

# **The role of smoking and diet in explaining educational inequalities in lung cancer incidence**

Gwenn Menvielle <sup>1 2 3 \*</sup>, Hendriek Boshuizen <sup>3</sup>, Anton E. Kunst <sup>1</sup>, Susanne O. Dalton <sup>4</sup>, Paolo Vineis <sup>5</sup>, Manuela M. Bergmann <sup>6</sup>, Silke Hermann <sup>7</sup>, Pietro Ferrari <sup>8</sup>, Ole Raaschou-Nielsen <sup>4</sup>, Anne Tjønneland <sup>4</sup>, Rudolf Kaaks <sup>7</sup>, Jakob Linseisen <sup>7</sup>, Maria Kostí <sup>7</sup>, Antonia Trichopoulou <sup>7</sup>, Vardis Dilis <sup>7</sup>, Domenico Palli <sup>9</sup>, Vittorio Krogh <sup>10</sup>, Salvatore Panico <sup>11</sup>, Rosario Tumino <sup>12</sup>, Frederike L. B. üchner <sup>3</sup>, Carla H. Van Gils <sup>13</sup>, Petra H.M Peeters <sup>13</sup>, Tonje Braaten <sup>14</sup>, Inger T. Gram <sup>14</sup>, Eiliv Lund <sup>14</sup>, Laudina Rodriguez <sup>15</sup>, Antonio Agudo <sup>16</sup>, Maria-José Sánchez <sup>17 18</sup>, Maria-José Tormo <sup>19 18</sup>, Eva Ardanaz <sup>18 20</sup>, Jonas Manjer <sup>21</sup>, Elisabet Wirfält <sup>23 22</sup>, Göran Hallmans <sup>24</sup>, Torgny Rasmuson <sup>25</sup>, Sheila Bingham <sup>26</sup>, Kay-Tee Khaw <sup>26</sup>, Naomi Allen <sup>27</sup>, Tim Key <sup>27</sup>, Paolo Boffetta <sup>8</sup>, Eric J. Duell <sup>8</sup>, Nadia Slimani <sup>8</sup>, Valentina Gallo <sup>28</sup>, Elio Riboli <sup>28</sup>, H Bas Bueno-de-Mesquita <sup>3</sup>

<sup>1</sup> Department of Public Health University Medical Center Rotterdam , Erasmus MC Rotterdam,NL

<sup>2</sup> Santé publique et épidémiologie des déterminants professionnels et sociaux de la santé INSERM : U687 , IFR69 , Université Paris Sud - Paris XI , Université de Versailles-Saint Quentin en Yvelines , Hôpital Paul Brousse 16, av Paul Vaillant Couturier 94807 VILLEJUIF,FR

<sup>3</sup> RIVM, National Institute of Public Health and Environment National Institute of Public Health and Environment , Bilthoven,NL

<sup>4</sup> Institute of Cancer Epidemiology Danish Cancer Society , Copenhagen,DK

<sup>5</sup> University of Torino and CPO-Piemonte Università degli studi di Torino , Torino,IT

<sup>6</sup> Department of Epidemiology German Institute of Human Nutrition , Potsdam-Rehbrücke,DE

<sup>7</sup> Division of Cancer Epidemiology German Cancer Research Center , Heidelberg,DE

<sup>8</sup> Nutrition and hormon group International Agency for Research on Cancer , Lyon,FR

<sup>9</sup> Molecular and Nutritional Epidemiology Unit CSPO-Scientific Institute of Tuscany , Florence,IT

<sup>10</sup> Department of Preventive & Predictive Medicine Italian National Center Institute , Nutritional Epidemiology Unit Milan,FR

<sup>11</sup> Department of Clinical and Experimental Medicine Federico II University , Naples,IT

<sup>12</sup> Cancer Registry Azienda Ospedaliera "Civile M.P. Arezzo" , Ragusa,IT

<sup>13</sup> Julius Center for Health Sciences and Primary Care University Medical Center , Utrecht,NL

<sup>14</sup> Institute of Community Medicine University of Tromso , Tromso,NO

<sup>15</sup> Public Health Directorate Public Health Directorate , Asturias,ES

<sup>16</sup> Unit of Nutrition, Environment and Cancer Catalan Institute of Oncology , Barcelona,ES

<sup>17</sup> Andalusian School of Public Health Andalusian School of Public Health , Granada,ES

<sup>18</sup> CIBER Epidemiologia y Salud Publica CIBERESP , ES

<sup>19</sup> Department of Epidemiology Murcia Health Council , Murcia,ES

<sup>20</sup> Public Health Institute of Navarra Public Health Institute of Navarra , Pamplona,ES

<sup>21</sup> Department of surgery Malmö University Hospital , SE

<sup>22</sup> Department of Clinical Sciences Department of Clinical Sciences , Malmö,SE

<sup>23</sup> Nutrition Epidemiology Lund University , SE

<sup>24</sup> Department of Public Health and Clinical Medicine, Nutritional Research Umea University , Umea,SE

<sup>25</sup> Department of Radiation Sciences, Oncology Umea University , Umea,SE

<sup>26</sup> MRC Centre for Nutritional Epidemiology and Cancer Prevention and Survival University of Cambridge , Cambridge,GB

<sup>27</sup> Cancer Epidemiology Unit University of Oxford , GB

<sup>28</sup> Departement of Epidemiology and Public Health Imperial College London , GB

\* Correspondence should be addressed to: Gwenn Menvielle <gwenn.menvielle@inserm.fr >

## **Abstract**

### **Background**

**Studies in many countries have reported higher lung cancer incidence and mortality in individuals with lower socioeconomic status.**

### **Methods**

To investigate the role of smoking in these inequalities, we used data from 391,251 participants in the European Prospective Investigation into Cancer and Nutrition study, a cohort of individuals in 10 European countries. We collected information on smoking (history and quantity) and education through questionnaires at study entry and gathered data on lung cancer incidence for a mean of 8.4 years. Socioeconomic status was defined as the highest attained level of education, and participants were grouped by sex and region of residence (Northern Europe, Germany, or Southern Europe). Relative indices of inequality (RIIs) of lung cancer risk unadjusted and adjusted for smoking were estimated using Cox regression models. Additional analyses were performed by histologic type.

## Results

During the study period, 939 men and 692 women developed lung cancer. Inequalities in lung cancer risk ( $RII_{men} = 3.62$ , 95% confidence interval [CI] = 2.77 to 4.73, 117 vs 52 per 100,000 person-years for lowest vs highest education level;  $RII_{women} = 2.39$ , 95% CI = 1.77 to 3.21, 46 vs 25 per 100,000 person-years) decreased after adjustment for smoking but remained statistically significant ( $RII_{men} = 2.29$ , 95% CI = 1.75 to 3.01;  $RII_{women} = 1.59$ , 95% CI = 1.18 to 2.13). Large RIIs were observed among men and women in Northern European countries and among men in Germany, but inequalities in lung cancer risk were reverse (RIIs < 1) among women in Southern European countries. Inequalities differed by histologic type. Adjustment for smoking reduced inequalities similarly for all histologic types and among men and women in all regions.

