Multistage Carcinogenesis and Lung Cancer Mortality in Three Cohorts

William D. Hazelton,1 Mark S. Clements,2 and Suresh H. Moolgavkar1

1Fred Hutchinson Cancer Research Center, Public Health Sciences Division, Seattle, Washington and 2National Centre for Epidemiology and Population Health, Australian National University, Canberra, ACT, Australia

Abstract

Experimental evidence indicates that tobacco smoke acts both as an initiator and a promoter in lung carcinogenesis. We used the two-stage clonal expansion model incorporating the ideas of initiation, promotion, and malignant conversion to analyze lung cancer mortality in three large cohorts, the British Doctors’ cohort and the two American Cancer Society cohorts, to determine how smoking habits influence age-specific lung cancer rates via these mechanisms. Likelihood ratio tests indicate that smoking-related promotion is the dominant model mechanism associated with lung cancer mortality in all cohorts. Smoking-related initiation is less important than promotion but interacts synergistically with it.

Introduction

The primary cause of lung cancer is exposure to tobacco smoke (1-3) followed (not necessarily in order) by radon, environmental and workplace carcinogens, nutrition, genetic susceptibility, and unknown factors that may be expressed as period and birth cohort effects.

Numerous carcinogenic components have been identified in the particulate and gas phases of cigarette smoke. Animal experimentation and other assays have thus far identified ~40 human or animal carcinogens in the >4,000 chemical compounds that have been identified in tobacco smoke (3). Major lung carcinogens in smoke include some of the polycyclic aromatic hydrocarbons, such as benzo[a]pyrene, as well as some tobacco-specific N-nitrosamines. These chemicals, and perhaps other compounds that have not been identified, may act synergistically as initiators, promoters, or both (4, 5).

The goal of this study is to analyze lung cancer mortality in three cohorts, the British Doctors’ cohort and the two American Cancer Society cohorts, with detailed information on cigarette smoking habits. For the analyses, we used a multistage model for carcinogenesis that incorporates the biological notions of initiation, promotion, and malignant conversion. Smoking may influence any of the biological processes in this model. Maximum likelihood methods are used to optimize the models to the cohort data, determine which processes are influenced by smoking, and find the corresponding dose-response relationships for the affected processes. The estimated model variables are used to predict 10-year lung cancer risks for hypothetical smoking scenarios, including the risk among smokers who quit. Thus, the models can be thought of as analogues for lung cancer of the Gail model (6) for breast cancer.

Materials and Methods

Subjects. The two American Cancer Society Cancer Prevention Study (CPS) cohorts, CPS-I and CPS-II, each included >1 million individuals. The CPS-I cohort was followed for 12 years beginning in 1959. The CPS-I data used in this study consist of White male and female current and never smokers as reported in the Smoking and Tobacco Control, Monograph 8, Chapter 3, Appendix B (7). The data provide information by age and duration by 5-year bins and by smoking rate as “never smokers,” “1” to 9, 10 to 19, 20, 21 to 39, and ≥40 cigarettes per day bins. The analysis of CPS-I White males included 960,220 person-years with 215 lung cancers among never smokers and 1,669,126 person-years and 3,437 lung cancers among current smokers. The analysis of CPS-I White females included 4,030,263 person-years with 573 lung cancers among never smokers and 1,628,783 person-years with 215 lung cancers among never smokers and 1,669,126 person-years and 3,437 lung cancers among current smokers.

The analysis of CPS-II White males included 1,628,783 person-years with 600 lung cancers among never smokers and 179,689 person-years and 76 lung cancers among current smokers. The analysis of CPS-II women included 1,475,943 person-years and 734 lung cancers among current smokers. The CPS-II cohort and the two American Cancer Society cohorts, with detailed information on cigarette smoking habits, were included.) The available data are limited to individuals ages ≥50 years in 5-year age bins, including never smokers, and current smokers restricted to either 20 or 40 cigarettes per day and at least 30 years’ duration by 5-year bins. The analysis of CPS-II White males included 559,156 person-years and 76 lung cancers among never smokers and 179,689 person-years and 734 lung cancers among current smokers. The analysis of CPS-II White females included 1,475,943 person-years and 734 lung cancers among current smokers.

