

# Cigarette Smoking and Lung Cancer: Modeling Total Exposure and Intensity

Jay H. Lubin<sup>1</sup> and Neil E. Caporaso<sup>2</sup>

<sup>1</sup>Biostatistics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute and <sup>2</sup>Genetic Epidemiology Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, Maryland

## Abstract

Investigators typically analyze cigarette smoking using smoking duration and intensity (number of cigarettes smoked per day) as risk factors. However, odds ratios (OR) for categories of intensity either adjusted for, or jointly with, duration of smoking may be distorted by differences in total pack-years of exposure to cigarette smoke. To study effects of intensity, we apply a linear excess OR model to compare total exposure delivered at low intensity for a long period of time with an equal total exposure delivered at high intensity for a short period of time to data from a large case-control study of lung cancer. The excess OR per pack-year increases with intensity for subjects who smoke  $\leq 20$  cigarettes per day and decreases with intensity for subjects who smoke  $> 20$  cigarettes per day.

The intensity patterns are homogeneous by histologic type of lung cancer, suggesting that observed differences in risks by histologic type are related to total smoking exposure or smoking duration and not smoking intensity. At lower smoking intensities, there is an "exposure enhancement" effect such that for equal total exposure, the excess OR per pack-year increases with intensity. At higher smoking intensities, there is a "reduced potency" or "wasted exposure" effect such that for equal total exposure, the excess OR per pack-year decreases with intensity (i.e., smoking at a lower intensity for longer duration is more deleterious than smoking at a higher intensity for shorter duration). (Cancer Epidemiol Biomarkers Prev 2006;15(3):517–23)

## Introduction

Unraveling the molecular basis of the smoking and risk of lung cancer, including mechanisms of activation and detoxification of various constituents of tobacco smoke and the genetic basis of smoking persistence, is the focus of many current epidemiologic studies of lung cancer (1). To better understand the biological basis of host reaction to the many chemical compounds in tobacco smoke, it is important to clarify the effects of total cigarette smoke exposure, smoking intensity, and duration in relation to diverse biological end points.

Investigators typically assess the association between cancer occurrence and cigarette smoking using duration of smoking and smoking intensity, as measured by the number of cigarettes smoked per day (1, 2). Faced with the complexity of the association between duration and intensity and disease risk, investigators may apply nonparametric or semiparametric models, such as splines (3) or generalized additive models (4), or create a single comprehensive smoking index (5, 6). However, problems of interpreting multiple characteristics of an exposure remain (7, 8). Comparisons of intensity based on odds ratios (OR) for categories of cigarettes per day either adjusted for, or jointly with, duration of smoking are influenced by differences in total exposure to cigarette smoke. For example, to assess the effect of smoking intensity, one would typically compare ORs of smoking 30 years and smoking 40 years among individuals who smoke 20 cigarettes per day to the ORs of smoking 30 years and smoking 40 years among individuals who smoke 30 cigarettes per day. Differences in the patterns of these ORs for duration would then be ascribed to the effects of intensity. However, differences in

total exposure influence this comparison; i.e., 30 and 40 pack-years of total exposure for the 20 cigarettes per day individuals, respectively, compared with 45 and 60 pack-years for the 30 cigarettes per day individuals.

Studies of smoking and cancers of the lung and the bladder show a leveling of risk for smokers consuming more than 20 to 40 cigarettes per day (9, 10). The precise implication of this pattern is, however, uncertain. Heavy smokers inhaling less deeply, exposure-dependent biases, or behavioral factors are possible reasons for the leveling of risk at higher intensities. Another plausible explanation is a nonlinear relationship of risk and intensity of exposure. Using data from a large case-control study of lung cancer, we apply a model for total pack-years of exposure, which enables a more direct assessment of smoking intensity. Our primary focus is the delivery of total exposure; i.e., the risk associated with total exposure delivered at low intensity for a long period of time compared with the risk with the same total exposure delivered at high intensity for a short period of time. Our interest involves how intensity of smoking influences the association between pack-years and disease or, correspondingly, to what extent the disease-exposure relationship is modified by smoking intensity.

## Materials and Methods

**European Smoking and Health Study of Lung Cancer.** We analyze data from the European Smoking and Health Study, which was a large, multicenter, hospital-based case-control study of lung cancer conducted between 1976 and 1980 at hospitals in seven areas of Europe (Glasgow, Scotland; Hamburg and Heidelberg, Germany; Vienna, Austria; Paris, France; and Milan and Rome, Italy; refs. 11, 12). The study enrolled 7,804 lung cancer cases and 15,207 controls. Controls were frequency matched to cases by center, sex, and categories of age. For current analyses, we limit subjects to ages 50 to 74 years to reduce the effect of any genetic cancer predisposition in younger cases or diagnostic ambiguity in elderly cases. We include never smokers and cigarette-only smokers and exclude 712 subjects (mainly males) who smoked cigars and/or pipes either exclusively or together with cigarettes, 199 subjects who

Received 11/8/05; revised 12/28/05; accepted 1/18/06.

**Grant support:** Intramural Research Program of the National Cancer Institute, NIH.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

**Requests for reprints:** Jay H. Lubin, Biostatistics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, 6120 Executive Boulevard, Rockville, MD 20852. E-mail: lubinj@mail.nih.gov

Copyright © 2006 American Association for Cancer Research.

doi:10.1158/1055-9965.EPI-05-0863

started smoking after the age of 40, and 36 subjects with missing or inconsistent smoking data. Finally, we exclude 2,468 subjects who ceased smoking >5 years before enrollment and thereby limit analysis to never smokers and current or recent former smokers to eliminate the need to model effects of time since cessation of smoking. The final data set includes 4,625 cases (3,991 males and 634 females) and 7,884 controls (6,660 males and 1,224 females).

The primary measure of total exposure is pack-years, which is computed as the product of duration of smoking and mean number of packs (20 cigarettes) smoked per day.

This study was done under approval of all applicable institutional review boards.