## Conclusion

**Self reported smoking consistently explains approximately 50% of the inequalities in lung cancer risk due to differences in education.**

**MESH Keywords** Adult ; Aged ; Confounding Factors (Epidemiology) ; Educational Status ; Europe ; epidemiology ; Female ; Food Habits ; Fruit ; Humans ; Incidence ; Lung Neoplasms ; epidemiology ; etiology ; pathology ; Male ; Middle Aged ; Odds Ratio ; Proportional Hazards Models ; Prospective Studies ; Questionnaires ; Risk Assessment ; Risk Factors ; Sex Factors ; Smoking ; adverse effects ; epidemiology ; Social Class ; Vegetables

**Author Keywords** lung cancer incidence ; cohort study ; Europe ; EPIC ; men ; women ; smoking ; diet ; education

Socioeconomic inequalities in lung cancer incidence and mortality are consistently found in North America or in Europe; that is, higher incidence and mortality rates are observed among subjects with lower socioeconomic position (1–4). A better understanding of the mechanisms underlying these inequalities will help to define the most effective preventive policies for the social groups with the highest cancer incidence. As a first step to uncovering these mechanisms, it is important to identify the intermediate factors (mainly behavioral, biologic, or environmental) that explain these inequalities.

It has been suggested that inequalities in smoking could explain the socioeconomic inequalities in lung cancer incidence. However, the few studies conducted on this topic (1, 2, 5) found that smoking explained at most 40% of socioeconomic inequalities in lung cancer incidence. Two main explanations have been suggested for this finding. First, there might be residual confounding due to misclassification of smoking. Given the strength of the association between smoking and lung cancer, it is essential to conduct analyses that minimize any residual confounding due to imprecision in the measurement of smoking. Second, other risk factors, such as diet or occupational exposures, may explain the remaining inequalities.

Furthermore, the published studies consistently report larger socioeconomic inequalities in lung cancer incidence and mortality in Northern European countries when compared with Southern European countries (3, 5). Again, these differences in the degree of inequality have been linked to differences in smoking behavior between countries (3, 6), but no study has been conducted across Europe to test this hypothesis.

Lastly, the association between smoking and lung cancer risk differs by histological type (7, 8). Smoking is most strongly associated with the risk of small cell carcinoma, followed by squamous carcinoma, and a weaker association is observed with adenocarcinoma. The association between smoking cessation and reduced lung cancer risk is strongest for small cell carcinoma and weakest for adenocarcinoma. As a consequence, inequalities in lung cancer incidence and the role of smoking in explaining these inequalities may differ by histological type. This issue has, however, not been thoroughly investigated (9).

The main objective of this study was to investigate the role of smoking in explaining socioeconomic inequalities in lung cancer incidence in the European Prospective Investigation into Cancer and Nutrition (EPIC). Fruit and vegetable consumption has been found to be associated with reduced lung cancer risk, especially in smokers (10). Therefore we decided to assess the role of diet in explaining socioeconomic inequalities in lung cancer incidence. Analyses were also stratified by geographic region and by histological type.

## SUBJECTS AND METHODS

### Population

The EPIC cohort is a multicenter prospective cohort conducted in 23 centers in 10 European countries (France, Italy [Florence, Varese, Ragusa, Turin, and Naples], Spain [Asturias, Granada, Murcia, Navarra, and San Sebastian], Great Britain [Cambridge, Oxford], The Netherlands [Utrecht, Bilthoven], Greece, Germany [Postdam, Heidelberg], Sweden [Malmö, Umeå], Denmark [Copenhagen, Aarhus], and Norway). The study started at the beginning of the 1990s and included more than 500,000 persons mostly aged between 40 and 65 years. In most centers, subjects were recruited from the general population in a given geographic area (country, region, or city). The French cohort consists of members of the health insurance program for school and university employees, a large part of the Spanish and Italian centers include blood donors, the Utrecht cohort is based on participants in a mammography screening program, and the cohort in Florence also includes screening program participants. In Oxford, most of the cohort consists of 'health conscious' subjects (vegetarian volunteers or healthy eaters). The cohorts in France, Norway, Utrecht, and Naples include only women. All subjects completed a dietary and lifestyle questionnaire at the time of enrollment in the cohort.

Subjects with prevalent cancer at baseline (except nonmelanoma skin cancer) (n=20,866) or with length of follow-up equal to zero (n=341) were excluded from the analysis. We also excluded subjects with a ratio for energy intake versus energy expenditure in the top and bottom 1% (n=9674); subjects with missing information on smoking status, diet, or education (n=31,728); and subjects with missing information on date of diagnosis for an incident cancer before the incident lung cancer (n=12). The date of diagnosis was available for all lung cancer patients. Compared with other cohorts, the French cohort was a demographically very homogeneous population and thus was excluded from the analyses (n=61,704). The analysis was finally based on 391,251 participants, among whom 939 men and 692 women with lung cancer were identified.

### **End points**

Incident cases of lung cancer were identified by population-based cancer registries in Denmark, Italy, The Netherlands, Norway, Spain, Sweden, and the United Kingdom or by active follow-up in Germany, and Greece. The end of the follow-up period occurred between December 2002 and December 2006. The mean follow-up was 8.4 years.

The outcome variable was first primary lung cancer (ICD 10: C33-C34). Participants who developed a different primary cancer before lung cancer were censored at the date of diagnosis of the earlier cancer. We conducted analyses using all lung cancers combined and separate analyses for the four main histological types: adenocarcinoma (International Classification of Diseases for Oncology [ICD-O]2-M codes 8140, 8143, 8200, 8211, 8230, 8250–1, 8260, 8300, 8480–1, 8490, 8550, and 8310) (n=550), squamous cell carcinoma (ICDO 8052, 8070–3, 8075, and 8123) (n=351), small cell carcinoma (ICDO 8041–6) (n=276), and large cell carcinoma (ICDO 8012, 8020–1 and 8082) (n=137). A substantial number of incident lung cancers (n=317) could not be defined as one of the four main histological types because of lack of information.

### **Statistical analyses**

Information about the highest attained educational level was collected using a questionnaire specific to each country and classified according to four categories (primary education or less, vocational secondary education, other secondary education, and college or university).

When studying lung cancer, confounding due to shortcomings in adjustment for smoking is always an issue. The first step of the analyses was to search for the smoking-adjusted model that best fit the data. We took into account several aspects of tobacco consumption, including quantity and duration. The final model included smoking status at recruitment as a categorical variable (never, current, or former smoker), age at the start of, and duration, of smoking (in years) as continuous variables, a linear and a quadratic term for current quantity smoked (number of cigarettes per day), and two interaction terms between duration and quantity and between age at start and duration. In addition, we introduced, for each continuous smoking variable, a dummy variable, which was coded 1 when missing (0 otherwise). Former smokers were defined as self-declared former smokers of any type of tobacco.

