Published in final edited form as:

Risk Anal. 2012 July ; 32(Suppl 1): S99–S116. doi:10.1111/j.1539-6924.2011.01681.x.

The FHCRC Lung Cancer Model

William D. Hazelton^{1,*}, Jihyoun Jeon¹, Rafael Meza^{2,1}, and Suresh H. Moolgavkar^{1,*}

¹Fred Hutchinson Cancer Research Center, Public Health Science Division, 1100 Fairview Avenue North, M2-B500 Seattle, WA 98109, USA

²Department of Epidemiology, School of Public Health, University of Michigan, 1415 Washington Heights, 4647 SPH Tower, Ann Arbor, MI 48109, USA

Abstract

As a member of the Cancer Intervention and Surveillance Modeling Network (CISNET), the lung cancer group at Fred Hutchinson Cancer Research Center (FHCRC) developed a model for evaluating US lung cancer mortality trends and the impact of changing tobacco consumption. Model components include a biologically-based two-stage clonal expansion (TSCE) model; a smoking simulator to generate smoking histories and other cause mortality; and adjustments for period and birth-cohort to improve calibration to US lung cancer mortality. The TSCE model was first calibrated to five substantial cohorts: British doctors', American Cancer Society CPS-I and CPS-II, Health Professionals' Follow-Up Study (HPFS), and Nurses' Health Study (NHS). The NHS and HPFS cohorts included the most detailed smoking histories and were chosen to represent the effects of smoking on US lung cancer mortality. The calibrated TSCE model and smoking simulator were used to simulate US lung cancer mortality. Further adjustments were necessary to account for unknown factors. This provided excellent fits between simulated and observed US lung cancer mortality for ages 30–84 and calendar years 1975–2000. The FHCRC lung cancer model may be used to study the effects of public health information on US lung cancer trends and the impact of tobacco control policy. For example, we estimated that over 500,000 males and 200,000 females avoided lung cancer death between 1975–2000 due to increasing awareness since the mid-1950's of the harmful effects of smoking. We estimated that 1.1 million male and 0.6 million female lung cancer deaths were avoidable if smokers quit smoking in 1965.

Keywords

US lung cancer model; tobacco control; impact of public health information

1 Introduction

Smoking is recognized as one of the most significant risk factors for lung cancer (LC) mortality. To improve our understanding of the effects of smoking and other factors on LC mortality risk, we developed a model to evaluate US LC trends and estimate the impact of changes in tobacco consumption. Subsequently, we refer to this as the Fred Hutchinson Cancer Research Center LC (FHCRC-LC) model.

The FHCRC-LC model used a biologically-based two-stage model to represent the natural history of LC in relation to cigarette smoking history, age, and gender. The model was applied to a simulated US population consisting of individual smoking histories and time of other cause mortality generated by a smoking history generator provided by the National

*To whom correspondence should be addressed: hazelton@fhcrc.org, moolgavkar@gmail.com.

Cancer Institute (NCI). Several types of adjustment: 1) period and birth cohort, or 2) age, period, and birth cohort were introduced to account for historical changes in cigarette composition, and for other factors that influence LC mortality. While both types of adjustment provided excellent fits to US LC mortality, the first adjustment method was chosen as being more parsimonious. The FHCRC-LC model may be utilized to address questions about the impact of public health information on US LC trends and predict the hypothetical impact of alternative tobacco control policies.

Although cigarette smoking is recognized as the primary risk factor for lung cancer, LC mortality trends in the US are complex, and not fully understood. LC mortality risk increases more rapidly with smoking duration than with smoking intensity⁽¹⁻⁴⁾. Thus smoking pack-years do not provide a good measure of LC risk. A good LC mortality risk prediction model requires as inputs both dose and duration, or preferably detailed information on individual smoking histories⁽²⁾. Comparisons of LC mortality risk in different smoking cohorts suggest that regional and temporal differences may also influence risk.^(1,2)

The central component of the FHCRC-LC model is a two-stage clonal expansion (TSCE) model⁽⁵⁻⁹⁾ that models LC risk based on smoking history, age, and gender. The stochastic TSCE model represents cellular processes of initiation, promotion, and malignant conversion during carcinogenesis. The influence of smoking on these processes is described by (non-linear) dose-response relationships. Model calibration consists of estimating cell kinetic rates, including background and dose-response parameters for the effects of smoking, to best fit LC mortality (or incidence) in a cohort, given the detailed smoking histories of cohort subjects.

The FHCRC-LC model was developed in two distinct phases. The first phase consisted of calibrating the TSCE model to LC mortality data in five substantial smoking cohorts. The cohorts include the British doctors', the American Cancer Society Cancer Prevention Studies I and II (CPS-I and CPS-II) cohorts⁽¹⁾ (see Tables I, II), and the Health Professionals' Follow-up Study (HPFS), and Nurses' Health Study (NHS) cohorts (see Table III). The NHS and HPFS cohorts were also used to calibrate a TSCE model to LC incidence⁽²⁾. The TSCE model parameters (including background and smoking dose-response parameters) for these calibrations were shared with other CISNET LC modeling groups. These parameter estimates were used by modeling groups from Erasmus⁽¹⁰⁾, PIRE⁽¹¹⁾, and Yale⁽¹²⁾, as well as FHCRC. The Rice modeling group also used a TSCE model, but calibrated to another data source⁽¹³⁾. Similarities and differences between CISNET lung models are summarized by McMahon, et al.⁽¹⁴⁾.

The second phase of FHCRC-LC model development consisted of calibrating a model to US LC mortality. However, smoking histories are not generally known for individuals in the US population. Thus we made a micro-simulation of the full US population using simulated individual smoking histories and other cause mortality. The full set of smoking histories were input to the TSCE model to predict US LC mortality, as a first approximation. Then adjustments for period and cohort (or alternatively, age, period, and cohort) were introduced to improve the fit to observed US LC mortality. Together, these steps comprise the CISNET Lung Smoking Base Case.

The micro-simulation of the US population utilized simulated individual smoking histories^(15,16) and other cause of mortality⁽¹⁷⁾ provided by a smoking history generator⁽¹⁶⁾ (SHG) provided by NCI. The SHG generates birth-cohort, calendar-year, and gender specific smoking histories and time of other-cause death, based on multiple data sources as discussed by Jeon, et al.⁽¹⁶⁾. The probability of LC mortality for each simulated individual at

a given age was calculated using the TSCE parameters for the HPFS (male) and NHS (female) cohorts. Cause of simulated death was taken as the earlier of other-cause or LC death. Simulated LC deaths and person years for simulated individuals were tabulated by single year of age from 30–84, and calendar year from 1975–2000, with person-years in each cell matching US population data. A second calibration was then made to US LC mortality data using macro-simulation methods to estimate a) period, and birth cohort effects, or alternatively b) age, period and birth cohort effects. Thus the FHCRC LC group used a combination of micro- and macro-simulation methods.

