



Published in final edited form as:

Crit Rev Toxicol. 2020 March ; 50(3): 189–200. doi:10.1080/10408444.2020.1727410.

Smoking, air pollution, and lung cancer risk in the Nurses' Health Study cohort: time-dependent confounding and effect modification

Ellen T. Chang^{1,2}, Edmund C. Lau¹, Suresh H. Moolgavkar^{3,4}

¹Center for Health Sciences, Exponent, Inc., Menlo Park, CA

²Stanford Cancer Institute, Stanford, CA

³Center for Health Sciences, Exponent, Inc., Bellevue, WA

⁴Division of Public Health Sciences, Fred Hutchinson Cancer Research Center, Seattle, WA

Abstract

The proportional hazards (PH) model is commonly used in epidemiology despite the stringent assumption of proportionality of hazards over time. We previously showed, using detailed simulation data, that the impact of a modest risk factor cannot be estimated reliably using the PH model in the presence of confounding by a strong, time-dependent risk factor. Here, we examine the same and related issues using a real dataset. Among 97,303 women in the prospective Nurses' Health Study cohort from 1994 through 2010, we used PH regression to investigate how effect estimates for cigarette smoking are affected by increasingly detailed specification of time-dependent exposure characteristics. We also examined how effect estimates for a fine particulate matter (PM_{2.5}), a modest risk factor, are affected by finer control for time-dependent confounding by smoking. The objective of this analysis is not to present a credible estimate of the impact of PM_{2.5} on lung cancer risk, but to show that estimates based on the PH model are inherently unreliable. The best-fitting model for cigarette smoking and lung cancer included pack-years, duration, time since cessation, and an age-by-pack-years interaction, indicating that the hazard ratio (HR) for pack-years was significantly modified by age. In the fully adjusted best-fitting model for smoking including pack-years, the HR per 10- $\mu\text{g}/\text{m}^3$ increase in PM_{2.5} was 1.06 (95% confidence interval (CI): 0.90, 1.25); the HR for PM_{2.5} in the full cohort ranged between 1.02 and 1.10 in models with other smoking adjustments, indicating a residual confounding effect of smoking. The HR for PM_{2.5} was statistically significant only among former smokers when adjusting for smoking pack-years (HR = 1.35, 95% CI = 1.00, 1.82 in the best-fitting smoking model), but not in models adjusting for smoking duration and average packs (pack-years divided by duration). The association between cumulative smoking and lung cancer is modified by age, and improved model fit is obtained by including multiple time-varying components of smoking history. The association with PM_{2.5} is residually confounded by smoking and modified by smoking status. These findings underscore limitations of the PH model and emphasize the advantages of directly estimating hazard functions to characterize time-varying exposure and risk. The hazard function, not the relative hazard, is the fundamental measure of risk in a population. As a consequence, the use of time-dependent PH models does not address crucial issues introduced by temporal factors in epidemiological data.

Introduction

The Cox proportional hazards (PH) regression model (Cox 1972) is commonly used in epidemiology for the analysis of observational cohort studies. The main assumption of the original PH model is that the hazard ratio is constant over the time scale. However, this assumption may not hold in the context of epidemiological studies with diverse and time-varying exposures, numerous potential confounders, and long-term follow-up.

We recently conducted a simulation study in which we investigated how hazard ratio (HR) estimates using the PH model are biased when the HR is strongly modified by time and depends on temporal exposure characteristics, such as duration and time since cessation (Moolgavkar et al. 2018). In that study, we also evaluated how inadequate control of a strong time-dependent confounder affected HR estimates for a weaker risk factor. We found that in the presence of residual confounding by a strong, time-dependent risk factor, such as smoking, use of the PH model can result in biased estimates of association with a modest risk factor.

A limitation of our prior study is that it relied on simulated data, even though the cohort had realistic life histories generated by the U.S. National Cancer Institute's well-validated Smoking History Generator (Moolgavkar et al. 2012, Holford et al. 2014a, Holford et al. 2014b, National Cancer Institute). Accordingly, we sought to evaluate the issues of time-dependent confounding and effect modification using the PH model using a real dataset with detailed individual-level information on smoking habits over time. An ideal dataset for this purpose is the Nurses' Health Study (NHS) cohort, which is distinguished by its repeated collection of exposure and health information from a large number of subjects over decades of follow-up. Moreover, the NHS has information on air pollution, in the form of particulate matter (PM) of various sizes ($< 2.5 \mu\text{m}$, $< 10 \mu\text{m}$, or $2.5\text{--}10 \mu\text{m}$ in aerodynamic diameter), a weak risk factor for lung cancer (Puett et al. 2014) with which to test the impact of inadequate control for confounding by smoking. Therefore, we used the NHS cohort to investigate how the association between smoking and lung cancer changes with increasingly detailed specification of exposure, and how the association between PM and lung cancer is affected by the way in which smoking exposure is modeled. The purpose of this investigation is to explore the inherent unreliability of estimates based on the PH model in the presence of confounding by a strong, time-dependent risk factor, which may often exist in long-term epidemiological datasets.

Materials and Methods

Study population

The NHS is an ongoing prospective cohort of 121,700 female nurses who were enrolled in 1976 when they were between 30 and 55 years of age. Data from the NHS are available to research collaborators through an application process (<https://www.nurseshealthstudy.org/researchers>).

Participants were initially recruited from 11 states, but have resided in each of the 50 U.S. states since the mid-1990s. Information on potential risk factors and self-reported new

diagnoses of health outcomes is provided by nurses through mailed biennial questionnaires, for which response rates are > 90%. Vital status is ascertained through next of kin and the National Death Index (<http://www.cdc.gov/nchs/ndi.htm>), both of which have identified an estimated 98% of deaths in the cohort.

We used nearly the same analytic cohort as Puett et al. (2014) in their analysis of air pollution and lung cancer incidence in the NHS. Eligible women were those who were alive and did not have a prior diagnosis of cancer (except for non-melanoma skin cancer) as of 1994–1996, responded to the questionnaire in 1994–1996 or a later follow-up cycle, and had complete information on PM throughout follow-up. Unlike Puett et al. (2014), we also required that eligible women have non-missing information on smoking status. This cohort comprised 97,303 women with 1,402,829 person-years of follow-up. (By comparison, the cohort analyzed by Puett et al. (2014) included 103,650 women with 1,510,027 person-years of follow-up.)

This study was approved by the Institutional Review Board of Brigham and Women's Hospital. Informed consent was implied through return of the questionnaires. In addition, this study was approved by the Human Investigations Committee of the Connecticut Department of Public Health, from which certain data used in this publication were obtained.

Smoking exposure

From the biennial mailed questionnaires, information on cigarette smoking is collected to enable time-varying characterization of smoking habits, including duration, pack-years, and time since cessation (if applicable), at each follow-up cycle. When current smoking information is missing, previously reported smoking information can be carried forward from prior questionnaires.

For example, for women who reported never smoking in an earlier questionnaire, never-smoker status is carried forward into all subsequent follow-up cycles, because of the rarity of smoking initiation after the start of follow-up. For women who reported having quit smoking more than 10 years ago, former-smoker status is carried forward into all subsequent follow-up cycles, because of the rarity of smoking re-initiation after having quit for more than 10 years. For women who reported currently smoking or having quit fewer than 10 years ago, current-smoker or former-smoking status, respectively, is carried forward for only one additional cycle, after which smoking status is considered missing if subsequent questionnaires are not completed. Number of cigarettes smoked per day is carried forward for current smokers if it is not reported. Overall, 1.6% of women had one or more follow-up cycles skipped due to missing smoking status.

Air pollution exposure

The assessment of air pollution exposure based on residential address in the NHS cohort is described in detail by Puett et al. (2014). Briefly, residential addresses are updated through the biennial questionnaires, and all available addresses have been geocoded to latitude and longitude coordinates. Using previously validated spatiotemporal models (Yanosky et al. 2008, Yanosky et al. 2009, Weuve et al. 2012, Yanosky et al. 2014), predictions of PM_{2.5} and

PM₁₀ were generated by NHS investigators for all months between January 1988 and December 2007 for the continental U.S.

These models used monthly average PM_{2.5} and/or PM₁₀ data from the U.S. Environmental Protection Agency's (EPA) Air Quality System (U.S. EPA 2018), the IMPROVE network (IMPROVE 2019), and various other sources (Spengler et al. 1996, Suh et al. 1997). Generalized additive mixed models with monthly penalized spline smooth spatial terms, penalized spline smooth terms of geospatial predictors (i.e., distance to nearest A1–A4 roads, percent urban land use within 1 km, elevation, point sources of PM, county population density, census tract population density (only for PM₁₀), and meteorological predictors), and terms for time were used to create separate PM prediction surfaces for each month and each PM size fraction (Yanosky et al. 2009). PM_{2.5} levels prior to 1999 were modeled by NHS investigators using data on PM₁₀ (Yanosky et al. 2009); PM_{2.5–10} was derived by subtracting monthly PM_{2.5} from monthly PM₁₀ estimates. Cross-validation R^2 values were 0.59 for PM₁₀, 0.76 for pre-1999 PM_{2.5}, and 0.77 for post-1999 PM_{2.5}. For the primary analysis, we used 72-month cumulative average PM_{2.5} levels, as did Puett et al. (2014), who reported finding no substantive differences in their results based on 24-, 48-, 96-, or 120-month cumulative average PM_{2.5} levels. Following the same approach as Puett et al. (2014), we did not institute a lag for PM_{2.5} exposure (or any confounders), even though a latency period of decades would be anticipated to intervene between exposure to a carcinogen and the onset of lung cancer (Boffetta et al. 2015).

Lung cancer incidence

Lung cancers in the NHS cohort were identified initially through self-report by participants or their next of kin, or from death certificates; these reports were subsequently confirmed by physician review of medical records, with blinding to exposure status. Medical records were obtained for 83% of reported lung cancer cases; of these, 87% had primary lung cancer confirmed by pathology reports. Because of the high validity of self-reported lung cancer, we included all primary reports that were reconfirmed by the participant where pathological reports were not available.

Statistical approach

Like Puett et al. (2014), we used time-varying Cox PH models, stratified by biennial time period and age in months (except when evaluating age interactions, in which case age was entered as a covariate). These models were used to estimate HRs with corresponding 95% confidence intervals (CIs) for the associations of incident lung cancer with cigarette smoking and exposure to PM_{2.5} (per 10-ug/m³ increase in concentration). Person-months were calculated from 1994–1996 until the end of follow-up in June 2010, diagnosis of lung cancer, death from another cause, or loss to follow-up, whichever occurred first.

Puett et al. (2014) adjusted for cigarette smoking based on time-varying covariates for smoking status (never, former, or current), pack-years (continuous), and months since quitting for former smokers (continuous), with indicator variables for missing data. For our analysis, we used the Akaike information criterion (AIC) (Akaike 1974) to compare models

with various levels of detail in smoking characterization, including the following time-varying covariates:

- Smoking status only
- Smoking status and pack-years
- Smoking status, pack-years, and duration (years)
- Smoking status, average packs (calculated as total pack-years divided by smoking duration in years), and duration (years)
- Smoking status, pack-years, duration (years), and time since cessation (months)
- Smoking status, average packs, duration (years), and time since cessation (months)
- Smoking pack-years, duration (years), and time since cessation (months)
- Smoking average packs, duration (years), and time since cessation (months)
- Smoking pack-years, duration (years), time since cessation (months), and age*pack-year interactions, including age only, age and age squared, or age, age squared, and age cubed
- Smoking average packs, duration (years), time since cessation (months), and age*average-pack interactions, including age only, age and age squared, or age, age squared, and age cubed.

Models were adjusted for the same covariates as in Puett et al. (2014), i.e., age, time period, and geographic region (Northeast, South, Midwest, or West) in minimally adjusted models. Fully adjusted models additionally included time-varying body mass index (continuous kg/m²), alcohol consumption (none or > 0 g/day), physical activity (< 3, 3 to < 18, or 18 metabolic equivalent hours per week), overall diet quality (continuous Alternative Healthy Eating index (Chiuve et al. 2012)), and census tract median home value (continuous) and median income (continuous), as well as non-time-varying secondhand smoke exposure at home, at work, and during childhood as reported in 1982. Covariates included in the final models were selected by Puett et al. (2014) based on *a priori* consideration of factors previously associated with lung cancer or PM exposure in the NHS cohort, along with observation of a 10% or greater change in the HR estimate for PM_{2.5} and lung cancer. Because the objective of our analysis is not to provide a “valid” estimate of the association between PM_{2.5} and lung cancer, but rather to evaluate the reliability of the PH model, we did not explore alternative covariate adjustments. In any case, there is no *a priori* guidance to inform the selection of covariates on biological grounds. Puett et al. (2014) found that HR estimates for PM_{2.5} were not substantially affected by multivariable adjustment in the full cohort or among never smokers, but HRs were augmented (farther above 1.0) by multivariable adjustment among never and former smokers, and attenuated (from below 1.0) by adjustment among current smokers.

Like Puett et al. (2014), in models stratified by smoking status, we combined current smokers with former smokers who quit fewer than 10 years ago, and we analyzed former smokers who quit at least 10 years ago separately or combined with never smokers. Women

could contribute to multiple smoking-status categories throughout follow-up, with one status per biennial cycle. We used the same coding as Puett et al. (2014) for all variables, including smoking and PM.

We conducted secondary analyses using different PM_{2.5} averaging times (24 and 48 months instead of 72 months, using only shorter averaging times to evaluate more granular characterization of PM_{2.5} levels), different levels of baseline stratification by age group (1 year, 5 years, and 10 years instead of months to create larger, more robust baseline strata), and different PM size fractions (PM₁₀ and PM_{2.5-10} instead of PM_{2.5}).

P-values < 0.05 were considered statistically significant. All analyses were performed with SAS version 9 (SAS Institute Inc., Cary, NC).

Results

During follow-up, 0.1% of initial never smokers began smoking, 4% of former smokers took up smoking again, and 57% of current smokers quit smoking at least temporarily. Of those who quit smoking, 28% stopped for at least 10 years. Table 1 shows average smoking characteristics and average estimated ambient PM levels at study baseline in 1994 and over the duration of follow-up into 2010, stratified by baseline smoking status. In the total cohort, 72-month average ambient PM_{2.5} at baseline was 15.1 µg/m³ (standard deviation (SD) = 3.2), and 72-month average ambient PM₁₀ was 25.5 µg/m³ (SD = 6.5). Average ambient PM levels declined over the course of follow-up, with overall 72-month means of 13.2 µg/m³ (SD = 2.7) for PM_{2.5} and 21.7 µg/m³ (SD = 5.4) for PM₁₀ between 1994 and 2010. During follow-up, 1,992 incident cases of lung cancer were identified.

Characterization of smoking association

Table 2 shows estimated HRs for lung cancer risk and AICs from models of PM_{2.5} and various characterizations of smoking history based on cumulative pack-years in the overall cohort. Results are shown for minimally adjusted models accounting for time period, age, and geographic region, and fully adjusted models accounting for all covariates identified as potential confounders. Detailed results are provided in Appendix 1 (minimally adjusted models) and Appendix 2 (fully adjusted models). In the overall cohort, with or without adjustment for other covariates, model fit based on AIC generally improved with the inclusion of more characteristics of smoking history (i.e., duration and time since cessation, without smoking status) and an interaction between cumulative pack-years and linear age (Table 2). As expected, the HRs for each smoking variable generally decreased with the inclusion of additional smoking characteristics, due to the smaller proportion of variance explained by each variable. Additional interactions between cumulative pack-years and age squared or with both age squared and age cubed did not further improve model fit. For models including cumulative pack-years, the best-fitting model also included smoking duration, time since cessation, and an age*pack-years interaction (minimally adjusted model AIC = 18,509.50; fully adjusted model AIC = 18,491.76).

Table 3 shows the corresponding HRs for lung cancer risk and AICs from minimally and fully adjusted models of PM_{2.5} and various characterizations of smoking history based on

average packs in the overall cohort. Detailed results are provided in Appendix 1 and Appendix 2.

Again, model fit based on AIC generally improved with the inclusion of more characteristics of smoking history, but no significant interactions were detected between average packs and linear age, age squared, or age cubed in nested models (Table 3). Adding interactions with age rendered the estimates for average packs statistically unstable. For models including average packs, the best-fitting model also included smoking duration and time since cessation (minimally adjusted model AIC = 18,866.94; fully adjusted model AIC = 18,854.06).

Results stratified by smoking status are shown in Table 4, with detailed results in Appendix 1 and Appendix 2. The best-fitting models for cumulative pack-years or average packs in the overall cohort also were the best-fitting models among former smokers and current smokers. That is, after stratification by smoking status, the AIC among models for cumulative pack-years was lowest when including smoking duration, time since cessation, and an age*pack-years interaction, and the AIC among models for average packs was lowest when including smoking duration and time since cessation (other models shown in Appendix 1 and Appendix 2). Comparing these two models based on AIC, the fit was better with pack-years than average packs in the overall cohort, among former smokers, and among current smokers.

The improvement of model fit with an interaction between cumulative pack-years and age, as well as the statistical significance of the linear age interaction term ($p < 0.05$ in minimally and fully adjusted models for the full cohort, former smokers, and current smokers), demonstrates effect modification of the smoking-lung cancer association by age. This age-related effect modification is illustrated in Figures 1A (full cohort), 1B (former smokers), and 1C (current smokers) with full covariate adjustment; figures with minimal adjustment are provided in Appendix 1. These figures are based on models including interactions between cumulative pack-years and age, age squared, and age cubed, but results do not differ substantively from models including linear age only. For the full cohort, the figures show an increase in the HR with advancing age up to around 50–55 years, followed by a steady decline (Figure 1A). The HR is slightly later for former smokers (Figure 1B) and earlier for current smokers (Figure 1C). Results are unstable above age 85 years.

Effect modification and confounding of PM_{2.5} association by smoking

Most models, including all models of the total cohort, never smokers, and current smokers, showed no significant association between PM_{2.5} and lung cancer risk (Tables 2–4, Appendix 1 and Appendix 2). Statistically significant positive associations between PM_{2.5} and lung cancer risk were detected only in certain models adjusting for smoking pack-years (but not average packs) among former smokers who quit at least 10 years ago. In the fully adjusted best-fitting model including smoking pack-years, smoking duration, time since cessation, and an age*pack-years interaction, the HR per 10- $\mu\text{g}/\text{m}^3$ increase in PM_{2.5} was 1.06 (95% CI: 0.90, 1.25) in the overall cohort and 1.35 (95% CI = 1.00, 1.82) among former smokers. In the fully adjusted best-fitting model including average packs, smoking

duration, and time since cessation, the HR per 10- $\mu\text{g}/\text{m}^3$ increase in $\text{PM}_{2.5}$ was 1.02 (95% CI: 0.87, 1.20) in the overall cohort and 1.27 (95% CI = 0.95, 1.71) among former smokers.

In fully adjusted models including pack-years, the HR per 10- mg/m^3 increase in $\text{PM}_{2.5}$ ranged between 1.05 and 1.10 in the total cohort, a doubling of the parameter estimate from 5% to 10% (Table 2 and Appendix 2). Among former smokers, the HR ranged between 1.31 and 1.38 in these models. In fully adjusted models including average packs, the HR for $\text{PM}_{2.5}$ ranged between 1.02 and 1.07 (more than tripling the parameter estimate of 2%) in the total cohort, while it ranged between 1.26 and 1.27 among former smokers (Table 3 and Appendices 1 and 2).

For comparison, when contrasting results for the total cohort using the same models with minimal or full covariate adjustment, the HRs for $\text{PM}_{2.5}$ differed less than they did across various levels of smoking adjustment, except in models for $\text{PM}_{2.5}$ alone (Table 2 and Table 3). That is, confounding by all other covariates was less than residual confounding by varying adjustment for smoking.

Secondary analyses

As shown in Appendix 3 the results of secondary analyses of 24- and 48-month averaged $\text{PM}_{2.5}$ and baseline stratification by 1-year, 5-year, or 10-year age group were substantively similar to the primary analysis based on 72-month averaged $\text{PM}_{2.5}$ with baseline stratification by age in months. PM_{10} and $\text{PM}_{2.5-10}$ were not significantly associated with lung cancer risk in any models, including among former smokers.

Discussion

Our results based on real-world data from the NHS cohort largely confirm the findings from our analysis of a simulated cohort with realistic smoking-history data (Moolgavkar et al. 2018). In the simulated dataset, we found that relative risk of mortality from smoking was strongly modified by age, and that including time-dependent cumulative packs smoked, smoking duration, and time since cessation improved model fit. We also found that even after detailed control for time-dependent smoking history, residual confounding created a spurious modest association with a covariate that was correlated with smoking but not independently associated with the outcome.

Likewise, in the present study, we found strong evidence of effect modification of the association between cumulative smoking and lung cancer risk, and the best-fitting model for smoking, given the available data, included time-dependent smoking pack-years, smoking duration, time since smoking cessation, and an interaction between age and pack-years. An alternative model including average packs instead of pack-years also fit the data well, but did not exhibit significant effect modification by age. In the simulated cohort we found that interactions between cumulative smoking and age, age squared, and age cubed all were highly statistically significant (Moolgavkar et al. 2018), whereas in the NHS cohort we found a significant interaction only between cumulative smoking and linear age. Nevertheless, the age-specific pattern of lung cancer risk associated with smoking was similar to that in the simulated cohort as well as in the American Cancer Society Cancer

Prevention Study I cohort (Burns et al. 1997) (illustrated in Moolgavkar et al. 2018). In all cohorts, the age-specific relative risk due to smoking peaked between ages 50 and 60 years, followed by a continuous decline with older age. This pattern is unlikely to be detected by statistical tests for monotonic departures from proportionality of hazards.

Taken together, these results demonstrate that age-related effect modification needs to be addressed in any cohort analysis of smoking, that cumulative exposure alone is inadequate to capture the effect of smoking, and that the traditional Cox PH approach has fundamental limitations for addressing these temporal issues. The inability of a single exposure metric to characterize the impact of smoking, and the importance of characterizing exposure history over time, has been discussed previously (Knoke et al. 2004, Hazelton et al. 2005, Lubin et al. 2007, Tammemagi et al. 2011, Moolgavkar et al. 2012, Peto 2012, Thomas 2014, Vlaanderen et al. 2014).

These discussions underscore the inflexibility of estimating hazard ratios using the Cox PH model, as opposed to directly estimating hazard functions to allow explicit parameterization of temporal changes in exposure and resultant risk. Parametric models of cancer hazard functions, based on biological concepts of multistage carcinogenesis, have long provided a remedy to this problem (Peto 1977, Day and Brown 1980, Brown and Chu 1983, Thomas 1988, Knoke et al. 2004, Hazelton et al. 2005, Richardson 2009, Thomas 2009, Moolgavkar et al. 2012, Moolgavkar and Luebeck in press, Moolgavkar et al. 2015). Indeed, in the NHS cohort (combined with the Health Professionals Follow-Up Study), Meza et al. (2008) used multistage carcinogenesis models to demonstrate the strong dependence of lung cancer risk on temporal aspects of smoking, including a modifying effect of smoking duration on the association with smoking intensity, and decline in the smoking-related relative risk with increasing time since cessation.

The other main finding from the present analysis relates to the impact of effect modification and residual confounding by smoking on the observed association with $PM_{2.5}$ exposure. We found that improved control of confounding by smoking, through inclusion of additional time-dependent covariates in the multivariate model, had varying effects on the HR estimate for $PM_{2.5}$, sometimes moving it toward the null and sometimes away from the null. In the best-fitting model for pack-years, the HR estimate was attenuated relative to estimates in models that did not control for smoking, or that controlled for smoking status and pack-years, but it was comparable to the estimate in a model that controlled only for smoking status. In the best-fitting model for average packs, the HR estimate was also attenuated relative to estimates in models that did not control for smoking, or that controlled for smoking status alone, but it was comparable to estimates in other models that controlled for average packs and duration, with or without age interactions. $PM_{2.5}$ estimates were consistently weaker in models that controlled for average packs (and duration) than those that controlled for pack-years.

The magnitude of residual confounding by smoking was larger than the magnitude of confounding by all other covariates, including body mass index, alcohol consumption, physical activity, diet quality, area-level socioeconomic status, and secondhand smoke exposure. This contrast highlights the relevance of focusing on residual confounding by

smoking as one of the strongest risk factors for lung cancer and, accordingly, a major potential confounder of virtually any other estimated association with lung cancer risk.

Moreover, the NHS represents in many ways a best-case scenario for control of the time-dependent effects of smoking, because the investigators collected detailed information on smoking exposure repeatedly in biennial questionnaires. In nearly all other prominent cohorts for air pollution epidemiology, such as the American Cancer Society Cancer Prevention Study II cohort (Pope et al. 2002, Pope et al. 2009), the Harvard Six Cities cohort (Dockery et al. 1993, Laden et al. 2006), and the European Study of Cohorts for Air Pollution Effects (Beelen et al. 2014), smoking information was collected only once at study entry. In some cohorts, such as the Medicare Cohort Air Pollution Study (Zeger et al. 2008), no individual-level smoking information is available. Thus, the potential for residual confounding due to insufficient control for time-dependent smoking is greater in nearly all other cohort studies.

Besides acting as an important confounder of the $PM_{2.5}$ association, smoking also was a key effect modifier in the NHS cohort. Like Puett et al. (2014), we detected a statistically significant positive association with $PM_{2.5}$ in the combined group of never smokers and former smokers who quit at least 10 years ago. When we further distinguished between never and former smokers (an analysis not reported by Puett et al. (2014)), we found that the association was restricted to former smokers. This finding is not readily explained by the known toxicological effects of $PM_{2.5}$ (IARC 2016). For instance, if smoking increases susceptibility to ambient outdoor $PM_{2.5}$ through cellular injury and inflammation of the respiratory epithelium, then one might expect current smokers to exhibit a stronger positive association with $PM_{2.5}$ than former smokers. Conversely, if deposition of ambient outdoor $PM_{2.5}$ in the respiratory tract is more efficient in the absence of interference from smoking-related particles, then one might expect never smokers to exhibit a stronger association with $PM_{2.5}$ than former smokers. Thus, the observed pattern of smoking-related effect modification in the NHS cohort is puzzling, and may be due to chance.

Whether recent exposure to ambient $PM_{2.5}$ can plausibly contribute to the development of lung cancer, which has a latency of many years or even decades, is also questionable (Boffetta et al. 2015). In the absence of data on distant past $PM_{2.5}$ levels, reliance on recent levels (with no exposure lag) could result in considerable misclassification of the etiologically relevant exposure, with unpredictable bias in effect estimates. If current exposure levels are strongly correlated with past levels, then use of recent exposure data may be better justified, but will result in overestimated relative risks due to the declining air pollution levels in most industrialized countries.

