APPENDIX AVAILABLE ON REQUEST

Special Report

Reanalysis of the Harvard Six Cities Study and the American Cancer Society Study of Particulate Air Pollution and Mortality

Part II: Sensitivity Analyses

Appendix C. Flexible Modeling of the Effects of Fine Particles and Sulfate on Mortality

Daniel Krewski, Richard T. Burnett, Mark S. Goldberg, Kristin Hoover, Jack Siemiatycki, Michael Jerrett, Michal Abrahamowicz, Warren H. White, and Others

Correspondence may be addressed to Dr Daniel Krewski, Professor of Epidemiology and Statistics, Department of Epidemiology and Community Medicine, Room 3229C, 451 Smyth Road, University of Ottawa, Ottawa Ontario K1H 8M5, Canada

Although this document was produced with partial funding by the United States Environmental Protection Agency under Assistance Award R824835 to the Health Effects Institute, it has not been subjected to the Agency's peer and administrative review and therefore may not necessarily reflect the views of the Agency, and no official endorsement by it should be inferred. The contents of this document also have not been reviewed by private party institutions, including those that support the Health Effects Institute; therefore, it may not reflect the views or policies of these parties, and no endorsement by them should be inferred.

This document was reviewed by the HEI Health Review Committee but did not undergo the HEI scientific editing and production process.

UNIVERSITY OF OTTAWA

Faculty of Medicine **Faculty of Health Sciences**

Re-analysis of the Harvard Six-Cities Study and the American Cancer Society Study of Air Pollution and Mortality, Phase II: Sensitivity Analysis

Appendix A, B, C, D, E, F, G, H, and I

R. Samuel McLaughlin Center for Population Health Risk Assessment
Institute of Population Health
University of Ottawa

August, 2000

APPENDIX C: FLEXIBLE MODELING OF THE EFFECTS OF FINE PARTICLES AND SULFATE ON MORTALITY

INTRODUCTION

The original analyses of both the Six Cities Study data and the ACS Study data were based on the Cox proportional-hazards model, with the main exposure of interest being the long-term average of city-specific levels of fine particles (Dockery et al 1993; Pope et al 1995). This modeling strategy relies on two important assumptions. First, the Cox proportional-hazards (PH) assumption requires that the hazard ratio remain constant over the entire follow-up period. In the case of multiple covariates, the PH assumption must be satisfied by each covariate. The second assumption is relevant only for continuous predictors, and requires that the effect of increasing the value of the predictor on the logarithm of hazard can be described by a specific parametric function. In the original analyses of the Six Cities Study and the ACS Study, a linear function was used to describe the effect of fine particles, as well as the effects of other quantitative variables including body mass index (BMI) and number of pack-years of smoking, for either current or former-smokers.

Both assumptions are quite plausible in many applications of the Cox model, but they are not necessary a priori, and thus require empirical corroboration. Violation of the linearity assumption may occur, for example, if there is a threshold in the exposure-response relation. Accurate estimation of such nonlinear effects is essential in order to properly characterize the association between particulate air pollution and mortality. Accounting for non-proportional-hazards is also important, both to avoid biased estimates of relative risks, and to provide additional insights into the mechanisms by which long-term levels of air pollution affect mortality. Indeed, one of the more common reasons for the violation of the PH assumption in the Cox analyses with fixed-in-time covariates is that such analyses do not account for changes over time in the covariate values (Abrahamowicz et al 1996). If current risks depend mostly on the current value of the risk factor, then the hazard ratio, expressing the impact of a fixed-in-time measure of the risk factor, may vary over time. This consideration may be important in the Six Cities Study. Whereas the effects of fine particles are represented by a single (fixed-in-time) value, the annual average fine particle levels in the six cities tended to decrease during the study period.

The Original Investigators did not report on the assessment of consistency of these assumptions with the relevant data. One reason might be that simultaneous verification of both assumptions requires specialized software. Indeed, although several methods have been proposed for either (*i*) testing the PH assumption or (*ii*) testing the linearity assumption, they do not permit verification of both assumptions at the same time, for the same predictor. Mackenzie and Abrahamowicz (1999) show that failure to account for the fact that the effect of a continuous risk factor violates the PH assumption may produce spurious evidence of nonlinearity. Conversely, failure to account for nonlinearity may result in increased false-positive rates when testing the validity of the PH assumption.

OVERVIEW OF THE FLEXIBLE PRODUCT MODEL.

To address these issues, we reanalysed the Six Cities Study data and the ACS Study data using our regression spline generalization of the Cox model (Mackenzie and Abrahamowicz 1999) that extends our previous method for flexible modeling of time-dependent hazard ratios (Abrahamowicz et al 1996) to include simultaneous modeling of nonlinear effects of continuous covariates on the log hazard. The resulting model defines the log hazard ratio (HR) at time t, associated with the covariate value x, as the product of the two "marginal" functions: $f(t)$ and $g(x)$. The shape of $g(x)$ describes how, at any fixed point in time, the log hazard changes with increasing x values, with the special case of a linear function corresponding to the conventional linearity assumption in the Cox model. The estimate of $g(x)$ only permits an assessment of the relative impact of changing x values over different subintervals $[x_1,x_2]$ within its range, with steeper portions of the curve corresponding to greater impacts. To assess the absolute impact of such a change, the difference $g(x_1)-g(x_2)$ has to be multiplied by a function $f(t)$, since under the product model the impact of the covariate depends on t. Thus, $f(t)$ represents the time-dependence of the changes in the impact of x on the log hazard.

Both estimated functions are modeled using low-dimension quadratic regression splines, (ie, piecewise quadratic polynomials). Knots, representing points at which subsequent quadratic pieces of the estimated function join each other, are placed at equally spaced quantiles of the sample distribution of non-censored failure times or covariate values, respectively. Both functions may be conveniently represented by graphs of their point estimates, together with the corresponding pointwise 95% confidence intervals.

The product model offers two model-based likelihood ratio tests (LRT) that can be employed to simultaneously test the PH assumption and the linearity assumption, for the same continuous covariate. Each test is based on comparison of the log likelihood of the product model with one of the "marginal" models, in which the effect of the predictor is constrained either (*i*) to have a constant log hazard ratio $f(t)=C$, or (*ii*) to have a linear impact on the log hazard $g(x)=bx$. This is possible because constant and linear functions are included in the family of quadratic regression splines, so that the conventional models, corresponding to null hypotheses of the above LRT tests, are nested within the flexible product model.

Our method allows specification of how the effect of a particular covariate should be modeled. These a priori decisions are made independently for each covariate, based on knowledge of the covariate, goals of the analysis, and the amount of data available. For example, for some covariates the full product submodel may be entertained, implying flexible modeling of both time-dependent and nonlinear effects, although the effects of other covariates may be represented by a simpler submodel, with PH or linearity constraints imposed a priori.

In applications of the flexible product model, it is necessary to specify the required number of degrees-of-freedom (*df*) for this effect, that equals the number of quadratic components plus two. Thus, higher *df*s result in a more flexible model, reducing the risk of biased estimates, but also in a higher variance of the estimate and a risk of overfitting bias. MacKenzie and Abrahamowicz (1999) suggest the use of a 5 *df* model (3-piece quadratic spline) for time-dependent effect to achieve the best bias/variance trade-off. The use of 4 *df*'s for nonlinear effects was also suggested as a default option for large data, implying a 2-piece quadratic spline.

Specifying *df*s a priori is important with respect to the accuracy of the LRT, as selecting the "optimal" *df*s a posteriori and then carrying out tests based on such an "optimal" model inflates type I error rates (Abrahamowicz et al 1996). However, for exploratory purposes, fitting flexible models with different *df*s may be useful. This will help evaluating the robustness of the findings based on the default model, especially with respect to the shape of the estimated functions. Finally, results of flexible analyses may indicate that some or all of the assumptions being tested are well met by empirical data. In that case the initial flexible model may be simplified by removing those time-dependent and/or nonlinear effects specified a priori that are not significant.