**Models.** We first apply a simple linear model for the OR of lung cancer and pack-years of exposure,  $d$ ,

$$OR(d) = 1 + \beta d \tag{A}$$

where  $\beta$  represents the excess OR (EOR) for each pack-year of exposure. To evaluate departures from linearity we use the linear-exponential model,

$$OR(d) = 1 + \beta d \exp(\gamma d) \tag{B}$$

For nonsmokers,  $d = 0$  and  $OR(0) = 1$ . The  $\gamma$  parameter measures downward concavity ( $\gamma < 0$ ) or upward convexity ( $\gamma > 0$ ) with  $d$ , and thus the degree of departure from linearity. A test of the null hypothesis  $\gamma = 0$  is a test of no departure from the linear model in pack-years. We considered other variants for modeling nonlinearity, but Eq. (B) proved satisfactory.

In initial analyses, we find that model (A) fits the data poorly, unless intensity of smoking is included. We use two approaches to model intensity of smoking. The first approach defines  $I$  intensity categories and indicator variables,  $n_i$ ,  $i = 1, \dots, I$ , where  $n_i = 1$  if a subject's intensity level occurs within the  $i$ th category; 0 otherwise. The model is

$$OR(d) = 1 + \beta d \exp\{\sum \theta_i n_i\} \tag{C}$$

which specifies a different slope for each intensity category. With  $\theta_1$  set to zero for identifiability,  $\theta_2, \dots, \theta_I$  represent category-specific effects relative to the  $i = 1$  level. However, our primary interest is  $\beta \exp(\theta_i)$ , which represents the EOR/pack-year for the  $i$ th intensity category. For convenience in the fitting,  $\exp(\beta^*)$  replaces  $\beta$  in Eq. (C) to avoid range restrictions on the  $\beta$  parameter. With this reparametrization, SEs for the parameter estimates translate into multiplicative SEs on the original scale. The model is easily extended to other categorical factors. We assess variation in the EOR/pack-year parameter for continuous intensity ( $n$ ) with

$$OR(d) = 1 + \beta d(n) \tag{D}$$

using three functional forms for  $g(\cdot)$ , including  $g(n) = \exp\{\varphi_1 n + \varphi_2 n^2\}$ ;  $g(n) = \exp\{\varphi_1 \ln(n) + \varphi_2 n\}$ ; and  $g(n) = \exp\{\varphi_1 \ln(n) + \varphi_2 \ln(n)^2\}$ . We select these forms primarily for their flexibility and convenience in fitting.

All models stratify on center (seven levels), sex (two levels), and age (five levels: 50-54, 55-59, 60-64, 65-70, and 70-74). We use likelihood ratio tests for comparisons of nested models. For all modeling, we use the Epicure software package (13).

**Results**

**Models for Lung Cancer and Total Exposure.** Because we are interested in general risk patterns rather than specific ORs

and their precision, we specify 11 categories for duration of smoking and pack-years of smoking based on never smokers and deciles of exposure. Figure 1 displays graphically a typical analysis of the joint ORs of smoking duration within four categories of intensity. All ORs are relative to never smokers. Patterns of ORs with duration vary depending on smoking intensity and do not follow a simple model in duration. Comparison of the OR for 40 years of smoking for individuals who smoke <20 cigarettes per day (OR = 7.4) with the OR for 40 years of smoking for individuals who smoke  $\geq 40$  cigarettes per day (OR = 18.4) is complicated by differences in total exposure to tobacco smoke.

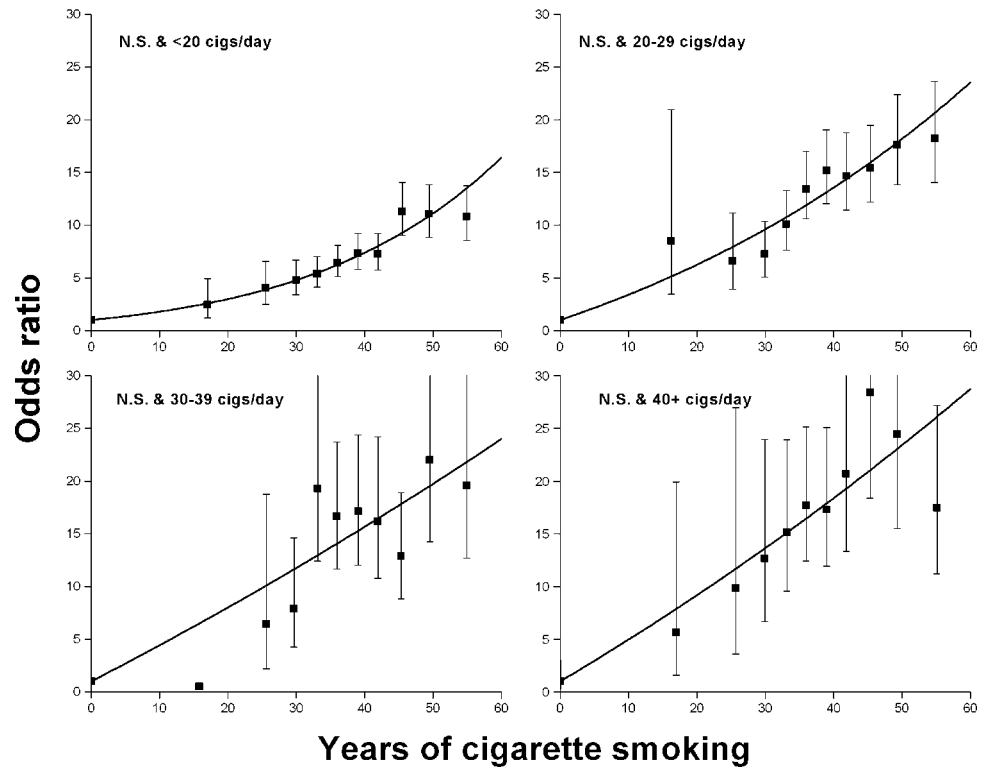
Figure 2 shows ORs for pack-years of smoking and the fitted model (A) within categories of smoking intensity. The relationships between the ORs for lung cancer and pack-years are approximately linear within each intensity category. Estimates of  $\beta$  (i.e., the EOR/pack-year) are smaller at higher intensities. For <20, 20-29, 30-39, and 40+ cigarettes per day, estimates of EOR/pack-year are 0.293, 0.315, 0.247, and 0.203, with  $P < 0.001$  for the test of homogeneity of slopes. Estimates of slope describe effects of intensity conditional on equal total exposure.

