




Article

Predicting the Epidemiological Dynamics of Lung Cancer in Japan

Takayuki Yamaguchi ^{1,2} and Hiroshi Nishiura ^{1,2,*} 

¹ Graduate School of Medicine, Hokkaido University, Kita 15 Jo Nishi 7 Chome, Kita-ku, Sapporo-shi, Hokkaido 060-8638, Japan; tyamaguchi@med.hokudai.ac.jp

² CREST, Japan Science and Technology Agency, Honcho 4-1-8, Kawaguchi, Saitama 332-0012, Japan

* Correspondence: nishiurah@med.hokudai.ac.jp; Tel.: +81-11-706-5066

Received: 1 February 2019; Accepted: 5 March 2019; Published: 8 March 2019



Abstract: While the prevalence of smoking has steadily declined over time, the absolute numbers of lung cancer cases and deaths have continued to increase in Japan. We employed a simple mathematical model that describes the relationship between demographic dynamics and smoking prevalence to predict future epidemiological trends of lung cancer by age and sex. Never-smokers, smokers, and ex-smokers were assumed to experience different hazard of lung cancer, and the model was parameterized using data from 2014 and before, as learning data, and a future forecast was obtained from 2015 onwards. The maximum numbers of lung cancer cases and deaths in men will be 76,978 (95% confidence interval (CI): 76,630–77,253) and 63,284 (95% CI: 62,991–63,507) in 2024, while those in women will be 42,838 (95% CI: 42,601–43,095) and 32,267 (95% CI: 32,063–32,460) in 2035 and 2036, respectively. Afterwards, the absolute numbers of cases and deaths are predicted to decrease monotonically. Our compartmental modeling approach is well suited to predicting lung cancer in Japan with dynamic ageing and drastic decline in smoking prevalence. The predicted burden is useful for anticipating demands for diagnosis, treatment, and care in the healthcare sector.

Keywords: tobacco; lung neoplasms; forecasting; simulations; epidemiology; Japan

1. Introduction

In Japan, the absolute numbers of cases of and deaths from lung cancer have been increasing for decades (Figure 1A,B) [1]. The numbers of newly diagnosed cases of lung cancer in men and women increased from 12,701 in 1975 to 77,617 in 2014 and from 4971 in 1975 to 36,933 in 2014, respectively. The increase is considered to have reflected demographic ageing and past exposure history to smoking among older adults (Figure 1C,D) [1]. Similarly, the numbers of lung cancer deaths in men and women increased from 10,711 in 1975 to 52,505 in 2014 and from 4048 in 1975 to 20,891 in 2014, respectively [1,2]. If diagnosed with lymph node or distant metastasis, the effectiveness of treatment for lung cancer is known to be limited, and the overall five-year survival rate of patients diagnosed in 2006–2008 is as small as 31.9% [2,3]. Pathologically, different types of lung cancers occur at different anatomical sites and have different characteristics. Squamous cell carcinoma and small cell carcinoma are common among men, especially smokers aged 35 years or older. They often arise centrally in larger bronchi, and small cell carcinoma is more rapidly metastatic than other types. Pulmonary adenocarcinoma is the most common type of lung cancer, seen in both smokers and lifelong nonsmokers, but most cases are associated with smoking, and it is common among women. These types of lung cancer share the common characteristic of an established causal link to tobacco smoking [3].

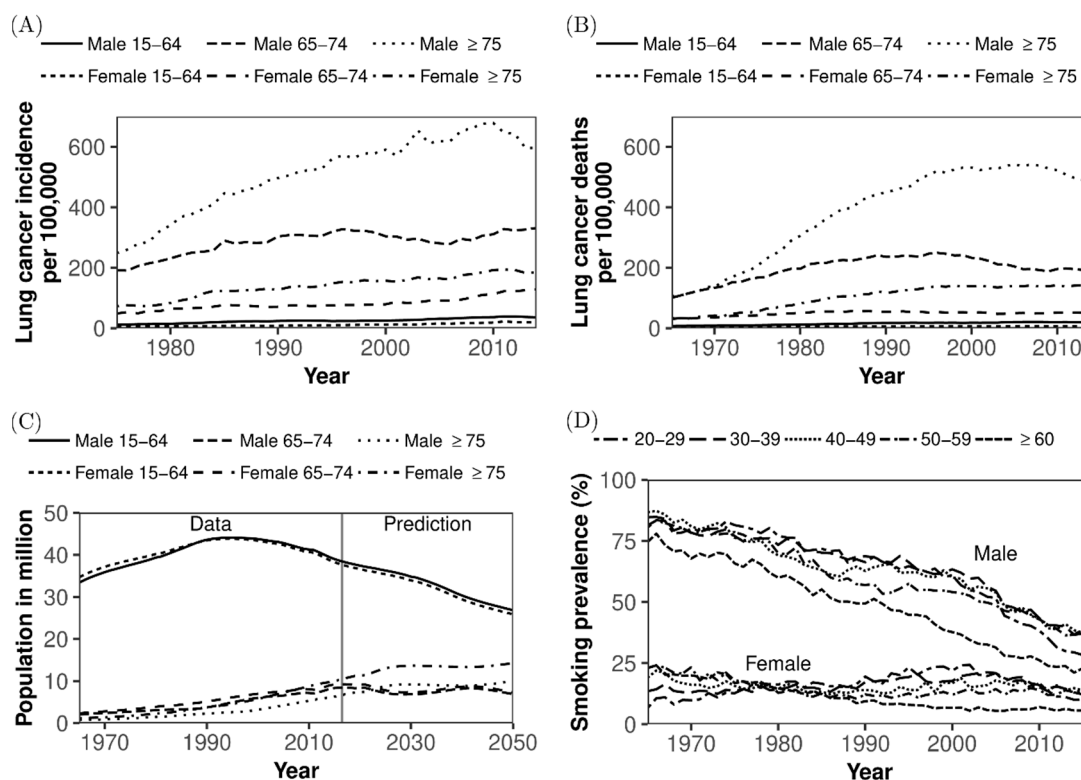


Figure 1. Incidence and mortality of lung cancer and the demographic dynamics and smoking prevalence in Japan. Published data of lung cancer incidence and mortality and the demographic dynamics and smoking prevalence in Japan. (A) Incidence of lung cancer by age group and sex per 100,000 population, (B) Mortality due to lung cancer by age group and sex per 100,000, (C) Age distribution of the population by sex, and (D) Prevalence of smokers by age group and sex.

In Japan, the population dynamics of smokers have dramatically changed over time (Figure 1D) [4]. In 1965, the smoking prevalence was 82.3% and 15.7% in men and women, respectively, and these figures have dropped to 29.7% and 9.7% in 2016, respectively [4]. In addition, interventions against secondhand smoking, including separation of smoking areas from public spaces and a total ban on smoking within buildings, are underway. Because of these drastic declines in exposure over time, it is anticipated that the epidemiological dynamics of lung cancer will also change dramatically in the future, reflecting the dynamic decline in the population at risk. However, the epidemiological dynamics of lung cancer are also regulated by demographic dynamics: Japan is a superaged nation, with a dependency ratio that will exceed 50% in the near future (Figure 1C) [5,6]. As the country still has ultralow fertility, the ageing is still expected to progress [6]. Then, the incidence of lung cancer per 100,000 persons could continue to increase, fueled by the inflow of older people who become increasingly aged and have an elevated risk of contracting lung cancer. Besides, as Japan’s population has started to decrease, the prevalence of smoking has also monotonically decreased, and thus, it is anticipated that the absolute number of lung cancer cases will eventually stop growing.

Considering these two contradicting dynamic features, i.e., decreasing prevalence of smokers and ageing in declining population, predicting the epidemiological dynamics of lung cancer in Japan’s future require us to account for detailed underlying population dynamics, including time-dependent changes in the population at high risk of lung cancer. While qualitative discussions on the eventual decline of lung cancer have taken place elsewhere [7], and although various modeling studies have been conducted in North America to explore the impact of anti-smoking measures on lung cancer epidemiology [8,9], we have yet to capture the dynamics in Japan or to fully grasp the outcomes using a mathematical model that explicitly describes the underlying population dynamics.

The purpose of the present study is to devise a simple mathematical model that allows us to describe both demographic dynamics and prevalence of smoking to appropriately predict future epidemiological trends of lung cancer by age and sex in Japan.

2. Materials and Methods

2.1. Epidemiological and Demographic Data

To accomplish the goal of this study, we reconstructed the demographic and epidemiological dynamics of smokers and patients with lung cancer. To do so, we extracted five pieces of information as the input data: (i) lung cancer incidence, (ii) lung cancer mortality, (iii) smoking prevalence, (iv) age- and sex-specific population size, and (v) mortality in general. For (i), we used the Cancer Registry and Statistics from the Cancer Information Service, National Cancer Center, Japan by 5-year age groups from 0 to 85 years and by sex from 1975 to 2014 [1]. The nationwide cancer registry system was fully set up in Japan only recently, and for this reason, the estimated cancer incidence from 32 population-based cancer registries was used as a substitute to represent the incidence of lung cancer in Japan [10]. For (ii) and (v), we used cause-specific mortality, given as the yearly number of deaths by cause, by 5-year age groups and sex from 1965 to 2014, from data publicly available from the Health, Labour and Welfare Statistics Association [5,11]. The time series of smoking prevalence (i.e., dataset (iii)) from 1965 to 2016 was derived from the report by Japan Tobacco, Inc. by 10-year age groups from 20 to 60 years and by sex [4,12]. Although the national survey (i.e., the National Health and Nutrition Survey in Japan) also recorded similar datasets of smoking prevalence, the national survey data have only been available since 1989, and because of the overall similarity between the two datasets, we decided to use the Japan Tobacco dataset on account of its longer time series. Moreover, we rigorously fitted our model to the age- and sex-specific population sizes in Japan (i.e., dataset (iv)) using census data from the Statistics Bureau, Ministry of Internal Affairs and Communications, Japan to obtain the population size by age (0–100 years) and sex [13]. Because the census is taken only every 5 years, and the population size in gap years was extrapolated by the Ministry of Health, Labour and Welfare, we used the compensated population estimates for every year. In addition, future population sizes from 2016 to 2065 were used to obtain future predictions of lung cancer, and for this purpose, the projected population by age and sex was extracted from the predictions of the National Institute of Population and Social Security, Japan [6].

