *Ulster Med J* 2006;75:120–5.
72. Fukuda Y, Nakamura K, Takano T. A combination of an extrapolation method and a benchmark method to develop quantitative health targets for Japan. *Health Policy* 2002;61:201–12.
73. Godlewski D, Wojtyś P, Antczak A. Predictions of cancer incidence in Wielkopolska in 2018. *Wo* 2012;1:38–43.
74. Hakulinen T, Dyba T. Precision of incidence predictions based on Poisson distributed observations. *Stat Med* 1994;13:1513–23.

75. Heinävaara S, Hakulinen T. Predicting the lung cancer burden: accounting for selection of the patients with respect to general population mortality. *Stat Med* 2006;25:2967–80.
76. Jee S*et al*. Projected mortality from lung cancer in South Korea, 1980-2004. *Int J Epidemiol* 1998;27:365–9.
77. Kuroishi T, Hirose K, Tominaga S, *et al*. Prediction of future cancer mortality in Japan. *Jpn J Clin Oncol* 1992;22:365–9.
78. Malvezzi M, Bertuccio P, Rosso T, *et al*. European cancer mortality predictions for the year 2015: does lung cancer have the highest death rate in Eu women? *Ann Oncol* 2015;26:779–86.
79. Martín-Sánchez JC, Clèries R, Lidón C, *et al*. Bayesian prediction of lung and breast cancer mortality among women in Spain (2014–2020). *Cancer Epidemiol* 2016;43:22–9.
80. Oberaigner W, Geiger-Gritsch S. Prediction of cancer incidence in Tyrol/Austria for year of diagnosis 2020. *Wien Klin Wochenschr* 2014;126:642–9.
81. Olajide OO, Field JK, Davies MMPA, *et al*. Lung cancer trend in England for the period of 2002 to 2011 and projections of future burden until 2020. *Int J Oncol* 2015;47:739–46.
82. O'Lorcain P, Comber H. Lung cancer mortality predictions for Ireland 2001–2015 and current trends in North Western Europe. *Lung Cancer* 2004;46:157–63.
83. Parsons NR, Somervaille L. Estimation and projection of population lung cancer trends (United Kingdom). *Cancer Causes Control* 2000;11:467–75.
84. Shamseddine A, Saleh A, Charafeddine M, *et al*. Cancer trends in Lebanon: a review of incidence rates for the period of 2003–2008 and projections until 2018. *Popul Health Metr* 2014;12:4.
85. Torres-Avilés F, Moraga T, Núñez L, *et al*. Lung cancer mortality trends in Chile and six-year projections using Bayesian dynamic linear models. *Cad Saude Publica* 2015;31:1975–82.
86. Tsoi KKF, Hirai HW, Chan FCH, *et al*. Cancer burden with ageing population in urban regions in China: projection on cancer registry data from World Health organization. *Br Med Bull* 2017;121:83–94.
87. Wiklund K, Hakulinen T, Sparén P. Prediction of cancer mortality in the Nordic countries in 2005: effects of various interventions. *Eur J Cancer Prev* 1992;1:247–58.
88. Yang L, Parkin DM, Ferlay J, *et al*. Estimates of cancer incidence in China for 2000 and projections for 2005. *Cancer Epidemiol Biomarkers Prev* 2005;14:243–50.
89. Yang L, Parkin DM, Li LD, *et al*. Estimation and projection of the National profile of cancer mortality in China: 1991–2005. *Br J Cancer* 2004;90:2157–66.
90. Baade PD, Meng X, Sinclair C, *et al*. Estimating the future burden of cancers preventable by better diet and physical activity in Australia. *Med J Aust* 2012;196:337–40.
91. D'Souza NDR, Murthy NS, Aras RY. Projection of burden of cancer mortality for India, 2011-2026. *Asian Pac J Cancer Prev* 2013;14:4387–92.
92. D'Souza NDR, Murthy NS, Aras RY. Projection of cancer incident cases for India -till 2026. *Asian Pac J Cancer Prev* 2013;14:4379–86.
93. Pearce A, Bradley C, Hanly P, *et al*. Projecting productivity losses for cancer-related mortality 2011 – 2030. *BMC Cancer* 2016;16:804.
94. Pisani P, Parkin DM, Ferlay J. Estimates of the worldwide mortality from eighteen major cancers in 1985. Implications for prevention and projections of future burden. *Int. J. Cancer* 1993;55:891–903.
95. Quante AS, Ming C, Rottmann M, *et al*. Projections of cancer incidence and cancer-related deaths in Germany by 2020 and 2030. *Cancer Med* 2016;5:2649–56.
96. Rahib L, Smith BD, Aizenberg R, *et al*. Projecting cancer incidence and deaths to 2030: the unexpected burden of thyroid, liver, and pancreas cancers in the United States. *Cancer Res* 2014;74:2913–21.
97. Smith BD, Smith GL, Hurria A, *et al*. Future of cancer incidence in the United States: burdens upon an aging, changing nation. *J Clin Oncol* 2009;27:2758–65.