We next define 13 categories of intensity based on never smokers and deciles for smokers, with the upper category split into three additional categories. As in Fig. 2, ORs for categories of pack-years of exposure within the expanded categorization of intensity are consistent with linearity (see Supplementary Table;  $P = 0.32$  for the 12 degrees of freedom global test of linearity; i.e., all  $\gamma_i = 0$  for model (B) applied simultaneously to all intensity categories, whereas only one of 12 individual tests of  $\gamma_i = 0$  is rejected at the 0.05 level,  $P = 0.04$ ). Estimates of EOR/pack-year increase for lower intensities, then decline at higher intensities (Fig. 3). We fit model (D) to the case-control data to characterize the effect of intensity on the relationship between pack-years and lung cancer risk (Table 1). Changes in deviance indicate that model M1 with  $g(n) = \exp\{\varphi_1 \log(n) + \varphi_2 \log(n)^2\}$  provides the best fit although its improvement in fit over model M2 is slight. The locus of points for model M1 closely follows the category-specific EOR/pack-year estimates (Fig. 3, *solid line*). The maximum EOR/pack-year based on model M1 occurs at  $\exp(-\varphi_1/2\varphi_2)$  or 17.6 cigarettes smoked per day. Below the maximum, there is a "direct exposure-rate effect"; i.e., lung cancer risk per pack-year of exposure increases with increasing intensity or, more specifically, a given total exposure imparted at a higher intensity is more harmful than the same exposure imparted at a lower intensity. Above the maximum, there is an "inverse exposure-rate effect"; i.e., for equal total exposure, the exposure-response is inversely related to intensity or, more specifically, total exposure imparted at a lower intensity for a longer period of time is more deleterious than the equivalent exposure imparted at a higher intensity for a shorter period of time. For categories 1-19, 20-29, 30-39, and 40+ cigarettes smoked per day, mean intensities are 13.0, 22.0, 32.6, and 47.3, respectively. The predicted EOR/pack-year based on model M1 closely corresponds to the observed patterns for the four intensity categories (Fig. 2, *dashed lines*).

The comparable analysis of pack-years within categories of duration of exposure,  $y$ , does not yield linear relationships between ORs and pack-years, and models with  $g(y)$  functions result in smaller deviances and poorer fits to the data (Table 1). We therefore limit analyses to pack-years and intensity.

**Histologic Type of Lung Cancer and Exposure.** We fit model M1 with cases restricted to a major histologic type, including squamous cell carcinoma, small-cell carcinoma, large-cell carcinoma, and adenocarcinoma using the same control group for each analysis (Table 2; Fig. 4, *solid lines*). The estimate of  $\beta$  varies with histologic type; however, tests of homogeneity (Table 2, *column 7*) indicate that intensity

Downloaded from http://aascjournal.org/cebip/article-pdf/15/3/517/2244608/517.pdf by guest on 06 October 2024

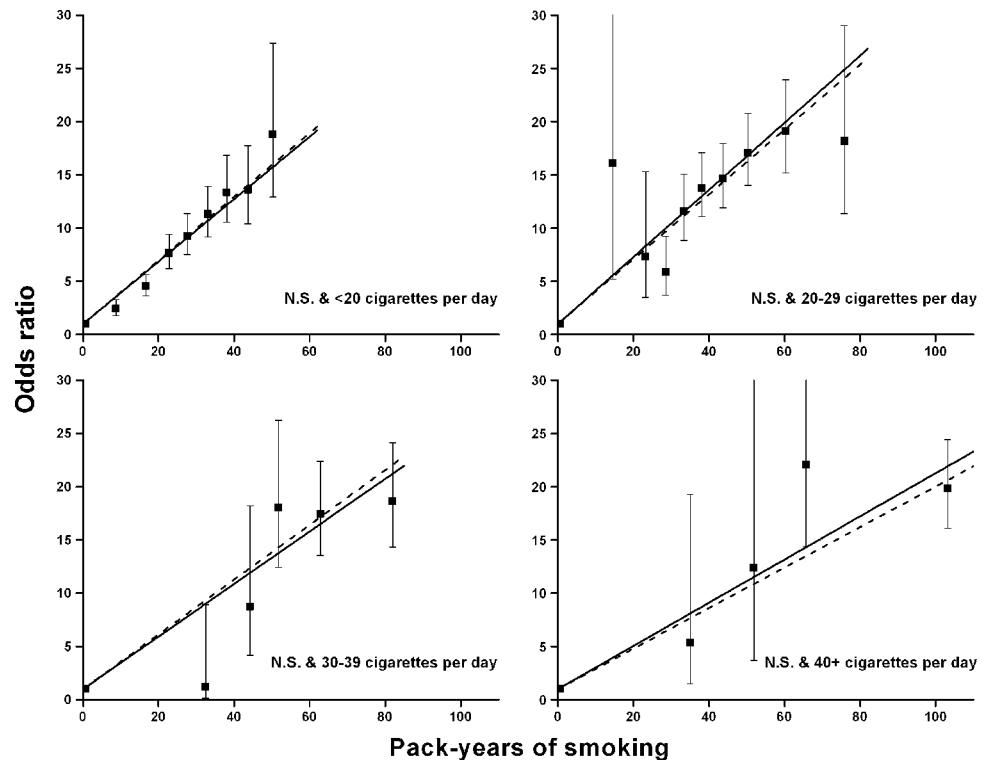


**Figure 1.** ORs for duration of smoking within categories of smoking intensity in cigarettes smoked per day, 95% CIs, and fitted linear-exponential OR models [Eq. (B)]. ORs relative to never-smokers and located at the category-specific mean duration.