2.2. Mathematical Model

We used a compartmental mathematical model that describes the natural history of smoking and the differential hazard rates of developing lung cancer according to different smoking statuses (Figure 2). Every person is born as a never-smoker, but people experience the hazard of becoming a smoker that depends on time and age. Smokers can quit smoking during the course of life to become ex-smokers. Let t and a represent time and age, respectively. The population's smoking status is expressed by $M(t, a)$, $S(t, a)$, and $E(t, a)$, representing the number of never-smokers, smokers, and ex-smokers, respectively, in year t at age a . Never-smokers begin smoking at a rate of $\sigma(t, a)$, and smokers cease smoking at a rate of $\delta(t, a)$. For simplicity, we assume that ex-smokers do not restart smoking again (i.e., we assume that purported ex-smokers who failed to quit smoking for a long period were not truly ex-smokers and that they remained in the smokers' compartment). All three compartments experience two different hazards: the hazard of lung cancer and the mortality hazard due to causes other than lung cancer. The hazard of lung cancer among never-smokers is time- and age-dependent, $\lambda(t, a)$, and smokers and ex-smokers experience k_S and k_E times greater hazard rates of lung cancer, which causes the individual enter the compartment of patients with lung cancer $L(t, a)$, who die at a rate of $\mu(t, a) + \nu(t, a)$, where $\mu(t, a)$ is the mortality rate of the general population due to any cause other than lung cancer and $\nu(t, a)$ is the excess mortality due to lung cancer. To represent the elevated risk of death by smoking status, we use $1 + q_S$ and $1 + q_E$ as the hazard ratios of death from

any cause other than lung cancer among smokers and ex-smokers, respectively, as compared with never-smokers. That is, never smokers, smokers, and ex-smokers die at the rates $\mu(a)$, $(1 + q_S)\mu(a)$, and $(1 + q_E)\mu(a)$, respectively.

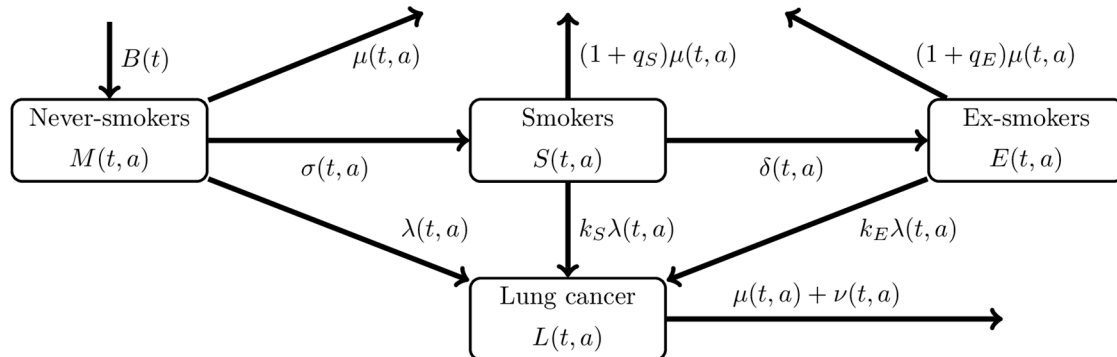


Figure 2. Compartmental model of lung cancer. Compartment diagram of the proposed model to predict lung cancer from the epidemiological, demographic, and smoking data. The population is categorized by smoking and lung cancer status. Never smokers begin smoking at a transition rate σ and move to the compartment of smokers. Smokers cease smoking at a transition rate δ and become ex-smokers. For simplicity and because of the absence of data over long time series, we assume that ex-smokers do not restart smoking. Never-smokers, smokers, and ex-smokers experience different hazard rates of lung cancer at λ , $k_S\lambda$, and $k_E\lambda$, respectively. Members of the population in each compartment die at given mortality rates.

The dynamics is governed by the following partial differential equations:

$$\begin{cases}
 \left(\frac{\partial}{\partial t} + \frac{\partial}{\partial a}\right)M(t, a) = -(\mu(t, a) + \sigma(t, a) + \lambda(t, a))M(t, a), \\
 \left(\frac{\partial}{\partial t} + \frac{\partial}{\partial a}\right)S(t, a) = \sigma(t, a)M(t, a) - ((1 + q_S)\mu(t, a) + \delta(t, a) + k_S\lambda(t, a))S(t, a), \\
 \left(\frac{\partial}{\partial t} + \frac{\partial}{\partial a}\right)E(t, a) = \delta(t, a)S(t, a) - ((1 + q_E)\mu(t, a) + k_E\lambda(t, a))E(t, a), \\
 \left(\frac{\partial}{\partial t} + \frac{\partial}{\partial a}\right)L(t, a) = \lambda(t, a)(M(t, a) + k_S S(t, a) + k_E E(t, a)) - (\mu(t, a) + \nu(t, a))L(t, a),
 \end{cases} \tag{1}$$

with the initial conditions and the boundary conditions

$$\begin{aligned}
 M(t_0, a) &= M_0(a), \\
 S(t_0, a) &= S_0(a), \\
 E(t_0, a) &= E_0(a), \\
 L(t_0, a) &= L_0(a), \\
 M(t, a_0) &= B(t), \\
 S(t, a_0) &= 0, \\
 E(t, a_0) &= 0, \\
 L(t, a_0) &= 0.
 \end{aligned}$$

where t_0 is the initial time, a_0 is the youngest age of the target population, $B(t)$ is the number of population members aged a_0 at time t , and $M_0(a)$, $S_0(a)$, $E_0(a)$, and $L_0(a)$ are the initial age distributions of the population. To fit the continuous model to discrete data, the abovementioned model was rewritten as ordinary differential equations, the details of which are available in the Appendix.

2.3. Statistical Estimation

We quantified the system (1) without imposing excessively unnecessary assumptions. Only parameters k_S , k_E , q_S , and q_E , i.e., the hazard ratios of lung cancer among smokers and ex-smokers and excess hazard ratios of mortality for reasons other than lung cancer among smokers

and ex-smokers, respectively, were extracted from the literature and fixed: $k_S = 4.94$, $k_E = 2.20$, $q_S = 0.49$, $q_E = 0.20$ for men and $k_S = 4.25$, $k_E = 2.19$, $q_S = 0.51$, $q_E = 0.46$ for women [14]. Other published studies from Japan yielded similar estimates [15–19]. All remaining rates were estimated via the following fitting procedure. As mentioned above, we used five different datasets, including the (i) incidence, (ii) mortality of lung cancer, (iii) smoking prevalence, (iv) population size, and (v) mortality other than lung cancer. Smoking prevalence was assumed to have been the result of binomial sampling, while other data-generating processes were each approximated as Poisson processes (see the Appendix for a detailed description of the likelihood function). We employed the maximum likelihood method to estimate parameters and derived the 95% confidence intervals (Cis) of model estimates (e.g., intervals of projected numbers of patients with lung cancer) using a parametric bootstrap method. During the bootstrapping exercise, model parameters were resampled from a multivariate normal distribution with mean $\bar{\theta}$ and variance-covariance matrix Σ , where $\bar{\theta}$ is the estimate obtained by maximum likelihood estimation and Σ is the inverse of the Hessian matrix of $-\log L(\theta)$ at $\theta = \bar{\theta}$. We obtained the model solution for each set of parameters and performed 1000 parameter resampling simulations to obtain a distribution for each model solution. We then computed the 95% CI by taking the 2.5th and 97.5th percentiles of the simulated distribution.

2.4. Ethical Considerations

The present study analyzed data that are publicly available. As such, the datasets used in our study were de-identified and fully anonymized in advance, and the analysis of publicly available data without identity information does not require ethical approval.

2.5. Data Sharing Policy

The major data source (i.e., the datasets on lung cancer incidence and mortality) is available in references [1,2] via the linked URLs.

3. Results

Figure 3 compares the observed and predicted incidence and mortality rates of lung cancer per 100,000 population along with forecasts for 2015–2050. Incidence and mortality were forecasted jointly by age and sex. Observed patterns were qualitatively captured very well for all the observed rates, indicating that both incidence and mortality rates per 100,000 population for the entire population as well as by gender will continue to steadily grow overall (also, see Appendix Figure A1). However, the incidence and mortality of older men aged 75 years or older are expected to have already peaked by 2015 (Figure 3A,C) and are predicted to decrease over time, whereas those among women will steadily increase over time in the future (Figure 3B,D). The predicted incidence and mortality among men aged 65–74 years will slightly decrease with time in men and will probably peak around 2045 in women. Long-term decline in the incidence and mortality in men's age groups other than those aged 75 years or older is smaller, because there is ageing within age groups, and thus, the risk of lung cancer is elevated over time within the same age group.

When the absolute numbers of lung cancer cases and deaths are examined by age group and sex (Figure 4), the predicted patterns are more counterintuitive than those of the rates per 100,000 population. The number of new lung cancer cases will start to wane starting in the mid-2020s in men and around 2035 in women (Figure 4A,B), and these overall patterns also hold for the predicted numbers of lung cancer deaths (Figure 4C,D). The maximum numbers of new cases and deaths due to lung cancer in men will be 76,978 (95% CI: 76,630–77,253) and 63,284 (95% CI: 62,991–63,507) in 2024, while those in women will be 42,838 (95% CI: 42,601–43,095) and 32,267 (95% CI: 32,063–32,460) in 2035 and 2036, respectively. The absolute numbers of lung cancer cases and deaths among older adults aged 65–74 years and 75 years or older will both fluctuate because of the underlying demographic dynamics (i.e., the visible impact of the baby boom in Japan), and for both men and women, the predicted numbers of cases and deaths will have mostly plateaued by 2030. Reflecting a decreased

population and its overall ageing, the numbers of lung cancer cases and deaths among those aged 15–64 years will steadily decrease among both men and women (see Appendix Figure A2).