The purpose in developing the FHCRC-LC model was to improve our understanding and predictive ability for US LC mortality risk, including projecting risk beyond the calibration range, and estimating the potential effects of alternative tobacco control policies. We used the FHCRC-LC model to estimate the effects of the awareness, since the mid-1950s, of the harmful effects of tobacco on LC mortality trends, and the numbers of life years lost (or gained) under alternative tobacco control scenarios.

2 Methods

The FHCRC-LC model includes a TSCE model to represent the natural history component for smoking related LC mortality in the US. The TSCE model is applied to simulated smoking histories to calculate age-specific probabilities of LC death, in competition with other-cause death, to simulate individuals in the full US population. Further adjustments (e.g. for period, and birth cohort) are then applied to improve the fit between simulated and observed LC mortality in the US population.

2.1 Approach/Model

The TSCE model is a stochastic model that represents the carcinogenic process in terms of the evolving probability distribution for the number of cells on the pathway to cancer.^(5–8) The model assumes that any of the normal stem cells X in the lungs may undergo, at random, a first mutation (called initiation) at rate ν during division to create an initiated cell. Initiated cells are assumed to divide at rate α and die/differentiate at rate β , forming pre-malignant clones. A random second mutation event may occur at rate μ during cell division, transforming an initiated cell into a malignant cell. A lag time, or lag time distribution, is used to represent the time between appearance of the first malignant cell and LC death (See Fig. 1).

Only particular combinations of the TSCE model parameters are mathematically identifiable⁽⁸⁾. To achieve model identifiability, we assumed a fixed number ($X = 10^7$) of normal stem cells in both lungs. We also assume that the first and second mutation rates are equal ($\nu = \mu$). The optimized model can be rescaled to relax these assumptions without changing the model fit, as the model likelihood remains invariant under this rescaling⁽¹⁸⁾.

The multistage process of cell birth, death/differentiation, and mutation gives rise naturally to a probability for LC mortality which increases with age. LC mortality risk depends also on smoking, modeled through dose-response relationships that change the cell kinetic rates. Because the model is stochastic, it allows for disease progression, or regression as premalignant clones may decrease in size or become extinct. Extinction is most likely when clones are small.

Exposure to tobacco smoke may alter one or more of the rates for initiation, promotion, or malignant conversion. We use likelihood based methods to estimate rates which are significantly changed by smoking, and dose-response relationships for these rates. The TSCE model has been used previously to relate LC incidence or mortality risk to individual

exposure patterns for smoking, radon, arsenic, fibers, and radiation.^(1,2,9,18–35) These analyses indicate that LC risk is strongly associated with smoking history, with risk increasing non-linearly with smoking duration. Analyses of cohorts with complex individual exposure histories, including smoking, radon, and arsenic, indicate that these different exposures show significant interactions.^(18,24–25)

2.2 Data

The FHCRC-LC model included multiple data sources as inputs: a) smoking cohorts used to calibrate the TSCE model, b) US population data by gender, year of age from 30–84, and calendar year from 1975–2000, that were used as a target for simulating the US population, c) a SHG provided by the NCI that was used to simulate individual smoking histories and other cause mortality in the US population, d) US LC mortality data by gender, age, and calendar year used in final calibration, e) data guiding hypothetical scenarios for modeling the effects of alternative tobacco control policies.

Smoking cohorts for TSCE model LC mortality calibration included CPS-I, CPS-II, British doctors', NHS, and HPFS cohorts. Individual smoking histories were updated every 2 years for NHS and HPFS (including both LC incidence and mortality outcomes)⁽²⁾. Individual histories in NHS and HPFS include: gender, smoking status, age at smoking initiation (if a smoker), beginning smoking rate (cigarettes per day), age at each change in smoking habit, rate during each of these periods, and age at quit smoking (if that occurs), age at study entry, and age at LC death or censoring age at end of follow-up.

2.3 Phase 1: Calibrating the TSCE model to data in five substantial smoking cohorts

The effects of smoking status, dose, duration, and age on LC mortality were investigated by calibrating a TSCE model to never and current smokers in the CPS-I, CPS-II, and British Doctors' (male) cohort⁽¹⁾ (see Tables I and II), and among never, current, and former smokers in the HPFS (male) and NHS (female) cohorts (see Table III).

The NHS and HPFS cohort data are the most detailed, including a variable number of non-smoking and smoking intervals, often at different smoking intensities. For each interval, the model relates smoking intensity to net cell proliferation and mutation rates through non-linear dose-response relationships. The TSCE model uses this sequence of piecewise constant cell kinetic rates to calculate the individual likelihood. Each individual likelihood is corrected for time of entry into the study, by dividing by the survival probability at age of entry.^(1,2)

We begin by assuming that initiation, promotion, and malignant conversion rates may be altered during periods of smoking exposure through flexible dose-response relationships,

$$\theta_c(t) = \theta_{c0}(1 + \theta_{c1}d(t)^{\theta_{c2}}), \quad (1)$$

where θ_c represents one of the identifiable biological parameters⁸ (combinations of α , β , μ , ν , and X), θ_{c0} is the background rate, θ_{c1} is a dose-response coefficient, θ_{c2} represents the non-linearity of the dose-response, and $d(t)$ is dose at time t .

The dose-response relationships relate the individual smoking history to the cell kinetic parameters in the TSCE model, and thus to the probability for LC mortality at age t for the individual. During calibration to smoking cohort data, the study follow-up times and the outcome for each individual (death from LC, or study censoring) are also required to formulate the individual likelihood.

The TSCE model hazard and survival probabilities are analytic functions that may be calculated recursively for piecewise constant model parameters⁽⁸⁾. Let the TSCE model survival at time u be represented $S_2(u)^{(8)}$. If we assume a constant lag time or a gamma-distributed lag time between appearance of the first malignant cell and LC death, the survival probability $S(t; d, \theta)$ at age t for an individual with smoking history d is

$$S(t; d, \theta) = \begin{cases} S_2(t - t_{lag}; d, \theta) & \text{if lag time is constant,} \\ 1 - \int_0^t (1 - S_2(u; d, \theta)) f(t - u; a, b) du & \text{if lag time is gamma distributed,} \end{cases} \quad (2)$$

where θ denotes the vector of identifiable model parameters $\{\theta_c\}$ and lag time parameters (t_{lag} for constant lag, or a and b for gamma distribution). Setting $\tau = t - u$, the gamma probability density function for lag time is $f(\tau; a, b) = \frac{1}{b\Gamma(a)} \tau^{a-1} e^{-\tau/b}$. The probability for LC mortality may be calculated for any single year of age, given the gender and the full smoking history of the individual.