parameters ( $\varphi_1$  and  $\varphi_2$ ) for each histologic type do not vary significantly and are consistent with estimates from all data combined (Fig. 4, *dashed lines*). Adjusted for intensity, EOR/pack-years are highest for cases with squamous cell carcinoma histology, followed by small-cell carcinoma, large-cell carcinoma, and adenocarcinoma histologies. For squamous cell carcinoma, small-cell carcinoma, large-cell carcinoma, and adenocarcinoma, fitted maxima occur at 16.4, 19.1, 13.9, and 19.1 cigarettes per day, respectively, which are similar to the 17.6 cigarettes per day for all data, resulting in an estimated

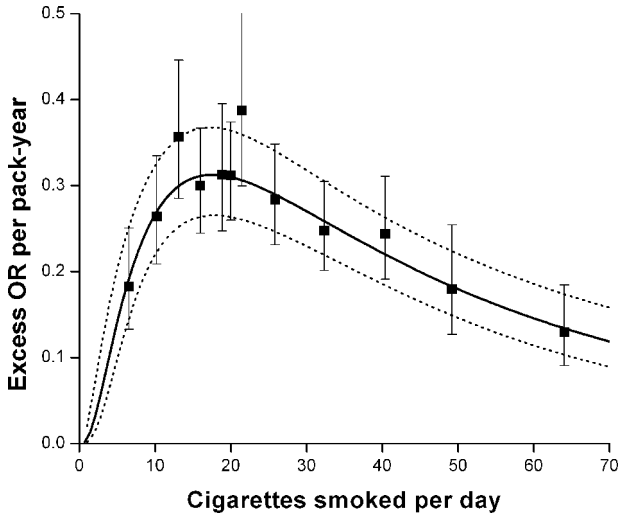
maximum EOR/pack-year estimates of 0.58, 0.38, 0.31, and 0.09/pack-year for the histologic types.

**Variations in Smoking Effects by Study Center.** The European Smoking and Health Study enrolled cases and controls from hospitals at seven centers in Europe. We evaluate model consistency by viewing each center as an independent replicate study. Overall, we find significant variations by center for overall effects of pack-years of exposure ( $\beta$ ;  $P < 0.001$ ) and smoking intensity ( $\varphi_1$  and  $\varphi_2$ ;  $P < 0.001$ ). However,



**Figure 2.** ORs for pack-years of smoking within categories of smoking intensity in cigarettes smoked per day, 95% CIs, fitted linear OR models (*solid line*), and predicted EOR based on model (D), with parameter estimates from Table 1, model M1 (*dash line*). ORs relative to never-smokers and located at the category-specific mean pack-years.

Downloaded from <http://aacrjournals.org/cebp/article-pdf/15/3/517/2244608/517.pdf> by guest on 06 October 2024



**Figure 3.** Estimated EOR per pack-year of smoking based on a linear OR model within categories of smoking intensity and fitted model (D) with parameter estimates from Table 1, model M1.

use of filter and nonfilter cigarettes varies considerably across study areas. After controlling for type of cigarette smoked and center, we find a significant interaction of the  $\beta$  parameter and center ( $P < 0.001$ ) but do not reject homogeneity of the intensity parameters by center [ $P = 0.15$  for the test of no interaction of  $g(n)$  and center].

**Discussion**

Our results show that the EOR/pack-year increases with intensity for subjects who smoke  $\leq 20$  cigarettes per day and decreases with intensity for subjects who smoke  $> 20$  cigarettes per day. At lower smoking intensities, the data support an "exposure enhancement" effect, such that for equal total exposure, the EOR/pack-year of smoking increases with intensity. At higher smoking intensities, data support a "reduced potency" or "wasted exposure" effect, such that for equal total exposure, smoking at a lower intensity for longer duration is more deleterious than smoking at a higher intensity for shorter duration. It is important to note that these patterns reflect effects of intensity and not lung cancer risk. For example, the inverse intensity effect implies that an increase in smoking intensity decreases risk per pack-year and does not

imply a decrease in the overall risk of lung cancer, which depends on both total pack-years of exposure and smoking intensity.

Results from epidemiologic studies indicate that risks of lung cancer, as well as bladder cancer, tend to level off with increased smoking intensity (10). The EOR/pack-year patterns observed in the European Smoking and Health Study data are consistent with these previous findings, at least at higher intensities, but precise comparisons are difficult because previous results do not control for total pack-years of exposure.

It remains to be determined whether and to what extent the patterns of variation of EOR/pack-year with intensity reflect nicotine dependency and its effect on smoking inhalation and smoking intensity or underlying biological processes, such as changes in activation and detoxification capacities for carcinogenic compounds in cigarette smoke or in DNA repair capacity, or biases in exposure assessment. Patterns of variation of the EOR/pack-year by intensity of smoking may reflect modulation of inhalation practices such that lower-intensity and higher-intensity smokers ingest relatively fewer carcinogens per cigarette smoked compared with moderate-intensity smokers. This would result in reduced risks at lower intensities and leveling or declining risks at higher intensities (10). Thus, variation in frequency or depth of inhalation is a possible explanation for observed intensity patterns. A recent study of 190 smokers found increased plasma cotinine and nicotine levels with increased cigarettes smoked per day, but a marginally significant ( $P = 0.08$ ) decline in "nicotine boost"; i.e., the increase in blood plasma nicotine after smoking one cigarette (14). We can directly evaluate the association between inhalation practices and intensity in the European Smoking and Health Study data. Table 3 shows percentages of smoking controls who inhale moderately or deeply or who inhale most or all of the time by categories of cigarettes per day and total exposure. Subjects with greater total exposure are more likely to inhale their cigarettes deeply or frequently; however, within categories of pack-years, there is little indication that depth or frequency of inhalation is related to smoking intensity.

