A recent analysis showed that the excess odds ratio (EOR) for lung cancer due to smoking can be modeled by a function which is linear in total pack-years and exponential in the logarithm of smoking intensity and its square. Below 15–20 cigarettes per day, the EOR/pack-year increased with intensity (direct exposure rate or enhanced potency effect), suggesting greater risk for a total exposure delivered at higher intensity (for a shorter duration) than for an equivalent exposure delivered at lower intensity. Above 20 cigarettes per day, the EOR/pack-year decreased with increasing intensity (inverse exposure rate or reduced potency effect), suggesting greater risk for a total exposure...
delivered at lower intensity (for a longer duration) than for an equivalent exposure delivered at higher intensity. The authors applied this model to data from 10 case-control studies of cancer, including cancers of the lung, bladder, oral cavity, pancreas, and esophagus. At lower intensities, there was enhanced potency for several cancer sites, but narrow ranges for pack-years increased uncertainty, precluding definitive conclusions. At higher intensities, there was a consistent reduced potency effect across studies. The intensity effects were statistically homogeneous, indicating that after accounting for risk from total pack-years, intensity patterns were comparable across the diverse cancer sites.

In studies of lung cancer and bladder cancer, odds ratios increase with smoking intensity, but often level off and even decline at high intensities (1, 2). This pattern suggests a changing impact of intensity on disease risk, although its precise implication is uncertain. Behavioral factors, such as high-intensity smokers’ inhaling less deeply, or exposure-dependent biases are possible explanations for the leveling off of risk, but the involvement of intensity-dependent molecular mechanisms in carcinogenesis is another plausible explanation. To better understand the biologic 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 smoking duration. In particular, we are interested in evaluating how the rate of delivery of cigarette smoke influences disease risk at a fixed total exposure.

Standard approaches for exploring the leveling off of odds ratios are problematic, since they depend on models for odds ratios that incorporate smoking duration and intensity; this can lead to difficulty in interpreting intensity effects due to changing total exposures. For example, in a logistic model, the intensity parameter represents the change in the natural logarithm of the odds ratio (OR) per unit of intensity (i.e., ln(OR) per cigarette per day) at a fixed duration. The duration parameter has a similar representation but at a fixed intensity. Therefore, odds ratios at two different intensities reflect not only the different intensities but also different total pack-years, since duration is fixed. For example, at 30 years’ duration, differences in odds ratios for 20 and 30 cigarettes per day embed differences in total exposure, that is, 30 and 45 pack-years, respectively. Thus, the intensity parameter does not represent a “pure” intensity effect but includes an additional effect of total pack-years.

In the European Smoking and Health Study (ESHS), Lubin et al. (3, 4) recently described a three-parameter model for the excess odds ratio (EOR) for lung cancer which was linear in pack-years and exponential in the logarithm of smoking intensity and its square. The model defined the impact of intensity on the EOR/pack-year, thereby isolating the contribution of intensity at a fixed total exposure. Under 15–20 cigarettes per day, the EOR/pack-year increased with intensity (direct exposure rate or an enhanced potency effect); that is, total exposure imparted at a higher intensity for a shorter duration induced greater risk than an equal total exposure imparted at a lower intensity for a longer duration.

At higher intensities, the EOR/pack-year decreased with increasing intensity (inverse exposure rate, or a reduced potency or “wasted” exposure effect); that is, total exposure imparted at a lower intensity for a longer duration induced greater risk than an equal exposure imparted at a higher intensity for a shorter duration.

In the current analysis, we applied the model to data from additional lung cancer studies and to studies of other smoking-related cancers, including cancers of the bladder, oral cavity, pancreas, and esophagus, and evaluated smoking intensity patterns across the diverse cancer sites.