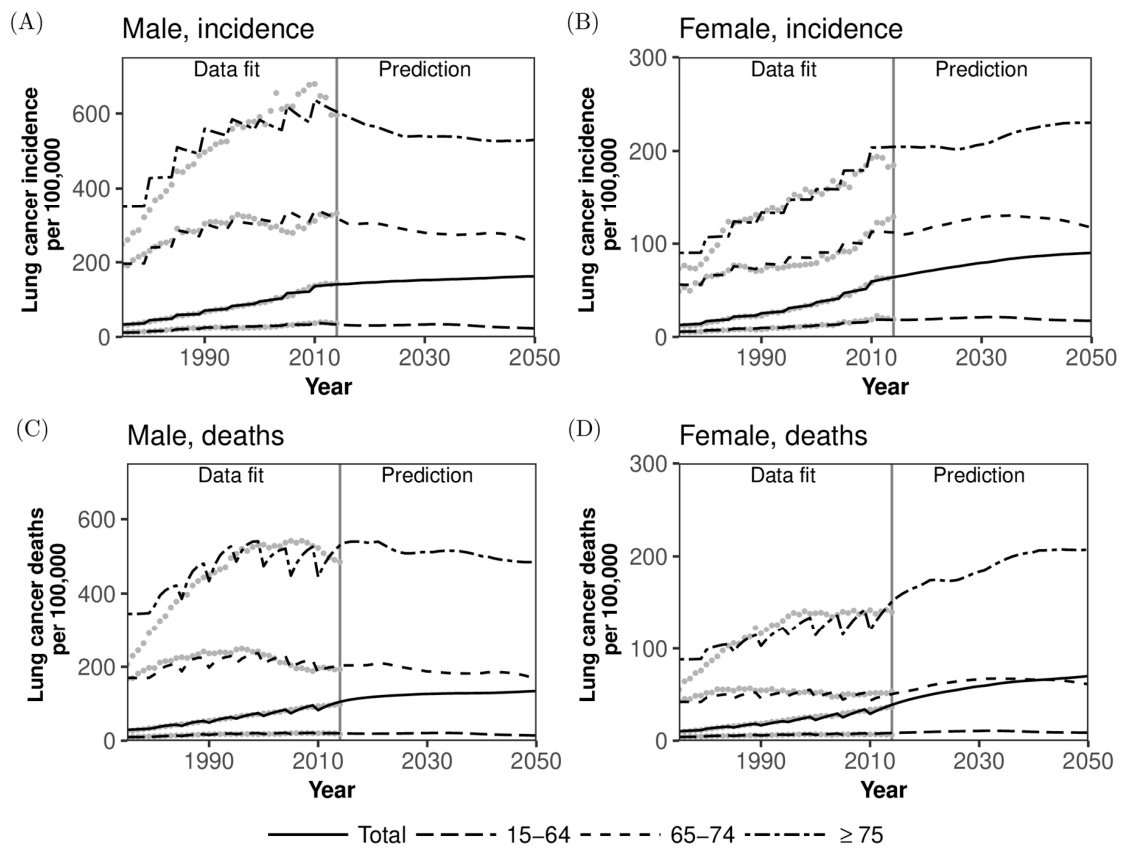


Figure 3. Comparison between observed and predicted lung cancer incidence and mortality per 100,000 population in Japan. Comparisons between observed and predicted lung cancer incidence and mortality per 100,000 population are made by age group and sex. (A) Lung cancer incidence in men, (B) Lung cancer incidence in women, (C) Mortality in men, and (D) Mortality in women. Bold dashed line represents working-age population aged 15–64 years, thin dashed line corresponds to those aged 65–74 years, uneven dashed line represents those aged 75 years or older, and solid line represents the entire population. Grey dots before 2014 were used to estimate model fitting parameters via the maximum likelihood method, and lines for 2015 and later are predictions calculated using maximum likelihood estimates at the latest time. The 95% confidence intervals are omitted from predictions because of the narrow uncertainty bounds.

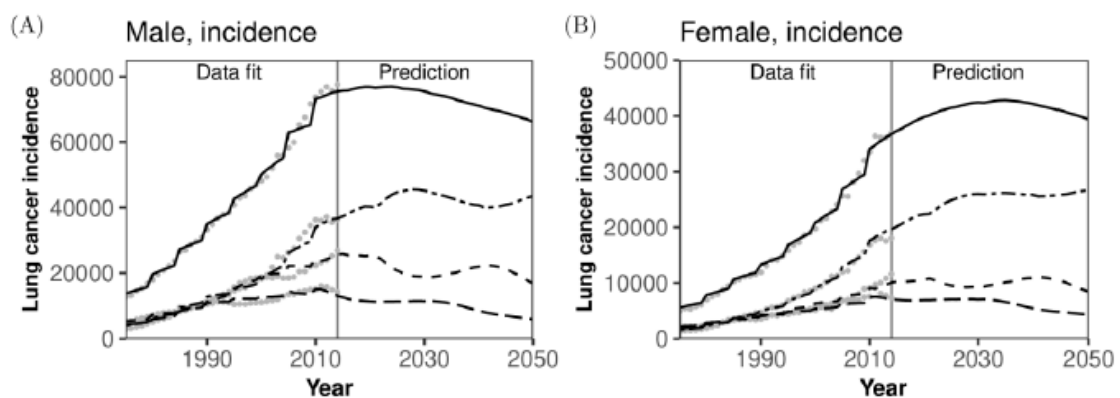


Figure 4. Cont.

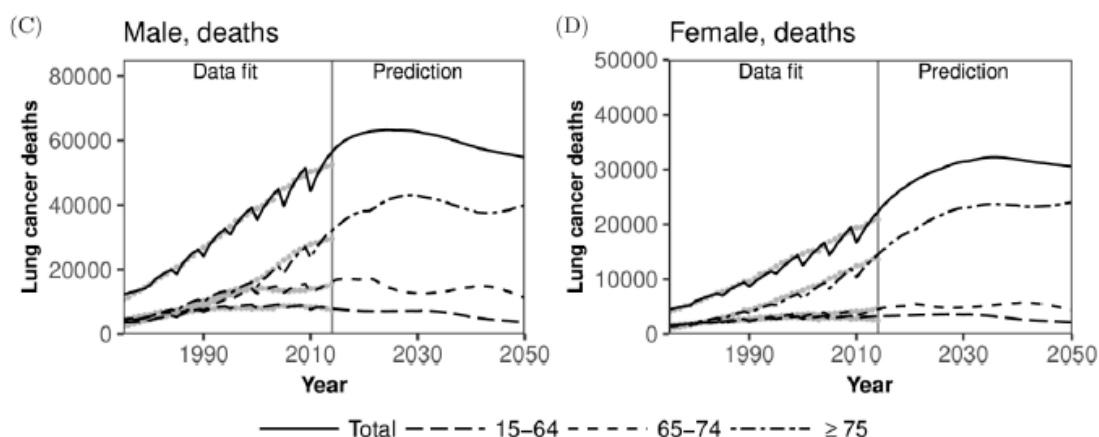


Figure 4. Comparisons between observed and predicted absolute numbers of lung cancer cases and deaths in Japan. Comparisons between observed and predicted absolute numbers of lung cancer cases and deaths are shown by age group and sex. (A) Lung cancer incidence in men, (B) Lung cancer incidence in women, (C) Mortality in men, and (D) Mortality in women. Bold dashed line represents working-age population aged 15–64 years, thin dashed line corresponds to those aged 65–74 years, uneven dashed line represents those aged 75 years or older, and solid line represents the entire population. Grey dots before 2014 were used to estimate model fitting parameters via the maximum likelihood method, and lines for 2015 and later indicate predictions calculated using maximum likelihood estimates at the latest time. The 95% confidence intervals are omitted from predictions because of the narrow uncertainty bounds.

Figure 5 shows the age-dependent dynamics and forecast of smoking prevalence. Smoking prevalence has drastically decreased in men over time, and the declining trend in the working age population is expected to continue until around 2030 (Figure 5A). Because of Japan’s quickening rate of demographic decline, the decrease in smoking prevalence is expected to plateau among those aged 15–64 years. The prevalence in women is lower than that of men, and women’s overall prevalence of smoking should eventually be close to or slightly lower than that of men (Figure 5B). As time passes, in terms of prevalence, ex-smokers will increasingly dominate the older adult population (Figure 5C,D). Especially, among men, more than 50% were predicted to be ex-smokers by 2015, and this increase will continue even after 2030. A similar increasing trend of ex-smokers is also expected among women by 2030.

Because the hazard ratio k_s of lung cancer among smokers compared with never-smokers is one of the most important but uncertain quantities that could influence our forecast, sensitivity analysis of the predicted values was conducted by varying the relative value of k_s (Figure 6). When we varied the hazard ratio from 0.5 to 1.5 times the original value, both lung cancer incidence and mortality per 100,000 population changed only slightly. Among men, the incidence and mortality increased with 1.5 times the original k_s , but we did not observe any abrupt decline by employing 0.5 times k_s , perhaps because of the greater impact of smokers and ex-smokers on the demographic dynamics of the population of patients with lung cancer (Figure 6A). Among women, a monotonic increase in both incidence and mortality was observed when the multiplier of the hazard ratio was increased from 0.5 to 1.5 (Figure 6B).

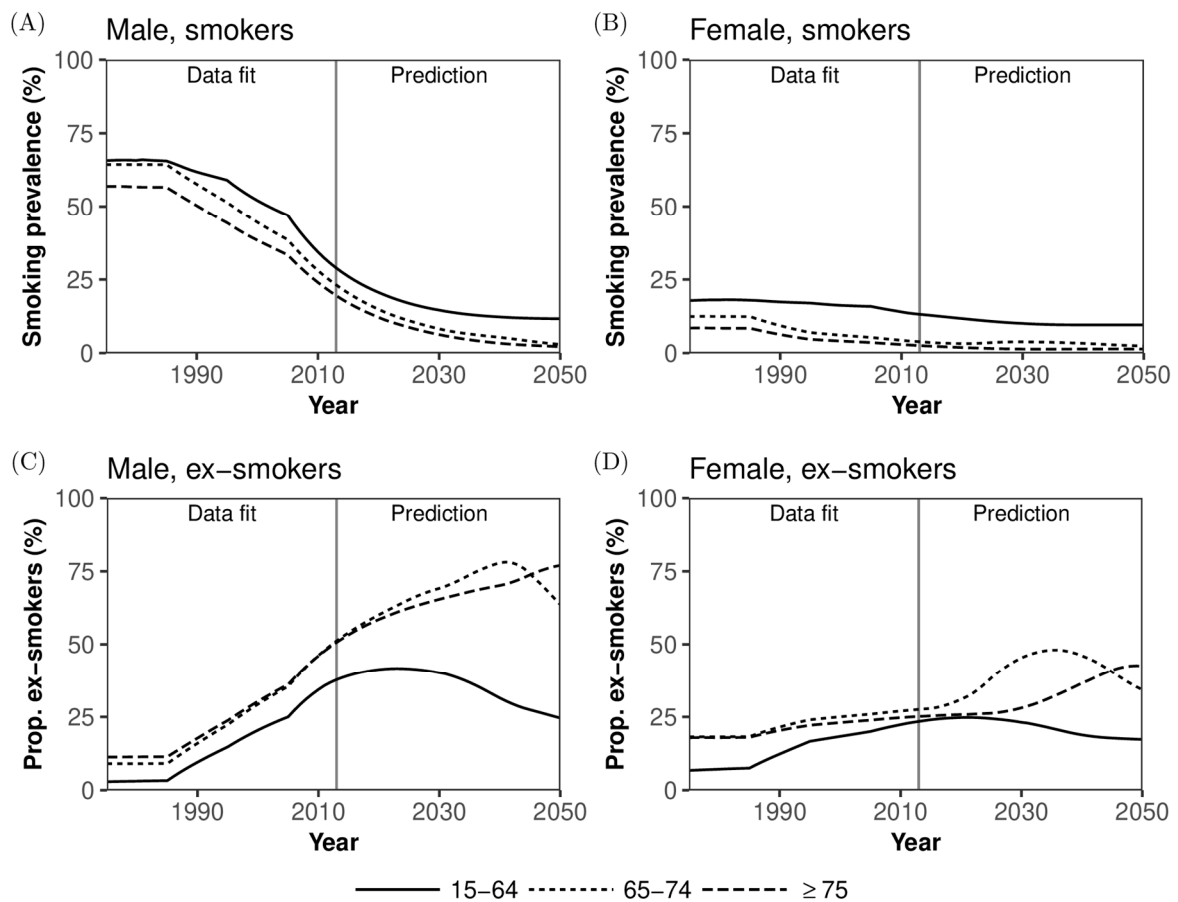


Figure 5. Smoking prevalence in Japan. Comparisons between observed and predicted smoking prevalence in Japan. (A) Smoking prevalence in men, (B) Smoking prevalence in women, (C) Ex-smokers’ prevalence in men, and (D) Ex-smokers’ prevalence in women. Solid line represents the estimated prevalence among those aged 15–64 years, dotted line represents the prevalence among those aged 65–74 years, and dashed line shows the estimate among people aged 75 years or older. Lines for 2015 and later indicate predictions calculated using maximum likelihood estimates at the latest time.