By using real-world observational data, this study overcomes the reliance on a simulated dataset in our prior analysis (Moolgavkar et al. 2018), and it illustrates the extent of potential residual confounding in an actual scenario that probably involves multivariate confounding and complex relationships between covariates. We did not explicitly consider how more detailed classification of other potential confounders, such as socioeconomic status, might further affect the residual confounding effect of smoking, nor did we evaluate possible non-

linear associations with smoking. We also did not compare the results of the Cox PH analysis directly with results from multistage carcinogenesis models.

This paper raises important ancillary issues regarding the choice of statistical models, the testing of null hypotheses, and the limits of standard methods for the analysis of epidemiological data. Many statistical approaches to model selection and model averaging are available. Here, we use the AIC as our statistical tool for model choice without meaning to suggest that it is the only or the best option. Other approaches to model selection would probably lead to similar conclusions about the suitability of the PH model overall. We caution, moreover, that the common but misguided practice of choosing statistical models to maximize effect estimates can lead to substantial bias when the effect estimates are small (Lumley and Sheppard 2000).

One of the critical problems with current standard approaches to the statistical analysis of epidemiological data is that whereas the hazard function, i.e., the rate at which a disease occurs in a previously disease-free population, would appear to be the fundamental measure of risk, current methods focus on estimating the relative hazard. When the PH model was introduced by D. R. Cox (1972), it was hailed as a major landmark in biostatistics, which indeed it was for analysis of clinical trials data, which are typically characterized by limited follow-up time and time-invariant covariates. The PH model was soon extended and generalized for analyses of epidemiological data with much longer follow-up times and with time-dependent covariates (e.g., Kalbfleisch and Prentice 2002). An unfortunate legacy of the great success of the PH model for analyses of clinical trials and observational epidemiology data is the virtual abandonment of parametric models for survival analysis. Because the PH model is semi-parametric, it was considered to be vastly superior to parametric models, which require making assumptions regarding the underlying hazard functions. However, the PH model makes assumptions of dubious biological validity, as we have noted here.

In particular, for epidemiological data, the assumption of constant proportionality of hazards over age and across populations is often biologically implausible. Why should the HR for a given exposure remain constant with age or across populations with different background hazards? Yet these are assumptions that are commonly made. Statistical tests can assess departures from constancy of the HR, where such constancy is treated as the statistical null hypothesis. It seems clear, however, that non-constancy of the HR should instead be the biologically expected null hypothesis. Similarly, biologically one would expect temporal factors such as ages at the start and end of exposure, duration of exposure, and time-varying intensities of exposure to be important determinants of the HR, yet the standard statistical null hypothesis is the constancy of the HR with respect to these factors.

In conclusion, this study builds on our prior findings (Moolgavkar et al. 2018) by using actual observational data from the uniquely suited NHS to illuminate some of the shortcomings of the Cox PH model for analyzing the time-dependent relationships of smoking and PM_{2.5} with each other and with lung cancer risk. The PH model is currently the mainstay for epidemiological analyses of cohort and, by extension, case-control data, yet violation of its basic assumptions can yield biased results (Hernán 2010). Estimates of

human health risk obtained by application of this model have been used to set national and state ambient air quality standards, and to estimate the public health benefits that would accrue from decreases in air pollution levels (U.S. EPA 2017). Because of the issues discussed here, these quantitative estimates cannot be considered to be reliable. The problem of unreliable estimates is compounded by the fact that several important datasets on which critical regulatory decisions are based are not made available to stakeholders for independent evaluation. Therefore, our results have regulatory and public-health implications and provide further impetus for the development of parametric alternatives to the PH model that explicitly estimate hazard functions to address the impact of time-varying exposure patterns on risk.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Declaration of interest

SHM and ETC have provided expert testimony and ECL has provided consulting support in litigation related to air pollution. All authors are employees of Exponent, Inc., an international science and engineering consulting company.

The Nurses' Health Study is funded by the National Cancer Institute (<https://www.cancer.gov/>) of the National Institutes of Health (grant UM1CA186107). We thank the participants and the researchers of the Nurses' Health Study.

The analysis described in this article was funded by the American Petroleum Institute (<https://www.api.org/>) through a contract with Exponent, Inc. The analytical approach, conduct of the analyses, interpretation of the results, and conclusions drawn are exclusively the professional work product of the authors and are not necessarily those of the organizations that funded the research. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript, and did not review this article prior to its submission, although staff at the American Petroleum Institute reviewed a written report that described provisional findings.

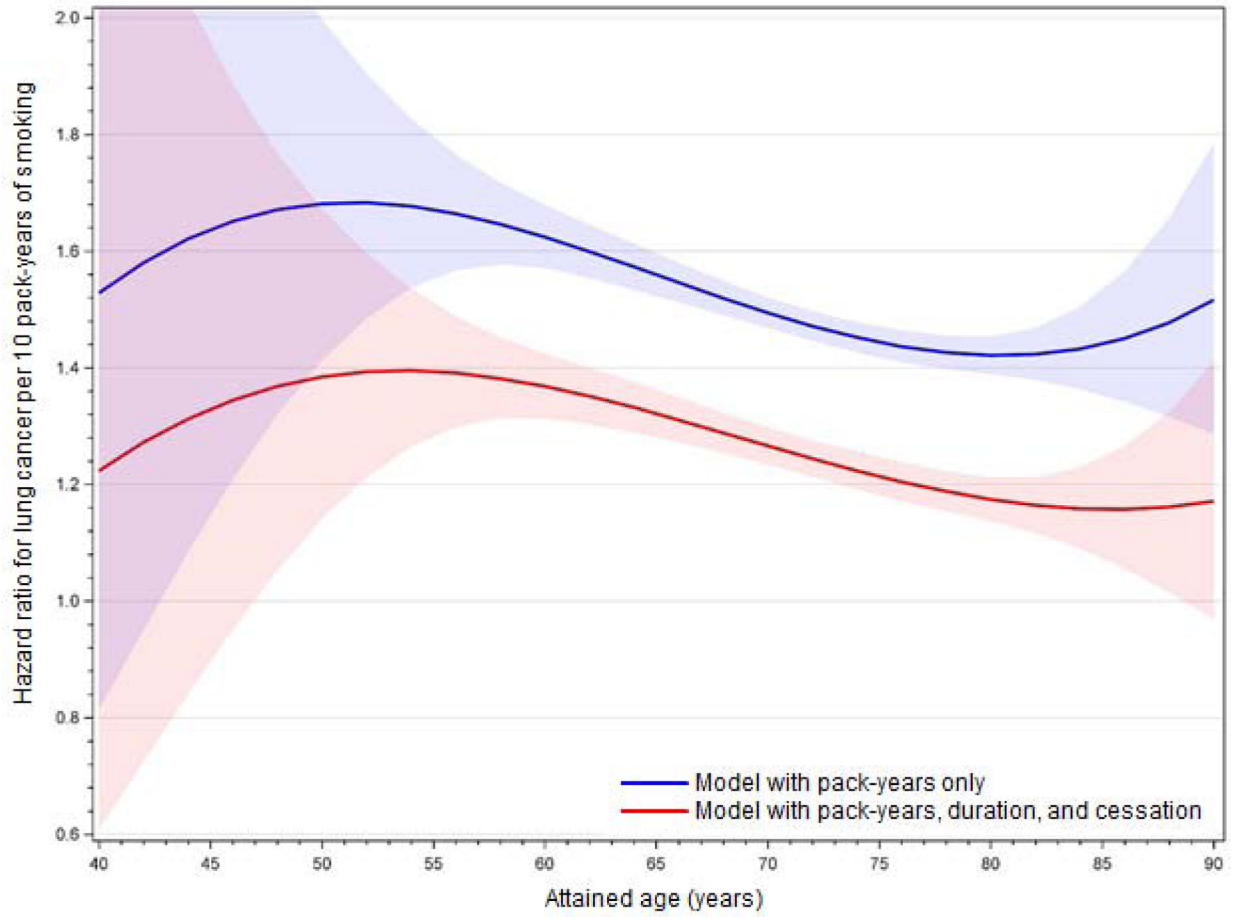
References

- Akaike H A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 1974;19(6):716–723.
- Beelen R, Raaschou-Nielsen O, Stafoggia M, et al. Effects of long-term exposure to air pollution on natural-cause mortality: an analysis of 22 European cohorts within the multicentre ESCAPE project. *Lancet* 2014;383(9919):785–795. [PubMed: 24332274]
- Boffetta P, La Vecchia C, Moolgavkar S. Chronic Effects of Air Pollution are Probably Overestimated. *Risk Anal* 2015;35(5):766–769. [PubMed: 25676287]
- Brown CC, Chu KC. Implications of the multistage theory of carcinogenesis applied to occupational arsenic exposure. *J Natl Cancer Inst* 1983;70(3):455–463. [PubMed: 6572736]
- Burns DM, Shanks TG, Choi W, Thun MJ, Heath CW Jr., Garfinkel L, 1997 Chapter 3. The American Cancer Society Cancer Prevention Study I: 12-year followup of 1 million men and women in: Burns DM, Garfinkel L, Samet JM (Eds.), *Monograph 8: Changes in Cigarette-Related Disease Risks and Their Implications for Prevention and Control*. U.S. Department of Health and Human Services, Public Health Service, National Institutes of Health, Bethesda, MD, p. 602.
- Chiuve SE, Fung TT, Rimm EB, et al. Alternative dietary indices both strongly predict risk of chronic disease. *J Nutr* 2012;142(6):1009–1018. [PubMed: 22513989]
- Cox DR. Regression models and life-tables. *J Roy Stat Soc Ser B* 1972;34(2):187–220.
- Day NE, Brown CC. Multistage models and primary prevention of cancer. *J Natl Cancer Inst* 1980;64(4):977–989. [PubMed: 6929006]