The CPS-II study began recruitment in 1982. We analyzed CPS-II data with 6-year follow-up as tabulated in the Smoking and Tobacco Control, Monograph 8, Chapter 4, Appendix 5 (7). The data for White male and female current and former smokers excluded persons with prevalent cancer other than nonmelanoma skin cancer. (This differed slightly from the CPS-I subcohort where all White males and females were included.) The available data are limited to individuals ages ≥50 years in 5-year age bins, including never smokers, and current smokers restricted to either 20 or 40 cigarettes per day and at least 30 years’ duration by 5-year bins. The analysis of CPS-II White males included 559,156 person-years and 76 lung cancers among never smokers and 179,689 person-years and 734 lung cancers among current smokers. The analysis of CPS-II White females included 1,475,943 person-years and 734 lung cancers among current smokers.
person-years and 192 lung cancer deaths among never smokers and 165,057 person-years and 374 lung cancers among current smokers. The British Doctors’ cohort (8, 9) consisted of 34,439 male doctors who replied to a postal questionnaire in 1951. This cohort has been followed by Doll et al. for 50 years (9). However, we only had access to categorical data presented at 20 years’ follow-up for mortality from lung cancer and other causes (8). The data are binned by 5-year age groups and smoking intensity for current smokers categorized as never smokers and current smokers of 1 to 4, 5 to 9, …, 30 to 34, and 35 to 40 cigarettes per day. As the age at initiation of smoking was unavailable for the British Doctors, we assumed it to be age 19 years as in Doll and Peto (8) and Moolgavkar et al. (10). The sensitivity of the model fit to this choice of initiation age was checked by systematically varying the age by 1-year increments. The consistency with CPS-I also suggests that this assumption is reasonable. The analysis of male British Doctors included 69,908 person-years and 6 lung cancers among never smokers and 109,353 person-years and 195 lung cancers among current smokers. The data were not pooled across cohorts in any of our analyses.

The three cohorts combined include 3,547,452 person-years of observation and 4,663 deaths from lung cancer among males and 7,300,046 person-years of observation and 1,739 deaths from lung cancer among females. The available CPS-I data are more informative than for the other cohorts because of more person-years factored into more extensive age, dose, and duration categories and longer follow-up than for CPS-II. The results presented here focus more on the male cohorts where variable estimates are more tightly constrained because of the larger fraction of smokers, higher average consumption per smoker, and greater number of lung cancers. We were also concerned that there might be more reporting bias for females than for males in the early cohorts when smoking by females was less socially acceptable.

A limitation of these data is that no information on ex-smokers is available. Despite this, we are able to make projections of lung cancer risk among smokers who quit as is described below.

Two-Stage Clonal Expansion Model. We use a model that explicitly acknowledges three distinct phases in the process of carcinogenesis. In the first phase (initiation), a susceptible stem cell acquires one or more mutations resulting in an increased population of cells that have already acquired some of the genetic alterations on the pathway to malignancy. Finally, in the last phase (malignant conversion), one of the initiated cells acquires the further genetic changes required to convert it into a malignant cell. There is considerable evidence that most human malignancies go through these three phases. Environmental agents, such as radiation and tobacco smoke, influence carcinogenesis via their effects on one or more phases of the process. The two-stage clonal expansion (TSCE) model is the simplest stochastic mathematical model of the initiation, promotion, and malignant conversion paradigm of carcinogenesis (11-16), with the “stages” corresponding to the “mutational” phases of initiation and malignant conversion.

The hazard (incidence) function generated by the TSCE model and its use for analyses of epidemiologic data have been described in several publications (12-16). In an earlier analysis of the British Doctors’ data (10), we used a mathematically tractable approximation to the hazard function. We believe now, however, that the approximate form of the hazard function can yield misleading results and should not be used. One specific property of the exact hazard function, not possessed by commonly used approximations (17-19), is particularly relevant to the analyses here and that is the following. After exposure to a carcinogen ends, for example, after smokers quit, the exact hazard function ultimately approaches the hazard function in the unexposed. That is, even without invoking tissue repair, the incidence of lung cancer among former smokers would be expected to approach the incidence in nonsmokers. It is this property of the hazard function that allows us to predict risk among ex-smokers, although in the cohorts analyzed here we have information only on nonsmokers and continuing smokers.

The key variables of the model are as follows: \( v \) is the rate of initiation, \( \alpha \) and \( \beta \) are the rates of cell division and apoptosis of initiated cells, respectively, and \( \mu \) is the rate of malignant conversion of initiated cells. In fitting the model to data, we use the variables \( v, \alpha, \beta, \) and \( \alpha - \beta \), the net rate of promotion (see Table 1). The effect of smoking on cancer rates is modeled via a dose-response function on each of these variables. Progression from the first malignant cell to death from lung cancer is modeled either as a constant lag time or as a two-variable \( I \) function to represent the distribution of lag times from first malignant cell to death from lung cancer in the population. Likelihood construction and variable identifiability are discussed in recent articles (16, 20). Likelihood ratio tests are used to determine the importance of individual variables. Profile likelihood-based 95% confidence intervals (95% CI) were

