Cigarette Smoking and Lung Cancer: Modeling Total Exposure and Intensity

Jay H. Lubin and Neil E. Caporaso

1Biostatistics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute and 2Genetic 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 ≤20 cigarettes per day and decreases with intensity for subjects who smoke >20 cigarettes per day.

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...
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, \(d\),

\[
\text{OR}(d) = 1 + \beta d
\]

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

\[
\text{OR}(d) = 1 + \beta d \exp(\gamma d)
\]

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, \(n_i, i = 1, \ldots, I\), where \(n_i = 1\) if a subject’s intensity level occurs within the \(i\)th category; 0 otherwise. The model is

\[
\text{OR}(d) = 1 + \beta d \exp(\Sigma \theta_i n_i)
\]

which specifies a different slope for each intensity category. With \(\theta_i\) set to zero for identifiability, \(\theta_2, \ldots, \theta_I\) 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 reparameterization, 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

\[
\text{OR}(d) = 1 + \beta(n) d
\]

using three functional forms for \(g()\), 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 ≥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. We use two approaches to model intensity of smoking. The first approach defines \(I\) intensity categories and indicator variables, \(n_i, i = 1, \ldots, I\), where \(n_i = 1\) if a subject’s intensity level occurs within the \(i\)th category; 0 otherwise. The model is

\[
\text{OR}(d) = 1 + \beta(n) \exp(\Sigma \theta_i n_i)
\]

which specifies a different slope for each intensity category. With \(\theta_i\) set to zero for identifiability, \(\theta_2, \ldots, \theta_I\) 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 reparameterization, 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

\[
\text{OR}(d) = 1 + \beta(n) d
\]

using three functional forms for \(g()\), 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 ≥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. We use two approaches to model intensity of smoking. The first approach defines \(I\) intensity categories and indicator variables, \(n_i, i = 1, \ldots, I\), where \(n_i = 1\) if a subject’s intensity level occurs within the \(i\)th category; 0 otherwise. The model is

\[
\text{OR}(d) = 1 + \beta(n) \exp(\Sigma \theta_i n_i)
\]

which specifies a different slope for each intensity category. With \(\theta_i\) set to zero for identifiability, \(\theta_2, \ldots, \theta_I\) 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 reparameterization, 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

\[
\text{OR}(d) = 1 + \beta(n) d
\]

using three functional forms for \(g()\), 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).

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
parameters ($\phi_1$ and $\phi_2$) 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 ($\phi_1$ and $\phi_2$; $P < 0.001$). However,
Our results show that the EOR/pack-year increases with modification by smoking duration and intensity for subjects who smoke >20 cigarettes per day or smoking duration. Parameter and estimates from Table 1, model M1.

Discussion

Our results show that the EOR/pack-year increases with intensity for subjects who smoke >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

<table>
<thead>
<tr>
<th>Model</th>
<th>Estimates</th>
<th>P*</th>
<th>Deviance †</th>
<th>P ‡</th>
</tr>
</thead>
<tbody>
<tr>
<td>M0: x₁ = 0; x₂ = 0</td>
<td>β = 0.282</td>
<td>&lt;0.001</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>Modification by smoking intensity</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M1: x₁ = ln(n); x₂ = ln(n); γ = 0</td>
<td>β = 0.00499; φ₁ = 2.89; φ₂ = -0.504</td>
<td>0.67</td>
<td>55.9</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>M2: x₁ = ln(n); x₂ = n; γ = 0</td>
<td>β = 0.0613; φ₁ = 0.830; φ₂ = -0.0431</td>
<td>0.58</td>
<td>54.0</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>M3: x₁ = n; x₂ = n; γ = 0</td>
<td>β = 0.257; φ₁ = 0.0166; φ₂ = -0.000429</td>
<td>0.45</td>
<td>38.7</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Modification by smoking duration</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>M4: x₁ = ln(y); x₂ = ln(y); γ = 0</td>
<td>β = 0.00000539; φ₁ = 5.57; φ₂ = -0.709; γ = 0</td>
<td>&lt;0.001</td>
<td>8.9</td>
<td>0.01</td>
</tr>
<tr>
<td>M5: x₁ = ln(y); x₂ = ln(y)</td>
<td>β = 0.0000130; φ₁ = 4.97; φ₂ = -0.594; γ = -0.00505</td>
<td>33.5</td>
<td>&lt;0.001</td>
<td></td>
</tr>
<tr>
<td>M6: x₁ = ln(y); x₂ = y</td>
<td>β = 0.000843; φ₁ = 2.63; φ₂ = -0.0355; γ = -0.00505</td>
<td>33.7</td>
<td>&lt;0.001</td>
<td></td>
</tr>
<tr>
<td>M7: x₁ = y; x₂ = y</td>
<td>β = 0.0563; φ₁ = 0.0756; φ₂ = -0.000722; γ = -0.00504</td>
<td>33.5</td>
<td>&lt;0.001</td>
<td></td>
</tr>
</tbody>
</table>

NOTE: OR = 1 + β × d × exp(φ₁ x₁ + φ₂ x₂ + γ); where d is total pack-years, x₁, and x₂ are modifying functions of cigarettes smoked per day (n) or duration of smoking in years (y), and β, φ₁, φ₂, and γ are unknown parameters. Models M0 to M4 fit with γ 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., γ = 0).