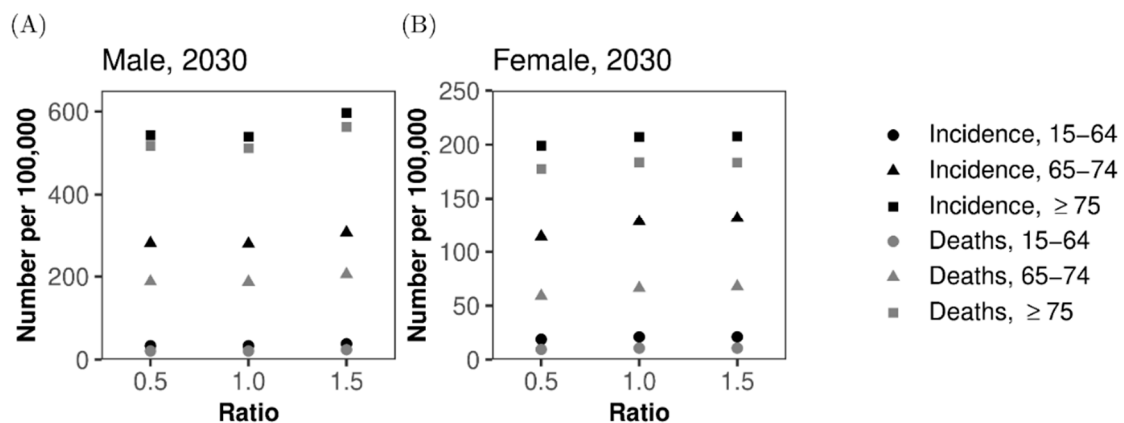


Figure 6. Sensitivity of lung cancer incidence and mortality to the relative hazard ratio among smokers. Predicted incidence and mortality of lung cancer are examined according to the different relative hazard ratios of lung cancer among smokers. (A) Incidence and mortality in men and (B) Incidence and mortality in women. Black marks show the incidence, while grey marks represent mortality. Circles, triangles, and squares represent those aged 15–64 years, 65–74 years, and 75 years or older, respectively.

4. Discussion

The present study parameterized a mathematical model that governs the demographic dynamics and population dynamics of smokers, allowing us to examine the time-dependent epidemiological dynamics of lung cancer in the future of Japan. The model was parameterized using data from 2014 and before as learning data and offered future predictions of lung cancer incidence and deaths, showing that the absolute numbers of new cases and deaths from lung cancer will increase until the mid-2020s in men and mid-2030s in women, respectively. Afterwards, the absolute numbers of cases and deaths were predicted to decrease monotonically, reflecting the decline in smoking prevalence and future population decline in Japan [20]. However, the incidence and mortality predictions per 100,000 persons continued to increase slightly, contradicting the overall decreasing trends in Western countries. The predicted burden of lung cancer is likely useful for anticipating future demands for diagnosis, treatment, and care in the healthcare sector. To our knowledge, except for individual risk-based prediction using cohort epidemiological data [21], the present study is the first to have offered an explicit prediction of lung cancer rates at a population level in Japan.

As an important lesson, we have successfully quantified the impact of decline in smoking rate and decreasing population on the epidemiology of lung cancer in an objective manner. Forecasted absolute numbers of cases and deaths are critical to plan future healthcare services, including estimations of required medical services (e.g., demand for screening) and hospital beds at facilities. In particular, we have shown that the dramatic increase in demand will peak in the mid-2030s among women, and subsequently, the demand is expected to wane drastically because of a decreasing population at high risk of lung cancer. Rather than the absolute value, the overall social burden of lung cancer might be better measured by the rate per 100,000 population. It is critical that the rate is expected to increase steadily over time, although breakthroughs in prevention or treatment of lung cancer might occur in the future. Our finding contradicts the ongoing epidemiological trends in other industrialized countries where both the incidence and mortality are in declining trend: the impact of rapid ageing has been very influential in Japan. The good news is that the absolute numbers of cases and deaths are expected to start to wane from the 2030s, and Japan may have already experienced its peak incidence rate of lung cancer in older men. Different trends between the absolute number and the estimate per 100,000 indicate a critical importance to account for the demographic dynamics (e.g., a flat or slight increase in the incidence among elderly aged 75 years or older in Figure 3C).

In addition to the value of overall prediction of epidemiological trends and medical demand related to lung cancer, the use of the proposed mathematical model has two remarkable advantages. First, our model appropriately offers predictions by age and sex, and thus would be useful for anticipating demands for diagnosis, treatment, and care. For instance, one of our long-term forecasts indicates that the absolute numbers of cases and deaths among those aged 75 years or older will increase again starting in the 2040s, while those among the working-age population will continue to decrease steadily (Figure 4). This is alarming for the treatment of patients over the next 30 years. Such fluctuating patterns were appropriately captured because our predictions were firmly supported by the tightly predicted demographic dynamics, including ageing, and this feature is well suited to the Japanese setting with its fast demographic changes over time. Second, the present study successfully obtained future trends of smoking prevalence using a simplistic structure of the natural history of smoking. Reflecting the trend of smoking prevalence, the decline in absolute number of cases in women was shown to be smaller than that in men (Figure 4). Moreover, smoking prevalence was firmly predicted to decrease monotonically in the future. However, the prevalence of ex-smokers will continue to grow, and this will be a remarkable characteristic among older adults in the future, putting a substantial number of people at risk of lung cancer.

By employing the proposed modeling scheme, public health insights into the effectiveness of anti-smoking measures could potentially be gained while appropriately accounting for the underlying demographic dynamics and smoking prevalence. Published modeling methods that employ statistical modeling techniques, including survival analysis and individual-based (micro)simulations, can also

be used for similar assessments [8,9,22–46], but manual control of the population at risk is required in these simplistic approaches. Our proposed model does not have to involve such manual input of the population size at risk because we explicitly model the demographic dynamics and examine the future course of smoking prevalence within our model directly when predicting the numbers of lung cancer cases [25,26,28,31,32,45]. Another direction of possible application would be to explore the effectiveness of additional screening, including regular or non-regular screening with chest X-rays or computed tomography, which is our area of ongoing future study [27,29–31,37,42,44].

While our model appears to be practically useful, the proposed modeling study involved a number of limitations that must be described here. First, our model does not account for future improvement in prevention and treatment. Presently, lung cancer remains one of the leading causes of death (i.e., it has a severe prognosis), but gene therapy for lung cancer and other novel approaches could drastically change case outcomes in the future. Second, risk factors other than smoking were not explicitly considered. Genetic predisposition and well-known carcinogenic exposures could be used to attain earlier diagnosis and improve prognosis [47,48]. In fact, the relative risk of smoking in causing lung cancer among Japanese is known as smaller than those in the United States and Europe, referred to as the smoking paradox [49]. Third, we did not have substantial datasets to incorporate the impact of smoking dose (e.g., pack-years) into our model, although the dose is known to be influential in clinical settings. Fourth, we ignored the impact of secondhand smoking on lung cancer, while it has been documented that secondhand exposure in households acts as a cause [50]. Incorporating secondhand smoking might require modeling and data at greater spatial precision (e.g., household-level data). Fifth, to obtain future forecasts, we have used the projected population in Japan and ignored the mortality rate and other demographic parameters [51]. If a drastic change happens in other diseases (e.g., certain types of cancers become not lethal at all someday) [52], the epidemiological structure of mortality would change and mortality due to lung cancer would also be influenced. Lastly, it has been pointed out that early initiation of smoking has become more and more common in Japan [20]. Although our model accounted for time-dependent hazard rate to start smoking, the future prediction had to assume a time-dependent hazard which was derived from the most recent observed time [53–55].

Despite these limitations, we believe that the present study successfully captured the demographic and population dynamics of smokers in Japan, yielding appropriate estimates and forecasts of lung cancer. We are open to criticisms or comments for future collaborations, hoping to expand the proposed modeling schemes to other types of cancer and extend the lung cancer model to scenario analysis of interventions.

5. Conclusions

The present study parameterized a mathematical model that governs the demographic and population dynamics of smokers, predicting the epidemiological dynamics of lung cancer in Japan. The model was parameterized using data from 2014 and before as learning data, and offered future predictions of lung cancer incidence and deaths, showing that the absolute numbers of new lung cancer cases and deaths will increase until the mid-2020s in men and the mid-2030s in women. The model can potentially be further applied to elucidate the effectiveness of anti-smoking measures and screening procedures at reducing lung cancer mortality.

Author Contributions: H.N. conceived the study. T.Y. and H.N. jointly conceptualized the study design, collected the data, and formulated the mathematical model. T.Y. performed statistical analyses, and H.N. gave advices. H.N. drafted the early version of the manuscript, and T.Y. and H.N. edited the early version. T.Y. drafted figures. All authors gave comments on the revised manuscript and approved the final version of the manuscript.

Funding: The present study was primarily supported by the Young Investigator Program, which is supported by the Smoking Research Foundation and the Japan Society for the Promotion of Science KAKENHI (18H04895). HN received funding support from the Japan Agency for Medical Research and Development (JP18fk0108050); the Japan Society for the Promotion of Science KAKENHI (Grant Numbers 16KT0130, 17H04701 and 17H05808); the Inamori Foundation, the Telecommunication Advancement Foundation; and the Japan Science and Technology Agency (JST) CREST program (JPMJCR1413). The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Acknowledgments: We thank Richard Lipkin, from Edanz Group (www.edanzediting.com/ac) for editing a draft of this manuscript.

Conflicts of Interest: H.N. is an editorial board member of *Journal of Clinical Medicine*. The authors report no other conflicts of interest.

List of Abbreviations: CI: Confidence interval.

Availability of Data and Materials: Collected cancer datasets are publicly available and can be retrieved from references [1,2].

Appendix A Derivation of Likelihood Function

To transform these partial differential equations to ordinary differential equations, we define

$$\begin{aligned} \overline{M}(t) &= M(t + t_0, t + a_0), \\ \overline{S}(t) &= S(t + t_0, t + a_0), \\ \overline{E}(t) &= E(t + t_0, t + a_0), \\ \overline{L}(t) &= L(t + t_0, t + a_0), \end{aligned}$$

where t_0 and a_0 are some fixed time and age. Then, we derive the following ordinary differential equations:

$$\begin{aligned} \frac{d\overline{M}}{dt}(t) &= -(\mu(t + t_0, t + a_0) + \sigma(t + t_0, t + a_0) + \lambda(t + t_0, t + a_0))\overline{M}(t), \\ \frac{d\overline{S}}{dt}(t) &= \sigma(t + t_0, t + a_0)\overline{M}(t) - ((1 + q_S)\mu(t + t_0, t + a_0) + \delta(t + t_0, t + a_0) + k_S\lambda(t + t_0, t + a_0))\overline{S}(t), \\ \frac{d\overline{E}}{dt}(t) &= \delta(t + t_0, t + a_0)\overline{S}(t) - ((1 + q_E)\mu(t + t_0, t + a_0) + k_E\lambda(t + t_0, t + a_0))\overline{E}(t), \\ \frac{d\overline{L}}{dt}(t) &= \lambda(t + t_0, t + a_0)(\overline{M}(t) + k_S\overline{S}(t) + k_E\overline{E}(t)) - (\mu(t + t_0, t + a_0) + \nu(t + t_0, t + a_0))\overline{E}(t). \end{aligned}$$

These ordinary differential equations describe the time evolution of a number of populations aged a_0 at t_0 .