**MATERIALS AND METHODS**

**Models**

For disease outcome C, where C = 1 denotes a case and C = 0 denotes a noncase, we fit a model for the odds of disease of the form

$$P(C = 1|x, d, n]/P(C = 0|x, d, n] = \exp(\beta^T x)\text{OR}(d, n),$$

where $x$ is a vector of adjustment variables, $T$ denotes vector transpose, and $\text{OR}(d, n)$ is the odds ratio model for $d$ total pack-years and $n$ cigarettes per day. For $I$ intensity categories, we define indicator variables $n_i$, $i = 1, \ldots, I$, where $n_i = 1$ for intensities within category $i$ and zero otherwise. We fit

$$\text{OR}(d, n) = 1 + \sum \gamma_i n_i d.$$  

Within category $i$, odds ratios are linear in $d$ (i.e., $\text{OR} = 1 + \gamma_i d$), where $\gamma_i$ is the EOR/pack-year. For calculating estimates and 95 percent confidence intervals, we replace $\gamma_i$ with $\exp(\gamma_i^{*})$ to avoid range restrictions. Factoring out $\gamma_i$, model 1 (equation 1) becomes

$$\text{OR}(d, n) = 1 + \gamma_1 d \sum (\gamma_i/\gamma_1)n_i.$$  

Note that $(\gamma_i/\gamma_1)$ represents the relative variation of the linear relation with categories of intensity. A natural extension for continuous intensity is

$$\text{OR}(d, n) = 1 + \beta d g(n),$$

where $\beta$ represents the EOR/pack-year at $g(n) = 1$, while $g(\cdot)$ defines the variation of the linear EOR/pack-year parameter with intensity. Model 2 (equation 2) is a simple EOR model in which the effects of pack-years and intensity

---

Abbreviations: EOR, excess odds ratio; ESHS, European Smoking and Health Study; OR, odds ratio.
multiply. We considered three forms for $g(n)$, including:

$$g(n) = \exp(\varphi_1 \ln(n) + \varphi_2 \ln(n)^2);$$

$$g(n) = \exp(\varphi_1 \ln(n) + \varphi_2 n);$$

and

$$g(n) = \exp(\varphi_1 n + \varphi_2 n^2),$$

with

$$g(n) = \exp(\varphi_1 \ln(n) + \varphi_2 \ln(n)^2)$$

providing the best fit. Fit was not substantially improved with further inclusion of $\ln(n) \times n$, $\ln(n)^3$, or $n^3$.

We evaluate departure from linearity by including an exponential factor in model 2,

$$\text{OR} \left( d,n \right) = 1 + \beta d \exp(\delta d) \ g(n),$$

where $\delta$ measures concavity ($\delta < 0$) or convexity ($\delta > 0$) in the EORs for pack-years. A test of the null hypothesis $\delta = 0$ is a test of no departure from linearity.

### Case-control studies of smoking and cancer

We conducted 10 individual analyses using data from nine case-control studies and one set of pooled data from multiple studies (table 1). Data sets included those from the ESHS (5, 6), the German Radon Study (7–9), the North American Radon Pooling Project (10), the Gansu Lung Cancer Study (11), the Colorado Plateau Uranium Miners Study (12, 13), the National Bladder Cancer Study (14), the Spanish Bladder Cancer Study (15), the Puerto Rico Oral Cancer Study (16), and the Investigation of Tumors that Occur Excessively among Blacks, which we designated the Population Health Study (17, 18). We included these studies because the cancers were smoking-related and data sets were readily available. For consistency, we established common inclusion criteria, limiting subjects to ages 50–74 years at enrollment and to never smokers, current smokers, or former smokers who had stopped smoking within 5 years of enrollment, except where noted. The age restriction reduced the impact of any genetic predisposition in younger cases or diagnostic ambiguity in elderly cases. The restriction to never, current, and recent former smokers eliminated any effect modification by time since cessation of smoking. We excluded subjects who had started smoking after age 40 years and subjects who smoked pipes or cigars, except where noted.

For the German Radon Study (7–9), we included 75-year-old subjects, since this was the upper age limit in that study, and we excluded subjects who paused their smoking for more than 5 years and males who smoked 1–8 cigarettes per day. The latter exclusion maintained consistency with the fitted model 2 based on continuous smoking intensity relative to never smoking, combining categories when data were sparse. Odds ratios for lung cancer by pack-years of smoking, relative to never smoking, were consistent with linearity within 12 intensity categories (figure 1). Homogeneity of the $\gamma_i$ estimates was rejected ($p < 0.001$). Since patterns were linear, the EOR/pack-year ($\gamma_i$) completely characterized odds ratios by pack-years within each intensity category.