We then searched for the smoking and diet adjusted model that best fit the data. We only considered dietary variables (continuous or coded in quintiles) that had been reported to be statistically significantly associated with lung cancer incidence—fruit, vegetable, meat, and egg consumption (10) (Linseisen J, personal communication)—and interaction terms between smoking status and dietary variables were considered. The final model included the smoking variables, total fruit and vegetable consumption (continuous), and the interaction between smoking status and the consumption of fruits and vegetables.

Analyses were conducted with Cox regression models that were stratified by center and age at baseline in 1-year age categories using follow-up as the time factor. The proportional hazards assumption was verified by visual inspection of log–log plots of survival. In addition to estimating hazard ratios (HRs), we computed relative indices of inequality (RIIs) using the highest educational level as the reference category (11). To calculate the RIIs, we used a relative measure of education. This is a ranked variable that is equal to, for each educational group, the mean proportion of the population with a higher level of education and was computed as follows. If the highest educational group is 20% of the population, this ranked variable is assigned a value of  $0.20/2=0.10$ . If the next highest educational group is

30% of the population, it is assigned a value of  $0.20+0.30/2=0.35$ , etc. We used a Cox regression model with cancer incidence as the outcome variable and this ranked variable as the explanatory variable. The RII corresponds to the estimate obtained for this ranked variable and quantifies the assumed linear effect of the relative level of education on lung cancer risk. Thus, the RII expresses inequality within the whole socioeconomic continuum and can be interpreted as the ratio of lung cancer incidence between the lowest educated (0<sup>th</sup> percentile) and the highest educated (100<sup>th</sup> percentile). Because the RII takes into account the size and relative position of each educational group, it is appropriate for comparing populations with different educational distributions. The ranked variable was computed separately for each stratum of sex, age category, and center. For the very small health conscious Oxford cohort (12), because of its very specific educational distribution, we assigned the distribution from the Cambridge cohort.

The following models were considered (all were stratified by center and age): 1) a model including only education, 2) a model adjusted for current smoking at recruitment only (without any information on duration), 3) a model fully adjusted for smoking, and 4) a model fully adjusted for smoking and adjusted for fruit and vegetable intake. We quantified the change in RII between model A and a further adjusted model B with the following formula:  $(RII_{\text{model A}} - RII_{\text{model B}}) / (RII_{\text{model A}} - 1) \times 100$ .

We tried to assess (in a crude way) the potential residual confounding due to smoking in different ways. In other words, we tried to answer the following question: to what extent are inequalities observed in lung cancer incidence when controlling for smoking due to residual confounding for smoking? First, we tested the interaction between smoking status and education and we conducted additional analyses stratified by smoking status. If remaining inequalities are explained by residual confounding, then education and lung cancer incidence should not be associated among never smokers. On the contrary, if inequalities are observed among never smokers, this is in favor of other risk factors involved in inequalities in lung cancer incidence and residual confounding is unlikely to be the only explanation for the remaining inequalities. We also compared the results of the second model (adjusted for current smoking at recruitment) with those of the third model (fully adjusted for smoking). If adjusting only for current smoking already substantially reduces the inequalities in lung cancer incidence and if these inequalities are comparatively little reduced in a model fully adjusted for smoking, this would suggest that crude tobacco related variables can already account for an important part of inequalities in lung cancer incidence and this would not be in favour of important residual confounding due to an imprecise measurement of smoking.

Analyses were conducted for all centers together and for the three defined geographic regions: Northern Europe (Norway, Sweden, Denmark, The Netherlands, and the UK), Southern Europe (Spain, Italy, and Greece), and Germany, and we tested for interaction between regions. This a priori grouping was based on previous work focusing on the smoking epidemic and socioeconomic inequalities in smoking (6, 13, 14). Previous publications clearly distinguish different smoking patterns across Europe. In Northern European countries, higher smoking prevalence is found among lower educated men and women of all ages. In Southern European countries, the association between education and tobacco consumption differs by age and sex: among older subjects, higher smoking prevalence is found among the higher educated men and women; in contrast, among younger subjects, generally higher smoking prevalence occurs among lower educated men and higher educated women.

To quantify the reduction in absolute inequality in lung cancer risk if the one major risk factor—smoking—could be eliminated, we computed age-standardized incidence rates by sex and education in a virtual population that would be the EPIC cohort in which all current smokers had stopped smoking and experienced the same lung cancer incidence rates as the ex-smokers. We then compared the rates in this virtual population with the observed rates in EPIC.

All statistical tests were two-sided. P values less than .05 were considered statistically significant.

## RESULTS

The education level of men and women was lower on average in Southern Europe than in Germany and Northern Europe (Table 1). Lung cancer incidence rates for men did not differ substantially between regions; however, rates for women were more than two times higher in Northern Europe than in Germany or in Southern Europe.

We found a higher proportion of current smokers and a lower proportion of never smokers among men and women with the lowest level of education in all geographic regions, except among women from Southern Europe (Table 2). In Germany, the highest percentage of women never smokers was observed in the group with the lowest level of education, whereas the percentage of current women smokers did not differ substantially across educational levels. We observed an increase in duration of smoking with decreasing educational level, except in Southern Europe where no association was found among women and the duration was longer than in other groups only among men with a primary or less education. The relationship was less clear for quantity of smoking (number of cigarettes smoked per day). The median fruit and vegetable consumption increased with education both in men and women in Northern Europe and Germany but not in Southern Europe (Table 2).

Lung cancer risk in men and women increased as educational level decreased ( $RII_{\text{men}} = 3.62$ , 95% confidence interval [CI] = 2.77 to 4.73, 117 vs 52 per 100,000 person-years for lowest vs highest education level;  $RII_{\text{women}} = 2.39$ , 95% CI = 1.77 to 3.21, 46 vs 25 per

100,000 person-years) (Table 3 ). After adjusting for smoking, the HRs remained statistically significant among men and women with primary education or less and among men with secondary vocational education. RIIs decreased by about 50% among men and 60% among women when models were adjusted for smoking but remained statistically significant ( $RII_{men}=2.29$ , 95% CI = 1.75 to 3.01;  $RII_{women}=1.59$ , 95% CI = 1.18 to 2.13). Comparison between model 2 (crude adjustment for smoking) and model 3 (refined adjustment for smoking) revealed that the refined adjustment lowered the estimates of RIIs but did not have impact on the statistical significance of any of these estimates. Only a marginal change was observed after further adjustment for fruit and vegetable intake.

The interaction between education and geographic region was statistically significant among men and women in model 1 (Table 4 ). Among men from all geographic regions and among women in Northern Europe, crude lung cancer risk was higher among lower educated participants. RIIs were nevertheless substantially lower among men from Southern Europe than Northern Europe and Germany. No statistically significant association was observed among women in Germany and among men in Southern Europe. Among women in Southern Europe, the RII was statistically significantly less than 1, which means that higher lung cancer risks were found among higher educated women. Adjustment for smoking moved all estimates toward unity. The interaction between education and geographic region remained statistically significant only among women when smoking was adjusted for. The RIIs remained statistically significantly greater than 1 among men and women in Northern Europe and among men in Germany. The comparison of models 2 (crude adjustment for smoking) and 3 (refined adjustment for smoking) leads to the same conclusion by region as that found among all participants namely that the refined adjustment lowered the estimates of RIIs but did not have impact on the statistical significance of any of these estimates. Additional adjustment for fruit and vegetable intake did not change the estimates in men and women.

