Effect Modification of the Association of Cumulative Exposure and Cancer Risk by Intensity of Exposure and Time Since Exposure Cessation: A Flexible Method Applied to Cigarette Smoking and Lung Cancer in the SYNERGY Study


Version: Post-Print/Accepted Manuscript


Publisher’s Statement: This is a pre-copyedited, author-produced PDF of an article accepted for publication in the American Journal of Epidemiology following peer review. The version of record Am J Epidemiol (2014) 179 (3): 290-298 is available online at: https://dx.doi.org/10.1093/aje/kwt273.

How to cite TSpace items

Always cite the published version, so the author(s) will receive recognition through services that track citation counts, e.g. Scopus. If you need to cite the page number of the TSpace version (original manuscript or accepted manuscript) because you cannot access the published version, then cite the TSpace version in addition to the published version using the permanent URI (handle) found on the record page.
Effect Modification of the Association of Cumulative Exposure and Cancer Risk by Intensity of Exposure and Time Since Exposure Cessation:

A Flexible Method Applied to Cigarette Smoking and Lung Cancer in the SYNERGY Study


Correspondence:
Jelle Vlaanderen, PhD
Universiteit Utrecht
Institute for Risk Assessment Sciences
P.O.Box 80178
3508 TD Utrecht
The Netherlands
email: j.j.vlaanderen@uu.nl
Tel: +31 (0)30 253 59 47
Fax: +31 (0)30 253 94 99

Word count abstract: 200
Word count main text: 3177

Abbreviations used in the manuscript:

efficiency odds ratio  (EOR)
life time average intensity of cigarette smoking (ICS)
pack-years of cigarette smoke (PCS)
time since smoking cessation (TSC)

running head: “Effect modification of cumulative exposure”
Abstract

The indiscriminate use of the cumulative exposure metric (product of intensity and duration of exposure) might bias reported associations between exposure to hazardous agents and cancer risk. To assess the independent effects of duration and intensity of exposure on cancer risk we explored effect modification of the association of cumulative exposure and cancer risk by intensity of exposure. We applied a flexible excess odds ratio model that is linear in cumulative exposure, but potentially non-linear in intensity of exposure to 15 case-control studies of cigarette smoking and lung cancer (1985-2009; SYNERGY). Our model accommodates modification of the excess odds ratio per pack-year of cigarette smoke by time since smoking cessation among former smokers. We observed negative effect modification of the association of pack-years of cigarette smoke and lung cancer by intensity of cigarette smoke for individuals that smoke more than 20-30 cigarettes per day. Patterns of effect modification were similar across individual studies and across major lung cancer subtypes. We observed strong negative effect modification by time since smoking cessation. Application of our method in this example of cigarette smoking and lung cancer demonstrated that reducing a complex exposure history to a metric such as cumulative exposure is too restrictive.

Keywords: cigarette smoke, cumulative exposure, effect modification, lung cancer, pooled analysis
Cumulative exposure (product of (average daily) exposure intensity and duration of exposure) is often the default exposure metric used in epidemiological cancer studies. However, the assumptions on which the use of cumulative exposure is based namely that “the cumulative probability of developing a disease is proportional to the sum of the daily probabilities of developing a disease, the daily probability of developing a disease increases monotonically with the concentration in the target tissue, the concentration in the target tissue is linearly related to the external exposure” are not always justified (1).

Pack-years of cigarette smoke (PCS) is calculated as the average packs of cigarettes smoked per day multiplied by the cumulative number of years of smoking. This example of the cumulative exposure metric is often used to evaluate smoking behavior in epidemiological analyses. There is considerable evidence that with a straightforward inclusion of PCS in epidemiological analyses of chronic health effects not all intensity and duration related aspects of smoking behavior are optimally characterized (2). For example, Doll and Peto (3) demonstrated that the absolute excess rate of lung cancer among smokers was related to at least the fourth power of smoking duration and only to the second power of smoking intensity. Because both smoking duration and baseline lung cancer rates vary with age, the relationship between the excess relative risk and duration and intensity of smoking is likely to be more complex.

One way to assess independent effects of duration and intensity of cigarette smoking on lung cancer risk is to explore effect modification of the association of PCS and lung cancer by intensity of cigarette smoking (4). Excess relative risk or excess odds ratio (EOR) models that are linear in total exposure and exponential in the intensity of exposure have been applied successfully to explore effect modification of cumulative exposure by intensity of exposure for a number of exposures (4–10). Such models have a general form $\text{OR}(d) = 1 + \beta_1 d \times \exp(\beta_2 n)$, where $\beta_1$ represents the EOR per unit of total exposure ($d$) (i.e. the EOR of disease changes in an additive fashion with total exposure), and $\beta_2$ represents the (multiplicative) modifying effect of intensity of
exposure (n). Formerly fitting these models needed use of specialized software, but they can now be fitted in standard software packages with relative ease (11).