Figure 2 (ESHs panel) shows the plot of the EOR/pack-year estimates by intensity. Estimates increased with intensities up to 15–20 cigarettes per day and then decreased, and closely conformed to the fitted model 2 based on continuous smoking variables (solid line). We included fixed offsets to enable unbiased estimation of smoking effects (25, 26).

For the Gansu Lung Cancer Study (11), we excluded females, since only 10 percent of them smoked and those who did smoke averaged only five cigarettes per day. Standard restrictions eliminated the 35 percent of men who smoked only Chinese long-stemmed pipes. In order to retain the 50 percent of smokers who smoked both cigarettes and long-stemmed pipes, we calculated a cigarette-equivalent variable using observed odds ratios. We did this by summing 1) cigarettes per day, 2) three times the liang (50 g) of tobacco used per month in hand-rolled cigarettes, and 3) 0.65 times the liang used per month in long-stemmed pipes. The Colorado Plateau Uranium Miners Study (12, 13) was a case-control study nested within an underground miners’ cohort. Smoking information came from four questionnaires administered between 1950 and 1960 and at other times through 1969. We assumed that smoking status at the most recent assessment was unchanged to the end of follow-up. The Population Health Study (17, 18) enrolled persons with four different types of cancers (pancreas, esophagus, and prostate cancer and multiple myeloma); however, we analyzed only cases with pancreatic and esophageal cancer and their controls. The latter site included only males. For the Population Health Study data, we extended the smoking cessation restriction to 10 years in order to increase sample size.

In all models, results were stratified by sex, study center, or location and were adjusted for age category and important study-specific factors (table 1). Models were fitted using the binary outcome module in the Epicure set of programs (27).

### RESULTS

Table 1 describes the studies analyzed and their selected smoking characteristics.

### Pack-years and smoking intensity

**Modeling lung cancer in the ESHS.** We illustrate the approach by expanding on the previous ESHS analysis (3). Since we were primarily interested in patterns rather than specific odds ratios, we computed odds ratios for the cross-classification of pack-years and intensity relative to never smoking, combining categories when data were sparse. Odds ratios for lung cancer by pack-years of smoking, relative to never smoking, were consistent with linearity within 12 intensity categories (figure 1). Homogeneity of the $\gamma_i$ estimates was rejected ($p < 0.001$). Since patterns were linear, the EOR/pack-year ($\gamma_i$) completely characterized odds ratios by pack-years within each intensity category. Figure 2 (ESHs panel) shows the plot of the EOR/pack-year estimates by intensity. Estimates increased with intensities up to 15–20 cigarettes per day and then decreased, and closely conformed to the fitted model 2 based on continuous smoking variables (solid line).

**Modeling cancer outcomes.** For each data set, odds ratios by pack-years were generally consistent with linearity within intensity categories. Across data sets, 105 intensity categories had finite $\gamma_i$ estimates. Eleven tests rejected
linearity at the $\alpha = 0.05$ level, with seven departures being convex and four concave. Adjusting for continuous intensity, there were no significant departures from linearity (table 2). EOR/pack-year estimates varied substantially within and across studies (figures 2 and 3). At higher intensities, EOR/pack-year estimates generally decreased with increasing intensity, except for the North American Radon Pooling Project, the Gansu Lung Cancer Study, and the Population Health Study (pancreas), where estimates were constant with intensity. The solid lines in figures 2 and 3 show fitted data from model 2 (parameter estimates in table 2). Smoking intensity significantly modified the EOR/pack-year at the 0.10 level in the ESHS, the German Radon Study, the North American Radon Pooling Project, the Colorado Plateau Uranium Miners Study, the National Bladder Cancer Study, the Spanish Bladder Cancer Study, and the Population Health Study (esophagus).

We considered several studies in greater detail. For the North American Radon Pooling Project, the fitted EOR function increased at low and high intensities.