Differences in risk for lung cancer by educational level according to histological type were found (Table 5 ). Among men, the RIIs were largest for small cell and squamous carcinoma in all analyses. However, the differences were statistically significant in model 1 (unadjusted for smoking) only (test for heterogeneity  $P=.04$ ). The RIIs were substantially smaller for adenocarcinoma and large cell carcinoma, and not statistically significant for the latter, except in unadjusted analyses. Among women, the RIIs increased for adenocarcinoma, large cell, small cell, and squamous cell carcinoma, in that order. After adjustment for smoking, the RIIs remained statistically significant among women only for squamous cell carcinoma and substantial but not statistically significant for small cell and large cell carcinoma. The association with education was no longer observed for adenocarcinoma. The RII was statistically significantly smaller for adenocarcinoma when compared with other histological types in model 1 only (test for heterogeneity  $P=.02$ ), the RIIs being borderline statistically significant in the other models (test for heterogeneity  $P=.06$ ). In all analyses, additional adjustment for fruit and vegetable intake did not influence the estimates. Analyses could be conducted by geographic region only for adenocarcinoma and squamous cell carcinoma (among men) because of the number of patients with the remaining histological types was too small (results not shown). Results suggested that the differences between histological types and education were stronger between regions than within regions.

Additional analyses were carried out to test the sensitivity of our results to various models with smoking that incorporated smoking. First, we tested the interaction between smoking status and education (in a crude model). We used different coding for smoking in this analysis: smoking status alone or with duration of smoking or current quantity smoked. There was no evidence of an interaction. We nevertheless conducted analyses stratified by smoking status (Figure 1 ). We observed inequalities among current and ex-smokers that were slightly smaller than among the whole population, and these were only slightly reduced after further adjusting for smoking characteristics. No statistically significant effect was found among never smokers, although the RII was substantially greater than 1 among men who never smoked. This analysis could unfortunately not be stratified by geographic region because of the small number of lung cancers. Partial results however suggested a weaker association between education and lung cancer in Southern Europe than in the other geographic regions, independent of smoking status.

We also estimated the reduction in lung cancer incidence rates and absolute inequality in lung cancer risk if smoking could be eliminated (Table 6 ). Under this assumption, the incidence rates would decrease dramatically, especially among subjects with low education but also among those with high education. The absolute inequality could be reduced substantially: from 65 to 29 per 100,000 person-years among men, and from 21 to 10 per 100,000 person-years among women.

## DISCUSSION

This study reveals that adjustment for smoking decreased relative educational differences in lung cancer incidence by 50–70%, most notably in countries where higher lung cancer incidence was observed among people with lower education (men and women in Northern European countries and men in Germany). These effects were substantially smaller in Southern European countries. This reduction was observed among men for all histological types. Among women, inequalities decreased by 70% for small cell carcinoma, 40–45% for squamous and large cell carcinoma, and no inequalities remained for adenocarcinoma. Further adjustment for relevant dietary factors (fruits and vegetable consumption) did not change the estimates.

### Validity of education as an indicator of the socioeconomic position

Education is an individual measure of socioeconomic position and allows classification of all individuals, including those who do not work. Nevertheless, the socioeconomic position of the nonsalaried participant may also be determined by the educational level of the salaried partner, an effect that may be sex dependent. Higher education may be associated with health through different pathways—subjects with higher education may be more receptive to prevention messages and may have a better ability to change their health behavior and to better use the health care system (15).

Although a common classification of education level has been used in all centers, we cannot rule out possible inconsistencies between centers. Moreover, the computation of the RII assumes a hierarchical order between all educational categories, and the hierarchy between the categories “vocational secondary education” and “other secondary education” is not always straightforward. However, we think that these limits would probably not affect the general patterns described at the international level. Our results are consistent with the available literature on this topic (3, 5).

### **Did we underestimate the role of smoking?**

In this study, smoking accounted for slightly more than half of the educational differences in lung cancer incidence. This percentage is, however, slightly higher than what is generally found in the literature (up to 40%) (1, 2, 4, 16). This difference may be due to a more precise adjustment for smoking in our study. However, because the model selection and the final model were performed using the same data, the standard errors of the estimated coefficients will be somewhat underestimated. It is unlikely that this will change the main conclusion of this study which is that smoking partly explains socioeconomic inequalities in lung cancer incidence. In studies of smoking, residual confounding can never be completely ruled out. Smoking rates may differ by education, and it is possible that this was not fully accounted for in our models. A crucial question is thus whether we underestimated the role of smoking in socioeconomic inequalities in lung cancer incidence to an important extent.

It is unlikely that we substantially underestimated the role of smoking in inequalities in lung cancer incidence. If residual confounding by smoking explained all remaining inequalities, it would mean that the effect of residual confounding is stronger than the effect that is due to the combined smoking variables we included in the model (which explain only slightly more than 50% of socioeconomic inequalities). Our results also showed that adjusting only for current smoking and dose substantially reduced the differences in lung cancer incidence associated with education and that a detailed measurement of past smoking (duration, age at starting) added comparatively little to the explanation.

However, several elements clearly point to substantial residual confounding due to smoking and consequently a possible underestimation of the weight of smoking in socioeconomic inequalities in lung cancer incidence. The differences observed between geographic regions in the level of educational differences after adjustment for smoking are consistent with the smoking epidemic (3, 6). After adjusting for smoking, inequalities were modest among men in Southern Europe, where the literature consistently suggests small educational differences in smoking among middle-aged men, and they are still present in Northern Europe, where the literature reports large educational inequalities in tobacco consumption (6, 13). The results by histological type are also consistent with residual confounding due to smoking. When smoking was adjusted for, the largest inequalities were still observed for the histological types having the strongest association with tobacco consumption, especially squamous cell carcinoma (7, 8).

Moreover, the analyses by smoking status may suggest that smoking is the main cause of inequalities in lung cancer incidence because the inequalities are larger among current and ex-smokers, and non-statistically significant among never smokers. Nevertheless, the RII was greater than 1 among all men who never smoked, but not statistically significant. The latter finding suggests that factors other than smoking play a role in socioeconomic inequalities in lung cancer incidence. However, this category of never smokers includes occasional smokers and also some light smokers or ex-smokers who stopped a long time ago, which may cause some residual confounding by smoking. This possible confounding may also account for part of the higher lung cancer risk found among men never smokers (17). Furthermore, the group of never smokers may have been exposed to passive smoking both at home and at work. We found smaller but statistically significant inequalities among current and ex-smokers compared with the entire population in the present study, which is consistent with a previous report (18).

