Lung Cancer 125 (2018) 68–76

Contents lists available at ScienceDirect

Lung Cancer

journal homepage: www.elsevier.com/locate/lungcan

# Lung cancer mortality in Australia: Projected outcomes to 2040

Qingwei Luo[a,b,*], Xue Qin Yu[a,b], Stephen Wade[a], Michael Caruana[a], Francesca Pesola[c], Karen Canfell[a,b], Dianne L. O'Connell[a,b,d]

[a] Cancer Research Division, Cancer Council NSW, Sydney, NSW, Australia
[b] The University of Sydney School of Public Health, Faculty of Medicine and Health, The University of Sydney, Sydney, Australia
[c] Faculty of Life Sciences & Medicine, School of Cancer & Pharmaceutical Sciences, Innovation Hub, Guys Cancer Centre, Guys Hospital, King's College London, London, UK
[d] School of Medicine and Public Health, University of Newcastle, Newcastle, NSW, Australia

ARTICLE INFO

*Keywords:*
Lung cancer mortality
Statistical projections
Australia
Tobacco consumption
Generalized linear models
Cohort effect
Population-based
Health service planning
Tobacco epidemic

ABSTRACT

*Objectives:* The aim was to develop and validate a statistical model which uses past trends for lung cancer mortality and historical and current data on tobacco consumption to project lung cancer mortality rates into the future for Australia.

*Methods:* We used generalized linear models (GLMs) with Poisson distribution including either age, birth cohort or period, and/or various measures of population tobacco exposure (considering cross-sectional smoking prevalence, cigarettes smoked and tar exposure per capita). Sex-specific models were fitted to data for 1956–2015 and age-standardized lung cancer mortality rates were projected forward to 2040. Possible lags of 20–30 years between tobacco exposure and lung cancer mortality were examined. The best model was selected using analysis of deviance. To validate the selected model, we temporarily re-fitted it to data for 1956–1990 and compared the projected rates to 2015 with the observed rates for 1991–2015.

*Results:* The best fitting model used information on age, birth cohort and tar exposure per capita; close concordance with the observed data was achieved in the validation. The forward projections for lung cancer mortality using this model indicate that male and female age-standardized rates will decline over the period 2011–2015 to 2036–2040 from 27.2 to 15.1 per 100,000, and 15.8 to 11.8 per 100,000, respectively. However, due to population growth and ageing the number of deaths will increase by 7.9% for males and 57.9% for females; from 41,040 (24,831 males, 16,209 females) in 2011–2015 to 52,403 (26,805 males, 25,598 females) in 2036–2040.

*Conclusion:* In the context of the mature tobacco epidemic with past peaks in tobacco consumption for both males and females, lung cancer mortality rates are expected to continually decline over the next 25 years. However, the number of lung cancer deaths will continue to be substantial, and to increase, in Australia's ageing population.

## 1. Introduction

Lung cancer has been the most common cause of cancer death over the last five decades in Australia [1]. Evidence of a strong causal association between tobacco smoking and lung cancer has been well established since the early 1950s [2]. Australia has successfully implemented many tobacco control interventions, and there have been subsequent reductions in lung cancer mortality [3]. However, despite the marked decline in smoking prevalence seen for Australian males since the 1950s and Australian females since the 1980s [3], it is still estimated that lung cancer will be the largest cause of cancer death in 2017 [4].

Cancer mortality is an important measure of the burden of cancer in a population, and can be of great use in informing the planning of health services and resource management [5,6]. As tobacco exposure is the most significant risk factor for lung cancer [2], it is well accepted that it is an important explanatory factor for lung cancer mortality [6,7]. The projected impact of tobacco control policies on future lung cancer mortality rates is crucial to understanding the success of such policies [5,7]. However, lung cancer mortality projections have

* Corresponding author at: Cancer Research Division, Cancer Council NSW, P.O. Box 572, Kings Cross, NSW, 1340, Australia.
*E-mail addresses:* qingweil@nswcc.org.au (Q. Luo), xueqiny@nswcc.org.au (X.Q. Yu), Stephen.Wade@nswcc.org.au (S. Wade), Michael.Caruana@nswcc.org.au (M. Caruana), francesca.pesola@kcl.ac.uk (F. Pesola), karen.canfell@nswcc.org.au (K. Canfell), dianneo@nswcc.org.au (D.L. O'Connell).

https://doi.org/10.1016/j.lungcan.2018.09.001
Received 5 July 2018; Received in revised form 30 August 2018; Accepted 3 September 2018

commonly been based solely on past mortality trends, due to the scarcity of data on historical smoking behaviour at the population level [8]. Currently, the most common method for projecting cancer incidence and mortality is to use age-period-cohort (APC) models. The basis of APC models is well explained in the published literature [9,10], but in brief, APC models describe the rate of an event as a function of age, period and cohort effects. There is, however, a non-identifiability problem inherent in APC models due to the linear relationship between age, period and cohort [9]. While there have been some developments in methods to overcome the non-identifiability problem [10–16], there is no way to distinguish the period effect and the cohort effect, and the parameter estimates obtained can be sensitive to the choice of constraints placed on the period and cohort factors [6,17].