### TABLE 1. Study design, summary statistics,* and factors used for adjustment in case-control studies of the relation between cigarette smoking and cancer

<table>
<thead>
<tr>
<th>Study (ref. no.)</th>
<th>Cancer site</th>
<th>Study design</th>
<th>Location of study</th>
<th>Dates of study</th>
<th>No. of cases</th>
<th>No. of controls</th>
<th>Adjustment factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>European Smoking and Health Study (6)</td>
<td>Lung</td>
<td>Hospital-based</td>
<td>Seven European study centers</td>
<td>1976–1980</td>
<td>7,804</td>
<td>15,207</td>
<td>Age, sex, and study center</td>
</tr>
<tr>
<td>German Radon Study (7)</td>
<td>Lung</td>
<td>Population-based</td>
<td>23 regions in Germany</td>
<td>1990–1997</td>
<td>4,071</td>
<td>4,628</td>
<td>Age, sex, and region</td>
</tr>
<tr>
<td>North American Radon Pooling Project (10)</td>
<td>Lung</td>
<td>Pooled data from seven population-based radon studies</td>
<td>United States and Canada</td>
<td>1982–1996</td>
<td>3,662</td>
<td>4,966</td>
<td>Age, sex, study, and radon</td>
</tr>
<tr>
<td>Gansu Lung Cancer Study (11)</td>
<td>Lung</td>
<td>Population-based</td>
<td>Two prefectures in Gansu Province, China</td>
<td>1994–1998</td>
<td>563 males</td>
<td>1,232 males</td>
<td>Age, prefecture, radon, and socioeconomic factors</td>
</tr>
<tr>
<td>Colorado Plateau Uranium Miners Study (12)</td>
<td>Lung</td>
<td>Nested case-control study from a cohort of underground miners</td>
<td>Colorado Plateau, United States</td>
<td>Enrolled in 1950–1960, with follow-up through 1990</td>
<td>263</td>
<td>10,322</td>
<td>Age, year, and radon</td>
</tr>
<tr>
<td>National Bladder Cancer Study (14)</td>
<td>Bladder</td>
<td>Population-based</td>
<td>10 areas in the United States</td>
<td>Enrollment for 1 year, starting between December 1977 and March 1978 (depending on area)</td>
<td>2,982</td>
<td>5,782</td>
<td>Age, sex, study center, race, and family history of urinary tract cancer</td>
</tr>
<tr>
<td>Spanish Bladder Cancer Study (15)</td>
<td>Bladder</td>
<td>Hospital-based</td>
<td>Five areas in Spain</td>
<td>1998–2001</td>
<td>1,090</td>
<td>1,264</td>
<td>Age, sex, region, fruit and vegetable intake, and employment in high-risk occupations</td>
</tr>
<tr>
<td>Population Health Study (18)</td>
<td>Pancreas</td>
<td>Population-based</td>
<td>Three areas in the United States</td>
<td>1986–1989</td>
<td>481</td>
<td>2,135</td>
<td>Age, sex, study center, race, and alcohol intake</td>
</tr>
<tr>
<td>Population Health Study (17)</td>
<td>Esophagus</td>
<td>Population-based</td>
<td>Three areas in the United States</td>
<td>1986–1989</td>
<td>581 males</td>
<td>1,357 males</td>
<td>Age, sex, study center, race, and alcohol intake</td>
</tr>
</tbody>
</table>

Table continues
the fitted model underpredicted EOR/pack-year between 15 and 30 cigarettes per day and overpredicted EOR/pack-year above 30 cigarettes per day, suggesting that lower-intensity data were distorting the model. At low intensities, there were relatively few smokers, but they were highly influential. For 289 smokers (125 cases and 164 controls) who smoked fewer than 10 cigarettes per day, the median number of pack-years was 14.7 and the interquartile range was 10.6–19.6 pack-years, in comparison with 45 pack-years and 32.8–61.1 pack-years, respectively, for all North American Radon Pooling Project smokers. After omitting smokers who smoked fewer than 10 cigarettes per day, estimates were $\varphi_1 = 1.91$ and $\varphi_2 = -0.27$, which were similar to estimates in other data sets (table 2). The EOR function decreased at higher intensities and conformed closely to EOR/pack-year estimates (figure 2, NARnP panel, dashed-dotted line).