SELECTION OF DATA FOR ANALYSIS

The Six Cities Study

Complex modeling required by the product model described in the previous section imposes restrictions on the size and complexity of the data. At present, we were able to extend the program to handle data with up to about 3,000 observations, provided only few covariates are modeled simultaneously. To reduce the number of observations to a tractable level, while maximizing the amount of information extracted from the data, we employed two approaches. The first consists of simply splitting the entire cohort into 4 randomly selected disjoint and complementary subsets, each of about 2025 subjects. Each of the four subsets is then analyzed independently using the same series of models. Whereas the "random subsets" approach avoids numerical problems, statistical power of the LRTs is substantially reduced, as each subset contains only about 375 deaths (about 25% of the total number of deaths). However, the independence of the results for the four subsets allows us, in part, to resolve this problem. In fact, because the subsets are disjoint for any specific model, the log likelihoods of the four subsets can be summed up in order to approximate the log likelihood of the entire sample under an extended model that estimates separate regression coefficients for each of the four subsets. Accordingly, the sum of subset-specific log likelihoods can be considered as corresponding to the model with 4 times as many parameters as the model estimated for each subset. It follows that, under the null hypothesis, the sum of the LRT statistics from the four models will be distributed as a Chi-squared statistic with 4 times as many degrees-of-freedom as the "nominal" LRT for each subset.

This conjecture was validated in a computer simulation where we generated 10,000 random samples of 1,000 each, randomly split the samples into four subsets and analyzed each using a conventional Cox model. We then added the LRT of the four subsets in each sample. In the simulations, the individual survival times were independent of the values of any continuous covariate included in the Cox model, so that the null hypothesis was true. The empirical distribution of 10,000 values of the combined LRT was virtually the same as the theoretical distribution of the Chi-squared statistic with 4 degrees-of-freedom: particular quantiles of the empirical distribution and the corresponding theoretical values were almost identical. Based on the results of these simulations, we assume that the sum of the four 1-*df* LRT values (for testing the adjusted effect of a given variable under the PH assumption) represents the 4-*df* LRT, and, accordingly, we assume the sum of the four 4-*df* tests (of time-dependence of a given effect) represents a 16-*df* LRT.

However, each of the four subsets also yields a different spline estimate of a function of interest (for example, the time-dependent effect of fine particles). Because of the flexibility of 5-*df* estimates and the relatively small number of deaths in each subset, the subset-specific estimates may be unstable; hence, random variation may induce some spurious differences in their respective shapes. Yet, it is unclear how the four estimates could be aggregated. Specifically, knots (points where subsequent polynomial pieces of a spline join each other) are automatically placed at the terciles of the distribution of uncensored event times observed in a given subset. Therefore, there is variability from subset to subset, so that the estimated spline coefficients cannot be directly compared or averaged.

To avoid these difficulties and provide a reasonably stable and accurate estimate of the pattern of time-dependent changes of the adjusted effect of fine particles on mortality, we used a procedure similar to the case-cohort approach (Barlow et al 1999). First, we randomly sampled 1,500 subjects from about 8,100 participants of the Six Cities Study, using simple random sampling (that is, independently of survival and covariate values). These 1,500 subjects, including the ones who died during the follow-up, were considered a random subcohort and, in the analysis, were treated as if they were followed prospectively (Barlow et al 1999). We then added to the random subcohort all other subjects who died during the study. Accordingly, the final sample included all deaths (cases) observed in the Six Cities Study. In the analyses, however, we had to account for the fact that cases corresponding to deaths occurring outside the subcohort were identified retrospectively (that is, selected based on the observed outcome). Therefore, as recommended by Prentice and Breslow (1978), each case from outside the subcohort was included in only one risk set, corresponding to the time of death of the particular case. By contrast, cases occurring within the subcohort and controls were included in all risk sets until their time of death or censoring, respectively.

It is emphasized that this version of the case-cohort approach was used mainly to obtain unique and relatively stable spline estimates of the effects of interest. Note that the case-cohort approach would be inferior to the random subsets approach as far as hypothesis testing is concerned. First, the case-cohort uses less information than the combined analyses of four random subsets, which reduces statistical power. Second, although the value of the test statistics will depend on the numerical weights (assigned to weigh differentially cases within vs. those outside the subcohort), the choice of correct weights remains controversial (Barlow et al 1999).

Using the above approaches, we analyzed subsets of the participants of the Six Cities Study with a number of versions of the flexible product model, which varied with respect to the set of included covariates and model complexity. Separate analyses were carried out using either fine particles ($PM_{2.5}$) or sulfate particles as the main exposure variable. We attempted to follow the Original Investigators' approach with respect to the choice of potential confounders and stratification variables, while taking into account numerical limitations of the flexible modeling approach. Specifically, as in the original analysis by Dockery and collegues (1993), all primary analyses used two stratification variables: sex and age at the beginning of follow-up, categorized in 5-year intervals. Moreover, in all analyses the effect of exposure to air pollution was adjusted for the baseline value of BMI and for two variables related to smoking status, one representing current-smokers and another former-smokers. Two variations of the model were considered, the first adjusting for the two binary variables indicating, respectively, current and former-smokers, and the second adjusting for the two quantitative variables, representing cumulative exposure to smoking, in terms

of the estimated number of pack-years of smoking, separately for current and former-smokers.

This strategy was necessary because simultaneous inclusion of all four smoking variables resulted in numerical convergence problems, especially with highly flexible models. This resulted partly from the limitations on the size of the data, and partly from the high degree of multicollinearity among these four variables. In addition to obvious negative correlation between variables representing current *vs* former-smokers, the quantitative and binary variables were inherently positively correlated, as all subjects who had the lowest value (0) for current-smoker pack-years had, by definition, the lowest value (0) of the binary indicator of current smoking. This creates identifiability problems in flexible nonlinear modeling of the effect of the number of pack-years for current-smokers, adjusted for the binary indicator of current-smoker. Indeed, the contrast between the current-smokers and other subjects may be represented either by the log hazard ratio estimate for the binary variable, or by estimating the difference in log hazards between value 0 and higher values of the number of pack-years for current-smokers. In flexible modeling, the estimate of this difference is locally independent of the estimated effect of increasing the number of pack-years (x) among current-smokers, as the former concerns the interval $0 < x < 1$ and the latter $x > 1$.

ACS Study

In order to employ flexible product model to analyze the effects of fine particles and sulfate on mortality in the ACS Study, it was necessary to develop a different sampling strategy. Given that the current implementation of our flexible model limits the sample size to less than 3,000 observations it would be impractical to analyze all the ACS data as this would imply separate analyses of each of about 200 subsets. Instead, we chose to limit the analyses to 10 disjoint subsets, each with 2,200 individuals selected using simple random sampling from the entire ACS cohort. The main reason for fixing the number of subsets at 10 was to obtain statistical power similar to that of the random-subsets analyses of the Six Cities Study data. The power of significance tests in survival analyses is largely driven by the number of observed deaths. Our flexible analyses of the Six Cities Study were based on four random subsets yielding a total of about 8,100 individuals with more than 1,400 deaths. Given that the censoring level was substantially higher in the ACS Study than in the Six Cities Study, we had to increase the number of subsets in the ACS Study to achieve a comparable number of deaths. As a typical subset of 2,200 participants of the ACS Study contained about 170 deaths, 8 subsets would be necessary to obtained a number of deaths close to 1,400.

However, we had to take into account that, with the same total number of deaths, the power of the combined likelihood ratio test in the ACS analyses will be lower than in the Six Cities Study analyses. This difference in power occurs because of the difference in degrees-of-freedom, reflecting the differences in the number of subsets: with m subsets the combined LRT of time-dependence uses 4m degrees-of-freedom whereas the combined LRT of nonlinearity uses 2m degrees-of-freedom. To compensate for the resulting loss of power, we increased the number of deaths in the ACS analyses by about 25%, compared to the Six Cities Study analyses, which implied increasing the number of random subsets from 8 to 10. Thus, flexible analyses of the ACS cohort were based on 10 disjoint random subsets, consisting of the total of 22,000 study participants with about 1,700 deaths. Even if this represents only a small fraction of the original ACS cohort, we assumed that 1,700 deaths will be sufficient to ensure a satisfactory precision of

the estimates and more than adequate statistical power to detect any effect of clinical relevance.