- Dockery DW, Pope CA 3rd, Xu X, et al. An association between air pollution and mortality in six U.S. cities. *N Engl J Med* 1993;329(24):1753–1759. [PubMed: 8179653]
- Hazelton WD, Clements MS, Moolgavkar SH. Multistage carcinogenesis and lung cancer mortality in three cohorts. *Cancer Epidemiol Biomarkers Prev* 2005;14(5):1171–1181. [PubMed: 15894668]
- Hernán MA. The hazards of hazard ratios. *Epidemiology* 2010;21(1):13–15. [PubMed: 20010207]
- Holford TR, Levy DT, McKay LA, et al. Patterns of birth cohort-specific smoking histories, 1965–2009. *Am J Prev Med* 2014b;46(2):e31–37. [PubMed: 24439359]
- Holford TR, Meza R, Warner KE, et al. Tobacco control and the reduction in smoking-related premature deaths in the United States, 1964–2012. *JAMA* 2014a;311(2):164–171. [PubMed: 24399555]
- IARC. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans Volume 109: Outdoor Air Pollution. International Agency for Research on Cancer (IARC), Lyon, France, 2016.
- IMPROVE. Interagency Monitoring of Protected Visual Environments. Available: <http://vista.cira.colostate.edu/Improve/>. Accessed: 29 January 2020 2019.
- Kalbfleisch JD, Prentice RL. The Statistical Analysis of Failure Time Data 2nd Edition, in: Shewhart WA, Wilks SS (Eds.), Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., Hoboken, New Jersey, 2002.
- Knoke JD, Shanks TG, Vaughn JW, Thun MJ, Burns DM. Lung cancer mortality is related to age in addition to duration and intensity of cigarette smoking: an analysis of CPS-I data. *Cancer Epidemiol Biomarkers Prev* 2004;13(6):949–957. [PubMed: 15184251]
- Laden F, Schwartz J, Speizer FE, Dockery DW. Reduction in fine particulate air pollution and mortality: Extended follow-up of the Harvard Six Cities study. *Am J Respir Crit Care Med* 2006;173(6):667–672. [PubMed: 16424447]
- Lubin JH, Caporaso N, Wichmann HE, Schaffrath-Rosario A, Alavanja MC. Cigarette smoking and lung cancer: modeling effect modification of total exposure and intensity. *Epidemiology* 2007;18(5):639–648. [PubMed: 17700253]
- Lumley T, Sheppard L. Assessing seasonal confounding and model selection bias in air pollution epidemiology using positive and negative control analyses. *Environmetrics* 2000;11(6):705–717.
- Meza R, Hazelton WD, Colditz GA, Moolgavkar SH. Analysis of lung cancer incidence in the Nurses' Health and the Health Professionals' Follow-Up Studies using a multistage carcinogenesis model. *Cancer Causes Control* 2008;19(3):317–328. [PubMed: 18058248]
- Moolgavkar S, Luebeck G. In press Multistage carcinogenesis: a unified framework for cancer data analysis In: Almudevar AL, Hall WJ, Oakes D, editors. *Statistical Modeling for Biological Systems*. Basel, Switzerland: Springer International Publishing, 115–133.
- Moolgavkar SH, Chang ET, Luebeck G, et al. Diesel engine exhaust and lung cancer mortality: time-related factors in exposure and risk. *Risk Anal* 2015;35(4):663–675. [PubMed: 25683254]
- Moolgavkar SH, Chang ET, Watson HN, Lau EC. An Assessment of the Cox Proportional Hazards Regression Model for Epidemiologic Studies. *Risk Anal* 2018;38(4):777–794. [PubMed: 29168991]
- Moolgavkar SH, Holford TR, Levy DT, et al. Impact of reduced tobacco smoking on lung cancer mortality in the United States during 1975–2000. *J Natl Cancer Inst* 2012;104(7):541–548. [PubMed: 22423009]
- National Cancer Institute. CISNET Publication Support and Modeling Resources. <https://resources.cisnet.cancer.gov/projects/>. Accessed 29 January 2020 2018.
- Peto J That the effects of smoking should be measured in pack-years: misconceptions 4. *Br J Cancer* 2012;107(3):406–407. [PubMed: 22828655]
- Peto R, 1977 Epidemiology, multistage models, and short-term mutagenicity tests, in: Hiatt HH, Watson JD, Winsten JA (Eds.), *Origins of Human Cancer*. Book C: Human Risk Assessment. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, pp. 1403–1428.
- Pope CA 3rd, Burnett RT, Krewski D, et al. Cardiovascular mortality and exposure to airborne fine particulate matter and cigarette smoke: shape of the exposure-response relationship. *Circulation* 2009;120(11):941–948. [PubMed: 19720932]
- Pope CA 3rd, Burnett RT, Thun MJ, et al. Lung cancer, cardiopulmonary mortality, and long-term exposure to fine particulate air pollution. *JAMA* 2002;287(9):1132–1141. [PubMed: 11879110]

- Puett RC, Hart JE, Yanosky JD, et al. Particulate matter air pollution exposure, distance to road, and incident lung cancer in the Nurses' Health Study cohort. *Environ Health Perspect* 2014;122(9):926–932. [PubMed: 24911062]
- Richardson DB. Lung cancer in chrysotile asbestos workers: analyses based on the two-stage clonal expansion model. *Cancer Causes Control* 2009;20(6):917–923. [PubMed: 19184474]
- Spengler JD, Koutrakis P, Dockery DW, Raizenne M, Speizer FE. Health effects of acid aerosols on North American children: air pollution exposures. *Environ Health Perspect* 1996;104(5):492–499. [PubMed: 8743436]
- Suh HH, Nishioka Y, Allen GA, Koutrakis P, Burton RM. The metropolitan acid aerosol characterization study: results from the summer 1994 Washington, D.C. field study. *Environ Health Perspect* 1997;105(8):826–834. [PubMed: 9347898]
- Tammemagi CM, Pinsky PF, Caporaso NE, et al. Lung cancer risk prediction: Prostate, Lung, Colorectal And Ovarian Cancer Screening Trial models and validation. *J Natl Cancer Inst* 2011;103(13):1058–1068. [PubMed: 21606442]
- Thomas DC. Models for exposure-time-response relationships with applications to cancer epidemiology. *Annu Rev Public Health* 1988;9(451–482). [PubMed: 3288239]
- Thomas DC, 2009 *Statistical Models in Environmental Epidemiology*. Oxford University Press, Oxford, UK.
- Thomas DC. Invited commentary: is it time to retire the “pack-years” variable? Maybe not! *Am J Epidemiol* 2014;179(3):299–302. [PubMed: 24355333]
- U.S. EPA. Reviewing National Ambient Air Quality Standards (NAAQS): Scientific and Technical Information. Available: <https://www.epa.gov/naaqs>. Last updated: 16 October 2017 2017.
- U.S. EPA. Air Quality System (AQS). Available: <https://www.epa.gov/aqs>. Last updated: 2018 November 2 2018.
- Vlaanderen J, Portengen L, Schuz J, et al. 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. *Am J Epidemiol* 2014;179(3):290–298. [PubMed: 24355332]
- Weuve J, Puett RC, Schwartz J, Yanosky JD, Laden F, Grodstein F. Exposure to particulate air pollution and cognitive decline in older women. *Arch Intern Med* 2012;172(3):219–227. [PubMed: 22332151]
- Yanosky JD, Paciorek CJ, Laden F, et al. Spatio-temporal modeling of particulate air pollution in the conterminous United States using geographic and meteorological predictors. *Environ Health* 2014;13(63). [PubMed: 25097007]
- Yanosky JD, Paciorek CJ, Schwartz J, Laden F, Puett R, Suh HH. Spatio-temporal modeling of chronic PM10 exposure for the Nurses' Health Study. *Atmos Environ (1994)* 2008;42(18):4047–4062. [PubMed: 19584946]
- Yanosky JD, Paciorek CJ, Suh HH. Predicting chronic fine and coarse particulate exposures using spatiotemporal models for the Northeastern and Midwestern United States. *Environ Health Perspect* 2009;117(4):522–529. [PubMed: 19440489]
- Zeger SL, Dominici F, McDermott A, Samet JM. Mortality in the Medicare population and chronic exposure to fine particulate air pollution in urban centers (2000–2005). *Environ Health Perspect* 2008;116(12):1614–1619. [PubMed: 19079710]

A.