98. Vardanjani HM, Zeinali M, Radmerikhi S, *et al*. Lung cancer prevalence in Iran by histologic subtypes. *Adv Biomed Res* 2017;6:111.
99. Yabroff KR, Bradley CJ, Mariotto AB, *et al*. Estimates and projections of value of life lost from cancer deaths in the United States. *J Natl Cancer Inst* 2008;100:1755–62.
100. Malvezzi M, Carioli G, Bertuccio P, *et al*. European cancer mortality predictions for the year 2018 with focus on colorectal cancer. *Ann Oncol* 2018;29:1016–22.
101. Martín-Sánchez JC, Lunet N, González-Marrón A, *et al*. Projections in breast and lung cancer mortality among women: a Bayesian analysis of 52 countries worldwide. *Cancer Res* 2018;78:4436–42.
102. Davis VN, Lavender A, Bayakly R, *et al*. Using current smoking prevalence to project lung cancer morbidity and mortality in Georgia by 2020. *Prev Chronic Dis* 2013;10:E74.
103. Murray CJL, Lopez AD. Alternative projections of mortality and disability by cause 1990–2020: global burden of disease study. *Lancet* 1997;349:1498–504.
104. Ng N, Winkler V, Van Minh H, *et al*. Predicting lung cancer death in Africa and Asia: differences with who estimates. *Cancer Causes Control* 2009;20:721–30.
105. Pierce JP, Thurmond L, Rosbrook B. Projecting international lung cancer mortality rates: first approximations with tobacco-consumption data. *J Natl Cancer Inst Monogr* 1992;12:45–9.
106. Preston SH, Stokes A, Mehta NK, *et al*. Projecting the effect of changes in smoking and obesity on future life expectancy in the United States. *Demography* 2014;51:27–49.
107. Stoeldraijer L, Bonneux L, van Duin C, *et al*. The future of smoking-attributable mortality: the case of England & Wales, Denmark and the Netherlands. *Addiction* 2015;110:336–45.
108. Strong K, Mathers C, Epping-Jordan J, *et al*. Preventing cancer through tobacco and infection control: How many lives can we save in the next 10 years? *Eur J Cancer Prev* 2008;17:153–61.
109. Winkler V, Mangolo NJ, Becher H. Lung cancer in South Africa: a forecast to 2025 based on smoking prevalence data. *BMJ Open* 2015;5:e006993.
110. Arslanhan S, Caner A, Helvacioglu K, *et al*. An economic analysis of tobacco elimination policies in turkey. *Health Policy* 2012;106:149–60.
111. Martín-Sánchez JC, Bilal U, Clèries R, *et al*. Modelling lung cancer mortality rates from smoking prevalence: fill in the gap. *Cancer Epidemiol* 2017;49:19–23.
112. Vogt T, van Raalte A, Grigoriev P, *et al*. The German East-West mortality difference: two crossovers driven by smoking. *Demography* 2017;54:1051–71.
113. World Health Organization, Department of Information, Evidence and Research. Mortality database 2017.
114. Francis B, Green M, Payne C. *The GLIM system: release 4 manual*. Clarendon Press, 1993.
115. Spiegelhalter D, Thomas A, Best N. *WinBUGS version 1.2 user manual*. Cambridge: MRC Biostatistics Unit, 1999.
116. Martino S, Rue H. *Implementing approximate Bayesian inference using integrated nested Laplace approximation: a manual for the inla program*. Norway: Department of Mathematical Sciences, NTNU, 2009.
117. Schmid VJ, Held L. Bayesian age-period-cohort modeling and prediction - BAMP. *J Stat Softw* 2007;21:1–15.
118. *Joinpoint Regression Program, Version 4.7.0.0 - June 2019*. Statistical Methodology and Applications Branch, Surveillance Research Program, National Cancer Institute, 2017.
119. *Stata Statistical Software: Release 13*. College Station, TX: StataCorp LP, 2013
120. *SAS software® Release 9.4 SAS Institute Inc*. Cary, NC, USA, 2013
121. Kupper LL, Janis JM, Karmous A, *et al*. Statistical age-period-cohort analysis: a review and critique. *J Chronic Dis* 1985;38:811–30.
122. Rutherford M, Lambert P, Thompson J. Age–period–cohort modeling. *The Stata Journal* 2010;10:24.
123. Nielsen B. Apc: an R package for age-period-cohort analysis. *R J* 2015;7:13.
124. Sasieni PD. Age–Period–Cohort models in Stata. *Stata J* 2012;12:45–60.
125. Dyba T, Hakulinen T. Do cancer predictions work? *Eur J Cancer* 2008;44:448–53.
126. Lopez AD, Collishaw NE, Piha T. A descriptive model of the cigarette epidemic in developed countries. *Tob Control* 1994;3:242–7.
127. Thun M, Peto R, Boreham J, *et al*. Stages of the cigarette epidemic on entering its second century. *Tob Control* 2012;21:96–101.
128. Kim H-J, Fay MP, Feuer EJ, *et al*. Permutation tests for joinpoint regression with applications to cancer rates. *Stat Med* 2000;19:335–51.