In the case of analyses of fine particles sampling frame was restricted to the inhabitants of the 50 cities with available measurements whereas for the sulfate analyses we sampled from the inhabitants of all 151 cities. To facilitate statistical inference, in each case we relied on simple random sampling of all eligible individuals, independently of the city of residence and covariate values.

In addition, to increase numerical stability and precision of the estimated spline functions, representing various effects of interest, we adopted an approach similar in spirit to the case-cohort approach (see section 2.8). Given that about 40,000 deaths occurred in the ACS Study, the size of the dataset including all deaths would largely exceed the numerical capabilities of the current version of our program. Therefore, we first randomly selected a subcohort of 1,200 participants of the ACS Study and then added additional 1,300 deaths randomly selected from about 40,000 deaths that occurred outside this subcohort. The resulting sample of 2,500 participants, including about 1,370 deaths (comparable to the 1,430 deaths used in the case-cohort analyses of the Six Cities Study data) was then analyzed while manipulating time-at-risk so as to account for retrospective identification of those cases that were not included in the subcohort (see section 2.8).

All analyses were stratified by sex and 5-year age categories and the effects of air pollution were adjusted for education (dichotomized: college or higher vs lower), BMI and two smoking variables, one representing the information on current-smokers (at baseline) and the other related to former-smokers. In most analyses of the ACS data, the smoking effects were represented by quantitative variables corresponding to cumulative exposure in terms of pack-years, that were calculated as the product of smoking intensity and smoking duration. Replacing these quantitative variables by the corresponding binary indicators of current and past smoker, respectively, did not change materially the estimated effects of fine particles or sulfate.

As in the analyses of the Six Cities Study data, time-dependent effects were represented by 5-degrees-of-freedom (*df*) quadratic spline and tested using 4-*df* LRT. Possibly nonlinear effects of fine particles and sulfate, and of each of the three quantitative covariates were represented by a 3-*df* quadratic spline, and the nonlinearity was tested using LRT on 2 *df*.

MODELING THE SIX CITIES STUDY DATA

Effect of Fine Particles

Flexible analyses of the effect of fine particles, represented by a quantitative city-level variable, indicated that its adjusted effect meets the linearity assumption. Regardless of the subset of covariates included in the model, and regardless of whether the PH assumption was imposed on this effect or not, the *P* values for the test of nonlinearity for each of the four disjoint subsets were far above the nominal 0.05 level associated with statistical significance. In contrast, flexible analyses of the effect of fine particles suggested that the Cox proportional-hazards (PH) assumption may be violated. Using the 5-*df* (default) representation, and adjusting for the quantitative variables representing pack-years of cigarette smoking

for current and former-smokers, respectively, as well as for BMI, the time-dependent hazard ratio for the effect of fine particles was marginally significant for the product model, and remained so in a simpler model in which $g(x)$ was restricted to be linear. The P values for each of the four subsets (as well as for the combined test) are listed in Table C.1. The combined LRT with 16 *df* yielded a P value of .032. Replacing quantitative smoking variables by the corresponding binary indicators of current and former-smoker did not affect this finding appreciably as the time-dependent effect of fine particles remained statistically significant for the product and the restricted models ($P = 0.048$).

The solid curves in Figures C.1a through C.1d represent the 5-*df* estimates of the pattern of changes in the log HR for fine particles during the first 15 years of follow-up (obtained from the flexible product model applied to the four disjoint datasets) after having been adjusted for the two binary smoking status indicators and BMI. The dashed curves, representing pointwise 95% confidence intervals, help assess the precision of the log HR estimates at specific points during follow-up. The graphs in Figures C.1a through C.1d suggest that changes in the effect of air pollution may follow a non-monotone pattern over time.

All four panels of Figure C.1 show that the estimated impact of fine particles on mortality increases substantially at about 10 to 12 years of follow-up, and then decreases in the last few years. The four estimates are less consistent in the first few years of follow-up where, for two subsets, the effect of fine particles seems to decrease, whereas for two other subsets it seems to increase. However, wide confidence intervals around the estimated function during the first 3 years of follow-up (reflecting the relatively low number of deaths in the initial phase) indicate that these apparent discrepancies may be due largely to numerical instability of the subset-specific estimates.

To provide a more stable estimate of the pattern of changes over time in the effect of fine particles, we show in Figure C.2 the estimate obtained with the case-cohort approach, where information contained in all deaths is used. As expected, the confidence intervals here are much tighter than in Figure C.1. The estimate shows a modest decline in the effect of fine particles in the first four years, followed by the relatively constant and practically null effect in years four through eight, and finally by a substantial increase in the effect in years 10 to 12. This latter increase confirms the consistent findings of the four subset-specific analyses (Figures C.1a through C.1d). When the number of *df*s was reduced to 3 or 2, the time-dependent effect became non-significant, with the model fit being much worse than for the 5-*df* model. This shows that less flexible models were not able to represent the complex non-monotone pattern of changes over time in the log HR for fine particles.

This non-monotonicity of changes in the effect of fine particles during the follow-up explains why our test (based on flexible modeling) rejected the PH hypothesis, although the test proposed Grambsch and Therneau (1994) (as incorporated in S-Plus) failed to reject this hypothesis. The reason is that the latter test statistic is developed assuming the log-hazard ratio is a linear function of time. (We have previously shown, in the context of assessing prognostic factors for mortality in nasopharyngeal cancer, that when the actual pattern of change shows marked non-monotonicity, the test of Grambsch and Therneau has very low power [Rachet et al 1998]).

To assess the plausibility of such a complex pattern of changes, we compared our results in Figures C.1 and C.2 with those in Figure 1 of the original publication by Dockery and collegues (1993), which shows the patterns of changes in yearly average particle levels for each of the six cities. The main objective was to establish to what extent the decreasing and increasing portions of our time-dependent estimates of the fine particle effect correspond to periods of respectively decreasing and increasing between-cities variation in the yearly particulate levels. Unfortunately, fine particle levels were not measured prior to 1979, corresponding approximately to the first four to five years of follow-up (Dockery et al 1993). However, the upper graph of Figure 1 in that article shows that during the same time interval (1975–1980), total particles decreased considerably in the two cities with the highest pollution levels, resulting in an important reduction in between-cities variation from 1975 to 1980. As fine particles represent a fraction of total particles, it may be expected that the variance of fine particle levels across cities may also have been reduced during the initial 5 years. If so, then this decrease would coincide in time with the estimated decrease in the impact of fine particles on mortality suggested by the case-cohort estimate in Figure C.2. Moreover, the increase in the estimated time-dependent effect of fine particles, seen after about 11 years of follow-up in Figures C.1 and C.2, coincides with a sharp increase in fine particle levels in Steubenville, and an increasing between-city variation in these levels (middle graph in Figure 1 of the original publication).

In the previous paragraph, we argued that the periods of decrease/increase in the estimated time-dependent effect of fine particles seem to agree with concurrent decreases/increases in the between-city variation in particulate levels. We now attempt to explain why such an agreement may occur. In the Six Cities Study, the city-specific levels of fine particles are represented by a fixed-in-time variable, corresponding to the average within-city level during the relevant period (1980–1985). Thus, the estimated log hazard ratio represents the effect of increasing this long-term average by one unit of exposure (here corresponding to a range of $29.6 - 11.0 = 18.6$). Assume now that the impact of air pollution on mortality in a given year depends mostly on the current (yearly average) particulate levels rather than on their long-term average levels. In addition, assume that the actual between-city variance of particles levels is much higher in year one than in year five. Then, under the above assumptions, the observed differences in mortality among cities due to air pollution will also be much greater in year 1 than in year 5. In the Cox model, however, differences in mortality in both years will be attributed to the same magnitude of the differences in the long-term average pollution levels. As a consequence, because of the failure to account for yearly changes in particulate levels, the apparent impact of the same differences in their long-term average levels on mortality will appear much bigger in year one than in year five. In this situation, a flexible time-dependent model will detect these apparent temporal differences in the estimated effect, and will likely yield evidence of time-dependence in the effect of long-term average particle levels. From this perspective, the results of our analyses may suggest that the relation between fine particles and mortality may be refined by taking into account short term, for example annual, variation of particulate levels in individual cities.