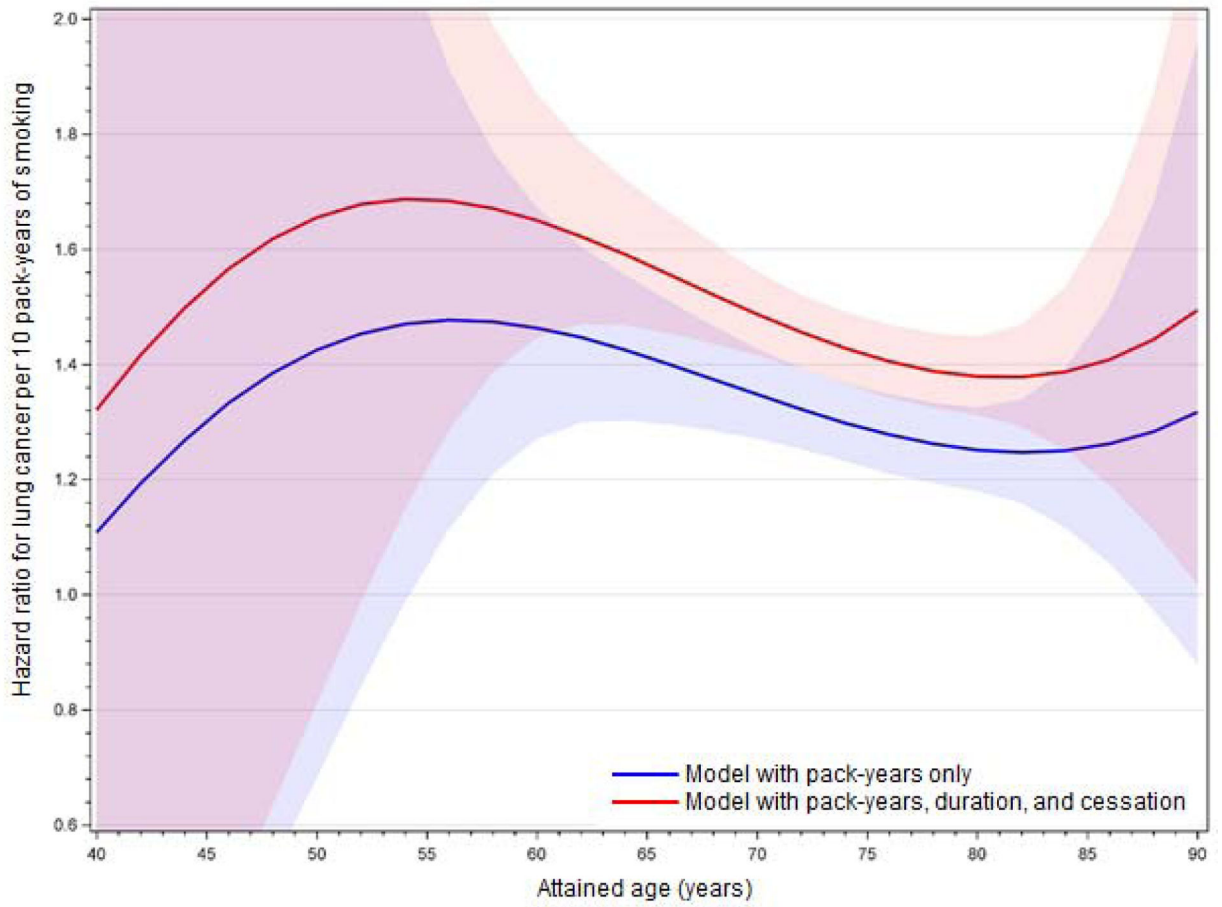
Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

B.



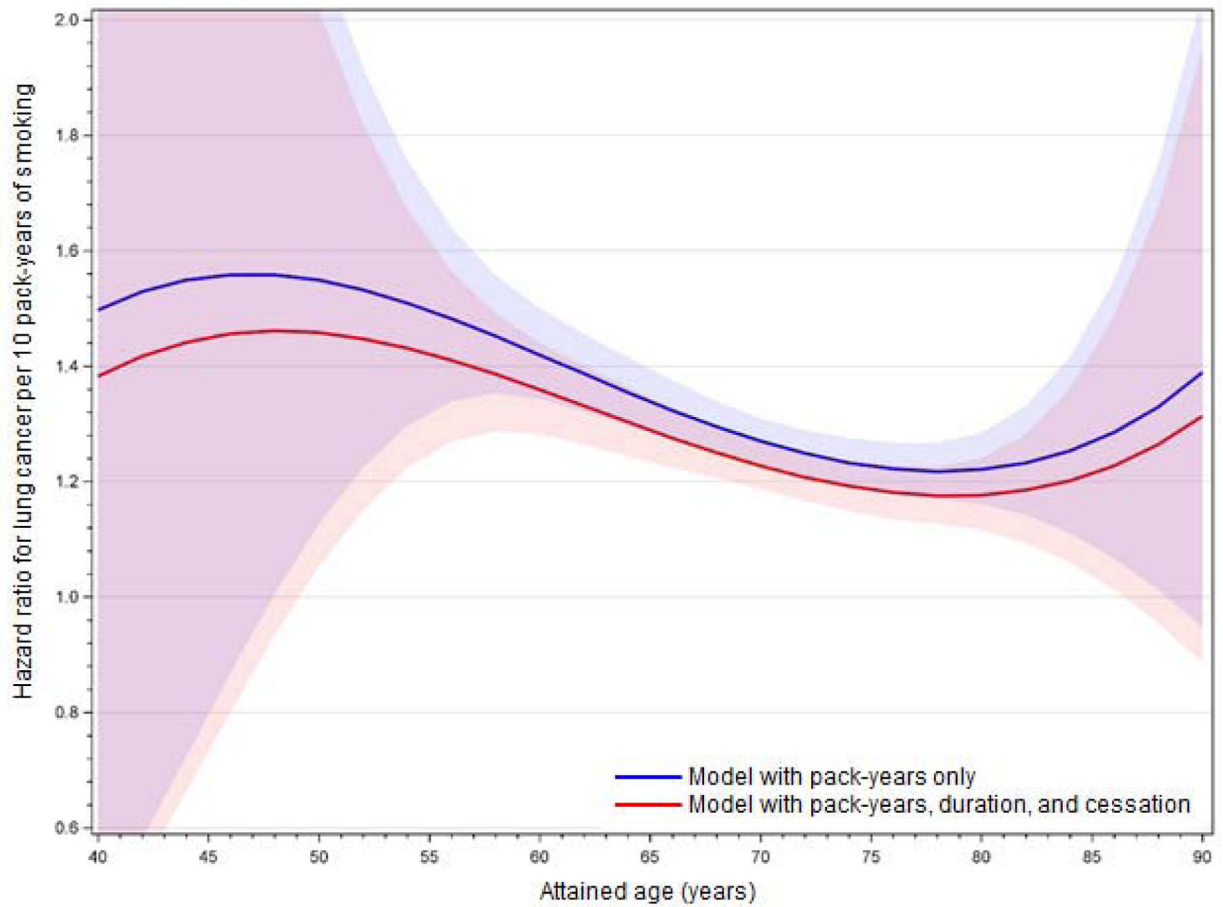
Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

C.

**Figures 1A–C.**

Age-related modification of cumulative pack-years in (A) the full cohort, (B) former smokers, and (C) current smokers, using fully adjusted models. Lines illustrate the hazard ratio per 10 pack-years of smoking, and shading illustrates 95% confidence intervals around the hazard ratio. Red denotes models including $PM_{2.5}$ and cumulative pack-years only (with age interactions), and blue denotes models including $PM_{2.5}$, cumulative pack-years (with age interactions), years of smoking, and months since cessation.

Smoking and air pollution characteristics of the Nurses' Health Study cohort at study baseline (1994) and during follow-up (1994–2010), stratified by smoking status at study baseline.

Table 1.