We explored the modification of the effect of PCS on the EOR for lung cancer by lifetime average intensity of cigarette smoking (ICS). For our analysis we modified a previously developed approach to model total exposure and exposure intensity (4). Unique about our analysis is that we were able to apply this model in a large dataset of 15 independently designed case-control studies with detailed smoking information (SYNERGY pooling project) (12–14). Furthermore, our approach differs from previous applications by including a direct assessment of the modification of the EOR per PCS by the time since smoking cessation (TSC), a strong predictor of lung cancer risk in former smokers (15), in the regression model, and we included a three-knot restricted cubic spline function for both ICS and TSC to allow a more flexible assessment of the shape of the modification of the EOR per PCS.
MATERIALS AND METHODS

Synergy dataset

We used data from the SYNERGY project (12–14). Our dataset includes data from 15 case-control studies from Canada (2), France (3), Germany (2), Italy (3), New Zealand (1), Spain (1), Sweden (1), The Netherlands (1), and a multi-center study conducted in Central and Eastern Europe and the UK (16). Controls were individually or frequency matched to cases by gender and age and recruited from the general population (82%) or hospitals (18%). Smoking information was predominantly collected by interviews with the subjects themselves (92% cases; 94% controls). Lung cancer subtypes were classified according to World Health Organization guidelines by the pathologists associated with the participating hospitals. The ethics committees of the individual studies approved the conduct of the study, as did the Institutional Review Board at the International Agency for Research on Cancer (IARC). We provide an overview of the characteristics of the studies included in the analysis in Web Table 1.

Smoking data

Information on cigarette smoking history included the number of cigarettes smoked per day in calendar-year periods and the age at smoking cessation for former smokers. We calculated continuous variables duration of smoking, ICS, and PCS based on the smoking history. A current smoker was defined as someone who had smoked more than one year and still smoked in the year of interview or in the year before. Former smokers were defined as having smoked for at least one year, but quit smoking at least two years before the date of the interview. Subjects who smoked less than one year were considered occasional smokers and were treated as never smokers in the analyses. All cases and controls with complete smoking data were included, without restriction on age or smoking status.
Statistical analysis

The model we use in this manuscript provides a balance between parsimony and model fit. The model falls within a more general framework for flexible modeling of exposure-time relations (17). Similar inferences were obtained using other model specifications within the more general framework (See Web Appendix).

Below we provide a description of the models that we applied in our study with an emphasis on the modifications that we made in comparison to the previously published models by Lubin et al. (4). Our models are linear in PCS and exponential in ICS and TSC (to force the modifying effect to be non-negative). We used two approaches to model ICS (expressed in cigarettes per day).

The first approach defines \( I \) intensity categories and indicator variables, \( n_i, i=1, \ldots, I \), where \( n_i = 1 \) if a subject's intensity level occurs within the \( i \)th category; and \( n_i = 0 \) otherwise.

The model is

\[
\text{OR}(d) = 1 + \beta_d \times \exp \{ \sum \theta_i n_i \} \quad (A)
\]

where \( \beta \) represents the EOR for each PCS, \( d \). The model specifies a different slope for each intensity category. With \( \theta_1 \) set to zero for identifiability, \( \theta_2, \ldots, \theta_i \) represent category-specific effects relative to the \( i=1 \) level. Model A has been published before (4,7,8) and was fitted to a subset of the data that is restricted to current and never smokers aged 50-75 to parallel the datasets used in previous publications.

We extended model A with a function for TSC to allow the inclusion of former smokers in our analysis.

\[
\text{OR}(d) = 1 + \beta_d \times \exp[g_1(t)] \times \exp \{ \sum \theta_i n_i \} \quad (B)
\]
where \( g_1(t) \) is a three-knot restricted cubic spline function for TSC (knots located at the 20\(^{th}\), 50\(^{th}\), and 80\(^{th}\) percentile of the distribution of TSC of all former smokers).

The variation in EOR per PCS by continuous ICS \((n)\) is assessed with three different models

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

where \( h(n) \) has the functional form \( h(n) = \exp(\Phi_1 \ln(n) + \Phi_2 \ln(n)^2) \). \( \Phi_1 \) and \( \Phi_2 \) are the parameters for the modifying function of continuous ICS. Model C has been published before (4,7,8).

We modified model C by including a flexible spline function for ICS.

\[
\text{OR}(d) = 1 + \beta d \times \exp(g_1(n))
\]  
(D)

where \( g_1 \) is a three-knot restricted cubic spline function of continuous ICS \((n)\) (knots located at the 20\(^{th}\), 50\(^{th}\), and 80\(^{th}\) percentile of the distribution of ICS of all smokers). Models C and D were fitted to a subset of the data that was restricted to current and never smokers aged 50-75.

Similar to model B, we extended model D with a function for TSC to allow the inclusion of former smokers in our analysis.

\[
\text{OR}(d) = 1 + \beta d \times \exp(g_1(n) + g_2(t))
\]  
(E)

where \( g_1 \) and \( g_2 \) are three-knot restricted cubic spline functions with knots located at the 20\(^{th}\), 50\(^{th}\), and 80\(^{th}\) percentile of the distribution of TSC of all former smokers. \( g_1 \) is a function of continuous ICS \((n)\) and \( g_2 \) is a function of continuous TSC \((t)\).
All models were fitted using the NLMIXED procedure in SAS v 9.2 (SAS Institute Inc., Cary, NC). Effects were adjusted for study center, age group (<45, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, ≥80 years), and sex by allowing for stratum-specific baseline odds. Analyses were also fully stratified by sex and study location and conducted for all lung cancer subtypes combined and for major lung cancer subtypes squamous cell carcinoma, small cell carcinoma, adenocarcinoma separately. We assessed the sensitivity of the restricted cubic spline functions for continuous ICS and TSC to alternative knot locations (10th, 50th, 90th percentile, and 5th, 50th, 95th percentile) in an analysis on all lung cancer subtypes combined among men and women. We observed marginal impact on both model fit (AIC) and model prediction. Therefore all analyses were conducted with the a priori specified knot locations (20th, 50th, 80th percentile). Bootstrapped 95% confidence intervals of the functions for ICS and TSC were estimated via 1000 bootstrap replications of the original data and taking the 2.5th and 97.5th percentiles of the resulting distribution. To avoid over-interpretation of patterns for regions where the data was extremely sparse, we excluded predictions for intensities less than the 1st percentile and higher than the 99th percentile of the distribution of the exposed individuals from all plots. The same approach was used for TSC. A likelihood ratio test (18) was used to compare differences in fit to the data between nested models. Because, conditional on attained age, age at smoking initiation is (i) multicollinear with duration of smoking and TSC, (ii) typically shows relatively little variation, and (iii) is not very strongly associated with cancer occurrence, we did not assess the effect of age at smoking initiation in our analysis (2).
RESULTS