Effect of Sulfate Particles

Similar analyses were carried out focusing on the effect of sulfate. Table C.2 shows the results of LRT testing of the PH hypothesis, indicating that sulfate also have a marginally significant time-dependent effect ($P = 0.0316$ for the combined 16-*df* LRT). Figures C.3a through C.3d present the estimated patterns of time-dependence of sulfate effects for the four random subsets. Figure C.4 shows the estimate

obtained using the case-cohort approach. The results in Figures C.3 and C.4 show that the estimated pattern of changes over time in the impact of sulfate on mortality is quite similar to that seen in Figures C.1 and C.2 for fine particles.

Effects of Selected Covariates

In addition to flexible modeling of the effects of exposure to particulate air pollution, we also considered flexible estimation of the effects of selected covariates, including smoking variables and BMI. For both binary smoking indicator variables, the time-dependent effects were nonsignificant and the corresponding log hazard estimates were very flat (graphs not shown). This further confirms that their impact on hazard remains constant during the follow-up period, and that the conventional Cox PH approach is adequate to represent the effects of these important risk factors.

When the binary smoking variables were replaced by the corresponding quantitative pack-years variables, flexible modeling yielded significant nonlinear effects of pack-years for current-smokers. The estimated effect of increasing the number of pack-years is presented in Figure C.5, which resembles a two-segment linear spline. The estimate indicates a large difference in risk between the zero and non-zero pack-years of cigarette smoking. Although the risk continues to increase gradually with increasing number of pack-years, the eventual slope is much smaller than the initial slope (it is between 0 and 1). This pattern has a clear interpretation. In the absence of the binary indicator for current smoking status, the flexible nonlinear modeling approach allowed us to detect a large gap in risks between current-smokers (pack-years > 0) and non-smokers (pack-years = 0), as reflected by the first, steep component in Figure C.5. The second component of this estimate shows that the actual effect of increasing the number of pack-years for current-smokers is almost perfectly linear. Although the nonlinear effect of the number of pack-years for former-smokers was not statistically significant, Figure C.6 shows that the pattern of nonlinearity is very similar to that for current-smokers, but the effects are weaker. Thus, the results of Figures C.5 and C.6 indicate that the flexibility of regression spline modeling may provide important insights in the role of continuous covariates and, in some cases, compensate for some aspects of model misspecification (due here to omitting the binary indicators of smoking status). The ability of our flexible model for pack-years variables to detect important difference in the risks between current-smokers and other subjects also explains the near collinearity problems discussed in section C.2. When the model included both flexible (3-*df*) representation of the continuous effect of pack-years and the corresponding binary smoking indicator variable, the difference in risks between smokers and nonsmokers could be represented by either variable, resulting in identifiability problems.

Finally, Figure C.7 presents the nonlinear estimate of the effect of BMI on the log hazard, which was statistically significant in all analyses. The relation between BMI and risk is non-monotone with the highest risks observed for either small or large BMI values. This pattern, which can be well approximated by a quadratic function, is consistent with the characterization of BMI in our extended model. Accordingly, in all analyses reported in this appendix, the effect of BMI was represented by a nonlinear 3-*df* estimate.

Modeling the ACS Study Data

Effects of Fine Particles

The left part of Table C.3 summarizes the results of testing of the time-dependence of the effect of fine particles on mortality, after adjusting for education, BMI and the two quantitative smoking variables, and taking into account the nonlinearity of the fine particles effect. The combined LRT on 40 *df* rejects the PH hypothesis ($P = 0.015$ in Table C.3), although the P values for individual subsets vary substantially and in only two among the 10 subsets fall below the conventional 0.05 cutoff.

Figure C.8 shows the estimated patterns of time-dependence in the 10 randomly selected subsets. The ten estimates show considerable variation so that no overall pattern of changes over time in the effects of fine particles emerges. An important portion of the differences between subset-specific estimates is likely due to sampling error. First, with only about 170 deaths per subset, the precision of the 5-*df* estimates is limited as indicated by relatively wide pointwise confidence intervals in Figure C.8. Moreover, in eight of the ten subsets represented in Figure C.8 the time-dependent effects of fine particles are statistically nonsignificant (P values between 0.12 and 0.58), in which case the flexible estimate may represent mostly the over-fitting bias (Abrahamowicz et al 1996).

It is also likely that the between-subsets variation in the pattern of changes in the impact of fine particles over time may be partly due to the random between-samples variation in the proportions of participants from different cities. Indeed, the results of the Six Cities Study analyses suggested that the observed changes over time in the estimated effect of the long-term average levels of fine particles may reflect the short-term changes in their levels in specific cities (section 3). On the other hand, it is likely that the patterns of changes in yearly fine particle levels may be quite different among the 50 cities with available fine particles measures in the ACS Study. If so, then the estimated patterns of changes, and their statistical significance, may vary considerably depending on which cities have relatively larger representation in a given subset. This might also explain why it is unlikely to observe one consistent overall pattern of changes over time.

To obtain a more stable estimate of the pattern of time-dependent changes in the impact of fine particles on mortality, Figure C.9 shows the results of the case-cohort approach, based on 2,500 individuals with more than 1,300 deaths. As expected, the pointwise confidence intervals in Figure C.8 are more narrow than in Figure C.9, reflecting enhanced precision, although it should be noted that the case-cohort-based intervals in Figure C.9 are only approximative, due to uncertainty about optimal weighting of cases (Barlow et al 1999). The most prominent feature of the time-dependent estimate of the log hazard ratio for fine particles in Figure C.9 is its flatness. This suggests that the impact of fine particles on mortality in the ACS Study remained remarkably stable across most of the follow-up period, corroborating the conclusion of the previous paragraph about the absence of any systematic changes.

The left column of Table C.4 shows the results of testing the nonlinearity of the adjusted effect of fine particles in the 10 subsets of the ACS Study. The combined LRT on 20 *df* yields a very significant evidence of nonlinearity ($P = 0.0002$), although subset-specific P values vary considerably. The fact that both time-dependent and nonlinear effects of fine particles are statistically significant underscores the importance of using the flexible product model that simultaneously accounts for both effects. (Accordingly

in this section, all reported tests and estimates that are related to nonlinearity of fine particles effects are adjusted for time-dependent effect and vice versa.) Figure C.10 shows the 3-*df* estimates of the nonlinear effects of fine particles for the 10 random subsets. It should be noted that the estimated functions show the log hazard ratio relative to the subset-specific mean fine particles value, at which the estimate is a priori fixed to 0 (Mackenzie and Abrahamowicz 1999). (This explains why the confidence interval shrinks to null at the sample mean and widens with increasing difference from the mean.) Whereas there is substantial variation between subset-specific shapes of the estimated dose-response curves, a lot of this variation may be due to sampling error and over-fitting bias, as indicated by wide confidence intervals. To obtain a more stable estimate, Figure C.11 shows the case-cohort-based estimate of the nonlinear effect of fine particles on mortality, adjusted for the estimate of their time-dependent effect (see Figure C.9). The estimate in Figure C.11 suggests that the impact of fine particles may be relatively stronger in the lower half of its range than in its upper half.

Effects of Sulfate

The combined 40-*df* LRT reported in the right column of Table C.3 shows a very significant time-dependence of the adjusted effect of sulfate on mortality ($P < 0.003$), based on 10 random subsets of the ACS cohort. Figure C.12 shows the corresponding estimates for the ten subsets. As for fine particles, there is considerable variation between subset-specific patterns of time-dependent changes in the effect of sulfate and this variation remains substantial even if the comparison is restricted to the subsets with significant or marginally non-significant time-dependence. This finding suggests that changes observed in individual subsets may reflect changes in sulfate levels in individual cities so that overall pattern of changes over time would be difficult to detect and to interpret. This is confirmed in Figure C.13 which shows that the case-cohort-based estimate of the time-dependent effect of sulfate particles is quite flat.