If the remaining effect of education is due to residual confounding by smoking, then it is important to determine the reason for this confounding. It is likely that the measurement of smoking was not optimal. We did not take into account all aspects of smoking, such as quantity smoked for ex-smokers, occasional smoking, former quantity smoked for present smokers, type of tobacco smoked, and inhalation, which may differ by education. Misclassification of smoking because of lack of exposure information during follow-up is an additional potential source of bias. Also, we could not adjust for exposure to environmental tobacco smoke at home because this information was available for only a few centers. Moreover, measuring smoking history retrospectively will inevitably lead to nondifferential misclassification, which in turn will lead to residual confounding. Measuring smoking history retrospectively may also introduce reporting errors that are differential according to educational level. Although the literature suggests that self-reported smoking status is accurate and does not differ or only slightly differs by education, these studies focus on self-reports of current smoking status and

do not provide any results on levels of consumption or history (number of cigarettes smoked or duration) (19 , 20 ). Information regarding dose and duration may be a more important issue than smoking status.

### **Other potential explanatory factors**

We also investigated the role of diet in socioeconomic inequalities in lung cancer incidence. Only fruit and vegetable consumption was statistically significantly associated with lung cancer incidence (when controlling for smoking); a statistically significant interaction between fruit and vegetable consumption and smoking has been reported previously (10 ). However, additional adjustment for fruit and vegetable consumption did not explain any further educational differences. The latter finding could have been expected, because we did not find any clear educational gradient in fruit and vegetable consumption, except among subjects in Northern Europe. It is nevertheless possible that we underestimated the role of diet in socioeconomic inequalities in lung cancer incidence. To our knowledge, no study on this topic has been published. One study adjusted for many risk factors, including smoking and diet, but did not specifically estimate the effect of each risk factor on inequalities in lung cancer risk (4 ). Previous studies consistently suggest a weak association between diet and lung cancer, both in studies using a refined adjustment for smoking (21 ) and those conducted among never smokers (22 ). Thus, because the potential effect of diet on socioeconomic inequalities is small, if it exists, it is difficult to observe. We may also have used an imprecise measure for diet, which may have resulted in underestimation of the strength of the association between diet and lung cancer and of the role of diet in socioeconomic inequalities in lung cancer incidence. Instead of analyzing daily consumption, biomarkers such as urinary measures of intakes of nitrogen or sodium may be the relevant indicators to take into account when studying the association between lung cancer and diet (23 , 24 ), but this topic is still under debate. Alcohol consumption was not adjusted for in the analyses. However, no association is reported between alcohol use and lung cancer incidence (25 ).

Apart from diet, exposure to radon at home and occupational exposures may also contribute to the residual inequalities. Some rough estimates suggest that approximately 50% of socioeconomic inequalities in lung cancer mortality could be attributable to occupational exposures (26 ), but there are few studies on this topic (27 ). In addition, other factors may play a role in inequalities in lung cancer incidence, such as environmental exposure to pollution (28 ), physical activity (29 ), and ethnicity (30 ).

### **Conclusion**

In summary, we investigated socioeconomic inequalities in lung cancer incidence and observed that smoking explained only slightly more than half of the excess of risk found among subjects with lower education. Our results do suggest residual confounding by smoking. However, because a substantial part of inequalities remained unexplained after adjustment for smoking, residual confounding may not be the only explanation, and this is supported by the finding that a socioeconomic gradient albeit not a statistically significant one in lung cancer incidence also existed in the never smoking population. In future studies, other risk factors should be considered, perhaps in relation with smoking. However, we also observed that removing smoking would reduce the population health burden that is associated with social inequality in lung cancer considerably, in terms of number of cancers avoided. Therefore, public health policies aiming at reducing smoking rates, especially among persons with low education, are still strongly needed.

### **Acknowledgements:**

**Funding:** Fondation pour la Recherche Médicale (SPE 20051105244 to GM). The European Commission, through the Eurocadet project (from the commission of the European communities research directorate-general, grant No EUROCADET:SP23-CT-2005-006528 to Dept Public Health, Erasmus MC, Rotterdam, NL). EPIC was supported by the European Commission: Public Health and Consumer Protection Directorate 1993–2004 and the Research Directorate-General 2005–2008. The EPIC study was funded by “Europe Against Cancer” Programme of the European Commission (SANCO); Ligue contre le Cancer (France); Société 3M (France); Mutuelle Générale de l’Education Nationale; Institut National de la Santé et de la Recherche Médicale; German Cancer Aid; German Cancer Research Center; German Federal Ministry of Education and Research; Danish Cancer Society; Red Temática de Investigación Cooperativa de Centros de Cáncer (C03/10); the participating regional governments and institutions of Murcia, Navarra, Asturias, País Vasco y Andalucía, Spain; Cancer Research UK; Medical Research Council, United Kingdom; Stroke Association, United Kingdom; British Heart Foundation; Department of Health, United Kingdom; Food Standards Agency, United Kingdom; The Wellcome Trust, United Kingdom; Greek Ministry of Education; Greek Ministry of Health and Social Solidarity; Hellenic Health Foundation; Italian Association for Research on Cancer; Dutch Ministry of Public Health, Welfare and Sports; Dutch Ministry of Health; Dutch Prevention Funds; LK Research Funds; Dutch Zorg Onderzoek Nederland; World Cancer Research Fund; Swedish Cancer Society; Swedish Scientific Council; Regional Government of Vasterbotten and Skane, Sweden; Norwegian Cancer Society; and Foundation to Promote Research into Functional Vitamin B12 Deficiency, Norway. Some authors are partners of Environmental Cancer Risk, Nutrition and Individual Susceptibility, a network of excellence of the European Commission (6FP contract 513943). Antonio Agudo and Paolo Vineis were supported by ECNIS.

### **Footnotes:**

The sponsors had no role in the study design, the data collection and analysis, the interpretation of the results, the preparation of the manuscript, or the decision to submit the manuscript for publication.