The pooled dataset consists of 17,975 cases (14,255 men and 3,720 women) and 22,353 controls (17,267 men and 5,086 women). Further details of the study population are provided in Web Table 1.

First we applied models A and C on a dataset restricted to current and never smokers, aged 50-75 years. Figure 1a shows the effect of ICS on the lung cancer EOR per PCS, estimated with models A (point estimates for deciles of ICS) and C (continuous line) among current, and never smokers (i.e. excluding former smokers). EORs per PCS estimated with model A increased with increasing ICS below 20 cigarettes (one pack) per day and slightly decreased with increasing ICS at intensities higher than 20 cigarettes per day. The continuous prediction of model C followed the pattern of the point estimates predicted with model A. Importantly, due to model specification, model C would predict EOR = 0 for zero ICS. Parameter estimates for $\Phi_1$ and $\Phi_2$ were 0.0258 (standard error (s.e.) 0.0062) and -0.0216 (s.e. 0.0052), respectively.

Next, we compared the effect of ICS predicted by model C (Figure 1b, grey line) to the effect of ICS predicted by a model that includes a flexible spline function for ICS (model D, Figure 1b, dashed line). Knots of the restricted cubic spline were located at 10, 19, and 26 cigarettes per day. A comparison based on the Akaike Information Criterion (AIC) (19), suggests that model D (AIC = 22677) had a better fit to the data than model C (AIC = 22687). For intensities higher than 20 cigarettes per day model D predicted a slightly stronger decrease in EOR per PCS with increasing ICS than model C. Furthermore, model D was not restricted to start at EOR/PCS = 0 for zero ICS, which resulted in a less pronounced increase in EOR per PCS with increasing ICS below 20 cigarettes per day.

The continuous black line in Figure 1b is the effect of ICS predicted by a model that was fitted on current, former, and never smokers of all ages and included flexible spline functions for both ICS and TSC (model E). The effect of ICS predicted by model E resembled closely the prediction of
model D. The effect of TSC predicted by model E is presented in Figure 1c. Knots of the
restricted cubic spline were located at 6, 15, and 28 years since smoking cessation. Model E
predicted a strong reduction (83%) in the EOR for lung cancer per PCS with increasing TSC.

In Table 1 we report the fit of model E to the study data. In model \( E_0 \) the functions for ICS and
TSC were set to zero, hence the effect of PCS on the OR for lung cancer was not modified.
Based on a likelihood ratio test, both model \( E_1 \) (where the effect of TSC on the EOR for lung
cancer per PCS was set to zero) and model \( E_2 \) (where the effect of ICS was set to zero) provided
a significantly better fit to the data than model \( E_0 \). Furthermore, model \( E_3 \), in which functions for
both ICS and TSC were estimated provided a significantly better fit to the study data than models
\( E_1 \) and \( E_2 \). Model \( E_3 \) (from here referred to as model E) was therefore selected for further
analyses of the dataset.

Similar patterns with ICS were observed in analyses for major subtypes squamous cell
carcinoma, small cell carcinoma, and adenocarcinoma (Figure 2). Analyses on men and women
separately resulted in patterns of the EOR per PCS comparable to the analyses on men and
women combined (Web Figure 1).

In Figure 3 we show study-specific patterns of modification of EOR per PCS by ICS. These
analyses were conducted on all subtypes combined. Predictions were generated by applying
model E to each individual study.

Wide confidence intervals in some of the panels of Figure 3 demonstrate that some studies had
limited statistical power to explore patterns in the EOR per PCS. Furthermore, model E did not
converge when applied to studies LUCA, PARIS, and MORGEN, three studies of modest sample
size (Web Table 1). In most studies for which we were able to observe a pattern, a downward
trend in EOR per PCS with increasing ICS was observed after reaching a maximum EOR per
PCS for ~20-30 cigarettes per day. Exceptions were studies LLP, LUCAS and INCO-Poland
which showed a flat or increasing in EOR per PCS. Confidence intervals for the predicted
patterns for these studies did not exclude a downward trend. Higher uncertainty within studies
and less consistency across studies was observed for patterns in effect modification by ICS for smoking intensities below 20 cigarettes per day. Among studies for which predictions were relatively precise upward patterns were generally observed. Absolute levels of the EOR per PCS varied considerably across study locations.
DISCUSSION

Application of our approach in the SYNERGY dataset provided insight into the consistency of patterns of modification of the effect of PCS on the EOR for lung cancer by ICS across a large number of independently designed case-control studies from Central and Eastern Europe, Canada, and New Zealand. We observed negative effect modification of the association of PCS and lung cancer by ICS for individuals that smoked more than 20-30 cigarettes per day. Patterns of effect modification were similar across major subtypes squamous cell carcinoma, small cell carcinoma, and adenocarcinoma. These findings corroborate analyses conducted on other datasets of smoking and lung cancer (4,7,8). Our analysis furthers existing knowledge by showing similar patterns of effect modification across a large number of independently designed studies, by allowing for and demonstrating strong effect modification by TSC, and by including semi-parametric spline functions that allow for flexible assessment of patterns of effect-modification.