The right column of Table C.4 shows the very significant nonlinearity of the adjusted effect of sulfate particles in the ACS Study, based on the combined LRT test ($P < 0.006$). Figure C.14 shows the 3-*df* nonlinear estimate for each of the ten subsets. A rather consistent pattern emerges from the ten panels of Figure C.14. With few exceptions, there is a clear dose-response relation in the upper half of the distribution of sulfate levels, but in the lower range there is little evidence of such a relation. This pattern is confirmed in Figure C.15 that shows the results of the case-cohort approach and, thus, provides a more stable estimate of the dose-response curve. Figure C.15 shows a clearly nonlinear though monotonic relation between sulfate levels and risks of death: increasing sulfate levels seems to have only very minor impact on log hazard ratio as long as their level does not exceed about 15 $\mu g/m^3$ but beyond this level the risks increase very sharply.

Effects of Selected Covariates

Figures C.16 and C.17 show the nonlinear 3-*df* estimates of the effects of pack-years of cigarette smoking for, respectively, current and former-smokers, obtained using the case-cohort approach. The dose-response curve in Figure C.16 shows a two-phase relation with the initial sharp increase in the risk representing the difference between subjects who did not smoke at baseline (0 pack-years) and current-smokers (1 or more pack-years) and the second segment showing that risks for non-smokers increase

gradually with increasing exposure. This pattern is quite similar to the estimate of the impact of pack-years for current-smokers in the Six Cities Study (see Figure C.5). The dose-response curve for past smokers is similar with the exception that the initial segment suggests that between 0 and 1 pack-years the risks tend to decrease rather than increase, although the estimated difference is minor (Figure C.17).

Finally, Figure C.18 confirms the U-shaped effect of BMI on mortality, with risks being the lowest in the middle of its range and increasing in both tails of the distribution. This pattern, which may be well represented by a quadratic function, is similar to that estimated for the Six Cities Study (see Figure C.7).

REFERENCES

Abrahamowicz M, MacKenzie T, Esdaile JM. 1996. Time-dependent hazard ratio: modeling and hypothesis testing with application in lupus nephritis. Journal of the American Statistical Association 91(436):1432–1439.

Barlow EB, Ichikawa L, Rosner D, Izumi S. 1999. Analysis of case cohort designs. Journal of Clinical Epidemiology. 52(12):1165–1172.

Dockery DW, Pope CA, Xu X, Spengler JD, Ware JH, Fay ME, Ferris BG, Speizer FE. 1993. An association between air pollution and mortality in six US cities. New England J of Medicine 329 (24):1753–1759.

Grambsch PM, Therneau TM. 1994. Proportional hazards tests and diagnostics based on weighted residuals. Biometrika. 81:515–526.

MacKenzie T, Abrahamowicz M. 1999. Flexible hazard ratio estimation: Non-proportional hazards with non-linearity of the log hazard ratio. Manuscript under revision.

Pope CA, Thun MJ, Namboodiri MM, Dockery DW, Evans JS, Speizer FE, Heath CW. 1995. Particulate air pollution as a predictor of mortality in a prospective study of US adults. Am J Respir & Crit Care Med 151:669–674.

Prentice RL, Breslow NE. 1978. Retrospective studies and failure time model. Biometrika. 65:153–158.

Rachet B, Sasco AJ, Abrahamowicz M, Benyamine D. 1998. Prognostic factors for mortality in nasopharyngeal cancer: accounting for time-dependence of relative risks. International Journal of Epidemiology. 27:772–780.

Table C.1 Test for the proportional hazards hypothesis for the effect of PM 2.5 on mortality in the Harvard Six-Cities Study: Results for the 4 random subsets.

Random Subset	Likelihood Ratio Test	Degrees of Freedom	P-value
1	7.10	4	.1303
2	6.06	4	.1944
3	8.48	4	.0756
4	6.32	4	.1767
All[a]	27.96	16	.03197

a. Combined likelihood ratio test.

Table C.2. Test for the proportional hazards hypothesis for the effect of Sulfates on mortality in the Harvard Six-Cities Study: Results for the 4 random subsets

Random Subset	Likelihood Ratio Test	Degrees of Freedom	P-value
1	7.86	4	.0971
2	5.32	4	.2572
3	6.76	4	.1487
4	8.06	4	.0896
All[a]	28.00	16	.0316

a. Combined likelihood ratio test.

Table C.3. Test for time dependence of the effects of fine and sulfate particles in the American Cancer Society Study.

Fine Particles				Sulfates			
Random Subset	LRT[1]	df[2]	p-value	Random Subset	LRT	df	p-value
1	7.38	4	0.12	1	5.72	4	0.22
2	4.36	4	0.36	2	4.80	4	0.31
3	2.86	4	0.58	3	10.10	4	0.04
4	12.42	4	0.01	4	7.08	4	0.13
5	3.16	4	0.53	5	5.28	4	0.26
6	6.04	4	0.20	6	8.14	4	0.09
7	4.52	4	0.34	7	4.62	4	0.33
8	3.78	4	0.44	8	4.28	4	0.37
9	11.90	4	0.02	9	16.50	4	0.00
10	5.36	4	0.25	10	3.08	4	0.55
Sum	61.78	40	0.02	Sum	69.60	40	0.00

1. Likelihood ratio test statistic
2. Degrees of freedom
3. Combined LRT

Table C.4. Test for non-linearity of the effects of fine particles and sulfate particles in the American Cancer Society Study.

Fine Particles				Sulfates			
Random Subset	LRT[1]	df[2]	p-value	Random Subset	LRT	df	p-value
1	13.56	2	0.00	1	4.06	2	0.13
2	0.68	2	0.71	2	1.12	2	0.57
3	0.02	2	1.00	3	5.96	2	0.05
4	12.08	2	0.00	4	6.64	2	0.04
5	3.00	2	0.22	5	2.08	2	0.35
6	3.72	2	0.16	6	0.12	2	0.94
7	1.26	2	0.53	7	2.06	2	0.36
8	2.34	2	0.31	8	3.10	2	0.21
9	4.48	2	0.11	9	13.28	2	0.00
10	9.22	2	0.01	10	1.04	2	0.59
Sum	50.36	20	0.00	Sum	39.46	20	0.01