Variations of ORs with smoking intensity after adjustment for total exposure parallel relationships observed in studies of biomarkers of smoking effects and exposure, and suggest saturation of activation and detoxification capacities (10, 15-17). Polycyclic aromatic hydrocarbons are a group of more than 100 chemicals which result from incomplete combustion of tobacco and other organic products, many of which are known carcinogens (see the Agency for Toxic Substances and Disease Registry web site <http://www.atsdr.cdc.gov/toxprofiles/tp69.html>). Polycyclic aromatic hydrocarbons

**Table 1. Results of modeling the variation of a linear OR model for pack-years and lung cancer with smoking intensity (cigarettes smoked per day) or smoking duration**

Model	Estimates	$P^*$	Deviance <sup>†</sup>	$P^\ddagger$
M0: $x_1 = 0; x_2 = 0$	$\beta = 0.282$	$<0.001$	—	
Modification by smoking intensity				
M1: $x_1 = \ln(n); x_2 = \ln(n)^2; \gamma = 0$	$\beta = 0.00499; \varphi_1 = 2.89; \varphi_2 = -0.504$	0.67	55.9	$<0.001$
M2: $x_1 = \ln(n); x_2 = n; \gamma = 0$	$\beta = 0.0613; \varphi_1 = 0.830; \varphi_2 = -0.0431$	0.58	54.0	$<0.001$
M3: $x_1 = n; x_2 = n^2; \gamma = 0$	$\beta = 0.257; \varphi_1 = 0.0166; \varphi_2 = -0.000429$	0.45	38.7	$<0.001$
Modification by smoking duration				
M4: $x_1 = \ln(y); x_2 = \ln(y)^2; \gamma = 0$	$\beta = 0.00000539; \varphi_1 = 5.57; \varphi_2 = -0.709; \gamma = 0$	$<0.001$	8.9	0.01
M5: $x_1 = \ln(y); x_2 = \ln(y)^2$	$\beta = 0.0000130; \varphi_1 = 4.97; \varphi_2 = -0.594; \gamma = -0.00505$		33.5	$<0.001$
M6: $x_1 = \ln(y); x_2 = y$	$\beta = 0.000843; \varphi_1 = 2.03; \varphi_2 = -0.0355; \gamma = -0.00505$		33.7	$<0.001$
M7: $x_1 = y; x_2 = y^2$	$\beta = 0.0563; \varphi_1 = 0.0756; \varphi_2 = -0.000722; \gamma = -0.00504$		33.5	$<0.001$

NOTE:  $OR = 1 + \beta \times d \times \exp(\varphi_1 x_1 + \varphi_2 x_2 + \gamma d)$ , where  $d$  is total pack-years,  $x_1$  and  $x_2$  are modifying functions of cigarettes smoked per day ( $n$ ) or duration of smoking in years ( $y$ ), and  $\beta, \varphi_1, \varphi_2$ , and  $\gamma$  are unknown parameters. Models M0 to M4 fit with  $\gamma$  fixed at zero. All models were adjusted for center, age, and sex.

\* $P$  value for test of linearity in the OR with pack-years (i.e.,  $\gamma = 0$ ).

<sup>†</sup>Change in deviance relative to model M0. Larger values indicate better model fit to the data.

<sup>‡</sup> $P$  value for 2 degrees of freedom (models M1-M4,  $\varphi_1 = 0, \varphi_2 = 0$ ) or 3 degrees of freedom (models M5-M7,  $\varphi_1 = 0, \varphi_2 = 0, \gamma = 0$ ) test of no effect modification for intensity or duration.

**Table 2. EOR per pack-year of smoking with adjustment for smoking intensity with case subjects restricted by histologic type**

Type* (cases)	$\beta$	$\varphi_1$	$\varphi_2$	$P^\dagger$	$\beta^\ddagger$	$P^\S$
SQ (2,411)	0.0247	2.26	-0.404	<0.001	0.00914	0.50
SM (795)	0.000395	4.65	-0.788	<0.001	0.00583	0.22
LG (336)	0.00365	3.38	-0.642	<0.001	0.00416	0.07
AD (562)	0.000297	3.87	-0.656	0.01	0.00141	0.81

NOTE: Model:  $EOR = \beta \times d \times \exp\{\varphi_1 \ln(n) + \varphi_2 \ln(n)^2\}$ , where  $d$  is pack-years of exposure and  $n$  is cigarettes smoked per day. All models were adjusted for center, age, and sex.

\*Histologic types include squamous cell carcinoma (SQ), small cell carcinoma (SM), large cell carcinoma (LG), and adenocarcinoma (AD).

$^\dagger P$  value for hypothesis test of no intensity effects,  $\varphi_1 = 0$  and  $\varphi_2 = 0$ .

$^\ddagger$  Estimated EOR/pack-year with  $\varphi_1$  and  $\varphi_2$  fixed at their values based on all cases and controls; i.e.,  $\varphi_1 = 2.89$  and  $\varphi_2 = -0.504$ .

$^\S P$  value for the 2 degree of freedom likelihood ratio test of fit with  $\beta$  estimated and  $\varphi_1$  and  $\varphi_2$  fixed at their values from all data.

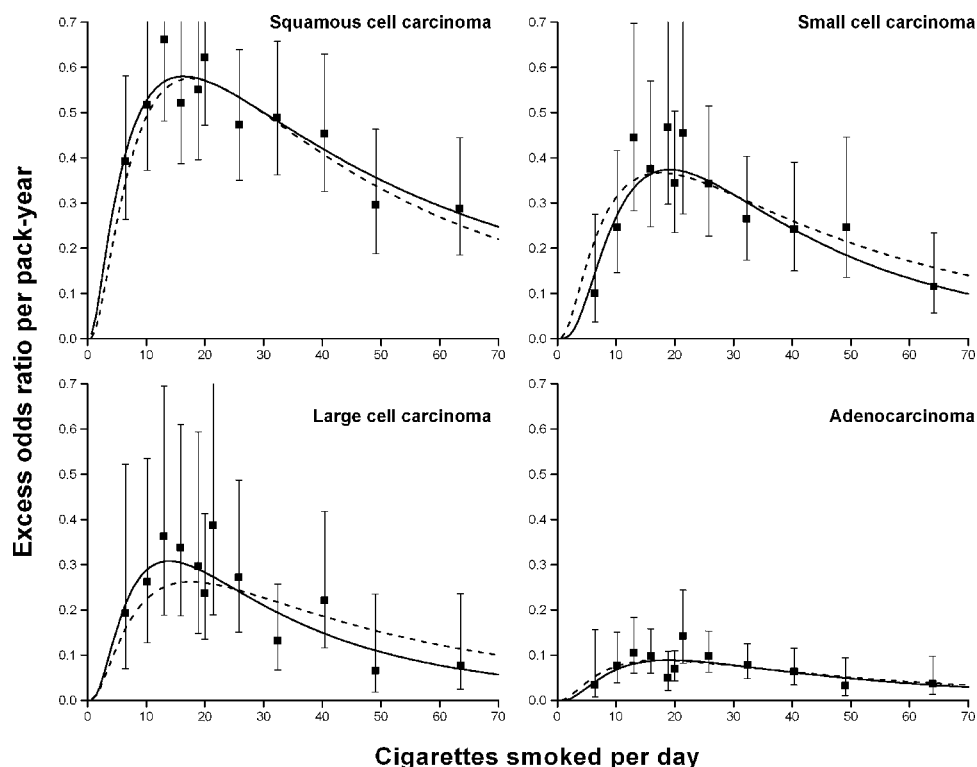
undergo metabolic activation and, as a first step in the carcinogenic process, can form DNA and protein adducts (18). Investigators reported that DNA adduct levels in WBC per unit exposure to polycyclic aromatic hydrocarbons were higher in individuals exposed at environmental levels than in workers exposed at high levels, thus suggesting reduced carcinogenic potency at high exposures (16); see ref. 18 for a review. Carbon monoxide is a combustion product formed when cigarettes are smoked and has an affinity for hemoglobin. Among never smokers and current smokers, Law et al. (19) found that the ratio of serum carboxyhemoglobin to the number of cigarettes smoked per day decreased with increasing smoking intensity.

Polycyclic aromatic hydrocarbons and DNA and protein adducts have limitations in analyses of tobacco effects because they can arise from sources other than smoking. The compound 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK) is a tobacco-specific carcinogen and levels of its metabolites directly represent markers of tobacco effects (20, 21). If the patterns of intensity effects in the European

Smoking and Health Study data derive from biological processes, then we would expect a nonlinear relationship between NNK metabolites and urinary cotinine levels, a marker of tobacco exposure, with relatively less NNK metabolites produced at high, as opposed to low, cotinine levels. Whereas this analysis has not formally been carried out, this pattern can be observed in Fig. 5 of Carmella et al. (20).

Reduced effects of smoking at higher intensities are also consistent with enhancement of DNA repair capacities. There is evidence that overall lung cancer case subjects have reduced DNA repair capacity as compared with controls (15, 22, 23) and that repair capacity is increased among subjects who are more heavily exposed to tobacco smoke (24-27).

The increasing EOR/pack-year at lower intensities and the decreasing EOR/pack-year at higher intensities could be due, at least in part, to misclassification of cigarettes smoked per day. Observed variations of the EOR/pack-year with intensity would require a very specific pattern of misclassification; i.e., a progressively increasing amount of misclassification with increasing intensity (above 15 to 20 cigarettes smoked per day), resulting in an increasing bias towards the null and a decreasing exposure-response relationship, and a decreasing amount of misclassification of intensity up to 15 to 20 cigarettes smoked per day, resulting in decreasing bias towards the null and an increasing exposure-response relationship. The latter pattern is plausible if an increasing intensity results in a smoker consuming a proportionally smaller fraction of each cigarette. Nonetheless, if low and high intensities were differentially misclassified compared with intermediate intensities, then pack-years of smoking would reflect the differential misclassification because we expect that duration of smoking is relatively accurate. We would expect to observe a progressively more concave relationship between the ORs for lung cancer and pack-years with increasing intensity and with decreasing intensity. However, exposure-response relationships for ORs and pack-years are consistent with linearity over the full range of intensities, suggesting that any differential misclassification by intensity is minimal and not likely to induce the observed patterns.



**Figure 4.** Estimated EOR per pack-year of smoking based on a linear OR model within categories of smoking intensity with cases restricted by histologic type of lung cancer, fitted model (D) for each histologic group (solid line), and fitted model (D) based on histologic-specific pack-years effect and fixed intensity parameters based on all data combined (dashed line).

**Table 3. Percentages and numbers of smoking controls who inhale moderately or deeply or who inhale most or all of the time**

Pack-years	Cigarettes smoked per day			
	<20	20-29	30-39	40+
<b>Inhale moderately or deeply</b>				
<20	64.2	66.7	—	—
20-34	78.3	78.2	75.0	75.0
35-49	80.8	83.6	80.0	80.0
50+	87.9	88.9	89.3	89.3
<b>Inhale most or all of the time</b>				
<20	65.9	66.7	—	—
20-34	80.5	77.4	75.0	75.0
35-49	84.6	85.1	60.0	60.0
50+	93.9	90.8	90.1	90.1
<b>Number of controls</b>				
<20	823	6	0	0
20-34	1,225	235	7	4
35-49	473	909	41	5
50+	33	489	378	352