Intensity of cigarette smoking

The observed variation in the EOR per PCS with ICS might be the result of biological processes such as saturation of metabolism or increasing DNA repair capacity with increasing ICS (4,20). Increasing misclassification of ICS with increasing ICS could also have contributed to the observed patterns. Studies of cigarette smoking and nicotine dependency have shown that an increase in the number of cigarettes smoked per day might be associated with reduced inhalation per cigarette and increasing misclassification in the reporting of the number of cigarettes smoked per day itself with increasing ICS is also conceivable (4,21). Studies using serum or urine cotinine levels as a marker of tobacco smoking intensity, have reported that lung cancer risks do not plateau at high exposure levels, suggesting that such patterns observed in studies using smoking behavior questionnaires were likely due to exposure misclassification (22,23). However cotinine levels only reflect smoking intensity over the last days, and should therefore not be considered as
an 'ideal' marker to estimate lifelong average smoking intensity (24). Our results suggest that the ICS patterns predicted for the low exposure range by our model E (which is not constrained to start at \( \beta = 0 \) at zero exposure) are highly variable in magnitude and direction across study locations. This is likely explained by the limited range of PCS at lower exposure intensities (4). We observed considerable variation across studies in the range of predicted EORs per PCS. Large heterogeneity of results might be associated with factors inherent to the studies, like design, response rates, and statistical power (25). Differential distribution of the relative occurrence of lung cancer subtypes, characteristics of smoking habits, and confounders and effect modifiers such as occupational exposures, indoor radon exposure and dietary components across study populations likely also played a role (14,25).

Time since smoking cessation

Our finding of a continuous decrease in the EOR per PCS with TSC corroborates findings from other studies. For example, Peto et al. (15) demonstrated that the ratio of lung cancer in former smokers compared to current smokers fell sharply with time since smoking cessation. Our analysis demonstrates that this effect remains when adjusted for PCS. Similar patterns with time since exposure cessation have been observed for exposure to benzene and leukemia (26), and exposure to radon and lung cancer (27).

Extension to other (time-varying) exposures

Through its flexible parameterization our model can accommodate various patterns of effect modification and is therefore a suitable tool to explore effect modification by intensity of exposure for a wide range of different exposures. Similar models have successfully been applied in studies arsenic as well as alcohol and smoking and a range of cancers (6,9,10). A limitation of these models (including ours) is that they ignore the possible variation in the EOR due to variation in exposure intensity over time. Using the general framework described in the Web Appendix as
starting point, our model can be extended to accommodate information on time varying exposure. Richardson et al. recently provided an example of such a model in a study of radon exposure and lung cancer (28). A further extension is to allow for more complicated patterns of effect modification by including tensor product splines as done by Berhane et al. (29), although these may come at the cost of reduced interpretability.

Our model and possible further extensions of it provide insight into whether the use of cumulative exposure in an epidemiological analysis is justified or whether reducing complex exposure history to a metric such as cumulative exposure is overly restrictive. Combining information on observed patterns of effect modification with mechanistic insights might contribute to the incorporation of biological hypotheses in the development of more biologically relevant exposure metrics (30).
ACKNOWLEDGEMENTS