1. Likelihood ratio test statistic
2. Degrees of freedom
3. Combined LRT

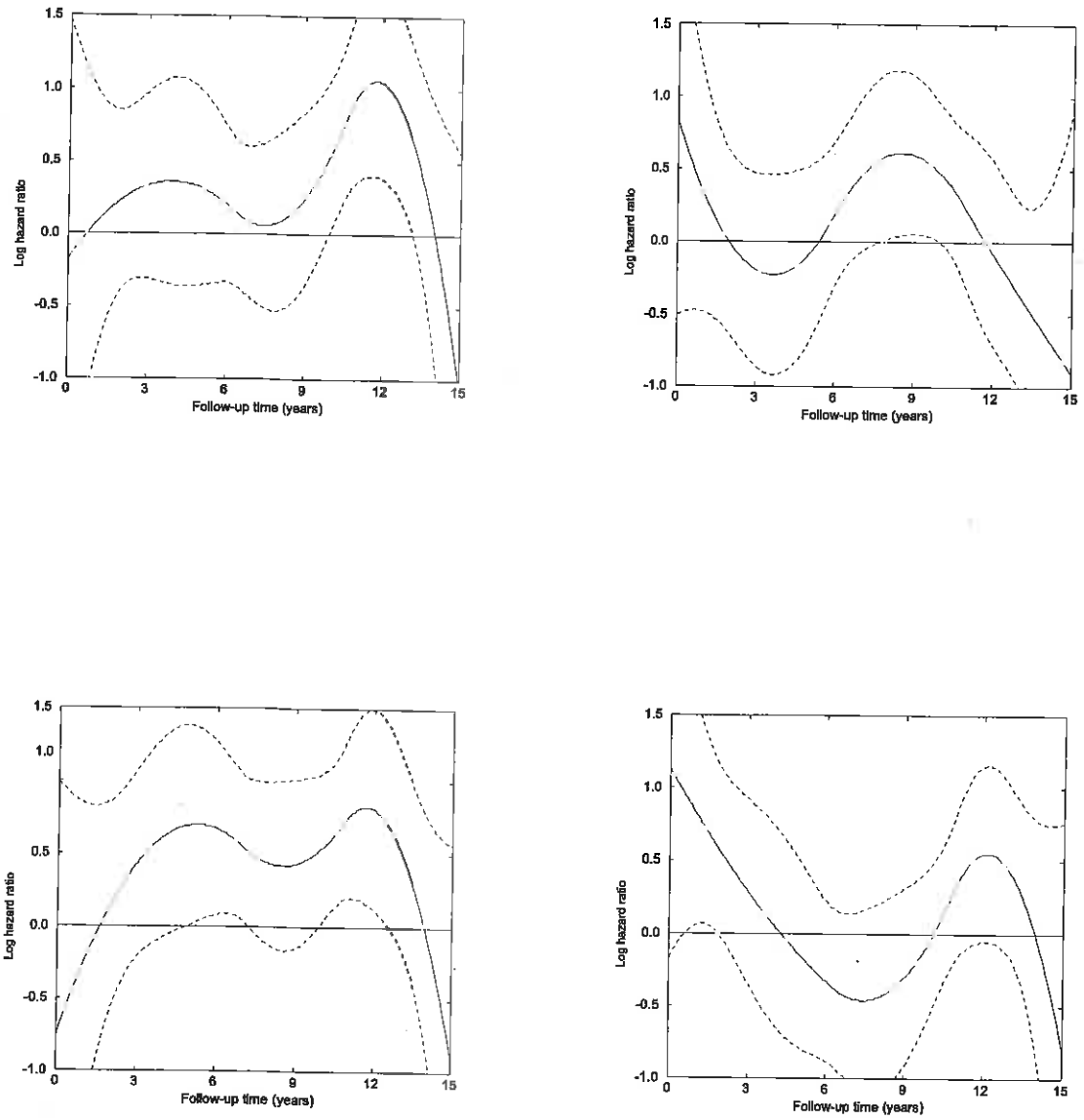

Figure C.1 Flexible quadratic spline estimate (5-df) of the time-dependent effect of fine particles on the log hazard of mortality of four disjoint subsets of the Harvard Six-Cities Study. The horizontal axes represent the follow-up time, and the vertical axes correspond to the log hazard ratio (HR) associated with increasing the fine particles level by 18.6 micrograms/m3, which correspond to the difference between the highest and the lowest of the city-specific levels. The solid curves represents the point estimates of the log HR and the dashed curves the pointwise 95% confidence intervals.

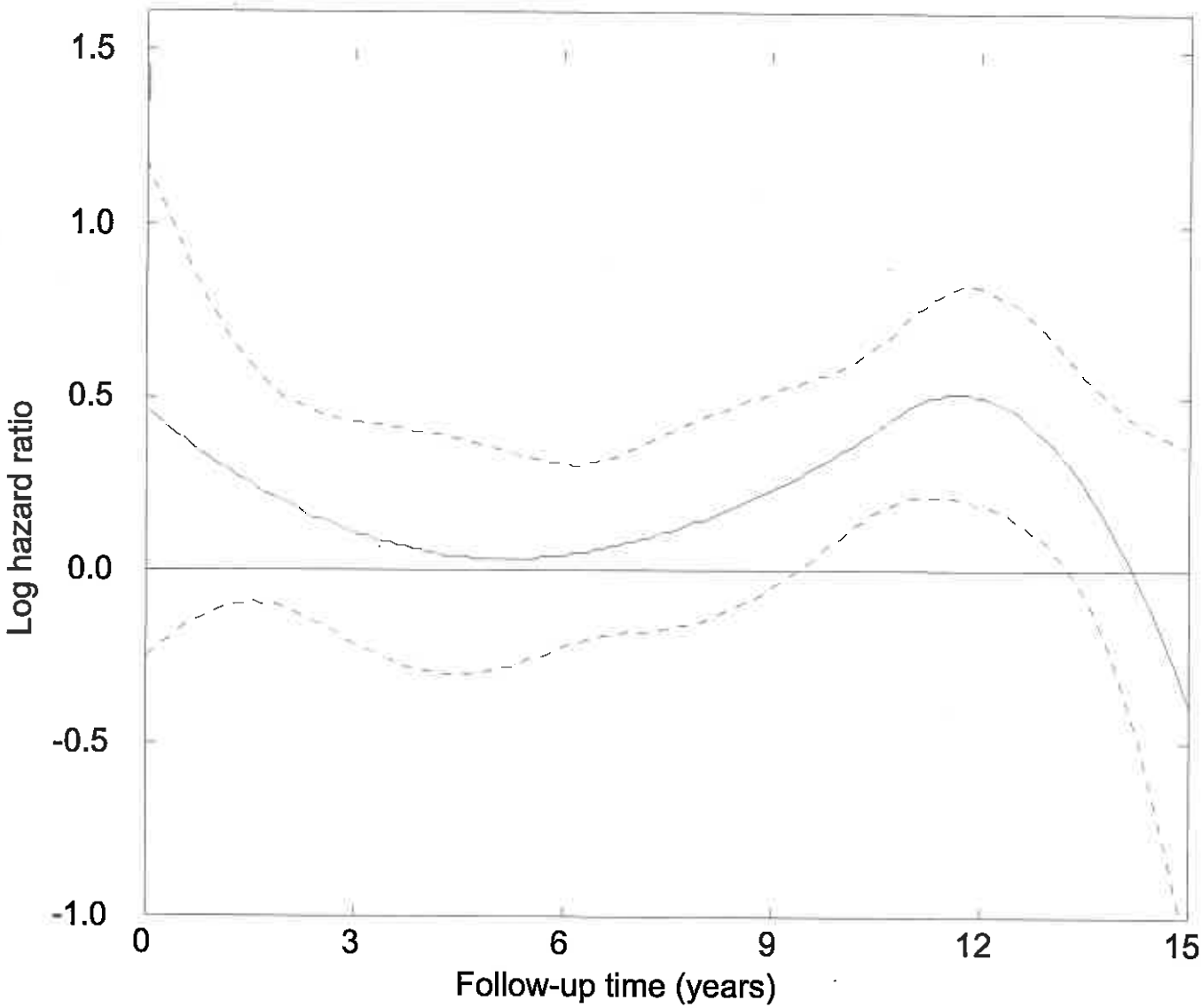

Figure C.2 Flexible quadratic spline estimate (5-df) of the time-dependent effect of fine particles on log hazard of mortality in a "case-cohort" type subset of the Harvard Six-Cities Study. The horizontal axis represents the follow-up time, and the vertical axis corresponds to the log hazard ratio (HR), associated with increasing the fine particles level by 18.6 micrograms/m3, which correspond to the difference between the highest and the lowest of the city-specific levels. The solid curve represents the point estimate of the log HR and the dashed curves the pointwise 95% confidence intervals.