	Never smokers n = 44,245		Former smokers n = 39,061		Current smokers n = 12,784		Total cohort* n = 96,090	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Age at baseline (years)	60.5	7.3	60.7	7.0	59.7	6.9	60.5	7.1
during follow-up	67.1	7.2	67.2	6.8	65.6	6.6	67.0	7.0
Smoking duration at baseline (years)	-	-	21.0	12.8	39.0	8.5	-	-
during follow-up	0.0	0.1	21.4	13.0	43.5	8.7	-	-
Cumulative smoking at baseline (pack-years)	-	-	18.8	17.6	42.2	20.6	-	-
during follow-up	0.0	0.0	19.1	17.8	46.0	21.9	-	-
Average packs at baseline (pack-years / duration)	-	-	0.3	0.3	0.1	0.1	-	-
during follow-up	0.0	0.0	0.3	0.3	0.1	0.1	-	-
Age at smoking initiation (years)	-	-	19.3	4.2	19.1	4.0	-	-
Age at smoking cessation (years)	-	-	41.7	13.1	-	-	-	-
Time since smoking cessation at baseline (months)	-	-	228.7	143.2	-	-	-	-
PM _{2.5} at baseline (µg/m ³)	15.2	3.3	15.0	3.2	15.2	3.1	15.1	3.2
during follow-up	13.2	2.8	13.0	2.7	13.4	2.7	13.2	2.7
PM ₁₀ at baseline (µg/m ³)	25.8	6.8	25.3	6.3	25.3	6.2	25.5	6.5
during follow-up	21.9	5.6	21.5	5.3	21.8	5.3	21.7	5.4
PM _{2.5-10} at baseline (µg/m ³)	10.6	5.0	10.3	4.6	10.1	4.5	10.4	4.7
during follow-up	8.6	4.1	8.5	3.8	8.5	3.8	8.5	4.0
	n	%	n	%	n	%	n	%
Geographic region at baseline								
Northeast	21,695	49%	21,204	54%	7,224	57%	50,123	52%
Midwest	8,779	20%	6,257	16%	2,178	17%	17,214	18%
West	6,254	14%	5,110	13%	1,288	10%	12,652	13%
South	7,517	17%	6,490	17%	2,094	16%	16,101	17%

* The total eligible cohort of 97,303 women is larger than the cohort of 96,090 with known smoking status at baseline in 1994, due to the availability of smoking data from an additional 1,213 women during follow-up.

PM: particulate matter
SD: standard deviation

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Model fit and associations between fine particulate matter and smoking characteristics with lung cancer incidence in the Nurses' Health Study, adjusted for age, time period, and geographic region only (minimally adjusted models) or additional covariates (fully adjusted models), with various classifications of smoking based on cumulative pack-years.

Table 2.

AIC, minimally adj.	AIC, fully adj.	Covariate	Minimally adjusted ^a		Fully adjusted ^b	
			HR	95% CI	HR	95% CI
21,796.11	21,096.64	PM _{2.5} (per 10 µg/m ³)	1.02	0.87 – 1.21	1.07	0.90 – 1.26
19,734.78	19,610.10	PM _{2.5} (per 10 µg/m ³)	1.06	0.90 – 1.24	1.06	0.90 – 1.25
		Current v. never smoker	25.85	21.91 – 30.51	19.25	16.18 – 22.91
		Former v. never smoker	6.45	5.50 – 7.55	5.80	4.93 – 6.82
18,721.24	18,705.14	PM _{2.5} (per 10 µg/m ³)	1.07	0.91 – 1.27	1.09	0.92 – 1.28
		Current v. never smoker	4.08	3.32 – 5.01	3.65	2.96 – 4.50
		Former v. never smoker	2.62	2.21 – 3.11	2.52	2.11 – 3.00
		Smoking pack-years	1.03	1.03 – 1.03	1.03	1.03 – 1.03
18,550.32	18,549.61	PM _{2.5} (per 10 µg/m ³)	1.05	0.89 – 1.23	1.06	0.89 – 1.25
		Current v. never smoker	1.20	0.91 – 1.59	1.14	0.86 – 1.51
		Former v. never smoker	1.14	0.91 – 1.43	1.12	0.90 – 1.41
		Smoking pack-years	1.02	1.02 – 1.02	1.02	1.02 – 1.02
		Smoking duration (years)	1.04	1.03 – 1.04	1.04	1.03 – 1.04
18,545.49	18,545.16	PM _{2.5} (per 10 µg/m ³)	1.05	0.89 – 1.23	1.06	0.89 – 1.25
		Current v. never smoker	2.06	1.27 – 3.33	1.92	1.18 – 3.12
		Former v. never smoker	2.03	1.26 – 3.28	1.97	1.22 – 3.19
		Smoking pack-years	1.02	1.02 – 1.02	1.02	1.02 – 1.02
		Smoking duration (years)	1.03	1.02 – 1.04	1.02	1.01 – 1.03
		Smoking cessation (months)	1.00	1.00 – 1.00	1.00	1.00 – 1.00
18,549.88	18,548.57	PM _{2.5} (per 10 µg/m ³)	1.04	0.89 – 1.23	1.05	0.89 – 1.25
		Smoking pack-years	1.02	1.02 – 1.02	1.02	1.02 – 1.02
		Smoking duration (years)	1.04	1.03 – 1.04	1.04	1.03 – 1.04
		Smoking cessation (months)	1.00	1.00 – 1.00	1.00	1.00 – 1.00
18,509.50	18,491.76	PM _{2.5} (per 10 µg/m ³)	1.05	0.89 – 1.24	1.06	0.90 – 1.25

AIC, minimally adj.	AIC, fully adj.	Covariate	Minimally adjusted ^a			Fully adjusted ^b		
			HR	95% CI	HR	95% CI	HR	95% CI
18,511.49	18,493.76	Smoking pack-years	1.08	1.06 – 1.10	1.08	1.06 – 1.09	1.08	1.06 – 1.09
		Smoking duration (years)	1.04	1.04 – 1.04	1.04	1.03 – 1.04	1.04	1.03 – 1.04
		Smoking cessation (months)	1.00	1.00 – 1.00	1.00	1.00 – 1.00	1.00	1.00 – 1.00
		Pack-years*age	1.00	1.00 – 1.00	1.00	1.00 – 1.00	1.00	1.00 – 1.00
		PM _{2.5} (per 10 µg/m ³)	1.05	0.89 – 1.24	1.06	0.90 – 1.25	1.06	0.90 – 1.25
		Smoking pack-years	1.08	0.95 – 1.24	1.08	0.94 – 1.23	1.08	0.94 – 1.23
		Smoking duration (years)	1.04	1.04 – 1.04	1.04	1.03 – 1.04	1.04	1.03 – 1.04
		Smoking cessation (months)	1.00	1.00 – 1.00	1.00	1.00 – 1.00	1.00	1.00 – 1.00
		Pack-years*age	1.00	1.00 – 1.00	1.00	1.00 – 1.00	1.00	1.00 – 1.00
		Pack-years*age squared	1.00	1.00 – 1.00	1.00	1.00 – 1.00	1.00	1.00 – 1.00
18,512.86	18,495.19	PM _{2.5} (per 10 µg/m ³)	1.05	0.89 – 1.24	1.06	0.90 – 1.25	1.06	0.90 – 1.25
		Smoking pack-years	0.74	0.29 – 1.90	0.75	0.29 – 1.92	0.75	0.29 – 1.92
		Smoking duration (years)	1.04	1.04 – 1.04	1.04	1.03 – 1.04	1.04	1.03 – 1.04
		Smoking cessation (months)	1.00	1.00 – 1.00	1.00	1.00 – 1.00	1.00	1.00 – 1.00
		Pack-years*age	1.02	0.98 – 1.06	1.02	0.97 – 1.06	1.02	0.97 – 1.06
		Pack-years*age squared	1.00	1.00 – 1.00	1.00	1.00 – 1.00	1.00	1.00 – 1.00
		Pack-years*age cubed	1.00	1.00 – 1.00	1.00	1.00 – 1.00	1.00	1.00 – 1.00
		PM _{2.5} (per 10 µg/m ³)	1.08	0.92 – 1.27	1.10	0.93 – 1.30	1.10	0.93 – 1.30
		Smoking pack-years	0.75	0.32 – 1.74	0.77	0.33 – 1.80	0.77	0.33 – 1.80
		Pack-years*age	1.02	0.98 – 1.05	1.02	0.98 – 1.05	1.02	0.98 – 1.05
18,878.51	18,810.61	Pack-years*age squared	1.00	1.00 – 1.00	1.00	1.00 – 1.00	1.00	1.00 – 1.00
		Pack-years*age cubed	1.00	1.00 – 1.00	1.00	1.00 – 1.00	1.00	1.00 – 1.00
		PM _{2.5} (per 10 µg/m ³)	1.08	0.92 – 1.27	1.10	0.93 – 1.30	1.10	0.93 – 1.30
		Smoking pack-years	0.75	0.32 – 1.74	0.77	0.33 – 1.80	0.77	0.33 – 1.80
		Pack-years*age	1.02	0.98 – 1.05	1.02	0.98 – 1.05	1.02	0.98 – 1.05
		Pack-years*age squared	1.00	1.00 – 1.00	1.00	1.00 – 1.00	1.00	1.00 – 1.00
		Pack-years*age cubed	1.00	1.00 – 1.00	1.00	1.00 – 1.00	1.00	1.00 – 1.00
		PM _{2.5} (per 10 µg/m ³)	1.08	0.92 – 1.27	1.10	0.93 – 1.30	1.10	0.93 – 1.30
		Smoking pack-years	0.75	0.32 – 1.74	0.77	0.33 – 1.80	0.77	0.33 – 1.80
		Pack-years*age	1.02	0.98 – 1.05	1.02	0.98 – 1.05	1.02	0.98 – 1.05

^a Stratified by time period and age in months (in absence of age interaction), and adjusted for geographic region; also adjusted for age in months (in presence of age interaction)

^b Stratified by time period and age in months (in absence of age interaction), and adjusted for geographic region, body mass index, alcohol consumption, physical activity, overall diet quality, census tract median house value, census tract median income, and secondhand smoke exposure at home, at work, and during childhood; also adjusted for age in months (in presence of age interaction)

AIC: Akaike information criterion; CI: confidence interval; HR: hazard ratio; PM_{2.5}: particulate matter < 2.5 µm in aerodynamic diameter

Model fit and associations between fine particulate matter and smoking characteristics with lung cancer incidence in the Nurses' Health Study, adjusted for age, time period, and geographic region only (minimally adjusted models) or additional covariates (fully adjusted models), with various classifications of smoking based on average packs.