The individual likelihood $\mathcal{L}_n(t_n, s_n; d_n, \theta)$ for subject n depends on time of entry into the study s_n , censoring or failure time t_n , detailed exposure histories d_n , and parameters θ for background and dose-response in the TSCE model, and constant lag or lag time distribution. The individual likelihoods for cases and survivors, including left truncation for age at entry, are given by

$$\mathcal{L}_n(t_n, s_n; d_n, \theta) = \begin{cases} -S'(t_n; d_n, \theta) / S(s_n; d_n, \theta) & \text{if death from LC,} \\ S(t_n; d_n, \theta) / S(s_n; d_n, \theta) & \text{otherwise,} \end{cases} \quad (3)$$

where the prime denotes derivative with respect to t .

Assuming independence between individuals, the cohort likelihood is the product of individual likelihoods,

$$\mathcal{L} = \prod_n \mathcal{L}_n(t_n, s_n; d_n, \theta). \quad (4)$$

Likelihood maximization was performed using a combination of modified Davidon-Fletcher-Powell gradient search, simplex, and Newton methods. Markov Chain Monte Carlo (MCMC) methods (Bhat FORTRAN software library)⁽³⁶⁾ were used for exploration of the likelihood surface and estimation of parameter confidence intervals.

2.4 Phase 2: The Lung Smoking Base Case

The goal of the CISNET Lung Smoking Base Case was to model the impact of cigarette smoking on LC mortality in the US between 1975–2000, and for ages 30 to 84. Secondary goals were to combine and compare results from different CISNET LC modeling groups to identify areas of common understanding and the range of uncertainties when different modeling approaches provide different estimates. At FHCRC, we used a TSCE model, calibrated to LC mortality in the NHS and HPFS cohorts, to represent the effects of smoking in the US LC mortality model (called the FHCRC-LC model). The NHS and HPFS cohorts were chosen for calibration because these data provide the most detailed smoking histories, and follow-up periods were closest (1976–2000 for NHS, 1986–2002 for HPFS) to the range chosen for the CISNET Lung Smoking Base Case.

The SHG provided by NCI was used to simulate US male and female smoking histories, birth-year, and age at death from causes other than LC. We simulated the time of LC death using the TSCE model, then combined the earlier of LC death or other-cause death generated by the SHG to generate a life history. These life histories were sampled to construct a synthetic population that matched the full US population by gender and single years of age 30–84 and calendar years 1975–2000. In this process, each simulated individual history generally contributes to multiple cells in the table until time of death. For example, beginning in 1975, a 35 year old male becomes age 36 in 1976 and so on, contributing to tables of person-years and LC deaths down a diagonal of cells indexed by age and calendar year until occurrence of death or censoring. As construction of the synthetic population nears completion, some histories were left or right truncated as required to not overfill cells in the population table.

The hazard function for each individual in the simulated US population was calculated using the TSCE model calibrated to LC mortality in the NHS and HPFS cohorts. However, the predicted LC deaths based on the TSCE model differed somewhat from observed US LC mortality by age and calendar year. Thus the FHCRC group introduced additional adjustments to fit the observed US LC mortality, using maximum likelihood methods to estimate parameters for age, period, and birth cohort effects.

Two approaches were taken to adjust for the differences between predicted LC mortality from the cohort-calibrated TSCE model applied to the simulated US population, and observed US LC mortality. We first introduced period, and birth cohort effects to represent LC mortality in the US population (called the TSCE-PC model, representing the age effects given by the TSCE natural history model, along with period and birth cohort effects). We also tried a second method (called the TSCE-APC model) which includes additional age effects to capture possible discrepancies between age effects from the TSCE model calibrated to the HPFS and NHS cohorts and in the US LC mortality, and to compensate for possible limitations of the TSCE model in representing the effects of tobacco on LC mortality.

The US LC mortality and population data are indexed by single year of age $i = 1 \dots I$ and calendar year $j = 1, \dots, J$. For age group i , the number of LC deaths during calendar year j is assumed to follow a Poisson distribution with mean $\Lambda_{i,j}$.

In the TSCE-PC model, $\Lambda_{i,j} = PY_{i,j} h(a_i) b_k c_j$, where a_i is the mean age of age group i , b_k and c_j are coefficients that adjust for birth cohort and period effects, respectively, $PY_{i,j}$ is the person years at risk, and $h(a_i)$ represents the hazard function of the TSCE model evaluated at age a_i . There are 16 birth cohorts ($k = 1, \dots, 16$, representing (1890 – 1894, 1895 – 1899, ..., 1960 – 1964, 1965). Each cell indexed by i, j maps to a particular k value.

In the TSCE-APC model, $\Lambda_{i,j} = PY_{i,j} h(a_i) e_i b_k c_j$, where e_i is a coefficient that adjusts for additional age effect for the age group i , with other variables the same as in the TSCE-PC model.

The overall likelihood $\mathcal{L}$ for the observed LC mortality in all age-calendar year groups is given by

$$\mathcal{L} = \prod_{i,j} \frac{\Lambda_{i,j}^{O_{i,j}} e^{-\Lambda_{i,j}}}{O_{i,j}!}, \quad (5)$$

where $O_{i,j}$ is the number of observed LC deaths for age group i and calendar year j .

3 Results

The FHCRC-LC model combines the TSCE model (Fig. 1 and Table I) calibrated to the LC mortality for males in the HPFS cohort and females in the NHS cohort (Table III) with US population-based adjustment for secular time trends (i.e., period and birth cohort in the TSCE-PC model (Fig. 4), or age, period and birth cohort in the TSCE-APC model).

3.1 Hypotheticals

Seven hypothetical individual smoking histories were selected as test cases for comparing estimates of individual LC mortality risk between modeling groups. These hypotheticals facilitated comparisons of the effects of different smoking intensities (20 or 40 cigarettes per day), quitting or continuing to smoke, or never smoking. Fig. 2 shows LC mortality risk between ages 30 and 84 for three hypotheticals consisting of a never smoker; beginning smoking at age 14 and continuing at 20 cigarettes/day until quitting at age 35; and starting smoking at age 14 and continuing to smoke at 20 cigarettes/day.

3.2 Calibration and Validation

Initial calibration of the TSCE model to LC mortality utilized data from five smoking cohorts. During this (Phase I) modeling, we assumed smoking may influence rates for promotion, initiation, and malignant conversion in each cohort model. This is called the full model. After optimization of the full model, we progressively eliminated parameters that did not contribute significantly to the cohort likelihood. This is called the reduced model (see Tables I–III).

The CPS-I and British doctors' cohorts were contemporaneous. Likelihood based testing indicated that both cohorts could be fit with all but one parameter in common. CPS-II cohort follow-up occurred approximately 20 years later (following significant changes in cigarette manufacture). CPS-II cohort dose-response differs from the earlier cohort by having no significant effect of cigarette smoking dose-response on initiation (perhaps associated with cigarette filters), but a larger dose-response on promotion (possibly associated with higher levels of tobacco specific nitrosamines due to flue-curing of tobacco) than found in the CPS-I cohort⁽¹⁾ (see Table II). HPFS and NHS cohorts have similar follow-up periods as CPS-II, and no significant effect of smoking dose-response on initiation, consistent with CPS-II. However, we found a significant smoking dose-response on promotion as well as malignant conversion in HPFS and NHS affecting LC mortality cohorts (see Table III). The ability to estimate effects of smoking on malignant conversion may be due to the availability of detailed smoking histories in HPFS and NHS.