Some previous studies have reported the inclusion of detailed smoking data as a factor in lung cancer mortality projections [6,7,18,19]. Brown and Kessler [6] using data for the United States of America (USA) and Shibuya et al. [7] using data for Australia, the USA, the United Kingdom and Canada, fitted a Generalized Linear Model (GLM) using Poisson regression for lung cancer mortality rates with terms for age, cohort and lagged sex-period-specific cigarette tar consumption [6,7]. One other study by Preston et al. used a log-linear binomial regression model for lung cancer mortality rates in the USA with terms for age and number of years of smoking prior to age 40 [18], and a study using Spanish data by Martin-Sanchez et al. used a linear regression model to predict lung cancer mortality rates based on smoking prevalence for two large age groups [19]. Here, we propose an alternative model to project lung cancer mortality rates using GLMs with a Poisson distribution including age and cohort, or age and period, together with sex-age-cohort-specific smoking-related variables. As with the methods used by Brown and Kessler [6], Shibuya et al. [7], and Preston et al. [18], our model does not suffer the non-identifiability problem. With this model, we are also able to examine the period or cohort effect after adjusting for smoking-related variables.

Due to delays in the impact of changes in smoking behaviour, lung cancer mortality rates to 2040 are expected to predominantly reflect tobacco exposure that has already occurred. The aim of this study was therefore to develop and validate a statistical model to project age-standardized lung cancer mortality rates and numbers of deaths from lung cancer for the period 2016–2040, based on past trends in lung cancer mortality and historical and current data on tobacco consumption for the Australian population.

## 2. Material and methods

### 2.1. Data sources

#### 2.1.1. Lung cancer mortality and population data

We obtained from the World Health Organization (WHO) Mortality Database (MDB) [20], national tabulated data on the numbers of deaths from lung cancer in Australia by sex, age and calendar year from 1956 to 2015 to allow for a minimum of 20 years of observed data before the peak in lung cancer mortality rates for males was reached in the early 1980s. The data available in the WHO MDB comprise deaths registered in national vital registration systems with underlying cause of death as coded by the relevant national authority in each country. Australia is one of the countries with near complete population coverage [20]. In Australia, it is a legal requirement for each state and territory to record all deaths in registries administered by the various state and territory Registrars of Births, Deaths and Marriages, and either a medical practitioner or a coroner is required to certify the cause of death [21]. The corresponding Australian population data by sex, 5-year age group and calendar year from 1956 to 2040 were obtained from the Australian Historical Population Statistics and Population Projections (Series B, based on medium population growth) produced by the Australian Bureau of Statistics (ABS) [22,23]. For the purposes of our analyses, we grouped the mortality data into 5-year age groups and 5-year periods. Deaths from lung cancer that occurred before the age of 30 were excluded, as death from lung cancer is rare for this age group [1].

#### 2.1.2. Data on smoking patterns in Australia

We obtained the data on smoking from two data sources: the International Smoking Statistics (ISS) Web Edition [24], and the National Drug Strategy Household Surveys (NDSHS) for 2007–2016 [25–28]. Integrated ISS and newly released NDSHS data are hereafter referred to as "ISS-NDSHS data". As data on smoking behaviour for pre-adolescents and young adolescents is very scarce, and is not included in the sales adjustment calculations in the ISS, we did not include smoking information for those under 15 years of age. Although smoking clearly does occur below this age, it is at a much lower level than for the adult population [25–28].

*2.1.2.1. ISS data.* The ISS database provides data from different surveys, and provides information on annual tobacco sales from 1920 to 2010, smoking prevalence by age group and sex from 1945 to 2004, and number of cigarettes consumed per person per day by age group and sex from 1972 to 2004 [24]. Data for men and women from nationally representative surveys were separately included. When multiple surveys were available, *a priori* defined selection criteria (as described in Appendix 1) were applied to determine which data sources to include in the current analysis for each calendar year.

*2.1.2.2. NDSHS data.* The NDSHS is one of the national surveys included in the ISS up to 2004, and has been conducted at three yearly intervals since 1985 [24]. The newly released NDSHS data for 2007–2016 were processed using the same methods previously described for the ISS data [24] to extend the smoking prevalence and tobacco consumption data to 2016. The Absolute Person Weight was used in the calculation of the smoking statistics to ensure that the sample is representative of the population [27].

### 2.2. Outcome and study variables

The outcomes of interest were the lung cancer mortality rate and the number of deaths from lung cancer by 5-year calendar period for men and women. Study variables used in this analysis included sex, 5-year age group, 5-year calendar period and 5-year birth cohorts, which were coded as the middle year of each five year period. Person-years at risk were approximated by the population estimates for the middle year of the five year period [6]. Several smoking-related variables were considered, including the smoking prevalence, number of cigarettes sold, number of cigarettes consumed per capita, and average tar per cigarette. Previous research has suggested that there is a considerable lag between the initiation of smoking and the development of lung cancer [29], but that this lag may vary considerably across populations and countries [6,7,30–33]. In this study, all sex-specific smoking-related variables were lagged by 20–30 years in each model for males and females separately. In order to allow for a 30-year lag, all smoking-related variables were reconstructed backwards to 1930 and projected forwards to 2020, and were aggregated into 5-year periods. Similar backward reconstruction techniques were used by Adair et al. and Shibuya et al. to estimate tobacco consumption data [3,7]. The data sources and the estimation process for each smoking-related variable are summarized and described in detail in Appendix 2.

### 2.3. Statistical analyses

Given the inherent non-identifiability problem in APC models due to the linear relationship between age, period and cohort [9], additional strategies are required to deal with this, and so these models need to be implemented using special software packages which often don't allow the user to modify the prediction of future cohort and period effects. We therefore chose to use GLMs with a Poisson distribution including age,