### Table 1. Reduced model dose-response applied to all cohorts

| Initial scale | Variable | Background or nonsmoking initiated cell division rate (per cell per year) | Background or nonsmoking net initiated cell promotion rate (per cell per year) | Background or nonsmoking initiation, malignant conversion (per cell per year) |
|---------------|----------|----------------------------------------------------------------------------|----------------------------------------------------------------------------|
| \( X = 10^7 \) | \( v_0 = \mu_0 \) | Assume \( 10^7 \) normal stem cells in both lungs | Equate background initiation, malignant conversion rates |
| Background variables for cohort | Rate | \( g_0 = \alpha_0 - \beta_0 - \mu_0 \) | \( p_1 \) |
| \( g_0 = \alpha_0 - \beta_0 - \mu_0 \) | \( p_2 \) |
| \( v_0 = \mu_0 \) | \( p_3 \) |
| Reduced model dose-response for cohort | | | | |
| \( v = v_0(1 + p_4) \) | Initiation rate (per cell per year) |
| \( g = g_0(1 + p_5) \) | Net initiated cell promotion rate (per cell per year) |
| \( \alpha = \alpha_0(1 + p_6) \) | Initiated cell division rate (per cell per year) |
| \( \mu = \mu_0 \) | Malignant conversion, no dose-response (per cell per year) |

**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: \( z(1 + bx^a) \). 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 \( xv \) 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 \( v \). A separate identifiability condition relates variables \( x, \mu, \) and \( v \). Thus, without loss of generality, we set \( \mu_0 = v_0 \). These variables may be rescaled without changing the model fit. See ref. 20.
consistent with those constructed by Markov chain Monte Carlo (MCMC) methods using the Metropolis-Hastings algorithm.

The TSCE model allows calculation of age-specific risk as a function of the smoking intensity (cigarettes per day) throughout life. The likelihood contribution made by each individual depends on the entire covariate history for that individual (i.e., age at start of smoking, daily intensity of smoking, and duration of smoking). If information on ex-smokers was available, the likelihood would depend in addition on age of quitting smoking and time since quitting smoking.

Ten-Year Risk Predictions. Models optimized for males and females in the CPS-I and CPS-II cohorts were used to predict 10-year risk estimates from dying of lung cancer. Competing causes of mortality were adjusted using standard actuarial methods for multiple decrement life tables. All cause annual risk estimates were extracted from the National Center for Health Statistics (21). We used the 1969 to 1971 age-specific and sex-specific life tables for CPS-I and the 1989 to 1991 life tables for CPS-II. 95% CIs were calculated by sampling model variables from a MCMC calculation using the Metropolis-Hastings algorithm.

Results

The male and female cohorts from CPS-I and CPS-II and the male British Doctors’ cohort were analyzed individually and jointly in all combinations. Each analysis began with a full model where the biological variables of the TSCE model were allowed to have a flexible dose-response pattern of the form: \(a(1 + bx^2)\). Likelihood ratio tests were used to eliminate variables that were not significant for any cohort, resulting in a reduced model, shown in Table 1.

For each cohort fitted separately, the reduced model has six variables, of which three determine background rates and three determine smoking dose-response. The six-variable reduced model provides good fits to lung cancer among never smokers. We also attempted to find a single model that would describe all three cohorts simultaneously; however, we were unable to do so. We did find, however, that a single model with a total of seven variables (three common dose-response and two common background variables; separate background \(\alpha\) in the two cohorts) did an excellent job of describing the contemporaneous British Doctors’ and CPS-I cohorts simultaneously. By way of comparison, a recent analysis of CPS-I and British Doctors’ data by Knöke et al. used what they call an extended Doll and Peto model requiring four variables for each cohort to describe smokers plus two variables to describe CPS-I nonsmokers separately, with no predictive ability to describe the risk for ex-smokers (22).

The British Doctors’ data, Knöke et al. fit lung cancer mortality only among continuing smokers. Thus, Knöke et al. require 10 variables to describe lung cancer among smokers in the two cohorts and lung cancer among CPS-I nonsmokers. Our combined model for the CPS-I and British Doctors’ cohorts requires only seven variables to describe lung cancer among both smokers and nonsmokers in both cohorts. Moreover, as shown in Fig. 3, we are able to make predictions of risk among ex-smokers, which are in good agreement with the risks estimated in other studies.

The model fits to the different cohorts for both genders were generally similar in showing a significant dose-response for tobacco affecting promotion among all cohorts, a significant dose-response for initiation only in the contemporaneous CPS-I and British Doctors’ cohorts, and a nonsignificant effect for malignant conversion in all cohorts. Most of the corresponding variable estimates were statistically indistinguishable in comparing the male CPS-I and British Doctors that were recruited around the same time. The male CPS-II cohort, recruited ~20 years later, required different variables than for the earlier cohorts, with significantly increased dose-response for promotion compared with the other two male cohorts and nonsignificant dose-response for initiation. There are significant gender differences in each cohort. Females were qualitatively similar to males in that the CPS-II female cohort had higher dose-response estimates for promotion and small dose-response estimates for initiation, whereas the earlier CPS-I female cohort had significant dose-response for both initiation and promotion. Table 2 shows variable estimates and MCMC 95% CI for reduced model fits to males and females in each cohort. We calculated 95% CI for model variables based on MCMC runs of a million cycles for each cohort.