Adjustments were applied to each of the cohort models to fit to US LC mortality during phase 2. Fig. 3 shows age-adjusted LC mortality rates predicted by each of these models, before and after adjustment for calendar year and birth cohort. Results after adjustment are similar for each model, with all adjusted models providing good fits to US LC mortality, as shown in the right hand panel of Fig 3.

The adjustment of each cohort model to US LC mortality required relatively large period and cohort effects, shown in Fig. 4. Period (calendar year) effects differ between cohort models, indicating a strong influence of period, region, cohort selection, and other factors that complicate the application of LC mortality risk from cohorts to LC mortality in the US population. Birth cohort effects for both gender generally increase (until about 1945 for NHS and HPFS) with a subsequent decline. The variable influence of these factors in different cohorts is indicated by different parameter estimates in Tables II and III. In each case, model calibration to the corresponding cohort data was excellent, while comparisons

of lung cancer risk between cohorts differ significantly prior to period and cohort adjustments.

We evaluated the sensitivity of the model fits to alternative procedures because a) smoking patterns for the oldest ages in the oldest birth cohort (representing the “bottom left corner” of the US population table) are questionable, and b) the age effects from the TSCE model calibration to the HPFS and NHS cohorts may not be the same as in the general US population.

To address the sensitivity of the fits to limited smoking information for the oldest ages and oldest birth cohorts, we compared two estimation approaches: fitting the model to ages 30–84 and calendar years 1975–2000 beginning with the 1900 birth cohort (leaving the bottom left corner of the table missing), and alternatively, extrapolating smoking patterns from comparable ages and nearby birth cohorts to fill in this missing corner of the table. We found no substantive differences in model fit or in projections to alternative tobacco control scenarios between these two estimation approaches. Thus all results shown subsequently are based on fits to the full calendar year range from 1975–2000, and ages 30–84. Predicted age adjusted rates by the NHS and HPFS cohorts, before adjustment for calendar year and birth cohort effects, are shown in Fig. 5.

To address sensitivity to possible differences in age effects associated with the HPFS/NHS calibration of the TSCE model and US population age effects, we compared secondary calibration using the Age-Age-Period-Cohort model (TSCE-APC model) and the Age-Period-Cohort model (TSCE-PC model). The TSCE-APC model provided better likelihood and closer detailed fits to the data when the data was divided into sub-age groups. However, the fit of the TSCE-PC model is also excellent, and requires significantly fewer parameters. Comparison of fits suggests the TSCE-APC model may over-fit the data. Thus, in the following, we present results from the TSCE-PC model.

Observed US LC mortality trends between 1975–2000 reflect many factors, including changes in cigarette consumption that may be attributable in part to effects of public health information provided through Surgeon General’s reports and other mechanisms, beginning in the mid-1950s. Predictions of US LC mortality by the FHCRC-LC model are presented for the (observed) US population, together with two alternative hypothetical scenarios consisting of alternate assumptions about tobacco consumption patterns.

3.3 Tobacco Control Scenarios

We first used the SHG to generate individual smoking histories approximating observed smoking trends in the US population, called the actual tobacco control (ATC) scenario.

To investigate the effects of tobacco control on US LC mortality risk, we used the SHG to generate comparative individual smoking histories by assuming that no tobacco control occurred, called the no tobacco control (NTC) scenario. To explore the maximum potential benefits from tobacco control on LC mortality, we used the SHG to generate a US population in which all smokers quit smoking in 1965 and no one began smoking after that, called the complete tobacco control (CTC) scenario.

Fig. 6 compares the predicted age-standardized US LC mortality rates from the TSCE-PC model for the three smoking scenarios; ATC, NTC, and CTC. The results in this figure indicate that the FHCRC-LC model provides good fits (solid lines) to the data (points) for age-adjusted US LC mortality for both males and females. We obtained similar results using the TSCE-APC model (results not shown).

We compared the FHCRC-LC model fits against US LC mortality by birth cohort and calendar year for both males and females (See Fig. 7). The LC mortality rates plotted by birth year (left) for older age groups show a pronounced peak, possibly attributable to increased smoking around the time of the second world war. The oldest female age groups show striking increases by calendar year. In all cases, the TSCE-PC model provided excellent fits to the data.

We also compared the TSCE-PC model predicted (and observed) number of US LC deaths given the historical ATC experience to the expected number of LC deaths under NTC scenario, calculated by the TSCE-PC model (Fig. 8). These results suggest that public health information about the risks of cigarette smoking may have contributed to a significant reduction in LC deaths. Our model estimates suggest that over 500,000 deaths were avoided among males, and over 200,000 deaths among females. To explore the potential maximum benefits possible by smoking intervention, we compared predictions of the TSCE-PC model for number of US LC deaths under ATC scenario and the expected number of LC deaths under the CTC scenario (Fig. 9). This comparison provides an estimate of the maximum number of avoidable LC deaths, under the (unrealistic) assumption that everyone quit smoking in 1965 and no one started smoking afterwards. According to the model estimates, approximately 1,100,000 deaths were potentially avoidable among males, and 600,000 among females. The TSCE-APC model provided similar results (See Figs. 10 and 11).

The population attributable fraction (PAF) (sometimes called the population attributable risk) is a useful tool to measure the risk attributable to a specific factor (such as LC deaths attributable to cigarette smoking) in a population.⁽³⁷⁾ The PAF measures the fraction of all LC deaths that would not have occurred if cigarette smoking had not occurred. Let D be a binary variable for LC death, and E for the exposure to cigarette smoking. Then the PAF is calculated by

$$PAF = \frac{P(D=1) - P(D=1|E=0)}{P(D=1)},$$

where $P(D=1)$, and $P(D=1|E=0)$ are the probability of LC deaths in the total population, and among never smokers, respectively. Using the TSCE-PC model results, we estimated the PAF to examine the contribution of cigarette smoking to LC deaths under NTC, ATC and CTC scenarios (See Fig. 12). Let f be the fraction of ever smokers in the population, and RR be the relative risk between ever smokers and never smokers, i.e., $f = P(E=1)$, and $RR = P(D=1|E=1)/P(D=1|E=0)$. Since $P(D=1) = P(E=1)P(D=1|E=1) + P(E=0)P(D=1|E=0)$, with simple algebra, we can show

$$PAF = \frac{f(RR-1)}{1+f(RR-1)}.$$

The estimated age-adjusted relative risk of LC death between ever smokers and never smokers under the NTC, ATC, and CTC scenarios are shown in Fig. 13.