## References:

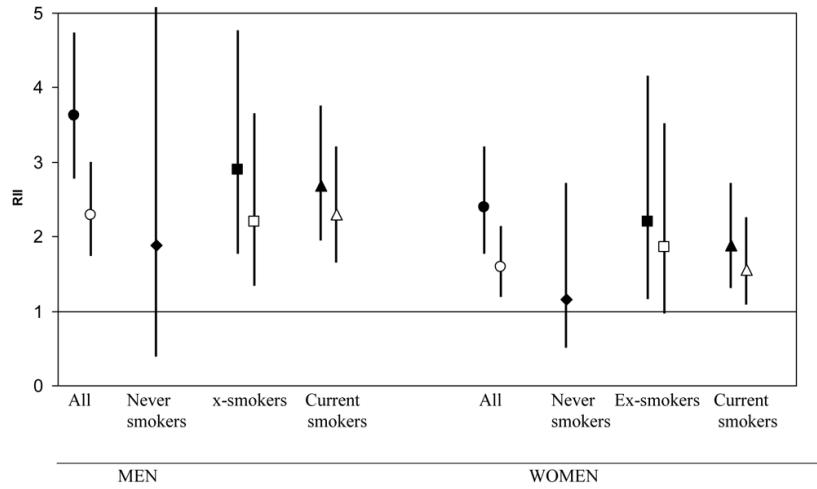
1. Louwman WJ, van Lenthe FJ, Coebergh JW, Mackenbach JP. Behaviour partly explains educational differences in cancer incidence in the south-eastern Netherlands: the longitudinal GLOBE study. *Eur J Cancer Prev*. 2004; Apr 13; (2) 119 - 25
2. Hart CL, Hole DJ, Gillis CR, Davey Smith G, Watt GC, Hawthorne VM. Social class differences in lung cancer mortality: risk factor explanations using two Scottish cohort studies. *Int J Epidemiol*. 2001; 30: (2) 268 - 74
3. Mackenbach JP, Huisman M, Andersen O. Inequalities in lung cancer mortality by the educational level in 10 European populations. *Eur J Cancer*. 2004; 40: (1) 126 - 35
4. Mao Y, Hu J, Ugnat AM, Semenciw R, Fincham S. Socioeconomic status and lung cancer risk in Canada. *Int J Epidemiol*. 2001; 30: (4) 809 - 17
5. Kunst AE, Groenhouf F, Mackenbach JP, Health EW. Occupational class and cause specific mortality in middle aged men in 11 European countries: comparison of population based studies. EU Working Group on Socioeconomic Inequalities in Health. *BMJ*. 1998; 316: (7145) 1636 - 42
6. Huisman M, Kunst AE, Mackenbach JP. Educational inequalities in smoking among men and women aged 16 years and older in 11 European countries. *Tob Control*. 2005; Apr 14; (2) 106 - 13
7. Barbone F, Bovenzi M, Cavallieri F, Stanta G. Cigarette smoking and histologic type of lung cancer in men. *Chest*. 1997; Dec 112: (6) 1474 - 9
8. Khuder SA, Dayal HH, Mutgi AB, Willey JC, Dayal G. Effect of cigarette smoking on major histological types of lung cancer in men. *Lung Cancer*. 1998; Oct 22: (1) 15 - 21
9. Bennett VA, Davies EA, Jack RH, Mak V, Moller H. Histological subtype of lung cancer in relation to socio-economic deprivation in South East England. *BMC Cancer*. 2008; 8: 139 -
10. Linseisen J, Rohrmann S, Miller AB. Fruit and vegetable consumption and lung cancer risk: Updated information from the European Prospective Investigation into Cancer and Nutrition (EPIC). *Int J Cancer*. 2007; Sep 1 121: (5) 1103 - 14
11. Mackenbach JP, Kunst AE. Measuring the magnitude of socio-economic inequalities in health: an overview of available measures illustrated with two examples from Europe. *Soc Sci Med*. 1997; 44: (6) 757 - 71
12. Riboli E, Hunt KJ, Slimani N. European Prospective Investigation into Cancer and Nutrition (EPIC): study populations and data collection. *Public Health Nutr*. 2002; 5: (6B) 1113 - 24
13. Cavelaars AE, Kunst AE, Geurts JJ. Educational differences in smoking: international comparison. *BMJ*. 2000; 320: (7242) 1102 - 7
14. Giskes K, Kunst AE, Benach J. Trends in smoking behaviour between 1985 and 2000 in nine European countries by education. *J Epidemiol Community Health*. 2005; 59: (5) 395 - 401
15. Galobardes B, Shaw M, Lawlor DA, Lynch JW, Davey Smith G. Indicators of socioeconomic position (part 1). *J Epidemiol Community Health*. 2006; Jan 60: (1) 7 - 12
16. van Loon AJ, Goldbohm RA, van den Brandt PA. Lung cancer: is there an association with socioeconomic status in The Netherlands?. *J Epidemiol Community Health*. 1995; 49: (1) 65 - 9
17. Bjerregaard BK, Raaschou-Nielsen O, Sorensen M. The effect of occasional smoking on smoking-related cancers: in the European Prospective Investigation into Cancer and Nutrition (EPIC). *Cancer Causes Control*. 2006; Dec 17: (10) 1305 - 9
18. Martikainen P, Lahelma E, Ripatti S, Albanes D, Virtamo J. Educational differences in lung cancer mortality in male smokers. *Int J Epidemiol*. 2001; 30: (2) 264 - 7
19. Suadicani P, Hein HO, Gyntelberg F. Serum validated tobacco use and social inequalities in risk of ischaemic heart disease. *Int J Epidemiol*. 1994; 23: (2) 293 - 300
20. Wagenknecht LE, Burke GL, Perkins LL, Haley NJ, Friedman GD. Misclassification of smoking status in the CARDIA study: a comparison of self-report with serum cotinine levels. *Am J Public Health*. 1992; 82: (1) 33 - 6
21. Skuladottir H, Tjoenneland A, Overvad K. Does insufficient adjustment for smoking explain the preventive effects of fruit and vegetables on lung cancer?. *Lung Cancer*. 2004; Jul 45: (1) 1 - 10
22. Brennan P, Fortes C, Butler J. A multicenter case-control study of diet and lung cancer among non-smokers. *Cancer Causes Control*. 2000; Jan 11: (1) 49 - 58
23. Day N, McKeown N, Wong M, Welch A, Bingham S. Epidemiological assessment of diet: a comparison of a 7-day diary with a food frequency questionnaire using urinary markers of nitrogen, potassium and sodium. *Int J Epidemiol*. 2001; Apr 30: (2) 309 - 17
24. Kipnis V, Subar AF, Midthune D. Structure of dietary measurement error: results of the OPEN biomarker study. *Am J Epidemiol*. 2003; Jul 1 158: (1) 14 - 21 discussion 2-6
25. Boffetta P, Hashibe M. Alcohol and cancer. *Lancet Oncol*. 2006; Feb 7: (2) 149 - 56
26. Boffetta P, Kogevinas M, Westerholm P, Saracci R. Exposure to occupational carcinogens and social class differences in cancer occurrence. *IARC Sci Publ*. 1997; 138: 331 - 41
27. van Loon AJ, Goldbohm RA, Kant IJ, Swaen GM, Kremer AM, van den Brandt PA. Socioeconomic status and lung cancer incidence in men in The Netherlands: is there a role for occupational exposure?. *J Epidemiol Community Health*. 1997; 51: (1) 24 - 9
28. Vineis P, Hoek G, Krzyzanowski M. Lung cancers attributable to environmental tobacco smoke and air pollution in non-smokers in different European countries: a prospective study. *Environ Health*. 2007; 6: 7 -
29. Steindorf K, Friedenreich C, Linseisen J. Physical activity and lung cancer risk in the European Prospective Investigation into Cancer and Nutrition Cohort. *Int J Cancer*. 2006; Nov 15 119: (10) 2389 - 97
30. Krieger N, Quesenberry C Jr, Peng T. Social class, race/ethnicity, and incidence of breast, cervix, colon, lung, and prostate cancer among Asian, Black, Hispanic, and White residents of the San Francisco Bay Area, 1988-92 (United States). *Cancer Causes Control*. 1999; Dec 10: (6) 525 - 37



**Figure 1**

Relative indices of inequality (RII) for education and their corresponding 95% confidence intervals (95% CIs) by smoking status and sex in the European Prospective Investigation into Cancer and Nutrition cohort (N=391,251)

(●◆▲, no adjustment; ○□△, adjusted for smoking) The confidence interval of the estimate among men never smokers is (0.39 to 8.99). All analyses are stratified for age at baseline and center. Among current smokers, the model includes age at starting (continuous), duration of smoking (continuous in years), current quantity (continuous), a quadratic term for current quantity, two interaction terms (quantity × duration and age at starting × duration). Among ex-smokers, the model includes age at starting (continuous), duration of smoking (continuous in years), and one interaction term (age at starting × duration).