We found smoking-related promotion to be the most significant mechanism associated with age-specific lung cancer mortality in all cohorts. Promotion consists of two components, the net cell proliferation rate and the division rate for initiated stem cells. In all cohorts, except for CPS-I females, both the net proliferation rate and the division rate of initiated cells were found to have significant nonlinear (concave
downward) dose-response as a function of increasing smoking rate. A joint model for males in CPS-I and the British Doctors using the same dose-response, but differing only in the background cell division rates, provided excellent descriptions of lung cancer mortality in each cohort as judged by fits to the data binned by age, dose, and duration. The higher background rate of division of initiated cells in the CPS-I cohort implies a higher extinction rate for initiated cells, leading to a lower asymptotic hazard, and a slower increase in risk several decades after smoking begins. Dose-response estimates for net cell proliferation rates for CPS-II are elevated significantly above those of the earlier cohorts over the full smoking range, but estimates of initiation are significantly lower, with no significant dose-response for initiation in CPS-II. These dose-response curves along with 95% MCMC CIs are shown in the Appendix.

Fits by dose, duration, and age and residual plots are shown for CPS-I White males in Figs. 1 and 2. The other cohorts were also fit well by the models as shown in the Appendix. The age-specific lung cancer risks binned by dose and duration for CPS-II are generally elevated above the risks for CPS-I at older ages.

Figure 1. CPS-I White male mortality rates for never and current smokers with 12 years’ follow-up by (A) dose and age and (B) duration and age. Solid line, MLE from joint reduced model fit to CPS-I and British Doctors; dashed lines, MCMC 95% CI; circles, ratio of observed lung cancer deaths to person-years at risk in 5-year age bins, with bars based on Poisson assumptions.
Figure 3 shows predictions of the effects of smoking cessation based on the fit of the combined model for CPS-I and the British Doctors’ cohorts. We assumed smoking began at age 19 years at either 20 or 40 cigarettes per day and compared the risk following stopping smoking at age 30 or 50 years to the risk for individuals continuing to smoke at the same rates. These results are seen to be in reasonably good agreement with relative risks for ex-smokers as a function of time since smoking cessation observed in Great Britain by Peto et al. (23) and in two large U.S. cohorts, the Nurses’ Health Study of women and Health Professionals Study of men (24).

Ten-year risk predictions with MCMC 95% CIs are shown in Table 3 for different smoking patterns among continuing smokers and for smokers who quit at the beginning of the 10-year risk-projection period. These calculations may overestimate the 10-year risk of dying from lung cancer for heavy smokers, because population-based annual life tables (21) were used to adjust for competing risk. (Separate life tables specific to different smoking levels were not available.) The calculation of risk for smokers who quit at the beginning of the 10-year interval was made by assuming dose-response functions return to background levels when smoking stops. Risk estimates have generally increased between CPS-I and CPS-II, with the largest increases seen among females. The risk estimates for CPS-II are somewhat less than recent estimates based on modeling using data from the Carotene and Retinol Efficacy Trial (25).

Discussion

Multistage Models. We are not suggesting that detailed biological mechanisms can be inferred from epidemiologic data using multistage carcinogenesis models. However, temporal patterns of risk are strongly influenced by whether mutation rates or rates of net cell proliferation are affected by an environmental agent. These temporal patterns, in turn, determine the maximum likelihood estimates (MLE) and therefore inferences regarding whether mutation rates or promotion or both are affected by an environmental agent. In our experience (26), both broad inferences regarding mechanism and the hazard functions are insensitive to choice of model, providing promotion is considered explicitly in the models.

For analyses of epidemiologic data, it is important to note that the exact hazard functions of rather general multistage models possess properties that are not reflected in the empirical hazard functions. We emphasize here that for these properties to hold the exact hazard function must be used; commonly used approximations, such as the power law approximation to the Armitage-Doll multistage model, do not possess these properties.

The first property is that the hazard tends to level off at old age. This is a prediction of the exact solution for any multistage model (17-19) and is in agreement with epidemiologic data for cancers of many types. Alternative explanations have been offered for the falloff in risk, including underascertainment of cancers among the elderly. Although ascertainment bias may contribute to an apparent leveling of risk at old age, the leveling is predicted by the multistage models. A second property is that the hazard following cessation of exposure to a carcinogenic agent will eventually approach the hazard for individuals never exposed to the agent. Thus, the risk for ex-smokers should eventually approach that of never smokers, consistent with observed trends for ex-smokers. This prediction is based on the assumption that mutation and
promotion rates return to background levels after smoking stops. The reversion of the hazard function for lung cancer among ex-smokers to nonsmoker levels is often attributed to repair of lung lesions. Repair may well play a role; however, this behavior of the hazard function is predicted by multistage carcinogenesis (17-19).

A third property is that carcinogenic agents that affect different steps in the multistage process lead to distinct patterns of risk following exposure. An agent that affects initiation, an early step, will show greater delay before an increase in risk than an agent that affects promotion or malignant conversion. Agents affecting promotion tend to give a prolonged “plateau” of increased risk. Agents that affect malignant conversion produce a relatively prompt increase in risk following exposure. Previous articles have examined the temporal evolution of risk following changes in intensity of exposure to carcinogens but only within the framework of the approximate form of the Armitage-Doll model (27-29).