4 Discussion

The FHCRC-LC model was used to evaluate the effects of alternative tobacco control policies on US LC mortality for ages 30–84 and calendar years 1975–2000. It is generally thought that increasing US LC mortality trends slowed, and even decreased, in response to

increasing public awareness about the dangers of cigarette smoking beginning in the 1960's.^(38,39) This increasing public awareness came from many sources, including statements by the US Surgeon General about the risk of cigarette smoking.^(40,41)

Among the most significant results of this modeling effort are estimates suggesting that approximately 500,000 fewer LC deaths may have occurred among US males, and 200,000 fewer LC deaths among females during the period from 1975–2000, compared with those expected without an increase in public awareness of the risks of cigarette smoking.

We also estimated the number of LC deaths that could have been avoided if all smokers quit smoking in 1965 and no one began smoking subsequently. According to the FHCRC-LC model estimates, approximately 1,100,000 among US males and 600,000 among US females would have avoided LC death between 1975–2000 under this complete tobacco control scenario.

Although US smoking rates have declined substantially since the 1950s, especially among males, the impact on LC mortality is much smaller than hypothetically achievable if all individuals stopped smoking in 1965 (See Fig. 11). It takes time for LC mortality risk to decrease significantly following smoking cessation. According to the TSCE model, there is a sustained influence of pre-malignant clones initiated much earlier in life that are promoted during periods of smoking. This slows the decline in risk after smoking cessation. Benefits of quitting smoking should increase in later years.

The FHCRC-LC model combines a biologically-based natural history model (TSCE model) that is calibrated to individual smoking histories in substantial LC mortality cohorts and combined with additional period and birth cohort adjustments. These adjustments improve the calibration to US population and LC mortality data. The TSCE model calibrations to the LC mortality cohorts include background, dose-response, and lag time parameters to capture essential features of smoking in relation to LC mortality. LC mortality risk increases more rapidly with smoking duration than with smoking rate.

The non-linear increase in LC mortality risk with smoking duration may be represented as direct and interaction effects between different smoking periods, as illustrated in Fig. 14. Fig. 14a) shows annual LC mortality risk for males in the CPS-I cohort smoking 20 cig/day due to direct effects and interaction between three periods of smoking: Early (ages 20–35), Mid (ages 35–50) and Late (ages 50–65). Fig. 14b) shows these interactions and direct contributions to LC mortality risk at age 70, represented as a Venn diagram.

We chose to use the TSCE model calibrated to the HPFS cohort for males and the NHS cohort for females since these cohorts included the most detailed smoking histories, and follow-up was similar to the range needed for the CISNET Lung Smoking Base Case for which US LC mortality trends between 1975–2000 were of primary interest. The NHS and HPFS cohorts included explicit information on changing smoking patterns among current and former smokers, and may reflect better the impact of quitting smoking on LC mortality. Detailed simulation of the US population was done by sampling individual histories generated by the SHG, which were used to fill up the full US population table by single year of ages 30–84 and calendar years 1975–2000. The calibrated TSCE model was applied to these histories to calculate expected numbers of LC deaths. However, there are significant differences between LC mortality in cohorts and the US population. Factors other than smoking may influence US LC mortality trends. Although the TSCE natural history model calibrated to cohort data accounts for most LC deaths observed in the US, additional period, and birth cohort factors were required to accurately represent LC mortality in the US population.

Fig. 3 shows predicted age-adjusted US LC mortality for males and females for each cohort model calibrated during phase I. Left and right panels of Fig. 3 show predicted age-adjusted LC mortality before and after adjusting each model for period and birth cohort. This figure shows the importance of period and birth cohort effects. The unadjusted TSCE models calibrated to data from the smoking cohorts do not adequately capture all of the factors that determine US LC mortality between 1975–2000. Factors may include changes in cigarette composition, second-hand smoke, radon, occupational exposures, and unknown factors. All models provided good fits to the US data after adjustment. The period and birth cohort adjustments differ between models, as shown in Fig. 4. This indicates the strong influence of period, geographic region, and cohort selection criteria on LC mortality rates.

There are many limitations inherent in modeling LC mortality in the US population. Model specification requires that complex biological and social processes are represented by a limited number of parameters. The TSCE model represents basic cellular processes that are generally hidden, as properties of individual cells in the lungs are not observable with current technology. Furthermore, the TSCE model represents a significant simplification of the biological processes associated with LC. The model assumes two rate-limiting mutation events, a single clonal expansion stage, and a lag time or lag time distribution from first occurrence of a malignant cell to death from LC. The effects of cigarette smoke are modeled using a constant dose rate during periods of smoking, and the smoking dose has separate non-linear (power-law) influences on initiation, promotion, and malignant conversion rates in the TSCE model (See Equation (1)). The model ignores the possibility of multiple pathways to cancer, and aggregates details of different patient outcomes and disease subtypes when calculating the probability of LC mortality. That being said, the TSCE model does provide a rigorous mathematical representation of processes that are considered as the rate-limiting events in carcinogenesis, and has provided excellent fits to individual and population data for many cancer types.⁽⁴²⁾

We believe that the FHCRC-LC model provides a useful tool for understanding the contribution of smoking-and non-smoking related factors associated with US LC mortality. The model was used to estimate the impact of changes in tobacco consumption on US LC mortality due to public awareness of the dangers of cigarette smoking, and hypothetical effects if everyone quit smoking in 1965. The model may be useful for examining other hypothetical tobacco control policies.

Acknowledgments

Supported by grants UO1 CA097415 and UO1 CA152956.

References

1. Hazelton WD, Clements MS, Moolgavkar SH. Multistage carcinogenesis and lung cancer mortality in three cohorts. *Cancer epidemiology, biomarkers & prevention*. 2005; 14(5):1171–1181.
2. Meza R, Hazelton WD, Colditz GA, Moolgavkar SH. Analysis of lung cancer incidence in the Nurses' Health and the Health Professionals' Follow-Up Studies using a multistage carcinogenesis model. *Cancer causes & control*. 2008; 19(3):317–328. [PubMed: 18058248]
3. Doll R, Peto R. Cigarette smoking and bronchial carcinoma: dose and time relationships among regular smokers and lifelong non-smokers. *Journal of epidemiology and community health*. 1978; 32(4):303–313. [PubMed: 744822]
4. Rachet B, Siemiatycki J, Abrahamowicz M, Leffondre K. A flexible modeling approach to estimating the component effects of smoking behavior on lung cancer. *Journal of clinical epidemiology*. 2004; 57(10):1076–1085. [PubMed: 15528059]
5. Moolgavkar SH, Venzon DJ. Two-event models for carcinogenesis: incidence curves for childhood and adult tumors. *Mathematical biosciences*. 1979; 47(1–2):55–77.