Author affiliations: International Agency for Research on Cancer, Lyon, France (Jelle Vlaanderen, Joachim Schüz, Ann Olsson, Kurt Straif); Institute for Risk Assessment Sciences, Utrecht, the Netherlands (Jelle Vlaanderen, Lützen Portengen, Hans Kromhout, Roel Vermeulen); Institute for Prevention and Occupational Medicine of the German Social Accident Insurance – Institute of the Ruhr-Universität Bochum (IPA), Germany (Benjamin Kendzia, Thomas Behrens, Beate Pesch, Thomas Brüning). Inserm, Centre for Research in Epidemiology and Population Health (CESP), U1018, Environmental Epidemiology of Cancer Team, F-94807, Villejuif, France (Isabelle Stücker, Florence Guida); Université Paris-Sud, Unité mixed de recherche en santé (UMRS) 1018, F-94807, Villejuif, France (Isabelle Stücker, Florence Guida); Institut für Epidemiologie I, Deutsches Forschungszentrum fur Gesundheit und Umwelt, Neuherberg, Germany (Irene Brüske, Heinz-Erich Wichmann); Unit of Epidemiology, Fondazione Istituto Di Ricovero e Cura a Carattere Scientifico (IRCCS) Ca´Granda – Ospedale Maggiore Policlinico, Milan, Italy (Dario Consonni); National Cancer Institute, Bethesda, USA (Maria Teresa Landi, Neil Caporaso); University of Montréal Hospital Research Center (CRCHUM), Montreal, Canada (Jack Siemiatycki, Lesley Richardson); Cancer Epidemiology Unit, Department of Medical Sciences, University of Turin, Turin, Italy (Franco Merletti, Dario Mirabelli, Lorenzo Richiardi); The Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden (Ann Olsson, Per Gustavsson, Nils Plato); Institute for Medical Informatics, Biometry and Epidemiology, University of Duisburg-Essen, Essen, Germany (Karl-Heinz Jöckel); Bremen Institute for Prevention Research and Social Medicine, Bremen, Germany (Wolfgang Ahrens, Hermann Pohlabeln); The Biomedical Research Centre Network for Epidemiology and Public Health (CIBERESP), University of Oviedo, Oviedo, Spain (Adonina Tardón); Russian Cancer Research Centre, Moscow, Russia (David Zardze); Roy Castle Lung Cancer Research Programme, Department of Molecular and Clinical Cancer Medicine, Cancer Research Centre, University of Liverpool, Liverpool, UK (John K. Field); Centre for Public Health Research, Massey University, Wellington, New Zealand (Andrea ‘t Mannetje, Neil Pearce);
Occupational Cancer Research Centre, Toronto, Canada (John McLaughlin, Paul Demers); The Nofer Institute of Occupational Medicine, Lodz, Poland (Neonila Szeszenia-Dabrowska); The M Skłodowska-Curie Cancer Center and Institute of Oncology, Warsaw, Poland (Jolanta Lissowska); National Institute of Environment Health, Budapest, Hungary (Peter Rudnai); Regional Authority of Public Health, Banska Bystrica, Slovakia (Eleonora Fabianova); Institute of Public Health, Bucharest, Romania (Rodica Stanescu Dumitru); Institute of Hygiene and Epidemiology, 1st Faculty of Medicine, Charles University, Prague, Czech Republic (Vladimir Bencko); Masaryk Memorial Cancer Institute, Brno, Czech Republic (Lenka Foretova); Palacky University, Faculty of Medicine, Olomouc, Czech Republic (Vladimir Janout); The Tisch Cancer Institute, Mount Sinai School of Medicine, New York, USA (Paolo Boffetta); International Prevention Research Institute, Lyon, France (Paolo Boffetta); Department of Epidemiology, ASL RomaE, Rome, Italy (Francesco Forastiere); The National Institute for Public Health and Environmental Protection (RIVM), Bilthoven, the Netherlands (Bas Bueno-de-Mesquita).

The SYNERGY project is funded by the German Social Accident Insurance (DGUV). The MONTREAL study was supported by the Canadian Institutes for Health Research and Guzzo-SRC Chair in Environment and Cancer. The Toronto study was funded by the National Cancer Institute of Canada with funds provided by the Canadian Cancer Society, and the occupational analysis was conducted by the Occupational Cancer Research Centre which was supported by the Workplace Safety and Insurance Board, the Canadian Cancer Society and Cancer Care Ontario. The ICARE study was supported by the French agency of health security (ANSES); the Fondation de France; the French National Research Agency (ANR); the National Institute of Cancer (INCA); the Fondation for Medical Research (FRM); The French Institute for Public Health Surveillance (InVS); The Health Ministry (DGS); the Organization for the Research on Cancer (ARC); and the French Ministry of work, solidarity and public function (DGT). The LUCA study was supported by the Fondation de France. The Paris study was funded by ARC. The AUT study in Germany was funded by the Federal Ministry of Education, Science, Research, and Technology grant no. 01 HK 173/0. The HdA study was funded by the Federal Ministry of Science (grant No. 01 HK 546/8) and the Ministry of Labour and Social Affairs (grant No. IIIb7-
The INCO study was supported by a grant from the European Commission’s INCO-COPERNICUS program (Contract No. IC15-CT96-0313). In Warsaw, the study was supported by a grant from the Polish State Committee for Scientific Research grant #: SPUB-M-COPERNICUS/P-05/DZ-30/99/2000. In Liverpool, the work was funded by the Roy Castle Foundation as part of the Liverpool Lung Project. The EAGLE study was funded by the Intramural Research Program of the National Institutes of Health, National Cancer Institute, Division of Cancer Epidemiology and Genetics, Bethesda, MD, USA; the Environmental Epidemiology Program of the Lombardy Region, Italy; and the Istituto Nazionale per l'Assicurazione contro gli Infortuni sul Lavoro, Rome, Italy. The TURIN study was supported by the Italian Association for Cancer Research, Region Piedmont, Compagnia di San Paolo. The study in Rome was conducted with a partial support from the European Union Nuclear Fission Safety Program, Grant number: F14P-CT96-0055 and from the Lazio Region. The MORGEN study was supported by the Dutch Ministry of Health, Welfare & Sports, National Institute of Public Health & the Environment, and the Europe Against Cancer Program. The CAPUA study was supported by the Instituto Universitario de Oncologia. Universidad de Oviedo. Asturias, the Fondo de Investigacion Sanitaria. FIS and the Ciber de Epidemiologia y Salud Publica, CIBERESP, Spain. The LUCAS study was supported by the Swedish Council for Work Life Research and the Swedish EPA. The OCANZ study was funded by the Health Research Council of New Zealand, the New Zealand Department of Labour, Lottery Health Research, and by the Cancer Society of New Zealand. The authors thank Ms. Veronique Luzon at IARC for pooling and managing the data.