Transition rates depending on t and a are assumed to be products of a step function of a and a step function of t . Splitting intervals of a and t to subintervals whose lengths are I_λ^A and I_λ^T for the transition rate $\lambda(t, a)$, we consider parameters λ_i^T for $i \in \mathbb{Z}$ and λ_j^A for $j \in \mathbb{Z}$ so that

$$\lambda(t, a) = \lambda_i^T \lambda_j^A,$$

when $t \in [t_0 + I_\lambda^T(i - 1), t_0 + I_\lambda^T i)$ and $a \in [a_0 + I_\lambda^A(j - 1), a_0 + I_\lambda^A j)$. Here, letting N_λ^T be a number of time intervals and N_λ^A be a number of age intervals, we also assume that $\lambda_i^T = \lambda_{N_\lambda^T}^T$ if $i > N_\lambda^T$ and $\lambda_j^A = \lambda_{N_\lambda^A}^A$ if $j > N_\lambda^A$.

Similarly, for interval lengths I_μ^T and I_μ^A and numbers of intervals N_μ^T and N_μ^A , we define parameters μ_i^T and μ_j^A for $i, j \in \mathbb{Z}$ so that

$$\mu(t, a) = \mu_i^T \mu_j^A,$$

when $t \in [t_0 + I_\mu^T(i - 1), t_0 + I_\mu^T i)$ and $a \in [a_0 + I_\mu^A(j - 1), a_0 + I_\mu^A j)$. Here, if $i > N_\mu^T$, then we let $\mu_i^T = \mu_{N_\mu^T}^T$, and if $j > N_\mu^A$, then we let $\mu_j^A = \mu_{N_\mu^A}^A$. For interval lengths I_ν^T and I_ν^A and numbers of intervals N_ν^T and N_ν^A , we define parameters ν_i^T and ν_j^A for $i, j \in \mathbb{Z}$ so that

$$\nu(t, a) = \nu_i^T \nu_j^A$$

when $t \in [t_0 + I_\nu^T(i - 1), t_0 + I_\nu^T i)$ and $a \in [a_0 + I_\nu^A(j - 1), a_0 + I_\nu^A j)$. Here, if $i > N_\nu^T$, then we let $\nu_i^T = \nu_{N_\nu^T}^T$, and if $j > N_\nu^A$, then we let $\nu_j^A = \nu_{N_\nu^A}^A$.

As for transition rates related to smoking status, we assume that people start smoking between ages 15 and 35 years and that people aged less than 20 years and those aged 20 years or older each start smoking at constant rates. We also assume that there is no smoking cessation among people aged

less than 20 years and that people aged 20 years or older stop smoking at a constant rate. Therefore, for a size of time interval I_s , $\sigma(t, a)$ and $\delta(t, a)$ are defined as

$$\sigma(t, a) = \begin{cases} \sigma^A \sigma_i & (a < 20) \\ \sigma_i & (20 \leq a < 35) \\ 0 & (a \geq 35) \end{cases}$$

and

$$\delta(t, a) = \begin{cases} 0 & (a < 20) \\ \delta_i & (a \geq 20) \end{cases}$$

when $t \in [t_0 + I_s i, t_0 + I_s(i + 1))$ for parameters σ^A , σ_i , and δ_i . Here, letting N_s be a number of time intervals, we define $\sigma_i = \sigma_{N_s}$ and $\delta_i = \delta_{N_s}$ if $i > N_s$.

Because of data availability, our model has the time range [1975, 2015] and the age range [15, 100]. We also let

$$\begin{aligned} I_u^v &= 5 \text{ for } u = \lambda, \mu, \nu \text{ and } v = A, T \\ N_u^T &= 8 \text{ for } u = \lambda, \mu, \nu \\ N_\lambda^A &= N_\nu^A = 15, N_\mu^T = 18, I_s = 10, N_s = 4. \end{aligned}$$

The ordinary differential equations are solved numerically by the fourth-order Runge-Kutta method. For $a_0 = 15$ and $t_0 = 1975, \dots, 2014$, the initial values are given by $\bar{M}(0) = B(t)$ and $\bar{S}(0) = \bar{E}(0) = \bar{L}(0) = 0$. For $a_0 = 16, \dots, 100$ and $t = 1975$, we need to determine the initial values $\bar{M}(0) = M_0(a_0)$, $\bar{S}(0) = S_0(a_0)$, $\bar{E}(0) = E_0(a_0)$, and $\bar{L}(0) = L_0(a_0)$. We assume that these initial values are derived from the situation where parameters at $t = 1975$ do not change for a long period. We consider the equations

$$\begin{aligned} \frac{d\tilde{M}}{da}(a) &= -(\mu(1975, a) + \sigma(1975, a) + \lambda(1975, a))\tilde{M}(a), \\ \frac{d\tilde{S}}{da}(a) &= \sigma(1975, a)\tilde{M}(a) - ((1 + q_S)\mu(1975, a) + \delta(1975, a) + k_S\lambda(1975, a))\tilde{S}(a), \\ \frac{d\tilde{E}}{da}(a) &= \delta(1975, a)\tilde{S}(a) - ((1 + q_E)\mu(1975, a) + k_E\lambda(1975, a))\tilde{E}(a) \\ \frac{d\tilde{L}}{da}(a) &= \lambda(1975, a)(\tilde{M}(a) + k_S\tilde{S}(a) + k_E\tilde{E}(a)) - (\mu(1975, a) + \nu(1975, a))\tilde{E}(a) \end{aligned}$$

with initial values $\tilde{M}(15) = 1$ and $\tilde{S}(15) = \tilde{E}(15) = \tilde{L}(15) = 0$. Let $n_{t,a}$ be the actual number of members of the population aged a at year t . From the solutions $\tilde{M}(a)$, $\tilde{S}(a)$, $\tilde{E}(a)$, and $\tilde{L}(a)$, we let

$$\begin{aligned} L_0(a) &= \frac{n_{1975,a}\tilde{L}(a)}{\tilde{M}(a) + \tilde{S}(a) + \tilde{E}(a) + \tilde{L}(a)} \\ M_0(a) &= \frac{(n_{1975,a} - L_0(a))\tilde{M}(a)}{\tilde{M}(a) + \tilde{S}(a) + \tilde{E}(a)} \\ S_0(a) &= \frac{(n_{1975,a} - L_0(a))\tilde{S}(a)}{\tilde{M}(a) + \tilde{S}(a) + \tilde{E}(a)} \\ E_0(a) &= \frac{(n_{1975,a} - L_0(a))\tilde{E}(a)}{\tilde{M}(a) + \tilde{S}(a) + \tilde{E}(a)}. \end{aligned}$$

We define the vector of unknown parameters

$$\theta = (\lambda_1^T, \dots, \lambda_{N_\lambda^T}^T, \lambda_1^A, \dots, \lambda_{N_\lambda^A}^A, \nu_1^T, \dots, \nu_{N_\nu^T}^T, \nu_1^A, \dots, \nu_{N_\nu^A}^A, \mu_1^T, \dots, \mu_{N_\mu^T}^T, \mu_1^A, \dots, \mu_{N_\mu^A}^A, \sigma^A, \sigma_1, \dots, \sigma_{N_s}, \delta_1, \dots, \delta_{N_s}),$$

which is estimated by maximum likelihood estimation.

To define the likelihood function, smoking prevalence is assumed to follow a binomial distribution. We assume that the total number of samples at each year is 10,000, and the number of samples for each age group is calculated by dividing the total number by the population of the age

group. The probabilities of the binomial distribution are calculated by our model. We assume that numbers of members of the population, deaths, lung cancer cases, and lung cancer deaths follow a Poisson distribution.

Let $N(t, a) = M(t, a) + S(t, a) + E(t, a) + L(t, a)$, which is the number of members of the population obtained from our model. The likelihood function for the number of members of the population is defined as follows.

$$l_1(\theta, t) = \prod_{a=15}^{100} \frac{N(t, a)^{n_{t,a}} e^{-N(t,a)}}{(n_{t,a})!}.$$

Note that $N(t, a)$ depends on θ . Let $P(t, a_1, a_2)$ be the smoking prevalence aged from a_1 to a_2 at year t calculated by our model, that is,

$$P(t, a_1, a_2) = \frac{\sum_{a=a_1}^{a_2} S(t, a)}{\sum_{a=a_1}^{a_2} (M(t, a) + S(t, a) + E(t, a))}$$

$L(t, a)$ is excluded from the calculation of $P(t, a_1, a_2)$ because $L(t, a)$ does not contain information about smoking status. The likelihood function for smoking prevalence is defined by

$$l_2(\theta, t) = \left(\frac{r_{t,60,100}}{q_{t,60,100}}\right) P(t, 60, 100)^{q_{t,60,100}} (1 - P(t, 60, 100))^{r_{t,60,100} - q_{t,60,100}} \\ \times \prod_{a \in \{20+10i \mid i=0, \dots, 3\}} \left(\frac{r_{t,a,a+9}}{q_{t,a,a+9}}\right) P(t, a, a+9)^{q_{t,a,a+9}} (1 - P(t, a, a+9))^{r_{t,a,a+9} - q_{t,a,a+9}}$$

where q_{t,a_1,a_2} is number of smokers among r_{t,a_1,a_2} samples from the population aged from a_1 to a_2 at year t .