6. Moolgavkar SH, Knudson AG Jr. Mutation and cancer: a model for human carcinogenesis. *Journal of the National Cancer Institute*. 1981; 66(6):1037–1052. [PubMed: 6941039]
7. Moolgavkar SH, Luebeck G. Two-event model for carcinogenesis: biological, mathematical, and statistical considerations. *Risk analysis*. 1990; 10(2):323–341. [PubMed: 2195604]
8. Heidenreich WF, Luebeck EG, Moolgavkar SH. Some properties of the hazard function of the two-mutation clonal expansion model. *Risk analysis*. 1997; 17(3):391–399. [PubMed: 9232020]
9. Castren O. Implications of a two-stage clonal expansion model to indoor radon risk assessment. *Health physics*. 1999; 76(4):393–397. [PubMed: 10086600]
10. Schultz FW, Boer R, de Koning HJ. Description of MISCAN-Lung, the Erasmus MC Lung Cancer Model for Evaluating Cancer Control Interventions. *Risk Analysis*. (chapter 7 of this monograph).
11. Levy D. 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 Model. *Risk Analysis*. (chapter 10 of this monograph).
12. Holford TR, Ebisu K, McKay L, Oh C, Zheng T. Yale Lung Cancer Model. *Risk Analysis*. (chapter 12 of this monograph).
13. Foy M, Deng L, Spitz M, Gorlova O, Kimmel M. Rice-MD Anderson Lung Cancer Model. *Risk Analysis*. (chapter 11 of this monograph).
14. McMahon P, Hazelton W, Kimmel M, Clarke L. CISNET lung models: Comparison of model assumptions and model structures. *Risk Analysis*. (chapter 13 of this monograph).
15. Anderson CM, Burns DM, Dodd KW, Feuer EJ. Birth Cohort Specific Estimates of Smoking Behaviors for the U.S. Population. *Risk Analysis*. (chapter 2 of this monograph).
16. Jeon J, Meza R, Krapcho M, Clarke L, Byrne J, Levy D. Actual and Counterfactual Smoking Prevalence Rates in the US population via Micro-simulation. *Risk Analysis*. (chapter 5 of this monograph).
17. Rosenberg MA, Feuer EJ, Yu B, Sun J, Henley J, Shanks T, Anderson C, McMahon PM, Thun M, Burns D. Cohort Life Tables By Smoking Status Removing Lung Cancer as a Cause of Death. *Risk Analysis*. (chapter 3 of this monograph).
18. Hazelton WD, Luebeck EG, Heidenreich WF, Moolgavkar SH. Analysis of a historical cohort of Chinese tin miners with arsenic, radon, cigarette smoke, and pipe smoke exposures using the biologically based two-stage clonal expansion model. *Radiation research*. 2001; 156(1):78–94. [PubMed: 11418076]
19. Hazelton WD, Moolgavkar SH, Curtis SB, Zielinski JM, Ashmore JP, Krewski D. Biologically based analysis of lung cancer incidence in a large Canadian occupational cohort with low-dose ionizing radiation exposure, and comparison with Japanese atomic bomb survivors. *Journal of toxicology and environmental health Part A*. 2006; 69(11):1013–1038. [PubMed: 16840251]
20. Haylock RG, Muirhead CR. Fitting the two-stage model of carcinogenesis to nested case-control data on the Colorado Plateau uranium miners: dependence on data assumptions. *Radiation and environmental biophysics*. 2004; 42(4):257–263. [PubMed: 14618344]
21. Heidenreich WF, Luebeck EG, Hazelton WD, Paretzke HG, Moolgavkar SH. Multistage models and the incidence of cancer in the cohort of atomic bomb survivors. *Radiation research*. 2002; 158(5):607–614. [PubMed: 12385638]
22. Kai M, Luebeck EG, Moolgavkar SH. Analysis of the incidence of solid cancer among atomic bomb survivors using a two-stage model of carcinogenesis. *Radiation research*. 1997; 148(4):348–358. [PubMed: 9339951]
23. Little MP, Haylock RG, Muirhead CR. Modeling lung tumour risk in radon-exposed uranium miners using generalizations of the two-mutation model of Moolgavkar, Venzon and Knudson. *International journal of radiation biology*. 2002; 78(1):49–68. [PubMed: 11747553]
24. Luebeck EG, Heidenreich WF, Hazelton WD, Paretzke HG, Moolgavkar SH. Biologically based analysis of the data for the Colorado uranium miners cohort: age, dose and dose-rate effects. *Radiation research*. 1999; 152(4):339–351. [PubMed: 10477911]
25. Luebeck EG, Curtis SB, Hazelton WD, Moolgavkar SH. The importance of promotion in lung carcinogenesis for protracted exposures to radon and radon progeny. *Radiation research*. 2000; 154(6):730. discussion 730–1. [PubMed: 11187018]

26. Moolgavkar SH, Dewanji A, Luebeck G. Cigarette smoking and lung cancer: reanalysis of the British doctors' data. *Journal of the National Cancer Institute*. 1989; 81(6):415–420. [PubMed: 2783979]
27. Moolgavkar SH, Luebeck EG, Krewski D, Zielinski JM. Radon, cigarette smoke, and lung cancer: a re-analysis of the Colorado Plateau uranium miners' data. *Epidemiology*. 1993; 4(3):204–217. [PubMed: 8512985]
28. Moolgavkar SH, Luebeck EG, Anderson EL. Estimation of unit risk for coke oven emissions. *Risk analysis*. 1998; 18(6):813–825. [PubMed: 9972584]
29. Moolgavkar SH, Luebeck EG, Turim J, Hanna L. Quantitative assessment of the risk of lung cancer associated with occupational exposure to refractory ceramic fibers. *Risk analysis*. 1999; 19(4):599–611. [PubMed: 10765424]
30. Moolgavkar SH, Luebeck EG, Turim J, Brown RC. Lung cancer risk associated with exposure to man-made fibers. *Drug and chemical toxicology*. 2000; 23(1):223–242. [PubMed: 10711399]
31. Moolgavkar SH, Brown RC, Turim J. Biopersistence, fiber length, and cancer risk assessment for inhaled fibers. *Inhalation toxicology*. 2001; 13(9):755–772. [PubMed: 11570360]
32. Moolgavkar SH, Turim J, Brown RC, Luebeck EG. Long man-made fibers and lung cancer risk. *Regulatory toxicology and pharmacology*. 2001; 33(2):138–146. [PubMed: 11350196]
33. Moolgavkar SH, Turim J, Brown RC. The power of the European Union protocol to test for carcinogenicity of inhaled fibers. *Regulatory toxicology and pharmacology*. 2001; 33(3):350–355. [PubMed: 11407937]
34. Stevens RG, Moolgavkar SH. Estimation of relative risk from vital data: smoking and cancers of the lung and bladder. *Journal of the National Cancer Institute*. 1979; 63(6):1351–1357. [PubMed: 292807]
35. Stevens RG, Moolgavkar SH. A cohort analysis of lung cancer and smoking in British males. *American Journal of Epidemiology*. 1984; 119(4):624–641. [PubMed: 6711550]
36. Luebeck, EG. 2009. <http://cran.r-project.org/web/packages/Bhat/Bhat.pdf>
37. Bruzzi P, Green SB, Byar DP, Brinton LA, Schairer C. Estimating the population attributable risk for multiple risk factors using case-control data. *American Journal of Epidemiology*. 1985; 122(5):904–914. [PubMed: 4050778]
38. Irvine L, Elliott L, Wallace H, Crombie IK. A review of major influences on current public health policy in developed countries in the second half of the 20th century. *The journal of the Royal Society for the Promotion of Health*. 2006; 126(2):73–78. [PubMed: 16562775]
39. Musk AW, de Klerk NH. History of tobacco and health. *Respirology*. 2003; 8(3):286–290. [PubMed: 14528877]
40. Parascandola M. Cigarettes and the US Public Health Service in the 1950s. *American Journal of Public Health*. 2001; 91(2):196–205. [PubMed: 11211627]
41. Parascandola M, Weed DL, Dasgupta A. Two Surgeon General's reports on smoking and cancer: a historical investigation of the practice of causal inference. *Emerging themes in epidemiology*. 2006; 3:1. [PubMed: 16403213]
42. Moolgavkar SH, Luebeck EG. Multistage carcinogenesis and the incidence of human cancer. *Genes, chromosomes & cancer*. 2003; 38(4):302–306. [PubMed: 14566848]