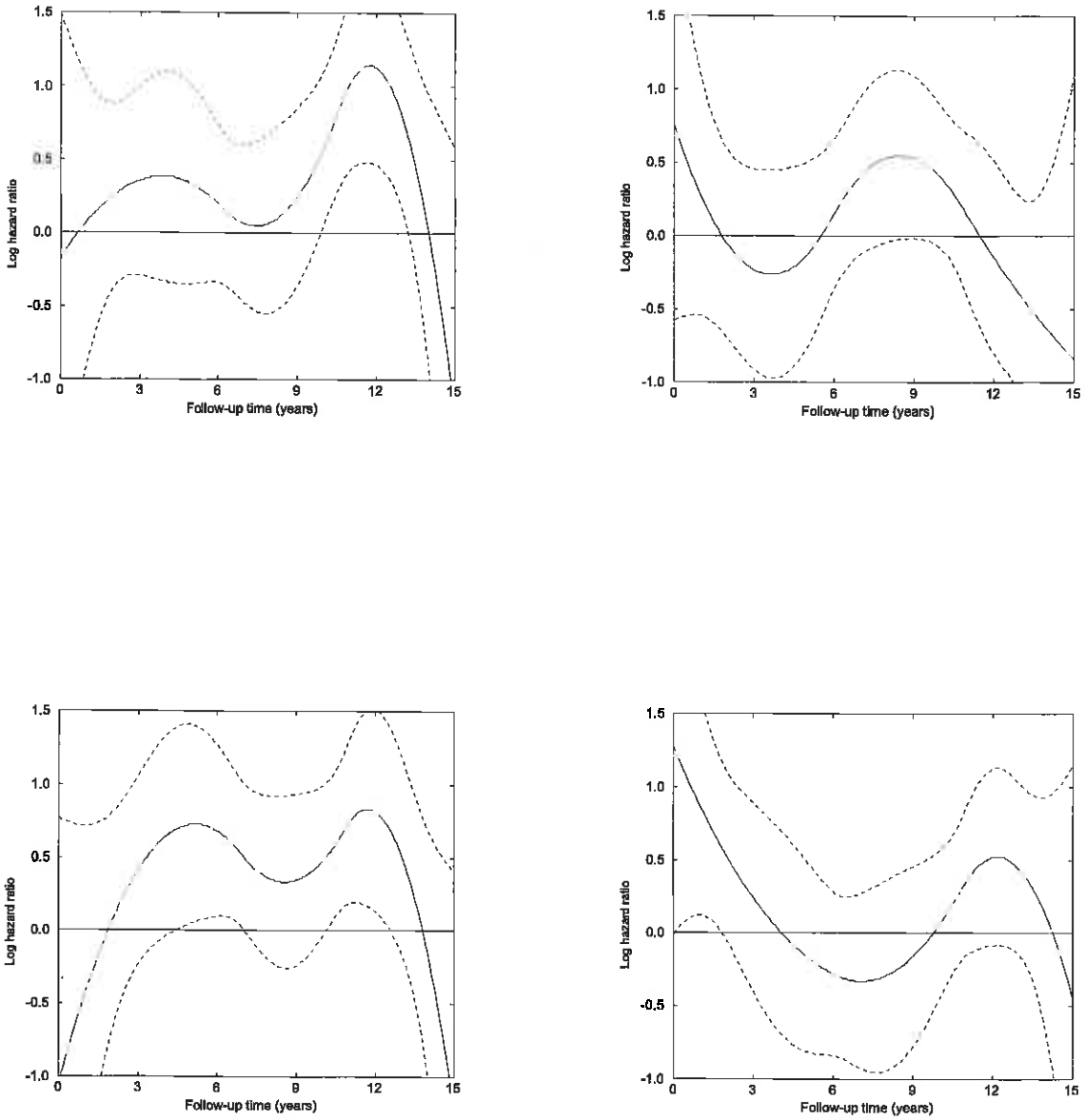

Figure C.3 Flexible quadratic spline estimate (5-df) of the time-dependent effect of sulfates on the log hazard of mortality on four disjoint subsets of the Harvard Six-Cities Study. The horizontal axes represent the follow-up time, and the vertical axes correspond to the log hazard ratio (HR) associated with increasing the fine particles level by 8.0 micrograms/m3, which correspond to the difference between the highest and the lowest of the city-specific levels. The solid curves represents the point estimates of the log HR and the dashed curves the pointwise 95% confidence intervals.

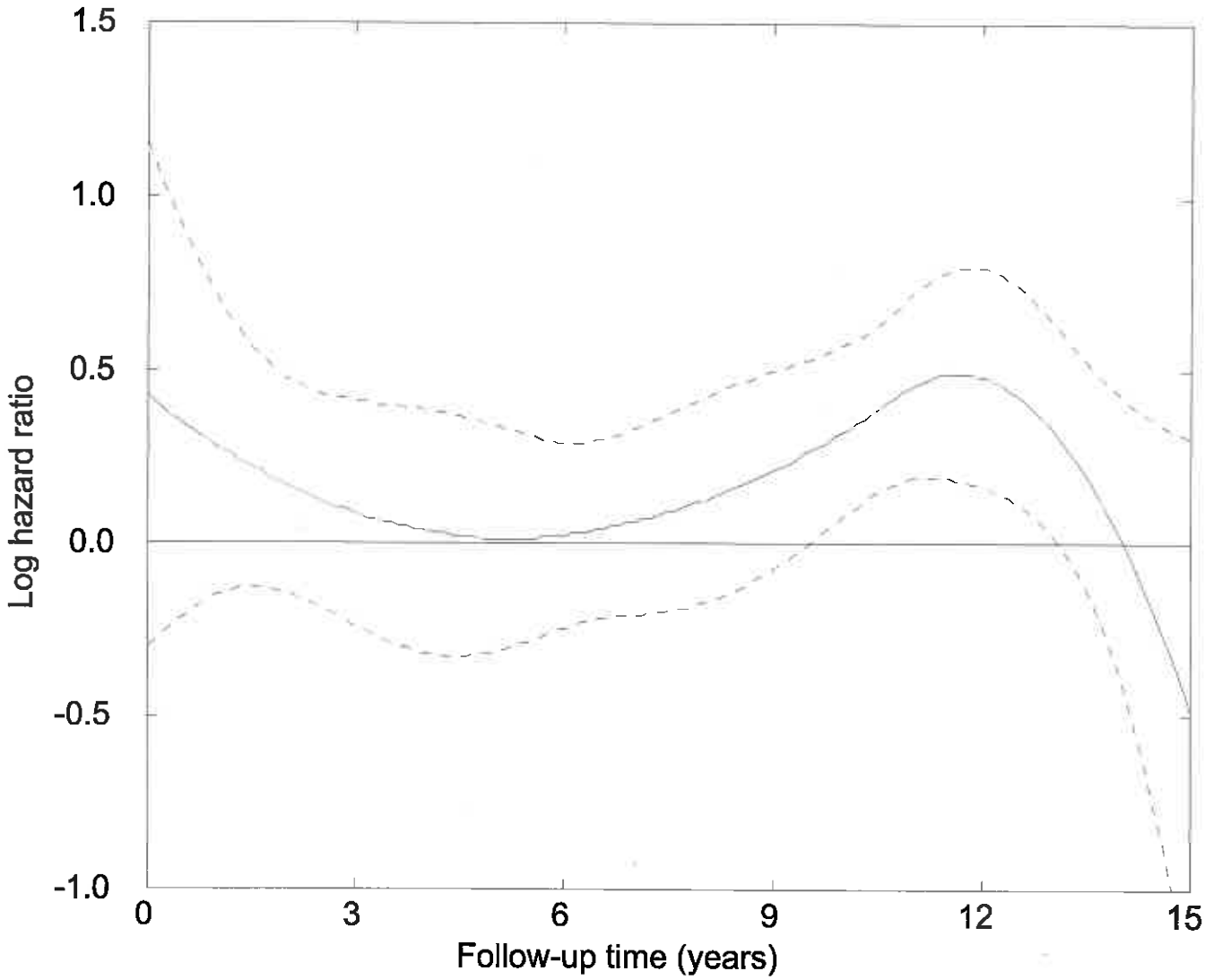

Figure C.4 Flexible quadratic spline estimate (4-df) of the time-dependent effect of sulfates on log hazard of mortality in a "case-cohort" type subset of the Harvard Six-Cities Study. The horizontal axis represents the follow-up time, and the vertical axis corresponds to the log hazard ratio (HR), associated with increasing the fine particles level by 8.0 micrograms/m3, which correspond to the difference between the highest and the lowest of the city-specific levels. The solid curve represents the point estimate of the log HR and the dashed curves the pointwise 95% confidence intervals.