Fitted EOR curves for the Population Health Study (pancreas) and Population Health Study (esophagus) increased sharply at low intensities. After omitting smokers who
smoked fewer than five cigarettes per day (four cases and 23 controls in the Population Health Study (pancreas) and seven cases and 14 controls in the Population Health Study (esophagus)), the EOR functions were similar to those of other studies (table 2; figure 3, dotted-dashed lines).

Estimates of $b$ varied widely, while estimates of $u_1$ and $u_2$ appeared similar across data sets (table 2). This suggests a general commonality of intensity patterns, with differences in the fitted curves in figures 2 and 3 being due primarily to differences in $b$. To examine this more formally, we computed summary estimates across studies of $u_1 = 2.72$ and $u_2 = 0.48$ using a random-effects approach (28, 29). Using model 2, we estimated $b$ for each data set with $u_1$ and $u_2$ fixed at the summary values (figures 2 and 3, dashed lines). Table 2 displays $p$ values for the 2-df test of the null hypothesis that study-specific intensity parameters equaled the summary estimates. Results indicated statistical consistency of intensity parameters for all studies except the North American Radon Pooling Project, the National Bladder Cancer Study, the Spanish Bladder Cancer Study, and the Population Health Study (esophagus). For the North American Radon Pooling Project, homogeneity was rejected after omission of smokers who smoked fewer than 10 cigarettes per day, although intensity estimates were similar to summary estimates. For the National Bladder Cancer Study and the Spanish Bladder Cancer Study, differences between study-specific and summary intensity parameters were due to initiation of smoking at age 12 years or younger ($p = 0.19$ and $p = 0.29$, respectively, in restricted data), and for the Population Health Study (esophagus), differences were due to smoking of very low intensity ($p = 0.29$, respectively).

Variations in smoking intensity effects by study center

Except for the Colorado Plateau Uranium Miners Study, each data set included multiple centers or independent studies. Assuming that centers or component studies represented independent replicates, there were significant variations in $b$ estimates only for ESHS centers ($p < 0.001$) and for the contributing studies of the North American Radon Pooling Project ($p = 0.02$). In contrast, intensity patterns (i.e., $u_1$ and $u_2$ estimates) were homogeneous across centers within each study, except for the ESHS. However, intensity effects in the ESHS were consistent with homogeneity ($p = 0.12$) after additional adjustment for type of cigarette smoked (filter, nonfilter, or mixed types).
DISCUSSION

Previous analysis of the ESHS identified a linear relation for pack-years of smoking and lung cancer risk, with a distinct pattern for the effects of smoking intensity (3). Our analysis of additional studies of lung cancer and cancers of the bladder, oral cavity, pancreas, and esophagus suggests that this relation is broadly consistent across diverse cancer sites. At lower intensities, variations in the EOR/pack-year were consistent with an increasing intensity effect (direct exposure rate or an enhanced potency effect); however, results are subject to uncertainty because of the relatively few case subjects and the increased variability due to limited ranges for pack-years. At higher intensities, variations in the EOR/pack-year with intensity were consistent with a decreasing intensity effect (inverse exposure rate, or a reduced potency or wasted exposure effect), with smoking at a lower intensity for a longer duration being more harmful than smoking at a higher intensity for a shorter duration.

We also found that intensity effects were generally consistent in magnitude across diverse smoking-related cancer sites. It may at first seem counterintuitive that effects of smoking intensity could be similar in pattern and magnitude across diverse cancer sites. However, the overall levels of risk from smoking are reflected in the diverse EOR/pack-year estimates, while intensity effects reflect only the modulation of site-specific disease risk.