Conflict of interest: none declared.
REFERENCES


Figure 1. Modification of the Excess Odds Ratio (EOR) for lung cancer per pack-year of smoking (all subtypes combined). In plot A point estimates and 95% confidence intervals from linear odds ratio models fitted within deciles of smoking intensity (model A) are combined with the effect of function $h$ from model C. All predictions in plot A are based on analysis of the dataset restricted to current and never smokers aged 50-75 years. In plot B the effect of function $h$ from model C (dash dotted line) and the effect of function $g_1$ from model D (dashed line) are combined with the effect of function $g_1$ from model E (continuous black line). Model E is fitted on a dataset that included former smokers. In plot C the modification of the EOR per pack-year of smoking by Time Since Smoking Cessation (TSC) (function $g_2$ in model E) was shown. Down-pointing triangles indicate the location of the knots of the restricted cubic splines. Bootstrapped 95% confidence intervals (dashed lines in plots A and C) are based on 1000 replications. See text for details on models.

Figure 2. Modification of the Excess Odds Ratio (EOR) for squamous cell carcinoma (plot A), small cell carcinoma (plot B), and adenocarcinoma (plot C) per pack-year of smoking by smoking intensity. Continuous line is the prediction from function $g_1$ of model E. Down-pointing triangles indicate the location of the knots of the restricted cubic splines. Bootstrapped 95% confidence intervals (dashed lines) are based on 1000 replications. Point estimates and associated 95% confidence intervals are from linear odds ratio models fitted within categories of smoking intensity are from model B. See text for details on models.

Figure 3. Modification of the Excess Odds Ratio (EOR) for lung cancer (all subtypes combined) per pack-year of smoking by smoking intensity stratified by study location. Plots show the effect of function $g_1$ from model E. Bootstrapped 95% confidence intervals (dashed lines) are based on 1000 replications. The following abbreviations were used: AUT-Munich, Arbeit und Technik; HdA, Humanisierung des Arbeitslebens; EAGLE, Environment and Genetics in Lung cancer Etiology; TURIN/VENETO, Population based case-control study of lung cancer in the city of Turin and in the Eastern part of Veneto Region; ROME, Rome lung cancer case-control study; LUCA, Lung cancer in France; PARIS, Lung cancer study in Paris; ICARE, Investigations Cancers Respiratoires et Environnement; CAPUA, Cancer de Pulmon en Asturias; MORGÉN, Monitoring van Risicofactoren en Gezondheid in Nederland; INCO, INCO Copernicus IARC multicenter case-control study of occupational, environment and lung cancer in Central and Eastern Europe; LLP, Liverpool Lung Project; LUCAS, Lung cancer in Stockholm; OCANZ, Occupational Cancer in New Zealand; MONTREAL, Montreal case-control study of environmental causes of lung cancer; TORONTO, Toronto lung cancer (case-control) study. Study details are described in Web Table 1. Predictions for studies LUCA, PARIS, and MORGÉN are not shown due to lack of model convergence. See text for details on models.
Table 1. Linear Odds Ratio Models for Total Cigarette Exposure and Three Lung Cancer Subtypes fitted on the SYNERGY data (1985-2009)\textsuperscript{a}.

<table>
<thead>
<tr>
<th>model</th>
<th>modification</th>
<th>EOR</th>
<th>Δdf</th>
<th>combined</th>
<th>squamous cell carcinoma</th>
<th>small cell carcinoma</th>
<th>adenocarcinoma</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(Δ) LL</td>
<td>(Δ) LL</td>
<td>(Δ) LL</td>
<td>(Δ) LL</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(Δ) LL</td>
<td>P value LRT</td>
<td></td>
</tr>
<tr>
<td>E\textsubscript{0}</td>
<td>not modified</td>
<td>45,828</td>
<td>25,099</td>
<td>14,220</td>
<td>23,001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E\textsubscript{1}</td>
<td>ICS</td>
<td>45,727</td>
<td>25,049</td>
<td>14,185</td>
<td>22,901</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E\textsubscript{1} vs. E\textsubscript{0}</td>
<td>2</td>
<td>101</td>
<td>50</td>
<td>&lt;0.0001</td>
<td>35</td>
<td>&lt;0.0001</td>
<td></td>
</tr>
<tr>
<td>E\textsubscript{2}</td>
<td>TSC</td>
<td>44,895</td>
<td>24,506</td>
<td>13,698</td>
<td>22,786</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E\textsubscript{2} vs. E\textsubscript{0}</td>
<td>2</td>
<td>933</td>
<td>593</td>
<td>&lt;0.0001</td>
<td>522</td>
<td>&lt;0.0001</td>
<td></td>
</tr>
<tr>
<td>E\textsubscript{3}</td>
<td>ICS &amp; TSC</td>
<td>44,843</td>
<td>24,487</td>
<td>13,686</td>
<td>22,714</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E\textsubscript{3} vs. E\textsubscript{0}</td>
<td>4</td>
<td>985</td>
<td>612</td>
<td>&lt;0.0001</td>
<td>534</td>
<td>&lt;0.0001</td>
<td></td>
</tr>
<tr>
<td>E\textsubscript{3} vs. E\textsubscript{1}</td>
<td>2</td>
<td>884</td>
<td>562</td>
<td>&lt;0.0001</td>
<td>499</td>
<td>&lt;0.0001</td>
<td></td>
</tr>
<tr>
<td>E\textsubscript{3} vs. E\textsubscript{2}</td>
<td>2</td>
<td>52</td>
<td>19</td>
<td>0.0001</td>
<td>12</td>
<td>&lt;0.0001</td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations used in table: df, degrees of freedom for likelihood ratio test; ICS, Intensity of cigarette smoke; LL, log likelihood estimate of the fitted model or difference in log likelihood (italic); LRT, Likelihood ratio test; TSC, Time since smoking cessation.