Finally, multistage modeling allows analysis of data with complex exposure patterns to multiple agents that would be difficult or impossible using standard statistical techniques. The entire covariate history can be explicitly considered for exposure to multiple carcinogenic agents. Thus, for each agent, age at start of exposure, intensity of exposure, age at stop of exposure, and time since stop of exposure are all used for construction of the likelihood. An example of analyses of lung cancer mortality with complex exposure patterns to three lung carcinogens is provided by our recent analysis of the Chinese Tin Miners’ data (26).

An understanding of the importance of different mechanisms in carcinogenesis may suggest better approaches to prevention, intervention, and treatment. This study identifies promotion as the dominant mechanism influencing tobacco-induced lung cancer risk in all cohorts. Based on this finding, it may be worth assessing if pharmacologic intervention could be targeted toward slowing cigarette-induced promotion among smokers. Recent analyses (13, 19) suggest that the antipromoting effects of cyclo-oxygenase-2 inhibitors in colon cancer is an efficient way to decrease risk even late in life.

It is certainly possible that some of the differences in risk we see between cohorts may be due to factors other than differences in reported smoking behavior or changing cigarette composition. For example, unreported differences in ramping-up during the uptake of cigarette smoking may have occurred between the earlier and the later cohorts. This could potentially influence risk patterns that we associate with a decreasing role of initiation and increasing role of promotion in conjunction with changes in cigarette composition.

**Previous Analyses Using Multistage Models.** The predictions in this article are based on a particular multistage model of carcinogenesis. Other forms of the model have been fitted to observational data for lung cancer, particularly the Armitage-Doll multistage model (29-35). Different models can offer similar statistical fits to observed data while having qualitatively different characteristics. Ideally, the selection of an epidemiologic model should be based on an understanding of the biology. For lung cancer, there is good evidence for initiation and promotion steps, suggesting the use of a model that includes clonal expansion. Further, we strongly advocate the use of exact solutions of whichever multistage model is used. As noted above and discussed in more detail elsewhere (16-19), commonly used approximations, such as the power law approximation to the Armitage-Doll model, have properties that are qualitatively quite different from those of the exact hazard functions.

In a recent article, Knoke et al. have suggested that the TSCE model provides a poor description of certain features of the data (22). There could be two reasons for their findings. First, they considered only the approximate version of the model, which was used by one of us in an article written ~15 years ago. Since then, we have emphasized the importance of using the exact solution in several recent articles, some of them analyzing other lung cancer data sets. Second, Knoke et al. assume linear dose-response relationships for the variables of the model. In fact, in this article, we find that the dose-response relationship for promotion is strongly nonlinear. Analyses of residuals show that the TSCE model provides excellent descriptions of all three data sets considered here (see Fig. 2).
Table 3. Ten-year risk projections for smokers who smoke for 25, 40, or 50 years and continue to smoke or quit at ages 55, 65, or 75 years based on models for White male and female smokers in the CPS-I and CPS-II cohorts [% risk (95% CI)]