Cigarette smoking is most strongly associated with lung cancer, and EOR curves for the ESHS, the German Radon Study, and the North American Radon Pooling Project generally lay above curves for other cancer sites (figure 2). The two- to threefold greater EORs/pack-year in the German Radon Study and the North American Radon Pooling Project, as compared with the ESHS, were unexpected. Reasons for the greater EOR/pack-year estimates are unknown. The ESHS was hospital-based, and smokers could have been overrepresented among control patients, although restricting control diseases did not change the results. Risk estimates for the Gansu Lung Cancer Study data were low and reflect the lower odds ratios for lung cancer and smoking found in many studies carried out in China (30–34), as compared with studies in Western countries. The lower smoking risk in the Colorado Plateau Uranium Miners Study has been noted previously and may be due to different smoking

**FIGURE 2.** Estimates of the excess odds ratio for lung cancer per pack-year of cigarette smoking in the European Smoking and Health Study (ESH) (1976–1980), the Colorado Plateau Uranium Miners Study (COUS) (1960–1990), the Gansu Lung Cancer Study (GLCS) (1994–1998), the German Radon Study (GRS) (1990–1997), and the North American Radon Pooling Project (NARnP), a combined data set from six studies (1982–1996). The figure shows results based on a linear odds ratio model within categories of cigarettes smoked per day (black squares; bars, 95% confidence interval) and the estimated model 2 (equation 2 in text) fitted to data from each study (solid line; dotted lines, pointwise 95% confidence interval). The figure also shows results based on model 2 with the pack-years parameter ($\beta$) estimated and the intensity parameters fixed at summary values for all studies combined ($\phi_1 = 2.72$ and $\phi_2 = -0.479$) (dashed line). Model 2 was also fitted to NARnP data that omitted smokers who smoked fewer than 10 cigarettes per day (dashed-dotted-dotted line).
practices among Native American miners, the low smoking rate in the study (smoking underground was banned in the 1950s), or misclassification of amount of smoking (35–39).

Smoking intensity effects were consistent within and across cancer sites, particularly after mild restrictions were imposed. We imposed restrictions rather than using a non-parametric generalized additive model (40) or a semiparametric spline model (13), which would have complicated comparisons across data sets. For completeness, we fitted cubic B-splines for intensity effects with one interior knot, and results were similar to those of the parametric models in figures 2 and 3. Confidence bounds were wide, particularly at low smoking intensities, highlighting the uncertainty of the direct intensity effect.

There is uncertainty about whether and to what extent intensity effects reflect smoking inhalation patterns and/or nicotine-related dependencies or fundamental biologic processes, such as modifications of activation pathways for carcinogens in cigarette smoke or DNA repair capacity. Patterns of variation in the EOR/pack-year by intensity could reflect inhalation practices if lower-intensity and higher-intensity smokers inhale fewer carcinogens per cigarette relative to moderate-intensity smokers. This would result in reduced risks at lower intensities and declining risks at higher intensities (1). Reduced frequency or reduced depth of inhalation is a plausible explanation; however, analysis of the ESHS data provided no indication that inhalation was related to smoking intensity after adjustment for total exposure (3).

The intensity effects may have been influenced by misclassification of the number of cigarettes smoked per day. However, consistency of effects for multiple diseases and for studies conducted in diverse locations suggests that an unusual pattern of misclassification would be required in order to fully explain the results. To induce an inverse intensity effect, misclassification would have to increase with increasing intensity above 15–20 cigarettes smoked per day, resulting in greater bias toward no effect of intensity and a declining exposure-response relation. To induce a direct intensity effect, misclassification would have to decrease with intensity up to 15–20 cigarettes smoked per day, resulting in less bias toward no effect and an increasing exposure-response relation. If only low and high intensities were differentially misclassified, then differential
TABLE 2. Parameter estimates obtained from fitting models of the cigarette smoking–cancer relation to data from several studies*