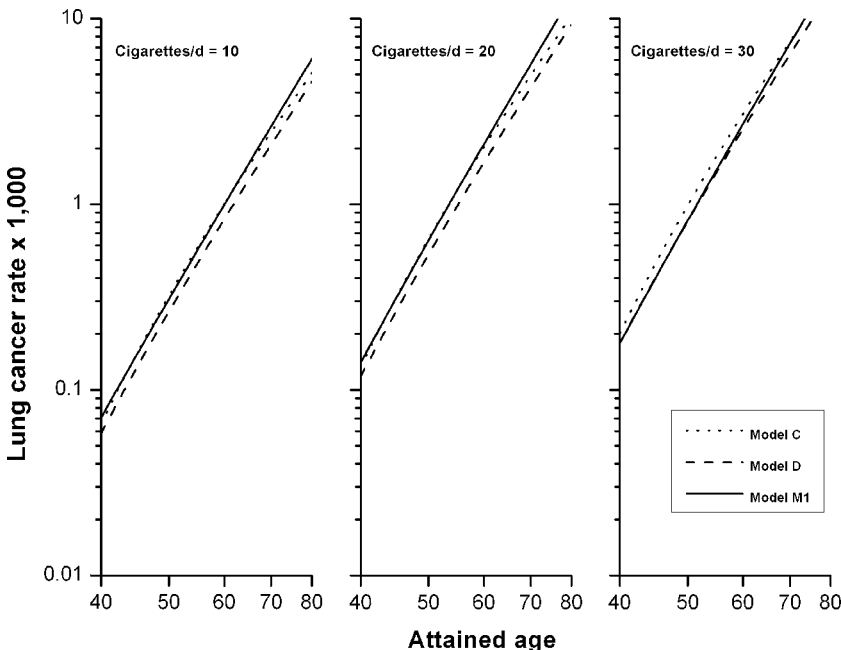
The best fitting model in Table 1 (model M1) involves an intensity function of the form  $g(n) = \exp\{\varphi_1 \ln(n) + \varphi_2 \ln(n)^2\}$ . Because pack-years,  $d$ , is  $y \times n$  and  $OR = 1 + \beta d \times g(n)$ , model M1 can be rewritten in terms of duration and intensity as  $OR = \beta y \times \exp\{\gamma \ln(n) + \varphi_2 \ln(n)^2\}$ , where  $\gamma = \varphi_1 + 1$ . Thus, model M1 with the above  $g(n)$  is equivalent to a model which is linear in duration for fixed intensity.

It is well recognized that the association between smoking and lung cancer varies with histologic type (1). Our analyses suggest that whereas levels of lung cancer risk vary by histologic classification, effects of smoking intensity (i.e., the curvatures) do not vary by histologic type. This suggests that stochastic factors which influence pathways that predispose towards a specific histology are not associated with smoking intensity but more closely linked with total exposure or smoking duration. The precise implication of this observation is unclear but may suggest that determinants of histologic type are more likely linked to early transformational processes.

The Armitage-Doll multistage model for carcinogenesis is based on the observation that the logarithm of the lung

cancer rate increases approximately linearly with the logarithm of age and is consistent with a multistage model for carcinogenesis, whereby a normal epithelial cell undergoes multiple, sequential, heritable, and nonreversible transformations to become malignant. Doll and Peto (28) used this model in their analysis of the British Doctors' Study and found that lung cancer rates in continuing smokers increased with the square of intensity plus 6 and the 4.5 power of duration minus 3.5. In an analysis of lung cancer mortality in the American Cancer Society's Cancer Prevention Study I, a cohort study of 1,078,894 subjects, Knoke et al. (29) found significant improvement in model fit, in comparison with the Doll and Peto model, by including a factor for either age at smoking initiation or attained age. The model for lung cancer mortality rate in smokers was  $rate = \alpha(n + 6)^\beta \times (y - 3.5)^\gamma \times f^\delta$ , where  $n$  is number cigarettes smoked per day,  $y$  is duration of smoking, and  $f$  is a factor representing either age at start of smoking or attained age (models C or D, respectively, in ref. 29), and where  $\alpha$ ,  $\beta$ ,  $\gamma$  and  $\delta$  are parameters. There was a separate model for lung cancer mortality rate in nonsmokers with the form:  $rate = \alpha(\text{age} - 3.5)^\beta$ . As in the Doll and Peto analysis, Knoke et al. added 6 cigarettes per day to account for the background risk in the absence of smoking and added 3.5 years to reflect a time lag from the appearance of a malignant cell to lung cancer mortality. To compare our model M1 to Knoke's models, we assume our OR model for lung cancer incidence approximates the relative risk for lung cancer mortality, and multiply model M1 by the Cancer Prevention Study I lung cancer mortality rate model for nonsmokers to obtain lung cancer (mortality) rates. For the comparison, we assume individuals start smoking at age 19 years. Figure 5 shows that predicted lung cancer mortality for males who smoke 10, 20, or 30 cigarettes per day based on our exposure and exposure-rate model are very similar to predictions based on models C and D in Knoke et al.

Finally, patterns of intensity effects are subject to substantial uncertainty, particularly at lower intensities. This uncertainty is due to the limited range of pack-years of exposure and an increased variability in estimating EOR/pack-years. For example, the median and interquartile range for total exposure are 6.7 pack-years and 4.9 to 8.4 pack-years, respectively, for subjects smoking under 5 cigarettes per day, and 14.4 pack-years and 11.4 to 18.2 pack-years for



**Figure 5.** Comparison of models for lung cancer mortality rate with smoking intensity, years of smoking and either age at start of smoking (model C) or attained age (model D) from ref. (29) based on data from the American Cancer Society's Cancer Prevention Study I cohort, and the current linear EOR model (Table 1, model M1) times the Cancer Prevention Study I model for nonsmokers.

subjects smoking 5 to 10 cigarettes per day, compared with 40 pack-years and 28 to 53.8 pack-years for all smokers. Thus, any conclusion about whether the estimated EOR/pack-year approaches zero remains constant or, indeed, increases for intensities below 5 to 8 cigarettes per day must be made cautiously. Results are subject to the additional uncertainty from use of the overall mean number of cigarettes smoked per day, rather than accounting for changes in smoking intensity throughout life.

In summary, our analysis of a large case-control study of lung cancer using a novel exposure and exposure-rate model reveals a direct intensity effect at low smoking intensities; i.e., an intensity enhancement effect, resulting in an increasing EOR/pack-year, and an inverse intensity-rate effect (i.e., reduced potency or wasted exposure effect), resulting in a decreasing EOR/pack-year at higher intensities. Our modeling approach is applicable in other epidemiologic settings where data on exposure and exposure rate are available.

## Acknowledgments

We thank Drs. Michael Alavanja, Montserrat Garcia-Closas, Michael Hauptmann, and Debra Silverman of the Division of Cancer Epidemiology and Genetics, National Cancer Institute, for useful discussions on this topic.