**Table 1**

Education by sex and geographic region and lung cancer incidence rate in the European Prospective Investigation into Cancer and Nutrition cohort (N=391,251)

Group	Person-years (%)					Incidence rate <sup>†</sup>
	All subjects	Education <sup>*</sup>				
		1 (lowest)	2	3	4 (highest)	
Men						
All	1173428	408620 (35)	289291 (24)	162315 (14)	313202 (27)	90
North <sup>‡</sup>	662008	181637 (28)	193539 (29)	101033 (15)	185799 (28)	85
Germany <sup>  </sup>	173643	41828 (24)	47790 (28)	9125 (5)	74900 (43)	95
South <sup>§</sup>	337777	185155 (55)	47963 (14)	52156 (15)	52503 (16)	93
Women						
All	2092298	733854 (35)	571903 (27)	360592 (17)	425949 (21)	42
North <sup>‡</sup>	1266153	286681 (23)	432827 (34)	254464 (20)	292180 (23)	56
Germany <sup>  </sup>	226640	52903 (23)	94177 (42)	18083 (8)	61478 (27)	22
South <sup>§</sup>	599505	394270 (66)	44898 (7)	88045 (15)	72291 (12)	17

\* The coding for education is as follows: 1=primary education or less, 2=vocational secondary education, 3=other secondary education, 4=college or university.

† Age-adjusted, including participants 50–69 years at baseline, per 100,000 person-years.

‡ Norway, Sweden, Denmark, UK, and The Netherlands.

§ Greece, Italy, and Spain.

|| The number of subjects by country is: Norway (33254 women), Sweden (21947 men and 26083 women), Denmark (26100 men and 28568 women), UK (19167 men and 42594 women), the Netherlands (9718 men and 26328 women), Germany (21521 men and 27843 women), Greece (10014 men and 14304 women), Italy (13644 men and 30449 women), Spain (15064 men and 24653 women)

**Table 2**

Baseline characteristics related to smoking and fruits and vegetables consumption by education, sex, and geographic region in the European Prospective Investigation into Cancer and Nutrition cohort (N=391,251)

Characteristic	Men Education level*				Women Education level*			
	1	2	3	4	1	2	3	4
Percentage of never smokers								
All	27	31	38	41	61	47	48	55
North†	27	33	46	45	42	46	49	58
Germany	27	28	29	39	62	54	52	56
South‡	28	25	26	32	74	49	46	45
Percentage of current smokers								
All	35	31	28	23	23	26	25	18
North†	34	30	24	22	32	26	23	15
Germany	29	28	30	20	19	20	19	15
South‡	38	40	35	34	17	30	32	32
No. of cigarettes smoked per day (among current smokers)								
All	15	15	15	13	13	13	13	12
North†	13	13	12	11	13	13	12	11
Germany	18	17	17	14	14	13	12	11
South‡	17	18	18	17	13	14	14	13
Duration of smoking, y (among ever smokers)								
All	29	25	23	22	25	22	20	18
North†	29	25	22	23	27	23	20	17
Germany	25	23	23	21	23	18	17	16
South‡	29	23	23	23	21	21	21	20
Fruit and vegetable consumption, median, g/d§								
All	446	360	396	401	482	385	436	472
North†	286	320	302	374	348	390	382	460
Germany	228	233	241	247	269	279	289	289
South‡	660	663	597	735	614	564	634	681

\* The coding for education is as follows: 1=primary education or less, 2=vocational secondary education, 3=other secondary education, 4=college or university.

† Norway, Sweden, Denmark, UK, and The Netherlands.

‡ Greece, Italy, and Spain.

§ Observed consumption.

**Table 3**

Hazard ratios (HRs) and relative indices of inequality (RIIs) for education and their corresponding 95% confidence intervals (95% CIs) for lung cancer by sex in the European Prospective Investigation into Cancer and Nutrition cohort (N=391,251)

Education	Model 1 Crude <sup>†</sup>		Model 2 Adjusted for current smoking <sup>‡</sup>		Model 3 Adjusted for smoking <sup>§</sup>		Model 4 Adjusted for smoking and diet <sup>  </sup>	
	N <sup>*</sup>	HR (95% CI)	HR	(95% CI)	HR	(95% CI)	HR	(95% CI)
<b>Men</b>								
Primary education or less	543	2.54 (2.06 to 3.14)	1.93	(1.56 to 2.38)	1.79	(1.45 to 2.21)	1.78	(1.44 to 2.20)
Vocational secondary education	213	1.77 (1.40 to 2.23)	1.46	(1.15 to 1.83)	1.39	(1.10 to 1.75)	1.38	(1.10 to 1.74)
Other secondary education	66	1.43 (1.05 to 1.94)	1.23	(0.90 to 1.68)	1.20	(0.88 to 1.63)	1.19	(0.87 to 1.62)
College or university	117	1.00 (Referent)	1.00	(Referent)	1.00	(Referent)	1.00	(Referent)
RII	939	3.62 (2.77 to 4.73)	2.54	(1.94 to 3.33)	2.29	(1.75 to 3.01)	2.27	(1.73 to 2.99)
<b>Women</b>								
Primary education or less	326	1.98 (1.50 to 2.61)	1.56	(1.18 to 2.06)	1.44	(1.09 to 1.90)	1.42	(1.07 to 1.88)
Vocational secondary education	207	1.35 (1.02 to 1.79)	1.21	(0.91 to 1.61)	1.17	(0.88 to 1.56)	1.16	(0.87 to 1.54)
Other secondary education	93	1.32 (0.96 to 1.83)	1.21	(0.88 to 1.68)	1.18	(0.85 to 1.64)	1.18	(0.85 to 1.63)
College or university	66	1.00 (Referent)	1.00	(Referent)	1.00	(Referent)	1.00	(Referent)
RII	692	2.39 (1.77 to 3.21)	1.75	(1.31 to 2.36)	1.59	(1.18 to 2.13)	1.55	(1.15 to 2.09)

\* Number of lung cancer patients.