<table>
<thead>
<tr>
<th>Cancer site</th>
<th>Study (ref. no.)</th>
<th>$\beta$</th>
<th>$\varphi_1$</th>
<th>$\varphi_2$</th>
<th>$p$ value for test of hypothesis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Linearity†</td>
</tr>
<tr>
<td>Analyses of unrestricted data</td>
<td>Lung</td>
<td>European Smoking and Health Study (6)</td>
<td>0.005</td>
<td>2.89</td>
<td>−0.51</td>
</tr>
<tr>
<td></td>
<td>Lung</td>
<td>German Radon Study (7)</td>
<td>0.007</td>
<td>3.14</td>
<td>−0.50</td>
</tr>
<tr>
<td></td>
<td>Lung</td>
<td>North American Radon Pooling Project (10)</td>
<td>2.45</td>
<td>−1.02</td>
<td>0.19</td>
</tr>
<tr>
<td></td>
<td>Lung</td>
<td>Gansu Lung Cancer Study (11)</td>
<td>0.0004</td>
<td>2.84</td>
<td>−0.38</td>
</tr>
<tr>
<td></td>
<td>Lung</td>
<td>Colorado Plateau Uranium Miners Study (12)</td>
<td>0.014</td>
<td>1.79</td>
<td>−0.43</td>
</tr>
<tr>
<td></td>
<td>Bladder</td>
<td>National Bladder Cancer Study (14)</td>
<td>0.001</td>
<td>3.06</td>
<td>−0.59</td>
</tr>
<tr>
<td></td>
<td>Bladder</td>
<td>Spanish Bladder Cancer Study (15)</td>
<td>0.147</td>
<td>1.10</td>
<td>−0.29</td>
</tr>
<tr>
<td></td>
<td>Oral cavity</td>
<td>Puerto Rico Oral Cancer Study (16)</td>
<td>0.032</td>
<td>1.59</td>
<td>−0.31</td>
</tr>
<tr>
<td></td>
<td>Pancreas</td>
<td>Population Health Study (18)</td>
<td>0.593</td>
<td>−1.97</td>
<td>0.31</td>
</tr>
<tr>
<td></td>
<td>Esophagus</td>
<td>Population Health Study (17)</td>
<td>2.46</td>
<td>−1.58</td>
<td>0.14</td>
</tr>
</tbody>
</table>

Analyses of restricted data

<table>
<thead>
<tr>
<th>Cancer site</th>
<th>Study (ref. no.)</th>
<th>$\beta$</th>
<th>$\varphi_1$</th>
<th>$\varphi_2$</th>
<th>$p$ value for test of hypothesis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Lung</td>
<td>North American Radon Pooling Project¶</td>
<td>0.017</td>
<td>1.91</td>
<td>−0.27</td>
</tr>
<tr>
<td></td>
<td>Bladder</td>
<td>National Bladder Cancer Study#</td>
<td>0.001</td>
<td>2.95</td>
<td>−0.57</td>
</tr>
<tr>
<td></td>
<td>Bladder</td>
<td>Spanish Bladder Cancer Study#</td>
<td>0.061</td>
<td>1.68</td>
<td>−0.36</td>
</tr>
<tr>
<td></td>
<td>Pancreas</td>
<td>Population Health Study**</td>
<td>0.001</td>
<td>2.16</td>
<td>−0.32</td>
</tr>
<tr>
<td></td>
<td>Esophagus</td>
<td>Population Health Study**</td>
<td>0.043</td>
<td>1.01</td>
<td>−0.27</td>
</tr>
</tbody>
</table>

* Odds ratio $= 1 + \beta \times d \times \exp(\varphi_1 \ln(n) + \varphi_2 \ln(n)^2)$, where $d$ is total pack-years of exposure, $n$ is the number of cigarettes smoked per day, and $\beta$, $\varphi_1$, and $\varphi_2$ are parameters. Adjustment factors and descriptions of the data are given in table 1.

† $p$ value for departure from linearity of the excess odds ratio ($\hat{d} = 0$ in model 3 (equation 3 in text)).

‡ $p$ value from a 2-df chi-squared test of no intensity effect ($\varphi_1 = 0$, $\varphi_2 = 0$).

§ $p$ value from a 2-df chi-squared test of homogeneity of intensity parameters with weighted mean of intensity parameters for all data sets ($\varphi_1 = 2.72$ and $\varphi_2 = −0.48$).

¶ Excludes smokers who smoked fewer than 10 cigarettes per day.

# Excludes smokers who started smoking at age 12 years or under.

** Excludes smokers who smoked fewer than five cigarettes per day.

misclassification would occur with pack-years of smoking, since duration is measured with relative accuracy. This would induce concavity in the relation between odds ratios and pack-years at low and high intensities. However, in the current study, relations between odds ratios and pack-years remained linear over the full range of intensities, suggesting that misclassification does not explain the observed patterns.