<table>
<thead>
<tr>
<th></th>
<th>25 y Quit</th>
<th>Still smoking</th>
<th>40 y Quit</th>
<th>Still smoking</th>
<th>50 y Quit</th>
<th>Still smoking</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPS-I males: 1 pack daily smokers</td>
<td>0.7 (0.6-0.7)</td>
<td>0.8 (0.7-0.8)</td>
<td>1.8 (1.7-1.9)</td>
<td>1.9 (1.9-2.0)</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>CPS-I males: 2 packs daily smokers</td>
<td>1.2 (1.1-1.2)</td>
<td>1.3 (1.2-1.4)</td>
<td>3.4 (3.2-3.6)</td>
<td>3.9 (3.6-4.1)</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>CPS-I females: 1 pack daily smokers</td>
<td>0.9 (0.7-1.0)</td>
<td>1.0 (0.9-1.2)</td>
<td>2.9 (2.5-3.2)</td>
<td>3.3 (2.9-3.7)</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>CPS-II males: 1 pack daily smokers</td>
<td>1.7 (1.5-1.9)</td>
<td>2.0 (1.7-2.2)</td>
<td>4.7 (4.2-5.2)</td>
<td>5.3 (4.7-5.8)</td>
<td>6.6 (5.7-7.6)</td>
<td>7.2 (6.3-8.3)</td>
</tr>
<tr>
<td>CPS-II males: 2 packs daily smokers</td>
<td>2.6 (2.4-3.1)</td>
<td>2.9 (2.4-3.4)</td>
<td>5.7 (4.9-6.4)</td>
<td>6.2 (5.3-7.0)</td>
<td>6.9 (5.9 8.0)</td>
<td>7.4 (6.3-8.5)</td>
</tr>
<tr>
<td>CPS-I females: 1 pack daily smokers</td>
<td>1.2 (1.1-1.4)</td>
<td>1.5 (1.3-1.7)</td>
<td>4.3 (3.7-4.8)</td>
<td>5.0 (4.3-5.6)</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>CPS-I females: 2 packs daily smokers</td>
<td>0.4 (0.3-0.4)</td>
<td>0.4 (0.4-0.5)</td>
<td>0.8 (0.7-0.9)</td>
<td>0.8 (0.7-0.9)</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>CPS-II females: 1 pack daily smokers</td>
<td>1.0 (0.8-1.2)</td>
<td>1.2 (0.9-1.4)</td>
<td>1.8 (1.4-2.2)</td>
<td>2.0 (1.5-2.5)</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>CPS-II females: 2 packs daily smokers</td>
<td>1.3 (1.1-1.6)</td>
<td>1.5 (1.2-1.8)</td>
<td>1.9 (1.4-2.4)</td>
<td>2.1 (1.6-2.7)</td>
<td>2.0 (1.3-2.8)</td>
<td>2.2 (1.5-3.1)</td>
</tr>
<tr>
<td></td>
<td>75</td>
<td></td>
<td>1.4 (1.0-1.7)</td>
<td>1.5 (1.1-1.9)</td>
<td>1.6 (1.2-2.2)</td>
<td>1.8 (1.3-2.4)</td>
</tr>
<tr>
<td></td>
<td>55</td>
<td></td>
<td>0.6 (0.5-0.7)</td>
<td>0.7 (0.6-0.8)</td>
<td>0.8 (0.6-0.9)</td>
<td>1.1 (0.9-1.3)</td>
</tr>
<tr>
<td></td>
<td>65</td>
<td></td>
<td>1.1 (0.9-1.2)</td>
<td>1.2 (1.0-1.4)</td>
<td>1.2 (1.0-1.4)</td>
<td>2.6 (2.2-3.0)</td>
</tr>
<tr>
<td></td>
<td>75</td>
<td></td>
<td>1.5 (1.2-1.9)</td>
<td>1.7 (1.3-2.1)</td>
<td>1.7 (1.3-2.1)</td>
<td>3.2 (2.4-3.9)</td>
</tr>
</tbody>
</table>

NOTE: Life tables are used to adjust for death from competing causes. Model-based 10-year risks are shown for each smoking pattern, with MCMC 95% CIs. Projections for individuals who quit smoking and continue to abstain for the following 10 years assume that model variables revert to background values following smoking cessation. Asterisks are placed in cells that correspond to unrealistically early ages for starting smoking. These rates are not generalizable and are probably lower than expected for the general population, as the cohorts were generally more educated and healthy, especially CPS-II where prevalent cancers were excluded.

Moreover, without information on ex-smokers, we are able to make predictions of risk among ex-smokers quitting at different ages that are in good qualitative agreement with the information available from other sources (refs. 9, 23, 24; see Fig. 3).

Mechanisms in Smoking-Related Lung Cancer. The analyses indicate that promotion is the most important mechanism (both in terms of statistical significance and attributable risk) in smoking-related lung cancer among males and females in all cohorts. Analysis of the earlier CPS-I and British Doctors’ cohorts indicates that initiation is also significant, but with a smaller effect on lung cancer mortality, gaining importance primarily through a synergistic interaction with promotion. Analyses of the CPS-II male and female cohorts show an increased dose-response for promotion, with nonsignificant estimates for initiation, although an effect of initiation is not excluded. The difference in dose-response for initiation between CPS-I and CPS-II is significant (P = 0.01), even when background initiation rates and all other variables are allowed to differ. The apparent increase in the role of promotion and decrease in the role of initiation in the later cohorts may be associated with changes in cigarette composition.

We ran several sensitivity analyses to test if the shorter (6-year) follow-up for CPS-II, or the exclusion of prevalent cancers, could contribute to decreased power to detect a dose-response effect for initiation. We analyzed 6-year follow-up data for CPS-I, finding a significant dose-response for both initiation and promotion consistent with estimates using the 12-year CPS-I follow-up. Estimates for initiation using 6-year follow-up data for CPS-I are significantly larger than for the 6-year follow-up data for CPS-II as found in the analysis of the 12-year CPS-I follow-up data. We also analyzed 6-year CPS-II data for males of all races, including prevalent cancers, finding a nonsignificant dose-response for initiation (although a reduced effect of initiation is not excluded) consistent with the CPS-II analysis for White males, excluding prevalent cancers.

Smoking seems to cause significant and remarkably consistent increases in promotion and initiation as functions of dose in the contemporaneous CPS-I and British Doctors’ cohorts. An influence on initiation is consistent with evidence for mutagenic effects of known carcinogens in tobacco, such as the nitrosamines and the polycyclic aromatic hydrocarbons, which are strong initiators (4). Some of these compounds also seem to be important as promoters (36-39). Other compounds in cigarette smoke condensate, such as catechol, are also co-carcinogens, affecting both initiation and promotion (4), but there seem to be many pure initiators and pure promoters.