† All analyses are stratified by center and age at baseline.

‡ The model includes smoking status (never smoker [reference category], current smoker, former smoker), current quantity (continuous), a quadratic term for current quantity, and a dummy variable for missing variables for current quantity.

§ The model includes smoking status (never smoker [reference category], current smoker, former smoker), age at starting (continuous), duration of smoking (continuous in years), current quantity (continuous), a quadratic term for current quantity, two interaction terms (quantity × duration and age at starting × duration), dummy variables for missing values for age at starting, duration of smoking, and current quantity.

|| The model includes the variables in model 3, fruits and vegetables consumption (continuous variable, per 100g), and an interaction term between smoking status and fruits and vegetables consumption.

**Table 4**

Relative indices of inequality (RIIs) for education and their corresponding 95% confidence intervals (95% CIs) by geographic region and sex in the European Prospective Investigation into Cancer and Nutrition cohort (N=391,251)

Geographic region	Model 1 Crude <sup>†</sup>		Model 2 Adjusted for current smoking <sup>‡</sup>		Model 3 Adjusted for smoking <sup>§</sup>		Model 4 Adjusted for smoking and diet <sup>  </sup>	
	N <sup>*</sup>	RII (95% CI)	RII	(95% CI)	RII	(95% CI)	RII	(95% CI)
Men								
North <sup>¶</sup>	530	5.42 (3.86 to 7.62)	3.49	(2.47 to 4.92)	2.87	(2.01 to 4.10)	2.84	(1.99 to 4.06)
Germany	145	4.10 (2.19 to 7.66)	2.33	(1.23 to 4.40)	2.17	(1.14 to 4.13)	2.17	(1.14 to 4.14)
South <sup>#</sup>	264	1.78 (1.02 to 3.11)	1.51	(0.87 to 2.64)	1.36	(0.78 to 2.38)	1.38	(0.78 to 2.38)
P <sub>interaction</sub> <sup>**</sup>		.009		.08		.11		.13
Women								
North <sup>¶</sup>	557	3.93 (2.86 to 5.40)	2.29	(1.67 to 3.15)	1.88	(1.35 to 2.62)	1.84	(1.32 to 2.57)
Germany	41	1.35 (0.43 to 4.29)	1.09	(0.34 to 3.44)	1.09	(0.34 to 3.47)	1.01	(0.32 to 3.21)
South <sup>#</sup>	94	0.30 (0.13 to 0.71)	0.52	(0.22 to 1.20)	0.54	(0.23 to 1.24)	0.53	(0.23 to 1.23)
P <sub>interaction</sub> <sup>**</sup>		<.001		.007		.02		.03

\* Number of lung cancer patients.

† All analyses are stratified by center and age at baseline.

‡ The model includes smoking status (never smoker [reference category], current smoker, former smoker), current quantity (continuous), a quadratic term for current quantity, and a dummy variable for missing variables for current quantity.

§ The model includes smoking status (never smoker [reference category], current smoker, former smoker), age at starting (continuous), duration of smoking (continuous in years), current quantity (continuous), a quadratic term for current quantity, two interaction terms (quantity × duration and age at starting × duration), dummy variables for missing values for age at starting, duration of smoking, and current quantity.

|| The model includes the variables in model 3, fruits and vegetables consumption (continuous variable, per 100g), and an interaction term between smoking status and fruits and vegetables consumption.

¶ Norway, Sweden, Denmark, UK, and The Netherlands.

# Greece, Italy, and Spain.

\*\* Test for interaction between geographic regions.

**Table 5**

Relative indices of inequality (RII) for education and their corresponding 95% confidence intervals (95% CI) by region, histological type and sex in the European Prospective Investigation into Cancer and Nutrition cohort (N=391,251)

Histologic type	Model 1 Crude <sup>†</sup>		Model 2 Adjusted for current smoking <sup>‡</sup>		Model 3 Adjusted for smoking <sup>§</sup>		Model 4 Adjusted for smoking and diet <sup>  </sup>	
	N <sup>*</sup>	RII (95% CI)	RII	(95% CI)	RII	(95% CI)	RII	(95% CI)
Men								
Adenocarcinoma	262	2.82 (1.73 to 4.62)	2.14	(1.31 to 3.51)	1.98	(1.21 to 3.26)	1.95	(1.19 to 3.22)
Small cell carcinoma	161	5.71 (2.93 to 11.14)	3.41	(1.73 to 6.70)	3.28	(1.66 to 6.49)	3.28	(1.66 to 6.48)
Squamous carcinoma	255	5.02 (2.96 to 8.51)	3.49	(2.05 to 5.95)	2.97	(1.73 to 5.08)	2.94	(1.72 to 5.04)
Large cell carcinoma	74	2.69 (1.04 to 6.94)	1.94	(0.75 to 5.01)	1.68	(0.65 to 4.34)	1.66	(0.64 to 4.32)
Women								
Adenocarcinoma	288	1.59 (1.02 to 2.49)	1.25	(0.80 to 1.96)	1.12	(0.72 to 1.76)	1.10	(0.70 to 1.72)
Small cell carcinoma	115	3.36 (1.66 to 6.96)	2.01	(0.97 to 4.14)	1.68	(0.81 to 3.49)	1.62	(0.78 to 3.38)
Squamous carcinoma	96	4.05 (1.77 to 9.27)	2.97	(1.30 to 6.82)	2.86	(1.24 to 6.61)	2.70	(1.17 to 6.24)
Large cell carcinoma	63	2.82 (1.02 to 7.74)	2.06	(0.75 to 5.60)	1.98	(0.72 to 5.45)	2.10	(0.75 to 5.71)

\* Number of lung cancer patients.

† All analyses are stratified by center and age at baseline.

‡ The model includes smoking status (never smoker [reference category], current smoker, former smoker), current quantity (continuous), a quadratic term for current quantity, and a dummy variable for missing variables for current quantity.

§ The model includes smoking status (never smoker [reference category], current smoker, former smoker), age at starting (continuous), duration of smoking (continuous in years), current quantity (continuous), a quadratic term for current quantity, two interaction terms (quantity × duration and age at starting × duration), and dummy variables for missing values for age at starting, duration of smoking, and current quantity.

|| The model includes the variables in model 3, fruits and vegetables consumption (continuous variable, per 100g), and an interaction term between smoking status and fruits and vegetables consumption.

**Table 6**

Age adjusted incidence rates for lung cancer by education and sex in the European Prospective Investigation into Cancer and Nutrition (EPIC) cohort (N=391,251) and in a fictive cohort in which smoking has been eliminated

Education	Incidence per 100,000 person-years	
	EPIC cohort	Fictive cohort
Men		
Primary education or less	117	49
Vocational secondary education	85	46
Other secondary education	73	38
College or university	52	20
Rate difference between the lowest and the highest education	65	29
Women		
Primary education or less	46	24
Vocational secondary education	46	28
Other secondary education	40	21
College or university	25	14
Rate difference between the lowest and the highest education level	21	10