\textsuperscript{a} Excess OR is modified by either a function for the intensity of exposure (E\textsubscript{1}), a function for the time since smoking cessation (E\textsubscript{2}), or both (E\textsubscript{3}).
A)

Excess OR per Pack-year

Cigarettes per Day

0 10 20 30 40 50 60

0.0 0.2 0.4 0.6 0.8 1.0 1.2 1.4
Web Appendix: General Framework for Modeling of Exposure-Time relations

The model described in this manuscript falls within a general framework for flexible modeling of the effects of intensity, duration, and time since exposure (1). In this framework the excess relative risk (ERR) is modeled as

$$\lambda [n(\cdot), t] = \lambda_0(t) \{1 + \text{ERR}[n(\cdot), t]\}$$

Where ERR() is of the form

$$\beta \int_{0}^{t} g[n(u)] w(t,u) du$$

$g[n(u)]$ is a function of exposure intensity $n$ at age $u$, $t$ is attained age, and $w(t,u)$ is a function of age at exposure and time since exposure that modifies each exposure effect.

In this framework, if function $g[n(u)]$ is set to $n(u)$ and $w(t,u)$ is set to 1, the ERR is modeled as a function of simple cumulative exposure. However $g[n(u)]$ and $w(t,u)$ can also be specified as (semi-)parametric functions, providing the flexibility to model the separate and combined effects of intensity, duration, and time since exposure.

In our manuscript we model the lung cancer ERR as a function of time weighted average intensity of smoking
Here we select the two special cases within the flexible framework described above and compare the impact of the specification of \( w(u, t) \) on model fit (AIC and BIC) and function...
## Web Table 1 Overview included SYNERGY studies

<table>
<thead>
<tr>
<th>Principal investigator</th>
<th>study acronym</th>
<th>country</th>
<th>men cases</th>
<th>women cases</th>
<th>recruitment period</th>
<th>ICS ( \text{a} )</th>
<th>duration ( \text{b} )</th>
<th>PCS ( \text{c} )</th>
<th>TSC ( \text{d} )</th>
<th>% current smokers ( \text{e} )</th>
<th>% former smokers ( \text{f} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>H-E Wichmann (2)</td>
<td>HDA</td>
<td>Germany</td>
<td>839</td>
<td>837</td>
<td>1988-1993</td>
<td>17</td>
<td>34</td>
<td>27</td>
<td>16</td>
<td>44.8</td>
<td>35.3</td>
</tr>
<tr>
<td>K-H Jöckel (3)</td>
<td>AUT-Munich</td>
<td>Germany</td>
<td>2659</td>
<td>2702</td>
<td>1990-1995</td>
<td>17</td>
<td>33</td>
<td>27</td>
<td>16</td>
<td>41.5</td>
<td>38.2</td>
</tr>
<tr>
<td>P. Gustavsson (4)</td>
<td>LUCAS</td>
<td>Sweden</td>
<td>1014</td>
<td>2307</td>
<td>1985-1990</td>
<td>16</td>
<td>40</td>
<td>29</td>
<td>15</td>
<td>38.7</td>
<td>31.6</td>
</tr>
<tr>
<td>V. Bencko (5)</td>
<td>INCO CZE(^3)</td>
<td>Czech Rep.</td>
<td>235</td>
<td>293</td>
<td>1998-2002</td>
<td>19</td>
<td>36</td>
<td>31</td>
<td>15</td>
<td>42.2</td>
<td>29.1</td>
</tr>
<tr>
<td>P. Rudnai (5)</td>
<td>INCO HUN(^5)</td>
<td>Hungary</td>
<td>313</td>
<td>248</td>
<td>1998-2001</td>
<td>19</td>
<td>36</td>
<td>33</td>
<td>12</td>
<td>52.3</td>
<td>27</td>
</tr>
<tr>
<td>J. Lissowska (5)</td>
<td>INCO POL(^5)</td>
<td>Poland</td>
<td>550</td>
<td>573</td>
<td>1999-2002</td>
<td>19</td>
<td>36</td>
<td>33</td>
<td>10</td>
<td>56</td>
<td>26.5</td>
</tr>
<tr>
<td>D. Mates (5)</td>
<td>INCO ROM(^5)</td>
<td>Romania</td>
<td>141</td>
<td>152</td>
<td>1998-2001</td>
<td>18</td>
<td>34</td>
<td>28</td>
<td>11</td>
<td>49.9</td>
<td>19.8</td>
</tr>
<tr>
<td>D. Zardize (5)</td>
<td>INCO RUS(^5)</td>
<td>Russia</td>
<td>521</td>
<td>503</td>
<td>1998-2000</td>
<td>15</td>
<td>41</td>
<td>28</td>
<td>19</td>
<td>61.6</td>
<td>19.3</td>
</tr>
<tr>
<td>E. Fabianova (5)</td>
<td>INCO SLO(^5)</td>
<td>Slovakia</td>
<td>287</td>
<td>236</td>
<td>1998-2002</td>
<td>19</td>
<td>36</td>
<td>33</td>
<td>11</td>
<td>46.7</td>
<td>31.6</td>
</tr>
<tr>
<td>J. Field (5,6)</td>
<td>LLP(^5)</td>
<td>United Kingdom</td>
<td>281</td>
<td>572</td>
<td>1998-2005</td>
<td>18</td>
<td>37</td>
<td>30</td>
<td>20</td>
<td>30.1</td>
<td>48.8</td>
</tr>
<tr>
<td>I. Stücker (7)</td>
<td>LUCA</td>
<td>France</td>
<td>304</td>
<td>299</td>
<td>1989-1992</td>
<td>20</td>
<td>35</td>
<td>33</td>
<td>12</td>
<td>48.3</td>
<td>42.3</td>
</tr>
<tr>
<td>F. Merletti (8)</td>
<td>TURIN/VENETO</td>
<td>Italy</td>
<td>955</td>
<td>1249</td>
<td>1990-1994</td>
<td>17</td>
<td>40</td>
<td>31</td>
<td>14</td>
<td>44</td>
<td>36.4</td>
</tr>
<tr>
<td>F. Forastiere (9)</td>
<td>ROME</td>
<td>Italy</td>
<td>294</td>
<td>264</td>
<td>1993-1996</td>
<td>20</td>
<td>42</td>
<td>45</td>
<td>13</td>
<td>50.6</td>
<td>30.6</td>
</tr>
<tr>
<td>J. Siemiatycki (10)</td>
<td>MONTREAL</td>
<td>Canada</td>
<td>714</td>
<td>894</td>
<td>1996-2002</td>
<td>25</td>
<td>40</td>
<td>51</td>
<td>16</td>
<td>42.5</td>
<td>37.9</td>
</tr>
<tr>
<td>N. Caporaso (11)</td>
<td>EAGLE</td>
<td>Italy</td>
<td>1536</td>
<td>1617</td>
<td>2002-2005</td>
<td>20</td>
<td>40</td>
<td>36</td>
<td>17</td>
<td>40.8</td>
<td>39.2</td>
</tr>
<tr>
<td>S. Benhamou (12)</td>
<td>PARIS</td>
<td>France</td>
<td>161</td>
<td>216</td>
<td>1988-1992</td>
<td>20</td>
<td>34</td>
<td>34</td>
<td>11</td>
<td>67.1</td>
<td>31.4</td>
</tr>
<tr>
<td>B. Bueno-de-Mesquita (13)</td>
<td>MORGEN</td>
<td>Netherlands</td>
<td>24</td>
<td>57</td>
<td>1993-1997</td>
<td>17</td>
<td>35</td>
<td>26</td>
<td>NA</td>
<td>60.1</td>
<td>0</td>
</tr>
<tr>
<td>I. Stücker (14)</td>
<td>ICARE</td>
<td>France</td>
<td>2243</td>
<td>2749</td>
<td>2001-2006</td>
<td>19</td>
<td>31</td>
<td>28</td>
<td>16</td>
<td>34.4</td>
<td>40.3</td>
</tr>
<tr>
<td>J. McLaughlin(15)</td>
<td>TORONTO</td>
<td>Canada</td>
<td>198</td>
<td>370</td>
<td>1997-2002</td>
<td>20</td>
<td>30</td>
<td>28</td>
<td>15</td>
<td>19.8</td>
<td>29.6</td>
</tr>
<tr>
<td>A. Tardón (16)</td>
<td>CAPUA</td>
<td>Spain</td>
<td>774</td>
<td>712</td>
<td>2000-2009</td>
<td>20</td>
<td>41</td>
<td>46</td>
<td>14</td>
<td>39.1</td>
<td>41.5</td>
</tr>
<tr>
<td>N. Pearce (17)</td>
<td>OCANZ</td>
<td>New Zealand</td>
<td>212</td>
<td>417</td>
<td>2007-2008</td>
<td>16</td>
<td>30</td>
<td>22</td>
<td>20</td>
<td>17.1</td>
<td>40.5</td>
</tr>
</tbody>
</table>