Promotion drives the clonal expansion of initiated cells that leads to recognizable premalignant lesions in lung cancer (38-40). Promotion includes two components: net cell proliferation and cell division of initiated cells. An increase in the net cell proliferation rate of initiated cells is associated with higher cancer risk, as this is just the mean rate at which initiated clones expand. However, given a fixed net cell proliferation rate, a decrease in the cell division rate implies a lower...
extinction rate for initiated clones and thus a higher asymptotic risk at older ages. This occurs for the British Doctors who have a three to four times lower estimate for the background rate of cell division than the estimate for males in CPS-I and have correspondingly higher risks at older ages.

The effects of dose-response on initiation and promotion are highly synergistic in the CPS-I and British Doctors’ cohorts (see Fig. 3S in the Appendix). With a dose-response only on initiation (at the rate estimated in the joint model), the excess risk would be very modest, rising slowly at older ages. The effect of dose-response only on promotion leads to a steep increase in risk around age 60 years followed by a plateau. The influence of promotion alone seems much more important than initiation, taken by itself. This is consistent with evidence cited above, indicating that the nitrosamines and the polycyclic aromatic hydrocarbons play major roles as promoters in carcinogenesis. Although the polycyclic aromatic hydrocarbons have been considered to be the main carcinogens in cigarettes, their concentration in cigarette condensate is far too low for the initiating effect to drive carcinogenesis (38). With dose-response affecting both initiation and promotion, the risk is increased much more than additively. This is consistent with the concept of the synergistic interaction between co-carcinogens, where initiation and promotion occur simultaneously (4).

As discussed in Materials and Methods, recruitment for CPS-I began in fall 1959, with most cohort members born between 1880 and 1919, whereas recruitment for CPS-II began in fall 1982 with most subjects born between 1900 and 1939. Over this period, there were significant changes in the availability of machine-manufactured cigarettes, cigarette composition, social norms, and other factors, such as nutrition, which could contribute to differences in the patterns of lung cancer mortality. Although the CPS-I and CPS-II studies used comparable techniques, there are some potentially important differences in regional coverage and follow-up. CPS-I enrolled subjects from 25 states, whereas CPS-II included subjects from all 50 states, the District of Columbia, Puerto Rico, and Guam. Reporting of smoking histories may differ across studies and calendar periods, possibly causing changes in variable estimates.

The models show qualitatively consistent changes between the earlier and the later cohorts for both males and females. Risks for a given smoking dose and duration seem to have increased, counter to what might be expected because of the introduction of low-tar and low-nicotine cigarettes. Indeed, the models indicate that initiation rates associated with smoking have significantly decreased, but increased promotion due to smoking has increased the overall risk.

The differences in dose-response may suggest an influence of changes in cigarette composition. An investigation of >170 types of cigarettes from the United States and Europe found no correlation between tar delivery and mainstream smoke concentration of nitrosamines (41). Tumor promoting activity, as measured by gap junctional intercellular communication, also does not closely reflect the tar yield of low-tar and high-tar cigarettes (42). Smokers of low-tar and low-nicotine cigarettes partially compensate by inhaling higher volumes of smoke (43). With enhanced smoke intake, smokers of low-tar and low-nicotine cigarettes may have higher intake of nitrosamines and/or enhanced tumor promotion compared with smokers of higher nicotine cigarettes.

The apparent reduction in smoking-related initiation may also be associated with the properties of low-tar cigarettes. A study of mutagenicity of tobacco smoke from a series of 16 low-tar (1-10 mg tar) cigarettes showed significantly reduced mutagenicity (by ~30%) by the Ames test for mainstream smoke from most of the low-tar cigarettes compared with high-tar (23 mg) cigarettes (44). Comparison of estimated initiation and promotion dose-response between CPS-I and CPS-II may be complicated by a shift from squamous cell carcinoma as the major subtype to increasing rates of adenocarcinoma.

**Projection of Risk.** The class of models used for data analyses in this article can be used to predict the risks of lung cancer in various subgroups. Because the models are fully parametric, it is clear how they can be used to project risks among nonsmokers and continuing smokers. It is less clear how the models can be used to predict risks among ex-smokers. This prediction depends on a property of the exact hazard functions generated by multistage models and which is not shared by the approximate solutions that have hitherto been widely used for analyses of data. When smokers quit, their lungs are no longer exposed to the initiating and promoting agents contained in tobacco smoke, and epidemiologic data indicate that lung cancer incidence rates among ex-smokers approach those among nonsmokers over a period of 15 to 20 years. Biologically, there are two possibilities, not mutually exclusive. (a) After exposure to the initiating and promoting agents in cigarette smoke stops, the rates of initiation and promotion revert to the background rates among nonsmokers and this reversion is sufficient to ultimately cause the incidence rates to revert to background rates. (b) Repair of damaged tissue occurs, for example, promoted lesions regress, and is responsible for the reversion of rates.