## References

1. U.S. Department of Health and Human Services. The health consequences of smoking: a report of the surgeon general. U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Chronic Disease Prevention and Health Promotion, Office on Smoking and Health, Superintendent of Documents. Washington (DC): U.S. Government Printing Office; 2004.
2. Leffondre K, Abrahamowicz M, Siemiatycki J, Rachet B. Modeling smoking history: a comparison of different approaches. *Am J Epidemiol* 2002;156:813–23.
3. Hauptmann M, Berhand K, Langholz B, Lubin JH. Using splines to analyse latency in the Colorado Plateau uranium miners cohort. *J Epidemiol Biostat* 2001;6:417–24.
4. Rachet B, Siemiatycki J, Abrahamowicz M, Leffondre K. A flexible modeling approach to estimating the component effects of smoking behavior on lung cancer. *J Clin Epidemiol* 2004;57:1076–85.
5. Dietrich T, Hoffmann KA. Comprehensive index for the modeling of smoking history in periodontal research. *J Dent Res* 2004;83:859–63.
6. Thurston S, Liu G, Miller D, Christiani D. Modeling lung cancer risk in case-control studies using a new dose metric of smoking. *Cancer Epidemiol Biomarkers Prev* 2005;14:2296–302.
7. Hoffmann K, Bergmann MM. Re: "Modeling smoking history: a comparison of different approaches." *Am J Epidemiol* 2003;158:393.
8. McKnight B, Cook LS, Weiss NS. Logistic regression analysis for more than one characteristic of exposure. *Am J Epidemiol* 1999;149:984–92.
9. Vineis P, Alavanja M, Garte S. Dose-response relationship in tobacco-related cancers of bladder and lung: a biochemical interpretation. *Int J Cancer* 2004;108:2–7.
10. Vineis P, Kogevinas M, Simonato L, Brennan P, Boffetta P. Levelling-off of the risk of lung and bladder cancer in heavy smokers: an analysis based on multicentric case-control studies and a metabolic interpretation. *Mutat Res-Fundam Mol Mech Mutagen* 2000;463:103–10.
11. Lubin JH, Blot WJ, Berrino F, et al. Patterns of lung cancer risk according to type of cigarette smoked. *Int J Cancer* 1984;33:569–76.
12. Lubin JH, Blot WJ. Assessment of lung cancer risk factors by histologic category. *J Natl Cancer Inst* 1984;73:383–9.
13. Preston DL, Lubin JH, Pierce DA, McConney ME. *Epicure user's guide*. Seattle (WA): HiroSoft International Corporation; 2002.
14. Patterson F, Benowitz N, Shields P, et al. Individual differences in nicotine intake per cigarette. *Cancer Epidemiol Biomarkers Prev* 2003;12:468–71.
15. Berwick M, Vineis P. Markers of DNA repair and susceptibility to cancer in humans: an epidemiologic review. *J Natl Cancer Inst* 2000;92:874–97.
16. Lewtas J, Walsh D, Williams R, Dobias L. Air pollution exposure DNA adduct dosimetry in humans and rodents: evidence for non-linearity at high doses. *Mutat Res-Fundam Mol Mech Mutagen* 1997;378:51–63.
17. Lutz WK. Dose-response relationships in chemical carcinogenesis: superposition of different mechanisms of action, resulting in linear-nonlinear curves, practical thresholds, J-shapes. *Mutat Res-Fundam Mol Mech Mutagen* 1998;405:117–24.
18. Phillips DH. Smoking-related DNA and protein adducts in human tissues. *Carcinogenesis* 2002;23:1979–2004.
19. Law MR, Morris JK, Watt HC, Wald NJ. The dose-response relationship between cigarette consumption, biochemical markers and risk of lung cancer. *Br J Cancer* 1997;75:1690–3.
20. Carmella SG, Akerkar SA, Richie JP, Hecht SS. Intraindividual and interindividual differences in metabolites of the tobacco-specific lung carcinogen 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK) in smokers urine. *Cancer Epidemiol Biomarkers Prev* 1995;4:635–42.
21. Anderson KE, Kliris J, Murphy L, et al. Metabolites of a tobacco-specific lung carcinogen in nonsmoking casino patrons. *Cancer Epidemiol Biomarkers Prev* 2003;12:1544–6.
22. Neumann AS, Sturgis EM, Wei QY. Nucleotide excision repair as a marker for susceptibility to tobacco-related cancers: a review of molecular epidemiological studies. *Mol Carcinog* 2005;42:65–92.
23. Wei QY, Spitz MR. The role of DNA repair capacity in susceptibility to lung cancer: a review. *Cancer Metastasis Rev* 1997;16:295–307.
24. Wei QY, Cheng L, Amos CI, et al. Repair of tobacco carcinogen-induced DNA adducts and lung cancer risk: a molecular epidemiologic study. *J Natl Cancer Inst* 2000;92:1764–72.
25. Spitz MR, Wei QY, Dong Q, Amos CI, Wu XF. Genetic susceptibility to lung cancer: the role of DNA damage and repair. *Cancer Epidemiol Biomarkers Prev* 2003;12:689–98.
26. Shen HB, Spitz MR, Qiao YW, et al. Smoking, DNA repair capacity and risk of nonsmall cell lung cancer. *Int J Cancer* 2003;107:84–8.
27. Hung RJ, Brennan P, Canzian F, et al. Large-scale investigation of base excision repair genetic polymorphisms and lung cancer risk in a multicenter study. *J Natl Cancer Inst* 2005;97:567–76.
28. Doll R, Peto R. Cigarette-smoking and bronchial-carcinoma—dose and time relationships among regular smokers and lifelong non-smokers. *J Epidemiol Community Health* 1978;32:303–13.
29. Knote JD, Shanks TG, Vaughn JW, Thun MJ, Burns DM. Lung cancer mortality is related to age in addition to duration and intensity of cigarette smoking: an analysis of Cancer Prevention Study I data. *Cancer Epidemiol Biomarkers Prev* 2004;13:949–57.