Two-stage clonal expansion model

Fig. 1.

The two-stage clonal expansion (TSCE) model is a stochastic model of cell events during carcinogenesis. The carcinogenic process may be thought of as consisting of three phases: initiation, promotion, and malignant conversion. Normal stem cells (labeled X) may mutate at rate ν to create an initiated cell (labeled I). An initiated cell may divide at rate α , die or differentiate at rate β , and mutate at rate μ to create a malignant cell. A lag time, or lag time distribution is used to represent the time from occurrence of the first malignant cell to LC death.

Fig. 2. Hypothetical lung cancer mortality risk between ages 30 and 84 for three groups of males (a) or three groups of females (b) who: never smoked (dotted line); who began smoking at age 14, continuing at 20 cigarettes/day until quitting at age 35 (dashed line); and who began smoking at age 14 and continued to smoke at 20 cigarettes/day (solid line).

Fig. 3.

Age adjusted LC mortality rates predicted by phase-I models calibrated to cohort data, before and after adjusting for period and birth cohort. The left panels show un-adjusted predictions of the HPFS, BD, CPS-II, and CPS-I cohort models in a) for males, and the NHS and CPS-II models in c) for females, applied to US population data. Each cohort model provided excellent fits to the corresponding cohort data (not shown), yet predictions differ significantly between cohort models, showing the importance of cohort selection, possible effects from changes in cigarette composition, regional issues such as differences in radon levels, and occupational exposures. In general, individuals selected for inclusion in the cohorts have lower LC mortality rates than observed in the US population (diamonds). The

effects of applying period and birth cohort adjustments to each model are shown in b) for males, and d) for females. These adjustments bring all model predictions into good agreement with observed US LC mortality.

Fig. 4.

Each of the cohort models calibrated during phase-I require different calendar year and birth cohort adjustments to fit US LC mortality. Period effects (calendar years 1975–2000) are shown for each model in a) for males and c) for females, and birth cohort effects (birth cohorts 1890 – 1894, 1895 – 1899, ..., 1960 – 1964, 1965) in b) and d). (Predicted US LC mortality rates before and after applying these adjustments are shown in Figure 3). The relatively large period and birth cohort effects reflect the importance of regional, temporal, and cohort selection factors.

Fig. 5.

TSCE-PC model results: predicted age-standardized LC mortality rates before adjusting for calendar year and birth cohort in the NHS and HPFS LC mortality models for a) males, and b) females. The 2000 US standard population, Census P25-1130 was used for age-standardization. (Figure 6 shows predictions after applying these adjustments). Dashed line are for no tobacco control (NTC), solid lines are for actual tobacco control (ATC), and dotted lines are for complete tobacco control (CTC). Diamond points represent age-standardized LC mortality rates for observed US LC mortality data.

Fig. 6.

TSCE-PC model results: predicted age-standardized LC mortality rates, including calendar year and birth cohort adjustments for a) males, and b) females. The 2000 US standard population, Census P25-1130 was used for age-standardization. Dashed line are for no tobacco control (NTC), solid lines are for actual tobacco control (ATC), and dotted lines are for complete tobacco control (CTC). Diamond points represent age-standardized LC mortality rates for observed US LC mortality data.

Fig. 7.

LC mortality rates by birth year and calendar year across 5-year age groups for males a) and b), respectively, and females c) and d), respectively. Symbols: observed US LC mortality rates. Lines: TSCE-PC model fits.

Fig. 8. TSCE-PC model results: avoided LC deaths resulted from the actual tobacco control over 1975–2000 for a) males, and b) females. The total number of avoided LC deaths = total number of LC deaths under no tobacco control (NTC) – total number of LC deaths under actual tobacco control (ATC).

Fig. 9.

TSCE-PC model results: avoidable LC deaths over 1975–2000 by assuming all smokers quit smoking in 1965 and no one started smoking afterwards. The total number of avoidable LC deaths = total number of LC deaths under actual tobacco control (ATC) – total number of LC deaths under complete tobacco control (CTC) for males a) and females b).

Fig. 10. TSCE-APC model results: avoided LC deaths resulted from the actual tobacco control over 1975–2000 for males a), and females b). These results adjusting for period, birth cohort, and additional age effects are similar to the results of adjusting only for period and birth cohort (See Fig. 8).

Fig. 11.

TSCE-APC model results: avoidable LC deaths over 1975–2000 by assuming all smokers quit smoking in 1965 and no one started smoking afterwards, for males a), and females b). These results adjusting for period, birth cohort, and additional age effects are similar to the results of adjusting only for period and birth cohort (See Fig. 9).

Fig. 12. TSCE-PC model results: predicted fraction of LC deaths attributable to cigarette smoking under three smoking scenarios; ATC: actual tobacco control, NTC: no tobacco control, CTC: complete tobacco control.

Fig. 13.

TSCE-PC model results: predicted age-adjusted relative risk of LC death between ever smokers and never smokers under three smoking scenarios; ATC: actual tobacco control, NTC: no tobacco control, CTC: complete tobacco control.

Fig. 14.