Table 3.

AIC, minimally adj.	AIC, fully adj.	Covariate	Minimally adjusted ^a		Fully adjusted ^b	
			HR	95% CI	HR	95% CI
21,796.11	21,096.64	PM _{2.5} (per 10 µg/m ³)	1.02	0.87 – 1.21	1.07	0.90 – 1.26
19,734.78	19,610.10	PM _{2.5} (per 10 µg/m ³)	1.06	0.90 – 1.24	1.06	0.90 – 1.25
		Current v. never smoker	25.91	21.96 – 30.57	19.25	16.18 – 22.91
		Former v. never smoker	6.45	5.50 – 7.55	5.80	4.93 – 6.82
18,872.12	18,859.40	PM _{2.5} (per 10 µg/m ³)	1.02	0.86 – 1.20	1.02	0.87 – 1.21
		Current v. never smoker	0.74	0.54 – 1.01	0.70	0.51 – 0.97
		Former v. never smoker	0.68	0.53 – 0.88	0.68	0.52 – 0.87
		Smoking average packs	2.74	2.08 – 3.60	2.75	2.08 – 3.63
		Smoking duration (years)	1.07	1.07 – 1.08	1.07	1.06 – 1.07
18,870.31	18,858.02	PM _{2.5} (per 10 µg/m ³)	1.01	0.86 – 1.19	1.02	0.87 – 1.20
		Current v. never smoker	1.09	0.66 – 1.77	1.01	0.62 – 1.66
		Former v. never smoker	1.03	0.64 – 1.67	1.00	0.61 – 1.62
		Smoking average packs	2.81	2.14 – 3.70	2.82	2.14 – 3.73
		Smoking duration (years)	1.06	1.05 – 1.07	1.06	1.05 – 1.07
		Smoking cessation (months)	1.00	1.00 – 1.00	1.00	1.00 – 1.00
18,866.94	18,854.06	PM _{2.5} (per 10 µg/m ³)	1.01	0.86 – 1.20	1.02	0.87 – 1.20
		Smoking average packs	2.89	2.22 – 3.76	2.84	2.17 – 3.71
		Smoking duration (years)	1.06	1.06 – 1.07	1.06	1.06 – 1.06
		Smoking cessation (months)	1.00	1.00 – 1.00	1.00	1.00 – 1.00
18,868.59	18,855.84	PM _{2.5} (per 10 µg/m ³)	1.02	0.86 – 1.20	1.02	0.87 – 1.20
		Smoking average packs	1.57	0.21 – 11.83	1.74	0.22 – 13.53
		Smoking duration (years)	1.06	1.06 – 1.07	1.06	1.06 – 1.06
		Smoking cessation (months)	1.00	1.00 – 1.00	1.00	1.00 – 1.00
		Ave. packs*age	1.01	0.98 – 1.04	1.01	0.98 – 1.04

AIC, minimally adj.	AIC, fully adj.	Covariate	Minimally adjusted ^a			Fully adjusted ^b		
			HR	95% CI	HR	95% CI	HR	95% CI
18,869.28	18,856.40	PM _{2.5} (per 10 µg/m ³)	1.02	0.86 – 1.20	1.02	0.87 – 1.21		
		Smoking average packs	25.51	0.15 – 4427	32.35	0.18 – 5710		
		Smoking duration (years)	1.06	1.06 – 1.07	1.06	1.06 – 1.06		
		Smoking cessation (months)	1.00	1.00 – 1.00	1.00	1.00 – 1.00		
		Ave. packs*age	0.93	0.80 – 1.08	0.92	0.79 – 1.07		
		Ave. packs*age squared	1.00	1.00 – 1.00	1.00	1.00 – 1.00		
18,870.50	18,857.65	PM _{2.5} (per 10 µg/m ³)	1.02	0.86 – 1.20	1.02	0.87 – 1.21		
		Smoking average packs	4.99	0.01 – 3670	6.21	0.01 – 4969		
		Smoking duration (years)	1.06	1.06 – 1.07	1.06	1.06 – 1.06		
		Smoking cessation (months)	1.00	1.00 – 1.00	1.00	1.00 – 1.00		
		Ave. packs*age	0.90	0.77 – 1.06	0.90	0.76 – 1.05		
		Ave. packs*age squared	1.00	1.00 – 1.01	1.00	1.00 – 1.01		
		Ave. packs*age cubed	1.00	1.00 – 1.00	1.00	1.00 – 1.00		

^a Stratified by time period and age in months (in absence of age interaction), and adjusted for geographic region; also adjusted for age in months (in presence of age interaction)

^b Stratified by time period and age in months (in absence of age interaction), and adjusted for geographic region, body mass index, alcohol consumption, physical activity, overall diet quality, census tract median house value, census tract median income, and secondhand smoke exposure at home, at work, and during childhood; also adjusted for age in months (in presence of age interaction)

AIC: Akaike information criterion; CI: confidence interval; HR: hazard ratio; PM_{2.5}: particulate matter < 2.5 µm in aerodynamic diameter

Stratified associations between fine particulate matter and smoking characteristics with lung cancer incidence in the Nurses' Health Study, adjusted for age, time period, and geographic region only (minimally adjusted models) or additional covariates (fully adjusted models), with various classifications of smoking based on cumulative pack-years or average packs, stratified by smoking status during follow-up, 1994–2010.

Table 4.

Stratum	Covariate	Minimally adjusted ^d		Fully adjusted ^b	
		HR	95% CI	HR	95% CI
Never smokers	PM _{2.5} (per 10 µg/m ³)	1.24	0.74 – 2.06	1.27	0.76 – 2.12
	PM _{2.5} (per 10 µg/m ³)	1.35	1.00 – 1.81	1.35	1.00 – 1.82
Former smokers	Smoking pack-years	1.08	1.04 – 1.13	1.09	1.04 – 1.13
	Smoking duration (years)	1.02	1.01 – 1.04	1.02	1.00 – 1.04
	Smoking cessation (months)	1.00	1.00 – 1.00	1.00	1.00 – 1.00
	Pack-years*age	1.00	1.00 – 1.00	1.00	1.00 – 1.00
	PM _{2.5} (per 10 µg/m ³)	1.27	0.95 – 1.71	1.27	0.95 – 1.71
	Smoking average packs	1.85	1.25 – 2.74	1.92	1.29 – 2.85
Current or recent former smokers	Smoking duration (years)	1.07	1.05 – 1.08	1.06	1.04 – 1.08
	Smoking cessation (months)	1.00	1.00 – 1.00	1.00	1.00 – 1.00
	PM _{2.5} (per 10 µg/m ³)	0.92	0.74 – 1.15	0.94	0.75 – 1.17
	Smoking pack-years	1.07	1.05 – 1.10	1.07	1.04 – 1.10
	Smoking duration (years)	1.03	1.02 – 1.05	1.03	1.02 – 1.05
	Smoking cessation (months)	1.00	1.00 – 1.00	1.00	1.00 – 1.00
Former smokers	Pack-years*age	1.00	1.00 – 1.00	1.00	1.00 – 1.00
	PM _{2.5} (per 10 µg/m ³)	0.90	0.73 – 1.12	0.92	0.74 – 1.14
	Smoking average packs	43.01	24.37 – 75.92	42.72	24.01 – 75.99
	Smoking duration (years)	1.09	1.08 – 1.11	1.09	1.07 – 1.11
	Smoking cessation (months)	1.00	1.00 – 1.00	1.00	1.00 – 1.00

^aStratified by time period and age in months (in absence of age interaction), and adjusted for geographic region; also adjusted for age in months (in presence of age interaction)

^bStratified by time period and age in months (in absence of age interaction), and adjusted for geographic region, body mass index, alcohol consumption, physical activity, overall diet quality, census tract median house value, census tract median income, and secondhand smoke exposure at home, at work, and during childhood; also adjusted for age in months (in presence of age interaction)

AIC: Akaike information criterion; CI: confidence interval; HR: hazard ratio; PM_{2.5}: particulate matter < 2.5 µm in aerodynamic diameter