The expected values for the rest of data are obtained from additional compartments that signify cumulative numbers of deaths and lung cancer cases. For given values of t_s and a_0 , the cumulative numbers of lung cancer cases $U_{a_0,t_0}(t)$, the lung cancer deaths $V_{a_0,t_0}(t)$, and deaths for reasons other than lung cancer $W_{a_0,t_0}(t)$ corresponding to $\bar{M}(t)$, $\bar{S}(t)$, $\bar{E}(t)$, and $\bar{L}(t)$ are determined by

$$\frac{dU_{a_0,t_0}}{dt}(t) = \lambda(t + t_0, t + a_0) (\bar{M}(t) + k_S \bar{S}(t) + k_E \bar{E}(t)) \\ \frac{dV_{a_0,t_0}}{dt}(t) = \nu(t + t_0, t + a_0) \bar{E}(t) \\ \frac{dW_{a_0,t_0}}{dt}(t) = \mu(t + t_0, t + a_0) (\bar{M}(t) + (1 + q_S) \bar{S}(t) + (1 + q_E) \bar{E}(t) + \bar{L}(t)) \\ U(0) = 0, V(0) = 0, W(0) = 0.$$

Therefore, the expected numbers of lung cancer cases $\bar{U}(t, a)$, lung cancer deaths $\bar{V}(t, a)$, and overall number of deaths $\bar{W}(t, a)$ among the population aged a in year t are calculated by

$$\bar{U}(t, a) = U_{a_0,t_0}(t - t_0 + 1) - U_{a_0,t_0}(t - t_0) \\ \bar{V}(t, a) = V_{a_0,t_0}(t - t_0 + 1) - V_{a_0,t_0}(t - t_0) \\ \bar{W}(t, a) = W_{a_0,t_0}(t - t_0 + 1) + V_{a_0,t_0}(t - t_0 + 1) - W_{a_0,t_0}(t - t_0) - V_{a_0,t_0}(t - t_0),$$

where $a = a_0 + t - t_0$. We note that $\bar{U}(t, a)$, $\bar{V}(t, a)$, and $\bar{W}(t, a)$ depend on θ . We introduce the following notations:

$$\tilde{U}(t, a_1, a_2) = \sum_{a=a_1}^{a_2} \bar{U}(t, a) \\ \tilde{V}(t, a_1, a_2) = \sum_{a=a_1}^{a_2} \bar{V}(t, a) \\ \tilde{W}(t, a_1, a_2) = \sum_{a=a_1}^{a_2} \bar{W}(t, a).$$

Let u_{t,a_1,a_2} , v_{t,a_1,a_2} , and w_{t,a_1,a_2} be the actual numbers of lung cancer cases, lung cancer deaths, and total deaths for the population aged from a_1 to a_2 at year t . The likelihood function for these data is defined by

$$\begin{aligned}
 l_3(\theta, t) &= \frac{\tilde{U}(t,85,100)^{u_{t,85,100}} e^{-\tilde{U}(t,85,100)}}{(u_{t,85,100})!} \prod_{a \in \{15+5i \mid i=0, \dots, 14\}} \frac{\tilde{U}(t,a,a+4)^{u_{t,a,a+4}} e^{-\tilde{U}(t,a,a+4)}}{(u_{t,a,a+4})!} \\
 l_4(\theta, t) &= \frac{\tilde{V}(t,100,100)^{v_{t,100,100}} e^{-\tilde{V}(t,100,100)}}{(v_{t,100,100})!} \prod_{a \in \{15+5i \mid i=0, \dots, 14\}} \frac{\tilde{V}(t,a,a+4)^{v_{t,a,a+4}} e^{-\tilde{V}(t,a,a+4)}}{(v_{t,a,a+4})!} \\
 l_5(\theta, t) &= \frac{\tilde{W}(t,100,100)^{w_{t,100,100}} e^{-\tilde{W}(t,100,100)}}{(w_{t,100,100})!} \prod_{a \in \{15+5i \mid i=0, \dots, 14\}} \frac{\tilde{W}(t,a,a+4)^{w_{t,a,a+4}} e^{-\tilde{W}(t,a,a+4)}}{(w_{t,a,a+4})!}.
 \end{aligned}$$

Finally, we have the likelihood function for all data

$$l(\theta) = \prod_t \left(\prod_{i=1}^5 l_i(\theta, t) \right).$$

Appendix B Population Size Adjustment

1. The population data [5] contain 1-year age groups up to age 85 years from 1975 to 1979, up to age 90 years from 1980 to 2004, and up to age 100 years from 2005 onward. A census was held every 5 years, and the census data contain 1-year age groups up to age 100 years from 1975 onward. We interpolated the population data by census data as follows.
2. Let T and $T + 5$ be years in which the census was held and $p(t, a)$ be the population size at year t and age a for $t = T, T + 5$, and $a = 0, \dots, 100$. Note that $p(t, 100)$ is the population size aged over 100 years. Let $q(t, a)$ be the population size at year t and age a for $t = T + 1, \dots, T + 4$ in which the census was not held. We assume that $A < 100$ so that $q(t, a)$ are given for $a = 0, \dots, A$ and $q(t, A)$ is the population size aged over A .
3. The population size $p(t, a)$ for $t = T + 1, \dots, T + 4$ and $a = 0, \dots, A - 1$ is $q(t, a)$. Distributing $q(t, A)$ into 1-year age groups using $p(T, a)$ and $p(T + 5, a)$, the population size $p(t, a)$ for $t = T + 1, \dots, T + 4$ and $a = A, \dots, 100$ was calculated by

$$\begin{aligned}
 r(t, a) &= \left(1 - \frac{t-T}{5}\right) \frac{p(T,a)}{\sum_{a=0}^{100} p(T,a)} + \frac{t-T}{5} \frac{p(T+5,a)}{\sum_{a=0}^{100} p(T+5,a)} \\
 p(t, a) &= q(t, A) \frac{r(t,a)}{\sum_{a=A}^{100} r(t,a)}.
 \end{aligned}$$

4. We obtained the interpolated population size $p(t, a)$ for all ages up to 100 years. Let $d(t, a_1, a_2)$ be mortality at year t and age from a_1 to a_2 . $d(t, 100, \infty)$ is assumed to be the mortality of the population aged over 100 years. Mortality values are given by 5-year age groups, that is, $a_1 = 5n$ and $a_2 = 5(n + 1)$ for $n \in \mathbb{Z}$. We corrected the population size so that the difference between $p(t, a)$ and $p(t - 1, a - 1)$ is equal to mortality. $s(t, a)$ is defined as

$$\begin{aligned}
 s(t, a) &= p(t, a) - p(t - 1, a - 1) \quad (0 \leq a < 100) \\
 s(t, 100) &= p(t, 100) - p(t - 1, 99) - p(t - 1, 100).
 \end{aligned}$$

5. The difference $u(t, a_1, a_2)$ between mortality and $s(t, a)$ is calculated by

$$\begin{aligned}
 u(t, a_1, a_2) &= d(t, a_1, a_2) - \sum_{a=a_1}^{a_2} s(t, a) \quad (a_1 < 100) \\
 u(t, 100, \infty) &= d(t, 100, \infty) - s(t, 100).
 \end{aligned}$$

6. If $u(t, a_1, a_2) > 0$, that is, the decline in population is less than mortality, we increase the population size at year $t - 1$. Let $\tilde{p}(t, a)$ represent a corrected 1-year age group. Distributing $u(t, a_1, a_2)$ according to the proportions of population sizes, we have

$$\tilde{p}(t-1, a) = p(t-1, a) + u(t, a_1, a_2) \frac{p(t-1, a)}{\sum_{a=a_1-1}^{a_2-1} p(t-1, a)}$$

7. for $a_1 - 1 \leq a < a_2$. If $u(t, a_1, a_2) < 0$, that is, the decline in population is larger than mortality, we decrease the population size at year $t - 1$. Distributing $u(t, a_1, a_2)$ according to the proportions of $s(t, a)$, we have

$$\tilde{p}(t-1, a) = p(t-1, a) + u(t, a_1, a_2) \frac{s(t-1, a)}{\sum_{a=a_1-1}^{a_2-1} s(t-1, a)}$$

8. for $a_1 - 1 \leq a < a_2$. For $a = 99$ and 100 ,

$$\begin{aligned} \tilde{p}(t-1, a) &= p(t-1, a) + u(t, 100, \infty) \frac{p(t-1, a)}{\sum_{a=99}^{100} p(t-1, a)} \text{ if } u(t, a_1, a_2) > 0 \\ \tilde{p}(t-1, a) &= p(t-1, a) + u(t, 100, \infty) \frac{s(t-1, a)}{\sum_{a=99}^{100} s(t-1, a)} \text{ if } u(t, a_1, a_2) < 0. \end{aligned}$$

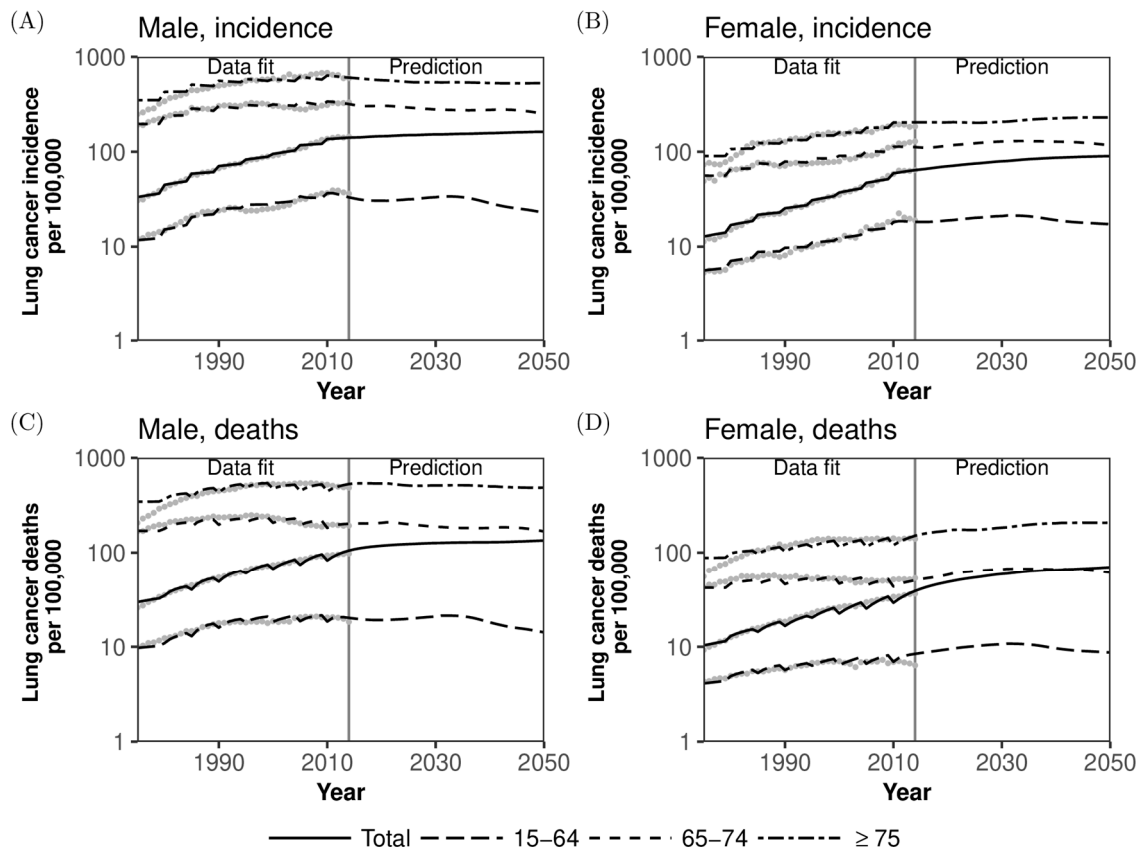


Figure A1. Comparison between observed and predicted lung cancer incidence and mortality per 100,000 population in Japan. Note that the vertical axis is in logarithmic scale. Comparisons between observed and predicted lung cancer incidence and mortality per 100,000 population are made by age group and sex. (A) Lung cancer incidence in men, (B) Lung cancer incidence in women, (C) Mortality in men, and (D) Mortality in women. Bold dashed line represents working-age population aged 15–64 years, thin dashed line corresponds to those aged 65–74 years, uneven dashed line represents those aged 75 years or older, and solid line represents the entire population. Grey dots before 2014 were used to estimate model fitting parameters via the maximum likelihood method, and lines for 2015 and later are predictions calculated using maximum likelihood estimates at the latest time. The 95% confidence intervals are omitted from predictions because of the narrow uncertainty bounds.

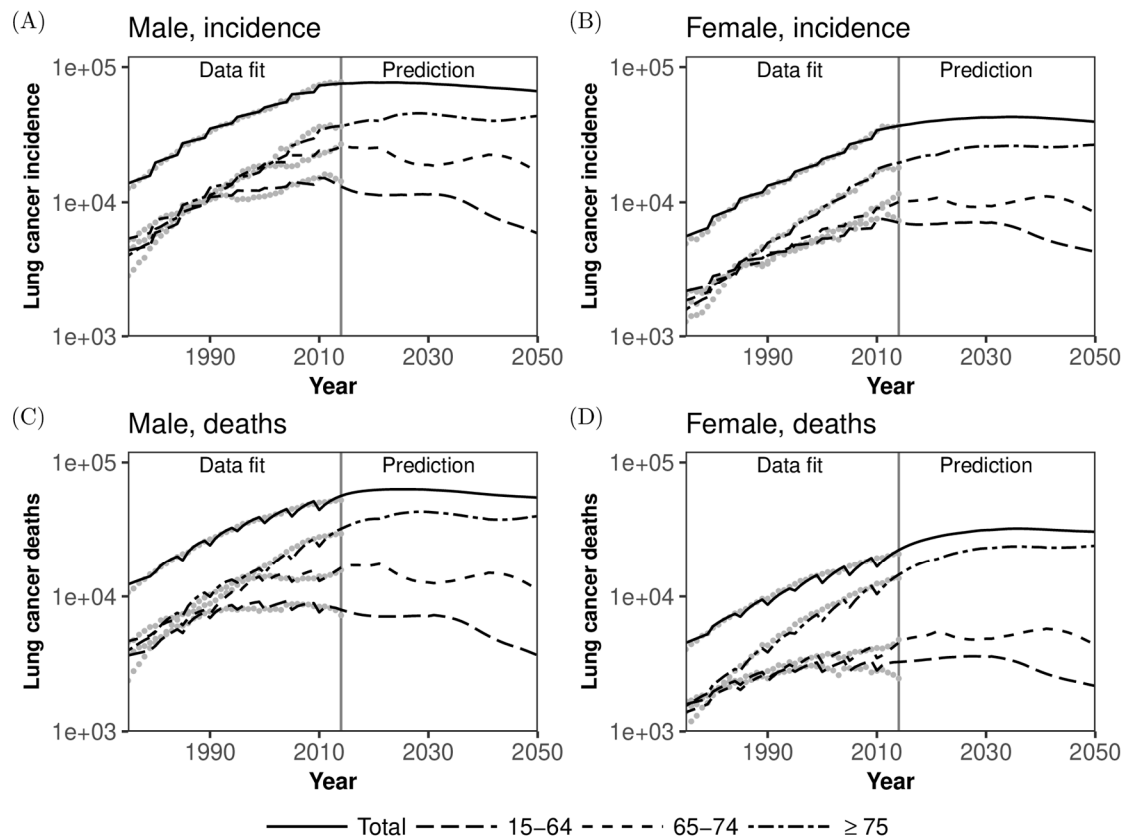


Figure A2. Comparisons between observed and predicted absolute numbers of lung cancer cases and deaths in Japan. Note that the vertical axis is in logarithmic scale. Comparisons between observed and predicted absolute numbers of lung cancer cases and deaths are shown by age group and sex. (A) Lung cancer incidence in men, (B) Lung cancer incidence in women, (C) Mortality in men, and (D) Mortality in women. Bold dashed line represents working-age population aged 15–64 years, thin dashed line corresponds to those aged 65–74 years, uneven dashed line represents those aged 75 years or older, and solid line represents the entire population. Grey dots before 2014 were used to estimate model fitting parameters via the maximum likelihood method, and lines for 2015 and later indicate predictions calculated using maximum likelihood estimates at the latest time. The 95% confidence intervals are omitted from predictions because of the narrow uncertainty bounds.

References

1. Cancer Information Service, National Cancer Center, Japan. Cancer Registry and Statistics. Available online: https://ganjoho.jp/en/professional/statistics/table_download.html (accessed on 27 December 2018).
2. Center for Cancer Control and Information Services, National Cancer Center. Monitoring of Cancer Incidence in Japan—Survival 2006–2008 Report. 2016. Available online: [https://ganjoho.jp/data/en/professional/statistics/files/cancer_survival\(1993-2008\)E.xls](https://ganjoho.jp/data/en/professional/statistics/files/cancer_survival(1993-2008)E.xls) (accessed on 27 December 2018).
3. Matsuda, T.; Ajiki, W.; Marugame, T.; Ioka, A.; Tsukuma, H.; Sobue, T.; Matsuda, T.; Shibata, A.; Mikami, H.; Okamoto, N.; et al. Population-based survival of cancer patients diagnosed between 1993 and 1999 in Japan: A chronological and international comparative study. *Jpn. J. Clin. Oncol.* **2011**, *41*, 40–51. [CrossRef] [PubMed]
4. Japan Tobacco, Inc. Nationwide Smoking Prevalence Survey. Available online: https://www.jti.co.jp/investors/library/press_releases/2018/pdf/20181213_01_appendix_01.pdf (accessed on 27 December 2018). (In Japanese)
5. Health, Labour and Welfare Statistics Association. Demographic Statistical Time Series Data 1899–2012. 2014. Available online: <https://www.mhlw.go.jp/toukei/list/81-1.html> (accessed on 27 December 2018). (In Japanese)

6. National Institute of Population and Social Security Research. *Population Projection for Japan: 2016–2065*; National Institute of Population and Social Security Research: Tokyo, Japan, 2017.
7. Loddenkemper, R.; Kreuter, M. (Eds.) *The Tobacco Epidemic*, 2nd ed.; Progress in Respiratory Research; Revised and Extended Edition; S. Karger AG: Basel, Switzerland, 2015; Volume 42.
8. McMahon, P.M.; Kong, C.Y.; Johnson, B.E.; Weinstein, M.C.; Weeks, J.C.; Tramontano, A.C.; Cipriano, L.E.; Bouzan, C.; Gazelle, G.S. Chapter 9: The MGH-HMS lung cancer policy model: Tobacco control versus screening. *Risk Anal.* **2012**, *32*, S117–S124. [[CrossRef](#)] [[PubMed](#)]
9. Levy, D.T.; Blackman, K.; Zaloshnja, E. Chapter 10: A macro-model of smoking and lung cancer: Examining aggregate trends in lung cancer rates using the CPS-I and CPS-II and two-stage clonal expansion models. *Risk Anal.* **2012**, *32*, S125–S141. [[CrossRef](#)] [[PubMed](#)]
10. Hori, M.; Matsuda, T.; Shibata, A.; Katanoda, K.; Sobue, T.; Nishimoto, H. Cancer incidence and incidence rates in Japan in 2009: A study of 32 population-based cancer registries for the Monitoring of Cancer Incidence in Japan (MCIJ) project. *Jpn. J. Clin. Oncol.* **2015**, *45*, 884–891. [[CrossRef](#)] [[PubMed](#)]
11. Health, Labour and Welfare Statistics Association. Demographic Statistical Time Series Data 2013–14. 2015. Available online: <https://www.mhlw.go.jp/toukei/list/81-1.html> (accessed on 27 December 2018). (In Japanese)
12. Sobue, T.; Yamamoto, S.; Hara, M.; Sasazuki, S.; Sasaki, S.; Tsugane, S.; JPHC study Group. Cigarette smoking and subsequent risk of lung cancer by histologic type in middle-aged Japanese men and women: The JPHC study. *Int. J. Cancer* **2002**, *99*, 245–251. [[CrossRef](#)] [[PubMed](#)]
13. Statistics Bureau, Ministry of Internal Affairs and Communications. Population Estimates. Available online: <https://www.stat.go.jp/data/jinsui/index.html> (accessed on 27 December 2018). (In Japanese)
14. Ozasa, K. Smoking and mortality in the Japan Collaborative Cohort Study for Evaluation of Cancer (JACC). *Asian Pac. J. Cancer Prev.* **2007**, *8*, 89–96. [[PubMed](#)]
15. Ando, M.; Wakai, K.; Seki, N.; Tamakoshi, A.; Suzuki, K.; Ito, Y.; Nishino, Y.; Kondo, T.; Watanabe, Y.; Ozasa, K.; et al. Attributable and absolute risk of lung cancer death by smoking status: Findings from the Japan collaborative cohort study. *Int. J. Cancer* **2003**, *105*, 249–254. [[CrossRef](#)] [[PubMed](#)]
16. Marugame, T.; Sobue, T.; Satoh, H.; Komatsu, S.; Nishino, Y.; Nakatsuka, H.; Nakayama, T.; Suzuki, T.; Takezaki, T.; Tajima, K.; et al. Lung cancer death rates by smoking status: Comparison of the three-prefecture cohort study in Japan to the Cancer Prevention Study II in the USA. *Cancer Sci.* **2005**, *96*, 120–126. [[CrossRef](#)] [[PubMed](#)]
17. Seki, T.; Nishino, Y.; Tanji, F.; Maemondo, M.; Takahashi, S.; Sato, I.; Kawai, M.; Minami, Y. Cigarette smoking and lung cancer risk according to histologic type in Japanese men and women. *Cancer Sci.* **2013**, *104*, 1515–1522. [[CrossRef](#)] [[PubMed](#)]
18. Wakai, K.; Inoue, M.; Mizoue, T.; Tanaka, K.; Tsuji, I.; Nagata, C.; Tsugane, S. Tobacco smoking and lung cancer risk: An evaluation based on a systematic review of epidemiological evidence among the Japanese population. *Jpn. J. Clin. Oncol.* **2006**, *36*, 309–324. [[CrossRef](#)]
19. Nakamura, K.; Ukawa, S.; Okada, E.; Hirata, M.; Nagai, A.; Yamagata, Z.; Ninomiya, T.; Muto, K.; Kiyohara, Y.; Matsuda, K.; et al. Characteristics and prognosis of Japanese male and female lung cancer patients: The BioBank Japan Project. *J. Epidemiol.* **2017**, *27*, S49–S57. [[CrossRef](#)] [[PubMed](#)]
20. Funatogawa, I.; Funatogawa, T.; Yano, E. Trends in smoking and lung cancer mortality in Japan, by birth cohort, 1949–2010. *Bull. World Health Organ.* **2013**, *91*, 332–340. [[CrossRef](#)] [[PubMed](#)]
21. Charvat, H.; Sasazuki, S.; Shimazu, T.; Budhathoki, S.; Inoue, M.; Iwasaki, M.; Sawada, N.; Yamaji, T.; Tsugane, S.; JPHC Study Group. Development of a risk prediction model for lung cancer: The Japan public health center-based prospective study. *Cancer Sci.* **2018**, *109*, 854–862. [[CrossRef](#)] [[PubMed](#)]
22. Clements, M.S.; Armstrong, B.K.; Moolgavkar, S.H. Lung cancer rate predictions using generalized additive models. *Biostatistics* **2005**, *6*, 576–589. [[CrossRef](#)] [[PubMed](#)]
23. Lubin, J.H.; Caporaso, N.; Wichmann, H.E.; Schaffrath-Rosario, A.; Alavanja, M.C.R. Cigarette smoking and lung cancer: Modeling effect modification of total exposure and intensity. *Epidemiology* **2007**, *18*, 639–648. [[CrossRef](#)] [[PubMed](#)]
24. Darby, S.C.; Pike, M.C. Lung cancer and passive smoking: Predicted effects from a mathematical model for cigarette smoking and lung cancer. *Br. J. Cancer* **1988**, *58*, 825–831. [[CrossRef](#)] [[PubMed](#)]

25. Chadeau-Hyam, M.; Tubert-Bitter, P.; Guihenneuc-Jouyaux, C.; Campanella, G.; Richardson, S.; Vermeulen, R.; De Iorio, M.; Galea, S.; Vineis, P. Dynamics of the risk of smoking-induced lung cancer: A compartmental hidden Markov model for longitudinal analysis. *Epidemiology* **2014**, *25*, 28–34. [[CrossRef](#)] [[PubMed](#)]
26. Jeon, J.; Holford, T.R.; Levy, D.T.; Feuer, E.J.; Cao, P.; Tam, J.; Clarke, L.; Clarke, J.; Kong, C.Y.; Meza, R. Smoking and lung cancer mortality in the United States from 2015 to 2065: A comparative modeling approach. *Ann. Intern. Med.* **2018**, *169*, 684–693. [[CrossRef](#)] [[PubMed](#)]
27. Han, S.S.; Erdogan, S.A.; Toumazis, I.; Leung, A.; Plevritis, S.K. Evaluating the impact of varied compliance to lung cancer screening recommendations using a microsimulation model. *Cancer Causes Control* **2017**, *28*, 947–958. [[CrossRef](#)] [[PubMed](#)]
28. Gauvreau, C.L.; Fitzgerald, N.R.; Memon, S.; Flanagan, W.M.; Nadeau, C.; Asakawa, K.; Garner, R.; Miller, A.B.; Evans, W.K.; Popadiuk, C.M.; et al. The OncoSim model: Development and use for better decision-making in Canadian cancer control. *Curr. Oncol.* **2017**, *24*, 401–406. [[CrossRef](#)] [[PubMed](#)]
29. Goffin, J.R.; Flanagan, W.M.; Miller, A.B.; Fitzgerald, N.R.; Memon, S.; Wolfson, M.C.; Evans, W.K. Biennial lung cancer screening in Canada with smoking cessation—outcomes and cost-effectiveness. *Lung Cancer* **2016**, *101*, 98–103. [[CrossRef](#)] [[PubMed](#)]
30. Ten Haaf, K.; Tammemägi, M.C.; Bondy, S.J.; van der Aalst, C.M.; Gu, S.; McGregor, S.E.; Nicholas, G.; de Koning, H.J.; Paszat, L.F. Performance and cost-effectiveness of computed tomography lung cancer screening scenarios in a population-based setting: A microsimulation modeling analysis in Ontario, Canada. *PLoS Med.* **2017**, *14*, 1–20. [[CrossRef](#)] [[PubMed](#)]
31. Meza, R.; ten Haaf, K.; Kong, C.Y.; Erdogan, A.; Black, W.C.; Tammemägi, M.C.; Choi, S.E.; Jeon, J.; Han, S.S.; Munshi, V.; et al. Comparative analysis of 5 lung cancer natural history and screening models that reproduce outcomes of the NLST and PLCO trials. *Cancer* **2014**, *120*, 1713–1724. [[CrossRef](#)] [[PubMed](#)]
32. Feuer, E.J.; Levy, D.T.; McCarthy, W.J. Chapter 1: The impact of the reduction in tobacco smoking on U.S. lung cancer mortality, 1975–2000: An introduction to the problem. *Risk Anal.* **2012**, *32*, S6–S13. [[CrossRef](#)] [[PubMed](#)]
33. Anderson, C.M.; Burns, D.M.; Dodd, K.W.; Feuer, E.J. Chapter 2: Birth-cohort-specific estimates of smoking behaviors for the U.S. population. *Risk Anal.* **2012**, *32*, S14–S24. [[CrossRef](#)] [[PubMed](#)]
34. Rosenberg, M.A.; Feuer, E.J.; Yu, B.; Sun, J.; Henley, S.J.; Shanks, T.G.; Anderson, C.M.; McMahon, P.M.; Thun, M.J.; Burns, D.M. Chapter 3: Cohort life tables by smoking status, removing lung cancer as a cause of death. *Risk Anal.* **2012**, *32*, S25–S38. [[CrossRef](#)] [[PubMed](#)]
35. Holford, T.R.; Clark, L. Chapter 4: Development of the counterfactual smoking histories used to assess the effects of tobacco control. *Risk Anal.* **2012**, *32*, S39–S50. [[CrossRef](#)] [[PubMed](#)]
36. Jeon, J.; Meza, R.; Krapcho, M.; Clarke, L.D.; Byrne, J.; Levy, D.T. Chapter 5: Actual and counterfactual smoking prevalence rates in the U.S. Population via microsimulation. *Risk Anal.* **2012**, *32*, S51–S68. [[CrossRef](#)] [[PubMed](#)]
37. Schultz, F.W.; Boer, R.; de Koning, H.J. Chapter 7: Description of MISCAN-Lung, the Erasmus MC lung cancer microsimulation model for evaluating cancer control interventions. *Risk Anal.* **2012**, *32*, S85–S98. [[CrossRef](#)] [[PubMed](#)]
38. Hazelton, W.D.; Jeon, J.; Meza, R.; Moolgavkar, S.H. Chapter 8: The FHCRC lung cancer model. *Risk Anal.* **2012**, *32*, S99–S116. [[CrossRef](#)] [[PubMed](#)]
39. Foy, M.; Deng, L.; Spitz, M.; Gorlova, O.; Kimmel, M. Chapter 11: Rice-MD Anderson lung cancer model. *Risk Anal.* **2012**, *32*, S142–S150. [[CrossRef](#)] [[PubMed](#)]
40. Holford, T.R.; Ebisu, K.; McKay, L.; Oh, C.; Zheng, T. Chapter 12: Yale lung cancer model. *Risk Anal.* **2012**, *32*, S151–S165. [[CrossRef](#)] [[PubMed](#)]
41. Holford, T.R.; Levy, D.T. Chapter 14: Comparing the adequacy of carcinogenesis models in estimating U.S. population rates for lung cancer mortality. *Risk Anal.* **2012**, *32*, S179–S189. [[CrossRef](#)] [[PubMed](#)]
42. Boer, R.; Moolgavkar, S.H.; Levy, D.T. Chapter 15: Impact of tobacco control on lung cancer mortality in the United States over the period 1975–2000—Summary and limitations. *Risk Anal.* **2012**, *32*, S190–S201. [[CrossRef](#)] [[PubMed](#)]
43. Brouwer, A.F.; Eisenberg, M.C.; Meza, R. Case studies of gastric, lung, and oral cancer connect etiologic agent prevalence to cancer incidence. *Cancer Res* **2018**, *78*, 3386–3396. [[PubMed](#)]

44. Criss, S.D.; Sheehan, D.F.; Palazzo, L.; Kong, C.Y. Population impact of lung cancer screening in the United States: Projections from a microsimulation model. *PLoS Med.* **2018**, *15*, 1–15. [[CrossRef](#)] [[PubMed](#)]
45. Levy, D.T.; Yuan, Z.; Li, Y. The US SimSmoke tobacco control policy model of smokeless tobacco and cigarette use. *BMC Public Health* **2018**, *18*, 696. [[CrossRef](#)] [[PubMed](#)]
46. Nautiyal, N.; Holford, T.R. A spatiotemporal back-calculation approach to estimate cancer incidence measures. *Stat. Med.* **2018**, *37*, 4472–4489. [[CrossRef](#)] [[PubMed](#)]
47. Kayano, T.; Lee, K.D.; Nishiura, H. Estimating the force of infection with *Helicobacter pylori* in Japan. *Can. J. Infect. Dis. Med. Microbiol.* **2019**, *2019*, 1451490. [[CrossRef](#)] [[PubMed](#)]
48. Nishiura, H. Estimating the incidence and diagnosed proportion of HIV infections in Japan: A statistical modeling study. *PeerJ* **2019**, *7*, e6275. [[CrossRef](#)] [[PubMed](#)]
49. Jung, K.J.; Jeon, C.; Jee, S.H. The effect of smoking on lung cancer: Ethnic differences and the smoking paradox. *Epidemiol Health* **2016**, *38*, e2016060. [[CrossRef](#)] [[PubMed](#)]
50. Hori, M.; Tanaka, H.; Wakai, K.; Sasazuki, S.; Katanoda, K. Secondhand smoke exposure and risk of lung cancer in Japan: A systematic review and meta-analysis of epidemiologic studies. *Jpn. J. Clin. Oncol.* **2016**, *46*, 942–951. [[CrossRef](#)]
51. Kishida, N.; Nishiura, H. Demographic supply-demand imbalance in industrial structure in the super-aged nation Japan. *Theor. Biol. Med. Model.* **2018**, *15*, 19. [[CrossRef](#)]
52. Jung, S.M.; Lee, H.; Nishiura, H. The impact of pneumococcal vaccination on pneumonia mortality among the elderly in Japan: A difference-in-difference study. *PeerJ* **2018**, *6*, e6085. [[CrossRef](#)]
53. Saito, M.M.; Nishiura, H.; Higuchi, T. Reconstructing the transmission dynamics of rubella in Japan, 2012–2013. *PLoS ONE* **2018**, *13*, e0205889. [[CrossRef](#)] [[PubMed](#)]
54. Hamaguchi, Y.; Yamaguchi, T.; Nishiura, H. Estimating the annual risk of tuberculosis infection in Japan from interferon-gamma release assay data. *J. Theor. Biol.* **2019**, *460*, 125–133. [[CrossRef](#)] [[PubMed](#)]
55. Akhmetzhanov, A.R.; Lee, H.; Jung, S.M.; Kinoshita, R.; Shimizu, K.; Yoshii, K.; Nishiura, H. Real time forecasting of measles using generation-dependent mathematical model in Japan, 2018. *PLoS Curr. Outbreaks* **2018**, *10*. [[CrossRef](#)] [[PubMed](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).