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

‡P value for 2 degrees of freedom (models M1-M4, β₁ = 0, β₂ = 0) or 3 degrees of freedom (models M5-M7, β₁ = 0, β₂ = 0, γ = 0) test of no effect modification for intensity or duration.
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 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. NNK is cigarettes smoked per day. All models were adjusted for center, age, and sex.

NOTE: Model: $EOR = \beta \cdot d \cdot \exp\{\varphi_1 \cdot \ln(n) + \varphi_2 \cdot \ln(n)^2\}$, 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).

<table>
<thead>
<tr>
<th>Type* (cases)</th>
<th>$\beta$</th>
<th>$\varphi_1$</th>
<th>$\varphi_2$</th>
<th>$P^1$</th>
<th>$\beta^2$</th>
<th>$P^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>SQ (2,411)</td>
<td>0.0247</td>
<td>2.26</td>
<td>-0.404</td>
<td>&lt;0.001</td>
<td>0.00914</td>
<td>0.50</td>
</tr>
<tr>
<td>SM (795)</td>
<td>0.003095</td>
<td>4.65</td>
<td>-0.788</td>
<td>&lt;0.001</td>
<td>0.00583</td>
<td>0.22</td>
</tr>
<tr>
<td>LG (336)</td>
<td>0.00365</td>
<td>3.38</td>
<td>-0.642</td>
<td>&lt;0.001</td>
<td>0.00416</td>
<td>0.07</td>
</tr>
<tr>
<td>AD (562)</td>
<td>0.000297</td>
<td>3.87</td>
<td>-0.656</td>
<td>0.01</td>
<td>0.00141</td>
<td>0.81</td>
</tr>
</tbody>
</table>

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

$^1P$ value for hypothesis test of no intensity effects, $\varphi_1 = 0$ and $\varphi_2 = 0$.

$^2$ 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 $\varphi_2 = -0.504$.

$^3P$ 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.

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 case subjects restricted by histologic type.
The best fitting model in Table 1 (model M1) involves an intensity function of the form \( \gamma(n) = \exp[\varphi_1 \ln(n) + \varphi_2 \ln(n)^2] \). Because pack-years, \( d \), is \( y \times n \) and \( \text{OR} = 1 + \beta d \times \gamma(n) \), model M1 can be rewritten in terms of duration and intensity as \( \text{OR} = 1 + \beta d \times \gamma(n) \). Thus, model M1 with the above \( \gamma(n) \) 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 = \( x (u + \delta)^6 \times (y - 3.5)^7 \times f^2 \), 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 = \( x (age - 3.5)^6 \). 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 rate in 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 an increased variability in estimating EOR/pack-years. The uncertainty is due to the limited range of pack-years of exposure and an increased variability in estimating EOR/pack-years.

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

<table>
<thead>
<tr>
<th>Pack-years</th>
<th>Cigarettes smoked per day</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>20-29</td>
</tr>
<tr>
<td>Inhale moderately or deeply</td>
<td></td>
</tr>
<tr>
<td>&lt;20</td>
<td>64.2</td>
</tr>
<tr>
<td>20-34</td>
<td>78.3</td>
</tr>
<tr>
<td>35-49</td>
<td>80.8</td>
</tr>
<tr>
<td>50+</td>
<td>87.9</td>
</tr>
<tr>
<td>Inhale most or all of the time</td>
<td></td>
</tr>
<tr>
<td>&lt;20</td>
<td>65.9</td>
</tr>
<tr>
<td>20-34</td>
<td>80.5</td>
</tr>
<tr>
<td>35-49</td>
<td>84.6</td>
</tr>
<tr>
<td>50+</td>
<td>93.9</td>
</tr>
<tr>
<td>Number of controls</td>
<td></td>
</tr>
<tr>
<td>&lt;20</td>
<td>823</td>
</tr>
<tr>
<td>20-34</td>
<td>1,225</td>
</tr>
<tr>
<td>35-49</td>
<td>473</td>
</tr>
<tr>
<td>50+</td>
<td>33</td>
</tr>
</tbody>
</table>

![Figure 5](image-url)
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