We also conducted a sensitivity analysis to explore the possibility that intensity patterns were due to nicotine satiation effects, whereby for high-intensity smokers, numbers of cigarettes smoked per day increasingly overestimate the “internal” exposure rate (4). Those analyses indicated that while some overestimation of internal exposure rate by the number of cigarettes smoked per day may have occurred, it is unlikely that nicotine satiation factors fully explain the inverse exposure rate pattern.

Intensity patterns, particularly those above 15–20 cigarettes per day, could reflect a differential “willingness to participate” for cases and controls—that is, intensity-related differential selection bias. This is unlikely, since cases’ diagnoses were incident, cases were typically interviewed within a few months of diagnosis, and participation proportions were generally high. In 13 of the 15 studies included in these analyses, participation proportions were similar in cases and controls and exceeded 70 percent, with participation in two studies (the Gansu Lung Cancer Study and the Spanish Bladder Cancer Study) exceeding 86 percent. There were two exceptions: the German Radon Study, with 76 percent and 44 percent participation among cases and controls, respectively, and the Population Health Study (pancreas), with 46 percent and 76 percent participation among cases and controls. In the latter study, the lower case proportion was due to the use of subject-only interviews. Intensity patterns were no different in these two studies. In addition, for differential enrollment to have influenced results, one would have to postulate that, conditional on equal total pack-years, the proportion of nonparticipating controls relative to nonparticipating cases was greater for intensities in the range of 15–25 cigarettes per day (resulting in too few controls and an increased EOR/pack-year) and then declined with increasing intensity (resulting in a decreased EOR/pack-year). In addition, the pattern of varying proportions of nonparticipation would have to be present across the diverse cancer sites.

The intensity function has the form $g(n) = \exp(\varphi_1 \ln(n) + \varphi_2 \ln(n)^2)$. Since pack-years is the product of duration of smoking, $y$, and intensity, $n$ (i.e., $d = y \times n$) and

Am J Epidemiol 2007;166:479–489
OR = 1 + β d g(n), equation 2 can be written in terms of duration and intensity as

\[ \text{OR} = 1 + \beta \gamma \exp\{\kappa \ln(n) + \varphi_2 \ln(n)^2\}, \]

where \( \kappa = \varphi_1 + 1 \). Thus, model 2 is comparable to a model that is linear in duration for a fixed intensity. However, interpretation of \( \beta \) and intensity effects \( \kappa \) and \( \varphi_2 \) is problematic when either \( \gamma \) or \( n \) is fixed, because of the changing total pack-years of exposure.

The relative simplicity of model 2 derives from the approximate linearity of the odds ratios for pack-years within levels of intensity and from the fact that a single parameter, \( \gamma_i \), fully characterizes linearity. Variations in \( \gamma_i \) (EOR/pack-year) then represent the modulating effects of intensity. In principle, we could develop a similar model for the EOR by pack-years with allowance for duration of exposure. However, odds ratios for pack-years within duration categories are not linear; that is, there is no comparable model 1 within duration categories. A simple characterization of changes in the EOR by pack-years with allowance for duration is not readily apparent.

Our analysis included never smokers, current smokers, and recent former smokers. The extent to which age, sex, time since stopping smoking, and other factors influenced estimates of EOR/pack-year and the shape of the intensity function is the subject of another analysis (4).

After adjustment for total smoking exposure, we found a general and specific consistency in the effects of smoking intensity across diverse smoking-related cancer sites, particularly decreasing effects of intensity above a level of 15–20 cigarettes smoked per day. Because of uncertainty at low intensities, it remains unclear whether enhanced intensity effects occur at smoking levels below 15 cigarettes per day, and results for low intensities should be interpreted cautiously. Further analyses are needed to verify these observations.

ACKNOWLEDGMENTS

This research was supported by the Intramural Research Program of the National Cancer Institute, National Institutes of Health (Bethesda, Maryland). Dr. Daniel Krewski is the NSERC/SSHRC/McLaughlin Chair in Population Health Risk Assessment at the University of Ottawa (Ottawa, Ontario, Canada).

Conflict of interest: none declared.

REFERENCES