Interaction between smoking periods on LC mortality risk for males in the CPS-I cohort, shown for ages 20–70 in a), and as a Venn diagram at age 70 in b). Smoking was assumed to occur at a rate of 20 cig/day during three periods: Early (ages 20–35), Mid (ages 35–50) and Late (ages 50–65). Background risk is calculated from the model by setting smoking rate equal to zero during all periods. Early, Mid, and Late Alone contributions to risk are calculated by setting smoking rate to 20 cig/day in each of these periods, respectively, with smoking rate set to zero in the other two periods, then subtracting the background rate. The Early + Mid Interaction is calculated by setting smoking to 20 cig/day in the Early and Mid periods, then subtracting Early Alone, Mid Alone, and Background contributions. This general procedure is continued to calculate the remaining attributable fractions, with total risk shown as the total height of the column at each age.

Table I

Full model and Reduced model dose-response for CPS-I, CPS-II, and British Doctors' LC mortality cohorts

Full model.		
Initial scale:		
$X=10^7$		Assume 10^7 normal stem cells in both lungs *
		Equate background initiation, malignant conversion rates *
Background variables for cohort		
Rate	Parameter	
α_0	p_1 (CPS-I)	Background or non-smoking cell division rate * (per cell per year)
	p_1' (Brit. docs.)	
$g_0=\alpha_0-\beta_0-\mu_0$	p_2	Background or non-smoking net cell promotion rate (per cell per year)
$\nu_0=\mu_0$	p_3	Background initiation, malignant conversion * (per cell per year)
Full model dose-response for cohort		
$\nu_i=\nu_0(1+p_4\ dose_i^{p_8})$		Initiation rate * (per cell per year)
$g_i=g_0(1+p_5\ dose_i^{p_6})$		Net initiated cell promotion rate (per cell per year)
$\alpha_i=\alpha_0(1+(p_5+p_9)\ dose_i^{p_6})$		Initiated cell division rate * (per cell per year)
$\mu_i=\mu_0(1+p_{10}\ dose_i^{p_{11}})$		Malignant conversion rate * (per cell per year)
Fixed lag:	Gamma distribution with random variable X as lag time:	
$t_{lag}=p_7$	$f(x, a, b)=\frac{1}{b\Gamma(a)}x^{a-1}e^{-x/b}$, $\mu_{lag}=p_{12}=a\ b$, $\sigma_{lag}=p_{13}=\sqrt{a\ b^2}$	
Reduced model.		
X and background rates ν_0 , α_0 , g_0 , and μ_0 are parameterized the same as in the full model above.		
Reduced model dose-response for the British Doctors', CPS-I and CPS-II cohorts		
$\nu_i=\ \nu_0(1+p_4)$, $p_4=0$ for nonsmokers		Initiation rate * (per cell per year)
$g_i=g_0(1+p_5\ dose_i^{p_6})$		Net initiated cell promotion rate (per cell per year)
$\alpha_i=\alpha_0(1+p_5\ dose_i^{p_6})$		Initiated cell division rate * (per cell per year)
$\mu_i=\mu_0$		Malignant conversion rate *, no dose-response (per cell per year)
Reduced model dose-response for the HPFS and NHS cohorts		
$\nu_i=\nu_0$		Initiation rate *, no dose-response (per cell per year)
$g_i=g_0(1+p_5\ dose_i^{p_6})$		Net initiated cell promotion rate (per cell per year)
$\alpha_i=\alpha_0(1+p_5\ dose_i^{p_6})$		Initiated cell division rate * (per cell per year)
$\mu_i=\mu_0(1+p_{10}\ dose_i^{p_{11}})$		Malignant conversion rate * (per cell per year)
Fixed lag:		

Full model.

$t_{lag} = 5 \text{ years}$

Note: A full model was first tested where initiation, promotion, and malignant conversion are assumed to be potentially modified by smoking through flexible dose-response parameterization of the form: $a(1 + bx^c)$. Likelihood ratio tests were used to eliminate variables that were not significant for any of the cohorts, resulting in the reduced model.

* Only the product νX is identifiable. Thus, without loss of generality, we assume $X = 10^7$ normal stem cells at risk. This number may be changed by reciprocal change in ν . A separate identifiability condition relates variables α , μ , and ν . Thus, without loss of generality, we set $\mu(0) = \nu(0)$. These variables may be rescaled without changing the model fit.

Table II

Reduced model parameter estimates for CPS-I, CPS-II, and British Doctors' LC mortality cohorts [MLE (MCMC 95% CI)]

Parameter	p_1	p_2	p_3	p_4	p_5	p_6
Cohort	Background cell division rate (per cell per year)	Background net cell proliferation rate (per cell per year)	Background initiation and malignant conversion rate (per cell per year)	Tobacco initiation rate	Tobacco promotion rate coefficient	Tobacco promotion rate power
CPS-I males	22.65(17.35,31.31)	0.075(0.065,0.085) *	1.40e-7(1.12e-7,1.76e-7) *	1.79(1.11,3.07) *	0.21(0.15,0.28) *	0.47(0.42,0.54) *
British Doctors	5.87(2.66,12.89)	0.075(0.065,0.085) *	1.40e-7(1.12e-7,1.76e-7) *	1.79(1.11,3.07) *	0.21(0.15,0.28) *	0.47(0.42,0.54) *
CPS-II males	7.70(6.45,12.99)	0.090(0.071,0.106)	7.16e-8(4.60e-8,1.21e-7)	~ 0.00(~ 0.00,1.76)	0.60(0.43,0.91)	0.22(0.12,0.30)
CPS-II females	15.82(13.39,42.12)	0.071(0.055,0.088)	1.07e-7(6.97e-8,1.62e-7)	0.02(~ 0.00,12.50)	0.50(0.27,0.86)	0.32(0.14,0.40)

Note: Estimates for net cell proliferation rates are tightly constrained, whereas cell division rates are not due to compensatory cell death.

* Shared variables in joint fit to White males in CPS-I and British Doctors' cohorts.

Table III
Reduced model estimates for the HPFS and NHS LC mortality cohorts [MLE (MCMC 95% CI)]

Parameter	p_1	p_2	p_3	p_5	p_6	p_{10}	p_{11}
Cohort	Background cell division rate (per cell per year)	Background net cell proliferation rate (per cell per year)	Background initiation and malignant conversion rate(per cell per year)	Tobacco promotion rate coefficient	Tobacco promotion rate power	Tobacco malignant conversion rate coefficient	Tobacco malignant conversion rate power
HPFS	3.000	0.106 (0.092,0.117)	7.311e-8 (5.303e-8,10.66e-8)	0.134 (0.082,0.220)	0.474 (0.342,0.603)	0.175 (0.049,0.0.940)	0.552 (0.043,0.878)
NHS	3.000	0.083 (0.066,0.095)	9.370e-8 (7.200e-8,12.94e-8)	0.189 (0.119,0.317)	0.491 (0.391,0.591)	0.087 (0.022,0.439)	0.711 (0.192,1.060)

Note: Cell division rates are not tightly constrained due to compensatory cell death, as discussed in the caption to Table II. Consistent with this, we found the likelihood surface essentially flat in response to changes in cell division rate (p_1). Thus p_1 was held constant at a value of 3,000 for both cohorts during subsequent optimization.