Previous analyses of data using multistage models by Freedman and Navidi have found that risk declines after smoking ceases (45). They suggested that repair or heterogeneity may underlie the declining risk but concluded that the first biological scenario above does not explain the decline in risk among ex-smokers. These analyses have relied, however, on approximate solutions to the hazard (incidence) functions. The exact solution, which has been used in this article and our other recent articles, makes entirely different predictions. It predicts that, after exposure to a carcinogen stops, the incidence rates approach the rates among the unexposed asymptotically. Although this result seems to be counterintuitive, a simple explanation is as follows. Because of the stochastic nature of the process of multistage carcinogenesis, there is considerable heterogeneity among any group of smokers with respect to the number and size distribution of initiated lesions. Some fortunate individuals may smoke for years without developing any initiated lesions. If the entire group now gives up smoking, the first individuals to develop lung cancer are those with the largest burden of initiated cells. Ultimately, over the course of many years, the only individuals remaining are those who, through great good fortune, had few or no initiated lesions as a result of their past smoking habits. These individuals are at the same risk as nonsmokers; therefore, the incidence rates now approach the rates among nonsmokers. The exact hazard function of multistage models captures this stochastic variation in risk among ex-smokers; the widely used approximations do not.

**Conclusion**

Promotion seems to be the dominant etiologic mechanism in smoking-related lung carcinogenesis. Significantly increased risk from cigarette smoking among both male and female smokers is seen in the CPS-II cohort when compared with the earlier CPS-I and British Doctors’ cohorts. The increased risk is associated with significant changes in dose-response, including a significant increase in promotion, and significant decrease in initiation for a given smoking intensity. Intervention and treatment strategies may benefit from targeting promotion as a primary mechanism in lung carcinogenesis.
Appendix

Figure S1: British doctors male mortality rates with 20 years follow-up by: (A) dose and age, (B) duration and age. Solid line, maximum likelihood estimate from joint reduced model fit to CPS-I and British doctors, dashed lines, MCMC 95% CI. Circles show the ratio of observed lung cancer deaths to person years at risk in 5-year age bins, with bars based on Poisson assumptions.

Acknowledgments
We thank Dr. Rob Boer and the Cancer Intervention and Surveillance Modeling Network Lung Cancer Group for useful suggestions.

References

Figure S2: Estimates of the dose-response for (A) the net cell proliferation rate, and (B) initiation for the joint reduced model for white male CPS-I and British doctors (solid lines), and for CPS-II (dashed lines). The initiation and promotion rates corresponding to a given smoking rate are assumed to continue for the smoking duration. Separate optimization of the British doctors gives (marginally significant) different dose-response for net cell proliferation (dotted lines). The British doctors initiation rate underlies the joint fit in (B). The symmetric cell division rates are scaled using different background parameters for CPS-I and the British doctors, but share the same dose-response shape as for net cell proliferation. Gray shaded regions, MCMC 95% CI for net cell proliferation and initiation rates as a function of smoking rate, calculated using the full model for CPS-I and the British doctors cohorts. Bars are located at central points in the categorical dose ranges, at 0, 5, 15, 20, 30 and 42 cigarettes per day for CPS-I (solid bars), and at 0, 20, and 40 cigarettes per day for CPS-II (dashed bars), based on MCMC calculations.


Figure S3: Contribution to hazard from background (solid lines), and dose-response on initiation (dotted lines), promotion (dashed lines), or both processes (dot-dashed lines), using the joint reduced model for white male CPS-I (thick lines) and the British doctors (thin lines) cohorts. The hazards are plotted for a hypothetical White male from CPS-I or British doctors’ cohort smoking 40 cigarettes per day beginning at age 15.

Figure S4: Comparison of joint reduced model for White male CPS-I and British doctors with model for White male CPS-II data with 6 years follow-up. (A and B) show rates aggregated by dose and age, and smoking duration and age, respectively. Dotted lines show the prediction of the final CPS-I and British doctors smoking model. Dashed line shows predictions using the same dose-response parameters as above, but with re-optimized background parameters. Solid line shows the predictions of a fully re-optimized final model fit to the CPS-II data.


Multistage Carcinogenesis and Lung Cancer Mortality in Three Cohorts

William D. Hazelton, Mark S. Clements and Suresh H. Moolgavkar


Updated version  Access the most recent version of this article at:
http://cebp.aacrjournals.org/content/14/5/1171

Cited articles  This article cites 33 articles, 7 of which you can access for free at:
http://cebp.aacrjournals.org/content/14/5/1171.full#ref-list-1

Citing articles  This article has been cited by 8 HighWire-hosted articles. Access the articles at:
http://cebp.aacrjournals.org/content/14/5/1171.full#related-urls

E-mail alerts  Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions  To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions  To request permission to re-use all or part of this article, use this link
http://cebp.aacrjournals.org/content/14/5/1171.
Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC)
Rightslink site.