Abbreviations used in table: AUT-Munich, Arbeit und Technik; Hda, Humanisierung des Arbeitslebens; EAGLE, Environment and Genetics in Lung cancer Etiology; TURIN/VENETO, Population based case-control study of lung cancer in the city of Turin and in the Eastern part of Veneto Region; ROME, Rome lung cancer case-control study; LUCA, Lung cancer in France; PARIS, Lung cancer study in Paris; ICARE, Investigations Cancers Respiratoires et Environnement; CAPUA, Cancer de Pulmon en Asturias; MORGEN, Monitoring van Risicofactoren en Gezondheid in Nederland; INCO, INCO Copernicus IARC multicenter case-control study of occupational, environmental and lung cancer in Central and Eastern Europe; LLP, Liverpool Lung Project (6) ; LUCAS, Lung cancer in Stockholm; OCANZ, Occupational Cancer in New Zealand; MONTREAL, Montreal case-control study of environmental causes of lung cancer; TORONTO; Toronto lung cancer (case-control) study.

\( \text{a} \) Lifetime average intensity of cigarette smoking (cigarettes per day) among active and former smokers
\( \text{b} \) Duration of cigarette smoking (years) among active and former smokers
\( \text{c} \) Pack-years of cigarette smoke (pack-years) among active and former smokers
\( \text{d} \) Smoking status at diagnosis
\( \text{e} \) % of current smokers
\( \text{f} \) % of former smokers
d Time since smoking cessation (years) among former smokers
e Percentage current smokers among cases and controls
f Percentage former smokers among cases and controls
g Part of INCO Copernicus study (18)
Web Figure 1

Modification of the EOR for lung cancer per pack-year of smoking (all subtypes combined) by the intensity of cigarette smoking for women (dashed line) and men (continuous line). Predictions based on the unique effect of function $g_1$ from model E. Bootstrapped 95% confidence intervals are based on 1000 replications. See text for details on model.
References:


