



Practice of Epidemiology

Exposure-Lag-Response in Longitudinal Studies: Application of Distributed-Lag Nonlinear Models in an Occupational Cohort

Andreas M. Neophytou*, Sally Picciotto, Daniel M. Brown, Lisa E. Gallagher, Harvey Checkoway, Ellen A. Eisen, and Sadie Costello

* Correspondence to Dr. Andreas M. Neophytou, Division of Environmental Health Sciences, School of Public Health, University of California, Berkeley, 50 University Hall #7360, Berkeley, CA 94720-7360 (e-mail: aneophytou@berkeley.edu).

Initially submitted September 5, 2017; accepted for publication January 25, 2018.

Prolonged exposures can have complex relationships with health outcomes, as timing, duration, and intensity of exposure are all potentially relevant. Summary measures such as cumulative exposure or average intensity of exposure may not fully capture these relationships. We applied penalized and unpenalized distributed-lag nonlinear models (DLNMs) with flexible exposure-response and lag-response functions in order to examine the association between crystalline silica exposure and mortality from lung cancer and nonmalignant respiratory disease in a cohort study of 2,342 California diatomaceous earth workers followed during 1942–2011. We also assessed associations using simple measures of cumulative exposure assuming linear exposure-response and constant lag-response. Measures of association from DLNMs were generally higher than those from simpler models. Rate ratios from penalized DLNMs corresponding to average daily exposures of 0.4 mg/m³ during lag years 31–50 prior to the age of observed cases were 1.47 (95% confidence interval (CI): 0.92, 2.35) for lung cancer mortality and 1.80 (95% CI: 1.14, 2.85) for nonmalignant respiratory disease mortality. Rate ratios from the simpler models for the same exposure scenario were 1.15 (95% CI: 0.89, 1.48) and 1.23 (95% CI: 1.03, 1.46), respectively. Longitudinal cohort studies of prolonged exposures and chronic health outcomes should explore methods allowing for flexibility and nonlinearities in the exposure-lag-response.

chronic disease; cohort studies; distributed-lag nonlinear models; exposure-lag-response; longitudinal studies; silica

Abbreviations: AIC, Akaike's Information Criterion; CI, confidence interval; DLNM, distributed-lag nonlinear model; NMRD, nonmalignant respiratory disease.

Long-term occupational exposure to crystalline silica has been associated with lung cancer and nonmalignant respiratory disease (NMRD) (1–5). However, little is understood about the relative importance of intensity, duration, and timing of exposure in relation to these diseases. Studies with quantitative exposure information most commonly use a cumulative exposure metric that bundles duration with intensity. Exposure metrics for average exposure remove the influence of duration, but they also lack information regarding the time-varying intensity of exposure. Duration of exposure, another common metric, does not include any quantitative information about exposure intensity, other than the minimum intensity required for classification as exposed. While cumulative exposure, average exposure, and duration of exposure are all attractive metrics in terms of implementation and interpretation, none take full advantage of what

are often rich work-history data sets with temporal resolution on the annual level.

Lung cancer and NMRD are characterized by latencies, in that both outcomes usually occur a long time after initial silica exposures (6–8). Latency periods for lung cancer are taken into account by lagging cumulative or average exposure, as is often done in studies of cancer outcomes in occupational epidemiology (9). Recent studies of lung cancer mortality with quantitative silica exposure have presented results with no lag (4, 10–13) and also for lags of 10 (4, 10, 11), 15 (4, 11–13), and 25 (12) years. Studies of NMRD mortality typically present results from unlagged analysis only (10, 14–16). However, in 2 recent papers in which results from lagged exposure analyses were presented, the associations between silica exposure and NMRD were strengthened with 10- (4) and 15- (4, 13) year lags, indicating that

timing of exposure may be an important consideration when assessing associations between silica and both malignant and non-malignant respiratory disease.

To explore the relative contributions of the intensity, duration, and timing of crystalline silica exposure to lung cancer and NMRD, we analyzed data from a cohort of California diatomaceous earth workers (3) using distributed-lag nonlinear models (DLNMs). In a recent study with applied examples of DLNMs, including an example using longitudinal occupational cohort data, Gasparrini (17) summarized the temporal relationships between exposure and risk of a health outcome as the “exposure-lag-response.” DLNMs allow for a nonlinear exposure-response, as well as a nonlinear lag-response, and they can use the participant’s full work history to estimate the exposure-lag-response. By comparing the results from exposure scenarios in which one aspect of intensity, duration, and timing of exposure is held constant while the others vary, the health effects of these 3 moving parts of protracted exposure can be disentangled. The method also allows for nonlinearities, which have been reported in relation to exposure-response in this cohort (18). In the current study, we aimed to parse out the complicated aspects of silica exposure with regard to both outcomes by using DLNMs as an estimation approach.

METHODS

Study population

Analyses were performed on data from a cohort mortality study of diatomaceous earth workers; the cohort is described in greater detail elsewhere (3). Briefly, the cohort consisted of 2,342 male workers from 2 diatomaceous earth plants in Lompoc, California. Inclusion criteria were cumulative employment for at least 1 year at either plant and having worked for at least 1 day between January 1, 1942, and December 31, 1987. Work histories and silica exposure assessments were available from the beginning of plant operations (1902 and 1946 for the 2 plants, respectively) through 1994. Mortality follow-up extended from January 1, 1942, to December 31, 2011, for a maximum follow-up time of 70 years, and was based on National Death Index data, state driver’s license bureaus, and commercial credit bureaus (4). Complete mortality follow-up was not available for 183 participants. These subjects contributed person-time until their last observed dates of employment. Demographic information on the cohort included work history data (year of hire, duration of employment at study sites, and dates of specific jobs held) and ethnicity. Information on smoking status (ever/never) was also available for 50% of the cohort ($n = 1,171$). Deaths due to lung cancer or NMRD were classified using the underlying cause of death according to *International Classification of Diseases* (Revisions 5–10) codes, as determined in the National Institute for Occupational Safety and Health’s Life Table Analysis System mortality program (19, 20).

Exposure assessment

Quantitative estimates of silica dust exposure were determined primarily from industrial air monitoring measurements made between 1962 and 1988, with archived company data providing some additional information for the period between 1948 and 1962 (21). Job-specific estimates of respirable crystalline silica concentrations (mg/m^3), mostly in the form of cristobalite, and

respirable dust exposure were generated on the basis of available measurements. Exposures incurred before 1948 were based on extrapolated job-specific exposures that accounted for interventions to reduce dust exposure and other changes over time (1). The estimates for crystalline silica were derived from the percentage of silica contained in a given diatomaceous earth product and the amount of time exposed to that product for a given job. Detailed work history information was available through 1994, by which time 88% of the cohort had left work (1, 21).

Because of 2 small operations involving chrysotile asbestos in the plants that occurred during the study period, asbestos exposures (fibers/mL) were also derived. Estimates were based on monitoring data and records of quantities of asbestos in mixed products from 1930 onwards, and extrapolated data were used to determine exposures for earlier years (1). Approximately 54% of cohort participants were exposed to asbestos at some point in their work history. In the current study, age was the time scale of interest, so exposures for each age (in 1-year periods) were estimated for both respirable crystalline silica and asbestos. The age-specific exposure values were estimated as a time-weighted sum of the time-varying, job-specific exposure values based on the job(s) held at different ages during employment for each participant. Age-specific exposures incurred during age periods without active employment were set to zero.

Statistical analyses

Cross-basis functions. Application of DLNMs relies on “cross-basis” functions, as described in more detail by Armstrong (22) and Gasparrini (17). Briefly, in a longitudinal cohort setting, derived exposure variables for each person-year i are generated based on nonlinear functions (such as spline functions) for the exposure and lag-response. These can be envisioned with a general notation of

$$\sum_{l=0}^L l_p x_{b,i-l},$$

where x_b is a basis function for the exposure of interest (e.g., a natural cubic spline with b degrees of freedom) and, similarly, l_p is a basis function for the lag (where $l = 0, \dots, L$). Thus, these derived exposure variables are the sum of products of the basis functions for exposure and lag, or “cross-basis” functions. For each person-year, the number of exposure variables derived is equal to the product of the degrees of freedom $b \times p$ of the basis functions for the exposure and lag-response. These cross-basis-derived variables (or the cross-basis exposure matrix) are entered into regression models, and the resulting coefficients and their standard errors can collectively be used to summarize the exposure-lag-response.

Unpenalized DLNMs. In the current study, the cross-basis exposure variables were derived from an initial exposure matrix consisting of annual exposures corresponding to lag $l = 1, \dots, 50$ for each person-year. We applied DLNMs using various combinations of exposure-response and lag-response functions, including a linear term for exposure-response, a constant lag-response, categorical terms for both lag- and exposure-response functions, and natural cubic splines for both functions. For applications of unpenalized DLNMs, Cox proportional hazards

models using age as the time axis were fitted. Analyses were performed in separate nested case-control samples for each of the 2 outcomes considered. Risk sets were created based on the failure times (measured as age in years) of the cases and a maximum number of time-matched eligible controls to restrict follow-up to a manageable number of observations (maximum of 300 controls per lung cancer case and 200 controls per NMRD case, resulting in similar numbers of total person-time observations in each nested case-control sample). Cross-basis exposure matrices were created (as described above) for each person-time observation comprising each of the created risk sets. Model fit for unpenalized DLNMs was assessed on the basis of Akaike's Information Criterion (AIC).

Penalized DLNMs. We also applied penalized DLNMs, as described in more detail by Gasparini et al. (23). In the case of the penalized DLNMs, the cross-basis matrix for exposure was created on the basis of penalized cubic regression splines within generalized additive model frameworks (24). A highly parameterized cross-basis exposure matrix based on natural cubic spline functions for the exposure and lag-response was created, and in addition, a penalty matrix was also defined. The penalties smooth the overall functions to avoid overfitting and effectively reduce the total number of degrees of freedom, with the degree of smoothness decided by maximizing the penalized log-likelihood of the fitted generalized additive model (23, 24). We varied the degree of penalization in the lag dimension of the exposure-lag-response by introducing additional penalties for more recent ($l = 1, \dots, 10$) and earlier ($l = 40, \dots, 50$) lags. These additional penalties correspond to specific assumptions about the nature of the lag-response, such as a diminished association or no association at more recent lags due to a latency of the outcome and diminished associations at earlier lags given survival since last exposure. For the penalized DLNMs, generalized additive Poisson models (approximating Cox proportional hazards models) were fitted, separately for each outcome of interest (lung cancer and NMRD mortality). We also repeated analyses using a pooled logistic model to approximate hazard ratios, by fitting generalized additive binomial models, as opposed to the additive Poisson models described above.

Cross-basis exposure matrices were entered into the regression models along with other covariates. All models were adjusted for calendar time (as a linear term), an indicator variable for Hispanic ethnicity, a categorical variable for smoking (ever smoking, never smoking, or missing data), and a cubic spline for baseline risk with age in the case of generalized additive Poisson or binomial models. We also controlled for asbestos exposure by modeling asbestos as a cross-basis matrix created using a natural cubic spline for the exposure-response and a 3-level category for the lag-response. All analyses were performed in R, version 3.3.3 (R Foundation for Statistical Computing, Vienna, Austria) using the "dlnm" package (25).

Measures of association under each model considered were summarized by predicting the association between different exposure scenarios (defining both exposure and lag values) and outcomes, as compared with a referent exposure value. Here we generated measures of association for various exposure scenarios, allowing for comparisons for different values of intensity, timing, or duration of exposure, with a common reference value of zero

(unexposed at all times). Figure 1 presents examples of such different exposure scenarios in a hypothetical open cohort, with values for intensity and duration of exposure within the ranges observed in the current study.

RESULTS

Summary statistics for the study population are presented in Table 1. The median duration of employment at the participating facilities was 5 years (range, 1–50), while the median duration of follow-up was 39 years (range, 1–70). The median age for lung cancer deaths was 66 years (interquartile range, 59–74), while the median age for NMRD deaths was 73 years (interquartile range, 64–79).

The complete list of unpenalized DLNMs considered, with corresponding AIC values, is presented in Table 2. Model fit based on the AIC suggested that the best-fitting unpenalized DLNMs were models with a constant lag-response and a natural spline for the exposure-response for lung cancer and natural splines from both the lag-response and exposure-response for NMRD. Models with categorical terms tended to have the worst fit according to the AIC.

Estimates of hazard ratios for mortality under different exposure scenarios (including the scenarios in Figure 1) obtained from selected models and from penalized DLNMs are shown in Tables 3 and 4 for lung cancer and NMRD, respectively. Specifically, we present results from 1) a "naive" model assuming a constant lag-response and a linear exposure-response (equivalent to a simple cumulative exposure metric); 2) models with categorical terms for the lag-response (with 10-year categories) and a natural spline for the exposure-response, as representatives of a more detailed "standard" analysis that considers lags; 3) models with natural spline terms for both the lag-response and the exposure-response; and, lastly, 4) penalized DLNMs, with both the lag-response and the exposure-response based on penalized cubic regression splines. In the case of penalized DLNMs, rate ratios from generalized additive Poisson models are presented in Tables 3 and 4 instead of hazard ratios. Estimates from generalized additive binomial models (results not shown) were similar to those from Poisson models in terms of the shape of the lag-response and exposure-response curves and in the quantitative measures of association for the different exposure scenarios considered.

Results from the simple "naive" models tended to underestimate associations for both outcomes as compared with the alternatives. For example, the hazard ratio for lung cancer mortality corresponding to 20 years of exposure (lag 31–50 years) at 0.40 mg/m^3 per year, resulting in 8 mg/m^3 -years of cumulative exposure from the simple model assuming a linear exposure-response and a constant lag-response, was 1.15 (95% confidence interval (CI): 0.89, 1.48), while the rate ratio from the penalized DLNM for the same exposure scenario was 1.47 (95% CI: 0.92, 2.35). With NMRD mortality as the outcome of interest, the hazard ratio corresponding to 20 years of exposure (lag 31–50 years) at 0.40 mg/m^3 per year from the simpler model was 1.23 (95% CI: 1.03, 1.46), while the rate ratio from the penalized DLNM corresponding to the same exposure scenario was 1.80

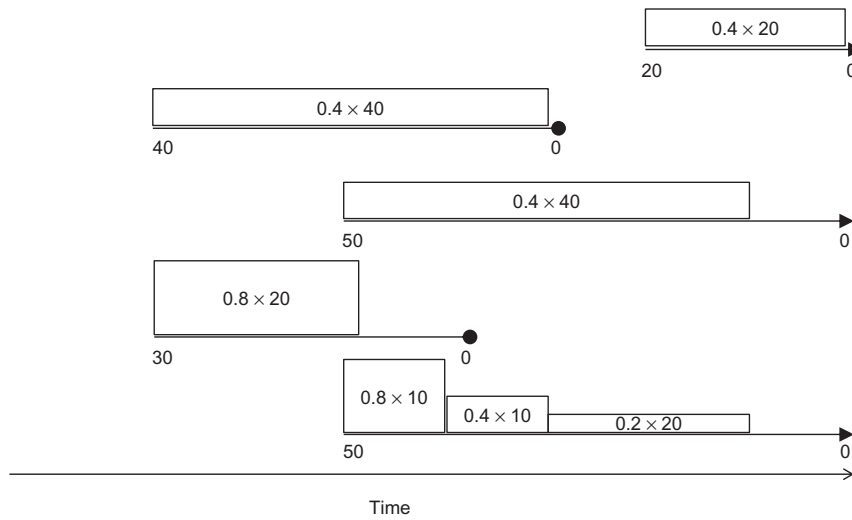


Figure 1. Representation of different exposure scenarios over time for different participants in a longitudinal cohort study. Each line represents a hypothetical cohort participant, with the lag dimension labeled on each line (with increasing numbers in the opposite direction of the follow-up) and the participant’s exposure represented by the rectangle(s) above each line. The height of each rectangle represents the intensity of exposure proportional to the decimal numeral inside the rectangle (e.g., in mg/m^3), and the width represents the duration proportional to the integer number inside the rectangle (in years), while the product of the two represents the cumulative exposure for each rectangle. Circles represent participants who are censored after experiencing an outcome of interest, and arrows represent participants who are still at risk at the administrative end of follow-up.

(95% CI: 1.14, 2.85). Results from penalized DLNMs also tended to result in less extreme values and narrower confidence intervals compared with models with a categorical lag-response or simple natural splines for the lag-response. Results from the

penalized DLNMs also maintained associations above the null for the entire range of lag and exposure.

Figure 2 shows a 3-dimensional representation of exposure-lag-response from penalized DLNMs for lung cancer (Figure 2A)

Table 1. Characteristics of a California Cohort of 2,342 Male Diatomaceous Earth Workers Exposed to Crystalline Silica and Followed for Mortality Between 1942 and 2011

Characteristic	No. of Men	%	Median (Range)	Mean (SD)
Hispanic ethnicity	546	23.3		
Ever smoking (yes) ^a	861	73.5		
Age at beginning of follow-up, years			27 (17–61)	
Year of hire			1952 (1908–1986)	
Year of birth			1927 (1881–1966)	
Duration of employment, years			5 (1–50)	
Duration of follow-up, years			39 (1–70)	
Mortality during follow-up				
All deaths	1,219	52.0		
Lung cancer deaths	113	4.8		
NMRD deaths	165	7.0		
Cumulative silica exposure, mg/m^3 -years ^b				2.15 (3.51)
Yearly silica exposure, mg/m^3 ^c				0.15 (0.21)
Cumulative asbestos exposure, fibers/mL-years ^b				1.40 (4.36)

Abbreviations: NMRD, nonmalignant respiratory disease; SD, standard deviation.

^a Smoking data were available for 1,171 participants. Number and percentage of ever smokers were based on this subset of participants.

^b Cumulative exposure at the end of follow-up.

^c Yearly silica exposure statistics were based on actively employed person-time only.

Table 2. Response Functions and AIC Values for Unpenalized Models Considered for Estimation of Lung Cancer and Nonmalignant Respiratory Disease Mortality According to Occupational Crystalline Silica Exposure in a Diatomaceous Earth Cohort, California, 1942–2011

Exposure-Response	Model		AIC	
	Lag-Response	df ^a	Lung Cancer	NMRD
Linear	Constant	1	1,254.0	1,777.1
Linear	Piecewise-constant ^b	4	1,259.8	1,776.5
Linear	Spline ^c	3	1,257.2	1,770.3
Piecewise-constant	Piecewise-constant	12	1,266.4	1,780.4
Spline	Constant	2	1,253.4	1,776.1
Spline	Piecewise-constant	8	1,261.4	1,780.3
Spline	Spline	6	1,259.4	1,770.3

Abbreviations: AIC, Akaike's Information Criterion; df, degrees of freedom; NMRD, nonmalignant respiratory disease.

^a Total number of df for the exposure-lag-response function.

^b Piecewise-constant functions for the lag-response were based on a categorical variable with a category per decade of lag, and for the exposure-response with categories based on quartiles of exposure.

^c Spline functions were based on natural cubic splines, with 2 df for the exposure-response (inner knot at the mean) and 3 df for the lag-response (inner knots at lags of 20 and 40 years).

and NMRD (Figure 2B) mortality in this cohort. The lag-response at various exposure intensities and the exposure-response at different lags from these models are summarized in Figure 3 for both outcomes. The lag-response and exposure-response from selected

unpenalized DLNMs are depicted in Web Figure 1 (available at <https://academic.oup.com/aje>), while Web Figure 2 shows the lag-response at various exposure intensities and the exposure-response at different lags from the best-fitting models according to the AIC.

Table 3. Hazard Ratios for Lung Cancer Mortality Associated With Different Scenarios of Occupational Crystalline Silica Exposure Using Models With Varying Exposure-Lag-Response Functions in a Diatomaceous Earth Cohort, California, 1942–2011

Scenario	Exposure			Model							
	Exposure Intensity, mg/m ³	Timing of Lag, years	Cumulative Exposure, mg/m ³ -years	Simple "Naive" Model ^a		Categorical Lag-Response ^b		Natural Spline Lag-Response ^c		Penalized DLNM ^d	
				HR	95% CI	HR	95% CI	HR	95% CI	RR	95% CI
1	0.2	1–20	4	1.07	0.94, 1.22	1.52	0.66, 3.50	1.27	0.76, 2.11	1.11	0.94, 1.31
2	0.2	1–40	8	1.15	0.89, 1.48	2.32	0.90, 6.01	1.83	0.93, 3.63	1.49	0.98, 2.27
3	0.2	11–50	8	1.15	0.89, 1.48	2.02	0.74, 5.54	2.58	0.93, 7.19	1.54	0.99, 2.40
4	0.4	11–30	8	1.15	0.89, 1.48	1.39	0.43, 4.41	1.81	0.72, 4.53	1.61	0.93, 2.79
5	0.4	31–50	8	1.15	0.89, 1.48	2.48	0.67, 9.12	2.73	0.66, 11.36	1.47	0.92, 2.35
6	0.8	31–40	8	1.15	0.89, 1.48	1.06	0.38, 2.95	1.87	0.89, 3.93	1.55	0.94, 2.53
7	1.0	31–40	10	1.19	0.86, 1.63	1.29	0.45, 3.71	2.08	0.95, 4.55	1.40	0.93, 2.10
8	1.0	41–50	10	1.19	0.86, 1.63	1.84	0.47, 7.22	1.04	0.22, 4.86	1.12	0.87, 1.44
9 ^e	0.2	11–30	8	1.15	0.89, 1.48	2.80	0.81, 9.71	2.78	0.69, 11.20	1.43	0.96, 2.14
	0.4	31–40									
	0.8	41–50									

Abbreviations: CI, confidence interval; df, degrees of freedom; DLNM, distributed-lag nonlinear model; HR, hazard ratio; RR, rate ratio.

^a The simple model was based on a constant lag-response and a linear exposure-response (1 df).

^b Unpenalized DLNM with a categorical lag-response with a category for each decade of lag and a natural spline function for the exposure-response (10 df combined).

^c Unpenalized DLNM with natural cubic spline functions for both the lag-response and the exposure-response (6 df combined).

^d Estimates from the penalized DLNMs are rate ratios from a Poisson generalized additive model aiming to approximate a Cox proportional hazards model. (The number of effective degrees of freedom for the exposure lag-response was 1.02.)

^e Scenario 9 represents time-varying exposure intensities: 0.2 mg/m³ during lag 11–30 years, 0.4 mg/m³ during lag 31–40 years, and 0.8 mg/m³ during lag 41–50 years.

Table 4. Hazard Ratios for Nonmalignant Respiratory Disease Mortality Associated With Different Scenarios of Occupational Crystalline Silica Exposure Using Models With Varying Exposure-Lag-Response Functions in a Diatomaceous Earth Cohort, California, 1942–2011

Scenario	Exposure			Model							
	Exposure Intensity, mg/m ³	Timing of Lag, years	Cumulative Exposure, mg/m ³ -years	Simple “Naive” Model ^a		Categorical Lag-Response ^b		Natural Spline Lag-Response ^c		Penalized DLNM ^d	
				HR	95% CI	HR	95% CI	HR	95% CI	RR	95% CI
1	0.2	1–20	4	1.11	1.02, 1.21	0.78	0.35, 1.75	1.11	0.77, 1.60	1.03	0.80, 1.34
2	0.2	1–40	8	1.23	1.03, 1.46	1.03	0.42, 2.54	1.73	0.97, 3.07	1.18	0.83, 1.68
3	0.2	11–50	8	1.23	1.03, 1.46	2.53	1.34, 4.79	2.09	1.18, 3.69	1.46	1.04, 2.06
4	0.4	11–30	8	1.23	1.03, 1.46	2.21	0.84, 5.85	1.63	0.76, 3.48	1.19	0.72, 1.97
5	0.4	31–50	8	1.23	1.03, 1.46	2.51	1.29, 4.91	2.42	1.27, 4.60	1.80	1.14, 2.85
6	0.8	31–40	8	1.23	1.03, 1.46	1.55	0.65, 3.70	1.59	0.83, 3.04	1.39	0.84, 2.30
7	1.0	31–40	10	1.29	1.04, 1.61	1.57	0.62, 4.01	1.39	0.70, 2.77	1.48	0.86, 2.55
8	1.0	41–50	10	1.29	1.04, 1.61	3.62	1.59, 8.24	3.61	1.63, 7.98	2.51	1.31, 4.81
9 ^e	0.2	11–30	8	1.23	1.03, 1.46	2.75	1.48, 5.14	2.31	1.26, 4.23	1.73	1.17, 2.55
	0.4	31–40									
	0.8	41–50									

Abbreviations: CI, confidence interval; df, degrees of freedom; DLNM, distributed-lag nonlinear model; HR, hazard ratio; NMRD, nonmalignant respiratory disease; RR, rate ratio.

^a The simple model was based on a constant lag-response and a linear exposure-response (1 df).

^b Unpenalized DLNM with a categorical lag-response with a category for each decade of lag and a natural spline function for the exposure-response (10 df combined).

^c Unpenalized DLNM with natural cubic spline functions for both the lag-response and the exposure-response (6 df combined).

^d Estimates from the penalized DLNMs are rate ratios from a Poisson generalized additive model aiming to approximate a Cox proportional hazards model. (The number of effective degrees of freedom for the exposure lag-response was 3.39.)

^e Scenario 9 represents time-varying exposure intensities: 0.2 mg/m³ during lag 11–30 years, 0.4 mg/m³ during lag 31–40 years, and 0.8 mg/m³ during lag 41–50 years.

DISCUSSION

We applied DLNMs to longitudinal data from a cohort of diatomaceous earth workers to examine the roles of intensity, duration, and timing of exposure to crystalline silica with respect to lung cancer and NMRD mortality. Exposure-lag-response estimates from DLNMs were generally higher for a variety of different exposure scenarios than for estimates from “naive” models assuming constant lag-response and linear exposure-response (equivalent to a simple linear term for cumulative exposure). Different exposure scenarios resulting in the same overall cumulative exposure, but with varying elements of timing and intensity of exposure, yielded varying association measures under DLNMs. These findings suggest that both timing and intensity of exposure are factors contributing to the overall relationship between crystalline silica exposure and mortality from malignant and nonmalignant respiratory disease. Nonlinearities of the exposure-response were also evident for both outcomes of interest.

Pathophysiological processes linking exposures to health outcomes are often complex mechanisms, with exposure effects persisting over varying time windows depending on the exposure-outcome relationship in question. It is also possible that latency of certain outcomes will correspond to delayed effects of exposure, with outcomes observed only after a certain amount of time has passed since initial exposures (9, 26, 27). These potential temporal aspects of exposure-response may complicate modeling of

relationships between an exposure and an outcome of interest in studies with time-varying exposure histories extending over long periods of follow-up, since both timing and intensity of exposure are relevant to disease risk. This is often the case in occupational epidemiologic studies in which detailed exposure histories are available, with exposure varying between people as well as within persons over time. Measures of cumulative exposure or average intensity are often used in studies of prolonged exposures and chronic health effects (28–30). There are strong implicit assumptions behind modeling of exposure-response as a function of either or even both of these measures: Cumulative exposures do not distinguish between exposure intensity and duration, while average intensity measures do not account for cumulative exposure. Even if both measures are used together, the time-varying nature of intensity is still not accounted for. (Note that in the current study we did not explicitly consider measures like average intensity of exposure in our comparisons.)

Timing of exposure is often considered in occupational epidemiology; lagging of exposure is common in order to account for disease latency (9). In most applications, exposure is lagged by the duration of the assumed minimum latency of the outcome. Previous applications of lagged exposures with respect to silica exposures and lung cancer or NMRD have used this simple lagging approach, considering various lengths of lags (4, 10–13). Although less common, other, more flexible approaches focusing on lag-response, including applications in occupational

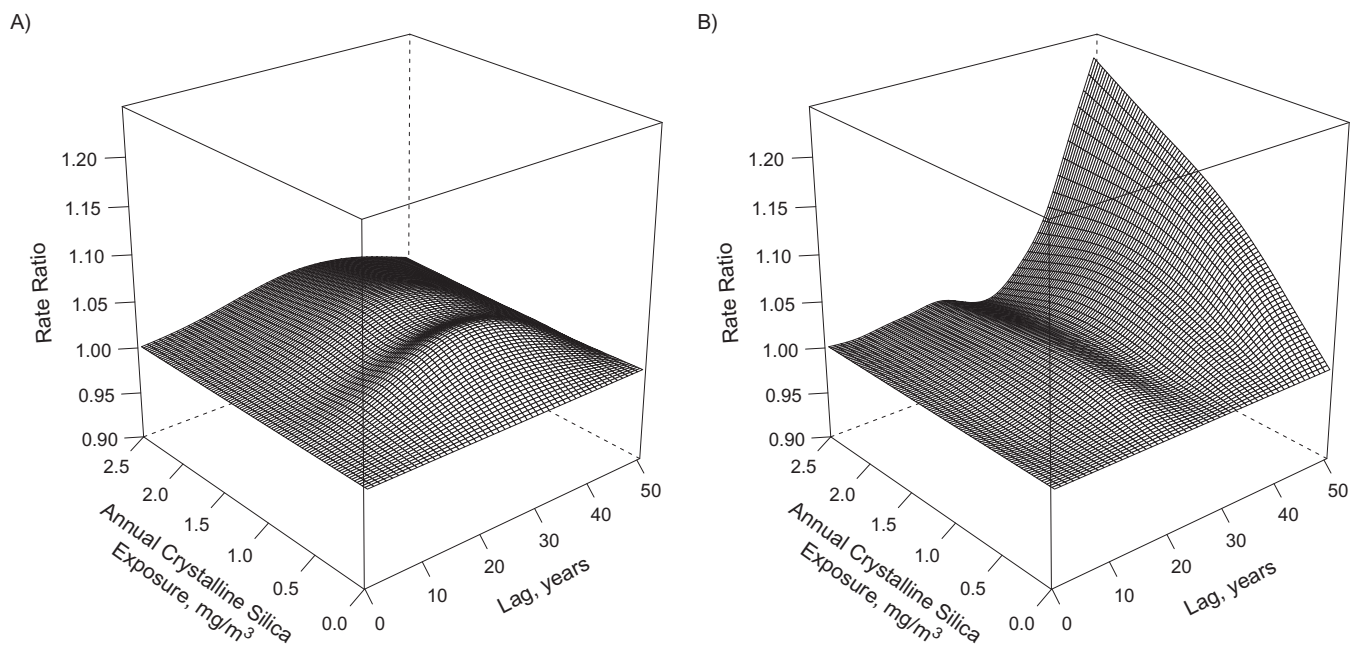


Figure 2. Three-dimensional representation of the exposure-lag-response for penalized distributed-lag nonlinear models of lung cancer mortality (A) and nonmalignant respiratory disease mortality (B) in California diatomaceous earth workers, 1942–2011.

epidemiology such as use of weighting functions for past exposures, have been described in the literature (26, 27, 31). There have also been approaches using flexible and smooth spline functions (32–34), as well as applications for time-to-event data (35). Richardson et al. (36) thoroughly summarize the limitations of the more traditional ways of modeling exposure-response and offer alternative options allowing for estimation of exposure-rate effects, thus evaluating cumulative effects of exposure in the presence of effect modification by exposure rate. They also describe different ways to deal with effects as they relate to time since exposure, including what are, essentially, distributed-lag functions.

Distributed-lag models have primarily been used in time-series studies of environmental exposures such as ambient air pollutants (22, 37–39). Recent studies have presented extensions of DLNMs beyond time-series designs, including survival data (17), as well as a penalized framework for DLNMs (23). These functions allow the joint estimation of exposure-response and lag-response, with potential for time-dependent nonlinear effects. Differences in exposure-response at different lags were evident when using both penalized and unpenalized DLNMs (both with categorical and spline terms for the lag-response). Models with categorical terms for the lag-response tended to have worse fit according to the AIC compared with models with natural cubic splines for the lag-response. In general, results from unpenalized DLNMs tended to be more prone to potential problems due to fewer observations at higher levels of exposure and/or lag periods compared with results from penalized DLNMs in our study. Results from penalized DLNMs maintained measures of association above or at the null for the entire windows of the lag and intensity dimensions, while confidence intervals were also more

stable compared with those from unpenalized DLNMs, which tended to result in less well-behaved functions with potentially unrealistic protective associations at certain ranges of exposure intensity and imprecise confidence intervals (as evidenced in Web Figures 1 and 2). Measures of association from unpenalized DLNMs for various exposure scenarios were also more extreme, depending on the values of intensity and lag selected, compared with penalized DLNMs.

Based on the results from the more flexible penalized DLNMs, models for lung cancer mortality were sensitive to the introduction of additional penalties. Associations for lung cancer mortality seemed to increase with increasing lag (indicative of a latency period), reaching an apparent maximum followed by a small decline in the association at lags greater than 30 years. The smaller association measures at more recent lags (1–10 years) compared with the maximum observed at greater lags were sensitive to additional penalties imposed for this section of the lag dimension, with association measures approaching null when those penalties were implemented. The decrease in the association at greater lags, compared with the maximum observed, would indicate a decline in risk of past exposures, conditional on survival for a given minimum duration. Decline of lung cancer risk in association with increased time since last exposure has been reported in the smoking literature. Decrease in lung cancer risk with years after smoking cessation has been reported in many studies, though the degree of reduction in risk and the durations associated with it vary across studies (40–45).

There were increases in associations with increasing lag for NMRD mortality as well, which were also sensitive to additional penalties at earlier lags. There did not appear to be

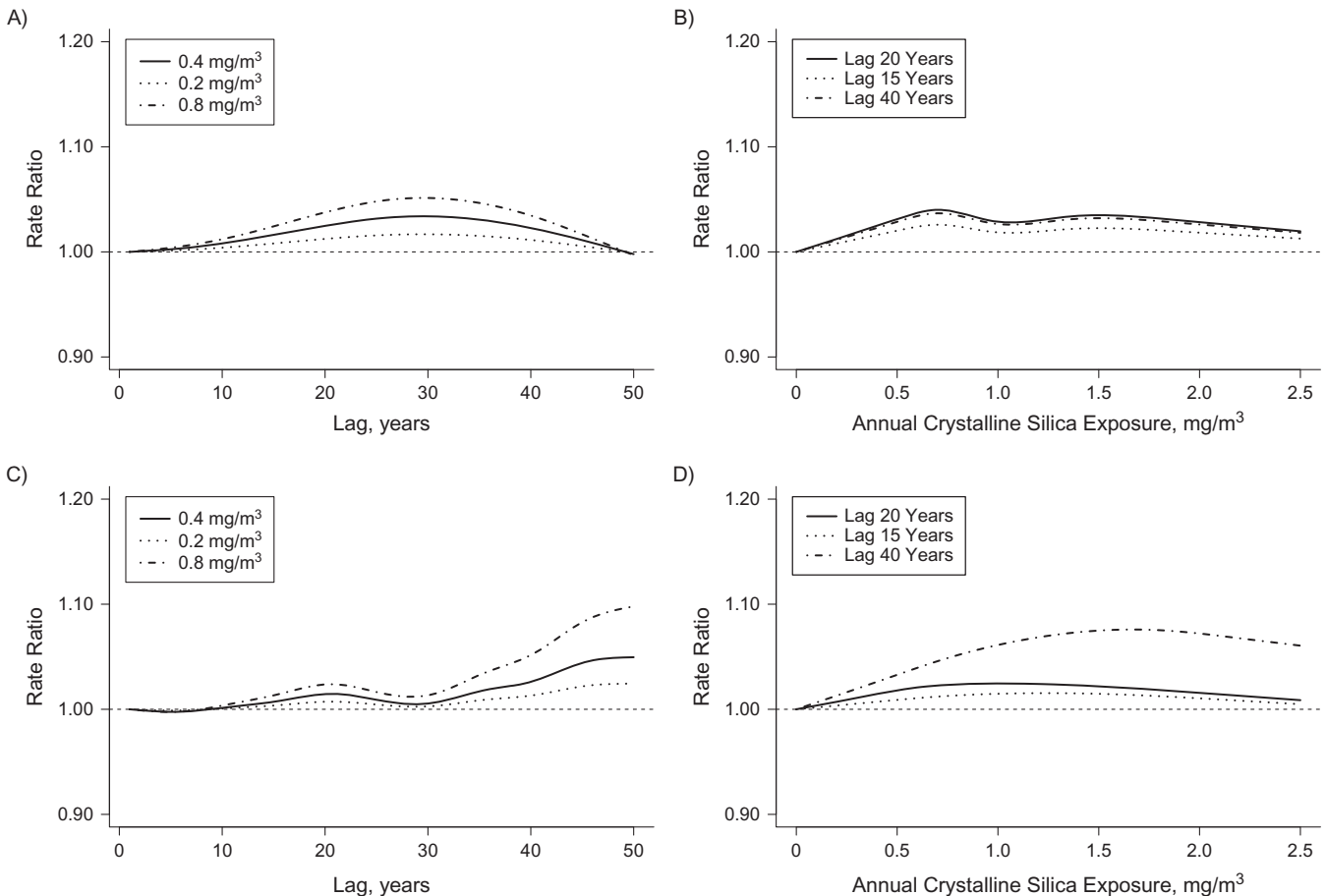


Figure 3. Rate ratios for lung cancer mortality (upper panels) and nonmalignant respiratory disease mortality (lower panels) according to occupational crystalline silica exposure, at lags of 0–50 years, among California diatomaceous earth workers, 1942–2011. A) Lag-response for lung cancer mortality at various (annual) exposure intensities; B) exposure-response of annual exposures at various lags from penalized distributed-lag nonlinear models; C) lag-response for nonmalignant respiratory disease mortality at various (annual) exposure intensities; D) exposure-response of annual exposures at various lags from penalized distributed-lag nonlinear models.

a comparable decrease associated with older exposures (greater lags) for NMRD, however; rather, association measures seemed to continue to increase with increasing lag in the window considered in this study. This may indicate initial and persisting lung damage, which eventually may lead to clinical manifestations and mortality. Potential trajectories of chronic obstructive pulmonary disease development suggest that lower lung function earlier in life, even when followed by normal rates of decline over time, is a more likely scenario for disease progression than a rapid decline later in life (46–48). This is consistent with the observed absence of a diminished risk of potentially harmful exposures at greater lags, as well as the observed latency at earlier lags.

Applications of DLNMs do have limitations, especially in the case of unpenalized DLNMs, where functions of both exposure-response and lag-response are fully defined a priori, with some strong parametric assumptions. Model fit assessment and selection from a variety of a priori–defined combinations of functions based on information criteria does not have a well-established theoretical basis (23). Penalized DLNMs better address this issue, as model selection is built in. The smooth functions through

penalized splines within generalized additive model frameworks are also part of an improved extension of the strong parametric forms of functions in unpenalized applications. General limitations do remain, however, and are summarized more thoroughly by Gasparrini et al. (23). In the current study, our choice of additional penalties on the lag dimension, to which the lag-response was particularly sensitive for lung cancer mortality, relied on assumptions that may be unverifiable (such as the length of a latency period and the timing of diminishing risk) about the nature of the lag-response. Furthermore, we lacked the statistical power to examine penalized spline functions with additional knots and/or penalty terms, as they tended to result in much simpler functions (usually resulting in linear lag-response).

In addition to allowing for disease latency, lagging exposures has been used to address healthy worker survivor bias (49). Lagging exposure is typically not enough to adequately address issues of time-varying confounding affected by previous exposure (such as healthy worker survivor bias) unless strict assumptions about the length of the minimum latency period and time from undetected early-stage disease to mortality are met (50, 51). However,

lagging exposure may help diminish the impact of the healthy worker survivor bias; thus, it is possible that the increasing associations observed with increasing lag of exposure may be partly due to reduced survivor bias rather than a true lag-response. Extending current methods for addressing time-varying confounding affected by prior exposure (52) to include a flexible exposure-lag-response estimation framework such as DLNMs would be a beneficial avenue for future research.

In summary, we applied DLNMs to assess aspects of crystalline silica exposure accrued over time in relation to both lung cancer and NMRD. Our findings indicate that intensity, timing, and duration are all potentially relevant aspects of exposure, and that approaches relying simply on cumulative exposure may underestimate associations as compared with more flexible DLNM approaches. Different lag-response shapes were observed for malignant respiratory disease mortality than for NMRD mortality, but our findings were suggestive of latency in the associations with exposure for both outcomes. Use of this flexible approach to model the exposure-lag-response can help researchers understand the relative contributions of exposure timing, intensity, and duration to the risk of chronic disease.

ACKNOWLEDGMENTS

Author affiliations: Division of Environmental Health Sciences, School of Public Health, University of California, Berkeley, Berkeley, California (Andreas M. Neophytou, Sally Picciotto, Daniel M. Brown, Ellen A. Eisen, Sadie Costello); Department of Epidemiology, School of Public Health, Boston University, Boston, Massachusetts (Lisa E. Gallagher); and Department of Family Medicine and Public Health, School of Medicine, University of California, San Diego, San Diego, California (Harvey Checkoway).

This work was supported by the Alpha Foundation for the Improvement of Mine Safety and Health (grant AFC215-31).

Conflict of interest: none declared.

REFERENCES

1. Checkoway H, Heyer NJ, Seixas NS, et al. Dose-response associations of silica with nonmalignant respiratory disease and lung cancer mortality in the diatomaceous earth industry. *Am J Epidemiol.* 1997;145(8):680–688.
2. Steenland K, Mannetje A, Boffetta P, et al. Pooled exposure-response analyses and risk assessment for lung cancer in 10 cohorts of silica-exposed workers: an IARC multicentre study. *Cancer Causes Control.* 2001;12(9):773–784.
3. Checkoway H, Heyer NJ, Demers PA, et al. Mortality among workers in the diatomaceous earth industry. *Br J Ind Med.* 1993;50(7):586–597.
4. Gallagher LG, Park RM, Checkoway H. Extended follow-up of lung cancer and non-malignant respiratory disease mortality among California diatomaceous earth workers. *Occup Environ Med.* 2015;72(5):360–365.
5. Pelucchi C, Pira E, Piolatto G, et al. Occupational silica exposure and lung cancer risk: a review of epidemiological studies 1996–2005. *Ann Oncol.* 2006;17(7):1039–1050.
6. Sun Y, Bochmann F. Lifetime risk of silicosis death for quartz exposed workers among German population [letter]. *Occup Environ Med.* 2004;61(4):374–375.
7. Cullinan P. Occupation and chronic obstructive pulmonary disease (COPD). *Br Med Bull.* 2012;104(1):143–161.
8. Smith DD. Re: “occupational exposure and lung cancer risk: a population-based case-referent study in Sweden” [letter]. *Am J Epidemiol.* 2001;153(10):1028–1029.
9. Checkoway H, Pearce N, Hickey JL, et al. Latency analysis in occupational epidemiology. *Arch Environ Health.* 1990;45(2):95–100.
10. Vacek PM, Verma DK, Graham WG, et al. Mortality in Vermont granite workers and its association with silica exposure. *Occup Environ Med.* 2011;68(5):312–318.
11. Sogl M, Taeger D, Pallapies D, et al. Quantitative relationship between silica exposure and lung cancer mortality in German uranium miners, 1946–2003. *Br J Cancer.* 2012;107(7):1188–1194.
12. Liu Y, Steenland K, Rong Y, et al. Exposure-response analysis and risk assessment for lung cancer in relationship to silica exposure: a 44-year cohort study of 34,018 workers. *Am J Epidemiol.* 2013;178(9):1424–1433.
13. Olsen GW, Andres KL, Johnson RA, et al. Cohort mortality study of roofing granule mine and mill workers. Part II. Epidemiologic analysis, 1945–2004. *J Occup Environ Hyg.* 2012;9(4):257–268.
14. Kreuzer M, Sogl M, Bröske I, et al. Silica dust, radon and death from non-malignant respiratory diseases in German uranium miners. *Occup Environ Med.* 2013;70(12):869–875.
15. Möhner M, Kersten N, Gellissen J. Chronic obstructive pulmonary disease and longitudinal changes in pulmonary function due to occupational exposure to respirable quartz. *Occup Environ Med.* 2013;70(1):9–14.
16. Cherry N, Harris J, McDonald C, et al. Mortality in a cohort of Staffordshire pottery workers: follow-up to December 2008. *Occup Environ Med.* 2013;70(3):149–155.
17. Gasparrini A. Modeling exposure-lag-response associations with distributed lag non-linear models. *Stat Med.* 2014;33(5):881–899.
18. Eisen EA, Agalliu I, Thurston SW, et al. Smoothing in occupational cohort studies: an illustration based on penalised splines. *Occup Environ Med.* 2004;61(10):854–860.
19. Robinson CF, Schnorr TM, Cassinelli RT 2nd, et al. Tenth revision US mortality rates for use with the NIOSH Life Table Analysis System. *J Occup Environ Med.* 2006;48(7):662–667.
20. Schubauer-Berigan MK, Hein MJ, Raudabaugh WM, et al. Update of the NIOSH Life Table Analysis System: a person-years analysis program for the Windows computing environment. *Am J Ind Med.* 2011;54(12):915–924.
21. Seixas NS, Heyer NJ, Welp EAE, et al. Quantification of historical dust exposures in the diatomaceous earth industry. *Ann Occup Hyg.* 1997;41(5):591–604.
22. Armstrong B. Models for the relationship between ambient temperature and daily mortality. *Epidemiology.* 2006;17(6):624–631.
23. Gasparrini A, Scheipl F, Armstrong B, et al. A penalized framework for distributed lag non-linear models. *Biometrics.* 2017;73(3):938–948.
24. Wood SN. *Generalized Additive Models: An Introduction With R.* Boca Raton, FL: Chapman & Hall/CRC; 2006.
25. Gasparrini A. Distributed lag linear and non-linear models in R: the package dlnm. *J Stat Softw.* 2011;43(8):1–20.
26. Langholz B, Thomas D, Xiang A, et al. Latency analysis in epidemiologic studies of occupational exposures: application

- to the Colorado Plateau uranium miners cohort. *Am J Ind Med*. 1999;35(3):246–256.
27. Richardson DB. Latency models for analyses of protracted exposures. *Epidemiology*. 2009;20(3):395–399.
 28. Lubin JH, Caporaso N, Wichmann HE, et al. Cigarette smoking and lung cancer: modeling effect modification of total exposure and intensity. *Epidemiology*. 2007;18(5):639–648.
 29. Lubin JH, Purdue M, Kelsey K, et al. Total exposure and exposure rate effects for alcohol and smoking and risk of head and neck cancer: a pooled analysis of case-control studies. *Am J Epidemiol*. 2009;170(8):937–947.
 30. Lubin JH, Moore LE, Fraumeni JF Jr, et al. Respiratory cancer and inhaled inorganic arsenic in copper smelters workers: a linear relationship with cumulative exposure that increases with concentration. *Environ Health Perspect*. 2008;116(12):1661–1665.
 31. Vacek PM. Assessing the effect of intensity when exposure varies over time. *Stat Med*. 1997;16(5):505–513.
 32. Thomas DC. Statistical methods for analyzing effects of temporal patterns of exposure on cancer risks. *Scand J Work Environ Health*. 1983;9(4):353–366.
 33. Hauptmann M, Wellmann J, Lubin JH, et al. Analysis of exposure-time-response relationships using a spline weight function. *Biometrics*. 2000;56(4):1105–1108.
 34. Hauptmann M, Pohlabein H, Lubin JH, et al. The exposure-time-response relationship between occupational asbestos exposure and lung cancer in two German case-control studies. *Am J Ind Med*. 2002;41(2):89–97.
 35. Abrahamowicz M, Beauchamp ME, Sylvestre MP. Comparison of alternative models for linking drug exposure with adverse effects. *Stat Med*. 2012;31(11-12):1014–1030.
 36. Richardson DB, Cole SR, Langholz B. Regression models for the effects of exposure rate and cumulative exposure. *Epidemiology*. 2012;23(6):892–899.
 37. Schwartz J. The distributed lag between air pollution and daily deaths. *Epidemiology*. 2000;11(3):320–326.
 38. Braga ALF, Zanobetti A, Schwartz J. The lag structure between particulate air pollution and respiratory and cardiovascular deaths in 10 US cities. *J Occup Environ Med*. 2001;43(11):927–933.
 39. Gasparini A, Armstrong B, Kenward MG. Distributed lag non-linear models. *Stat Med*. 2010;29(21):2224–2234.
 40. Office of the Surgeon General, Public Health Service, US Department of Health and Human Services. *The Health Benefits of Smoking Cessation: A Report of the Surgeon General*. Washington, DC: US Government Printing Office; 1990.
 41. Lubin JH, Blot WJ, Berrino F, et al. Modifying risk of developing lung cancer by changing habits of cigarette smoking. *Br Med J (Clin Res Ed)*. 1984;288(6435):1953–1956.
 42. Burns DM. Primary prevention, smoking, and smoking cessation: implications for future trends in lung cancer prevention. *Cancer*. 2000;89(11 suppl):2506–2509.
 43. Strauss GM, Dominioni L. Perception, paradox, paradigm: Alice in the wonderland of lung cancer prevention and early detection. *Cancer*. 2000;89(11 suppl):2422–2431.
 44. Speizer FE, Colditz GA, Hunter DJ, et al. Prospective study of smoking, antioxidant intake, and lung cancer in middle-aged women (USA). *Cancer Causes Control*. 1999;10(5):475–482.
 45. Ebbert JO, Yang P, Vachon CM, et al. Lung cancer risk reduction after smoking cessation: observations from a prospective cohort of women. *J Clin Oncol*. 2003;21(5):921–926.
 46. Lange P, Celli B, Agustí A, et al. Lung-function trajectories leading to chronic obstructive pulmonary disease. *N Engl J Med*. 2015;373(2):111–122.
 47. Martinez FD. Early life origins of chronic obstructive pulmonary disease. *N Engl J Med*. 2016;375(9):871–878.
 48. Sly PD, Bush A. From the cradle to the grave: the early life origins of chronic obstructive pulmonary disease. *Am J Respir Crit Care Med*. 2016;193(1):1–2.
 49. Gilbert ES. Some confounding factors in the study of mortality and occupational exposures. *Am J Epidemiol*. 1982;116(1):177–188.
 50. Robins JM. Causal models for estimating the effects of weight gain on mortality. *Int J Obes (Lond)*. 2008;32(suppl 3):S15–S41.
 51. Danaei G, Robins JM, Young JG, et al. Weight loss and coronary heart disease: sensitivity analysis for unmeasured confounding by undiagnosed disease. *Epidemiology*. 2016;27(2):302–310.
 52. Buckley JP, Keil AP, McGrath LJ, et al. Evolving methods for inference in the presence of healthy worker survivor bias. *Epidemiology*. 2015;26(2):204–212.