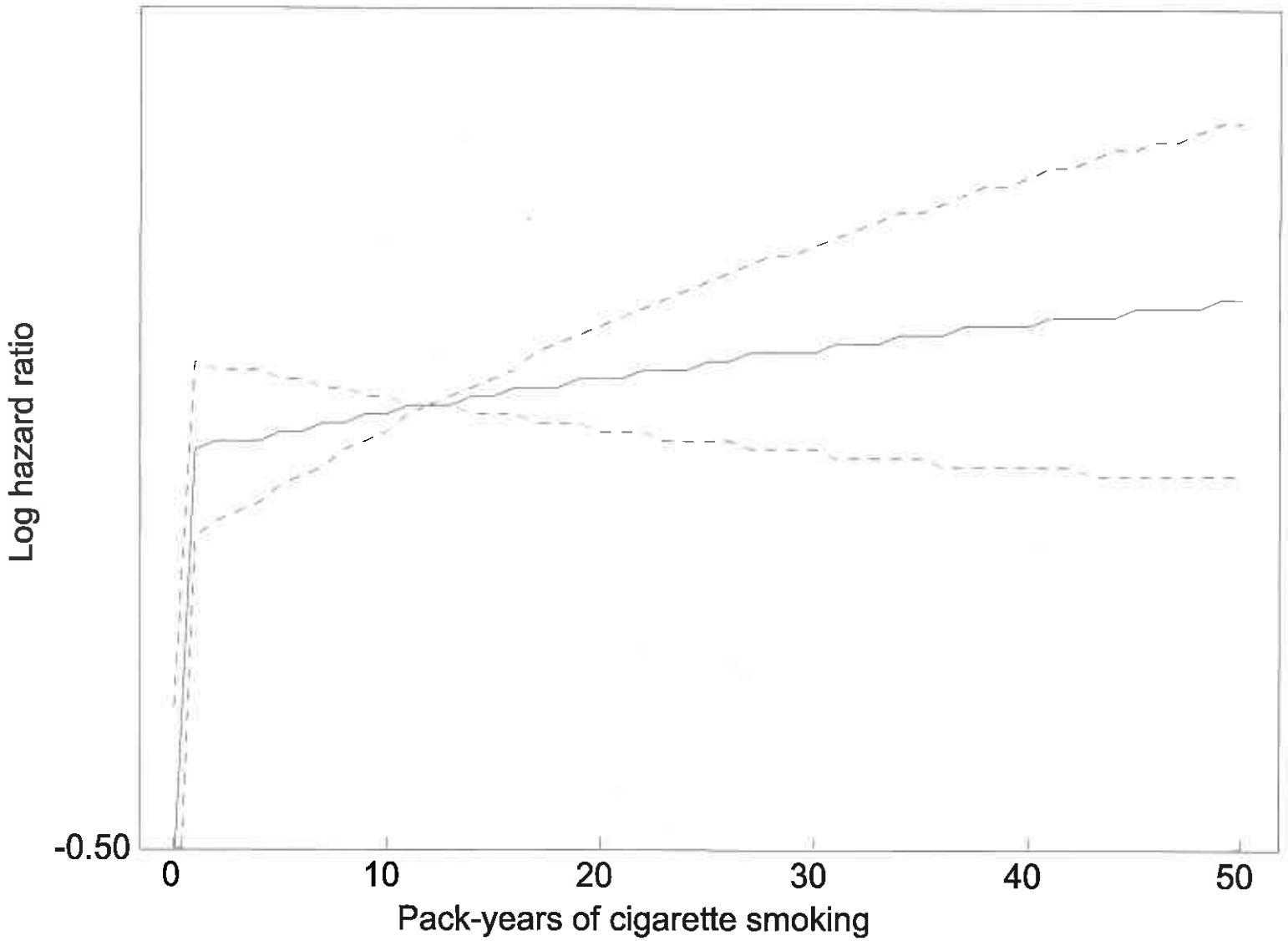

Figure C.5 Flexible quadratic spline (3-df) estimate of the non-linear effect of increasing the level of pack-years of cigarette smoking, for Current Smokers, on the log hazard of mortality in a subset of the Harvard Six-Cities Study. The horizontal axis represents the cumulative exposure to cigarette smoking, in pack-years, and the vertical axis corresponds to the log hazard ratio (HR), with respect to the mean pack-years value (reference value). The solid curve represents the point estimate of the log HR and the dashed curves the pointwise 95% confidence intervals.

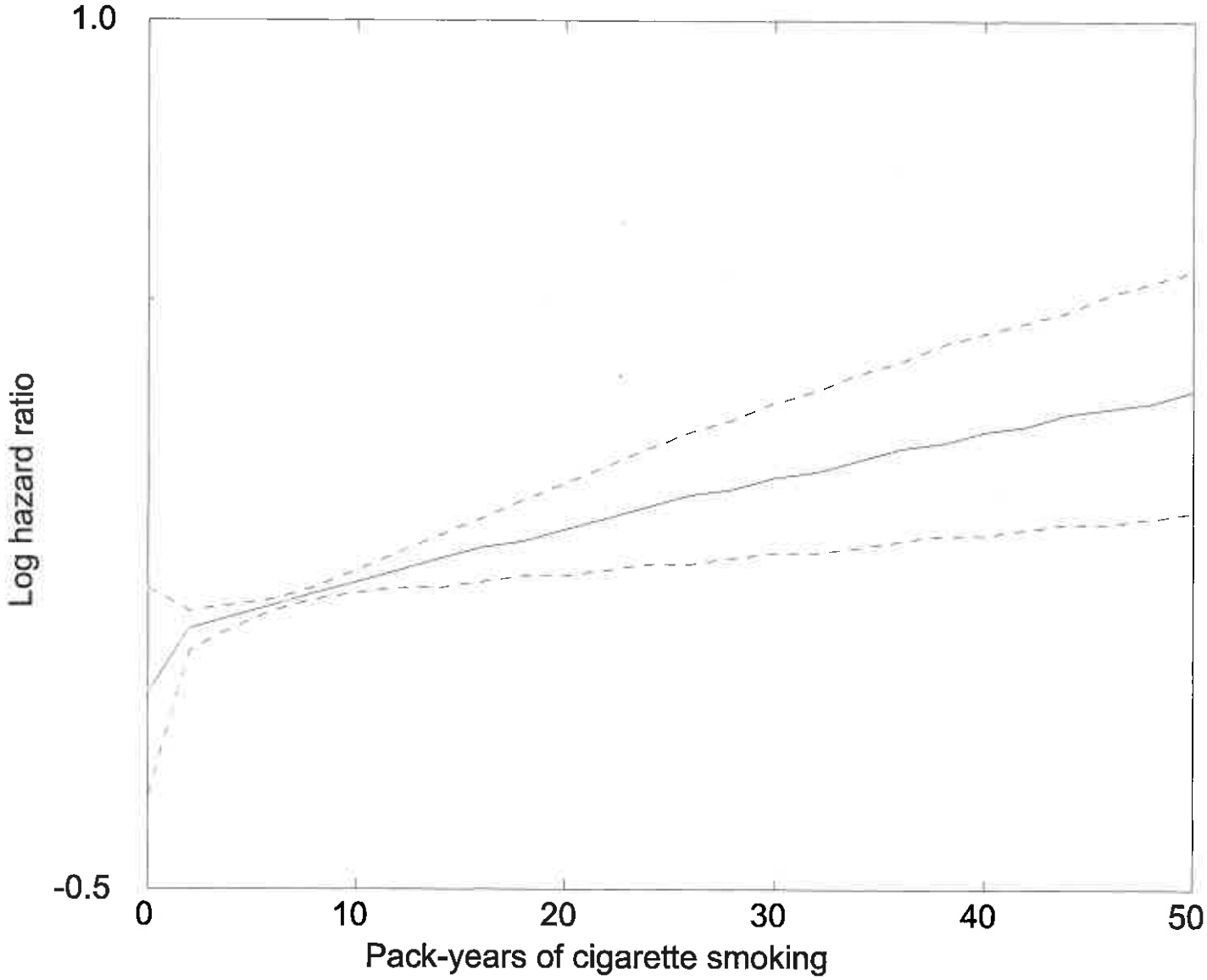

Figure C.6 Flexible quadratic spline (3-df) estimate of the non-linear effect of increasing the level of pack-years of cigarette smoking, for Former Smokers, on the log hazard of mortality in a subset of the Harvard Six-Cities Study. The horizontal axis represents the cumulative exposure to cigarette smoking, in pack-years, and the vertical axis corresponds to the log hazard ratio (HR), with respect to the mean pack-years value (reference value). The solid curve represents the point estimate of the log HR and the dashed curves the pointwise 95% confidence intervals.

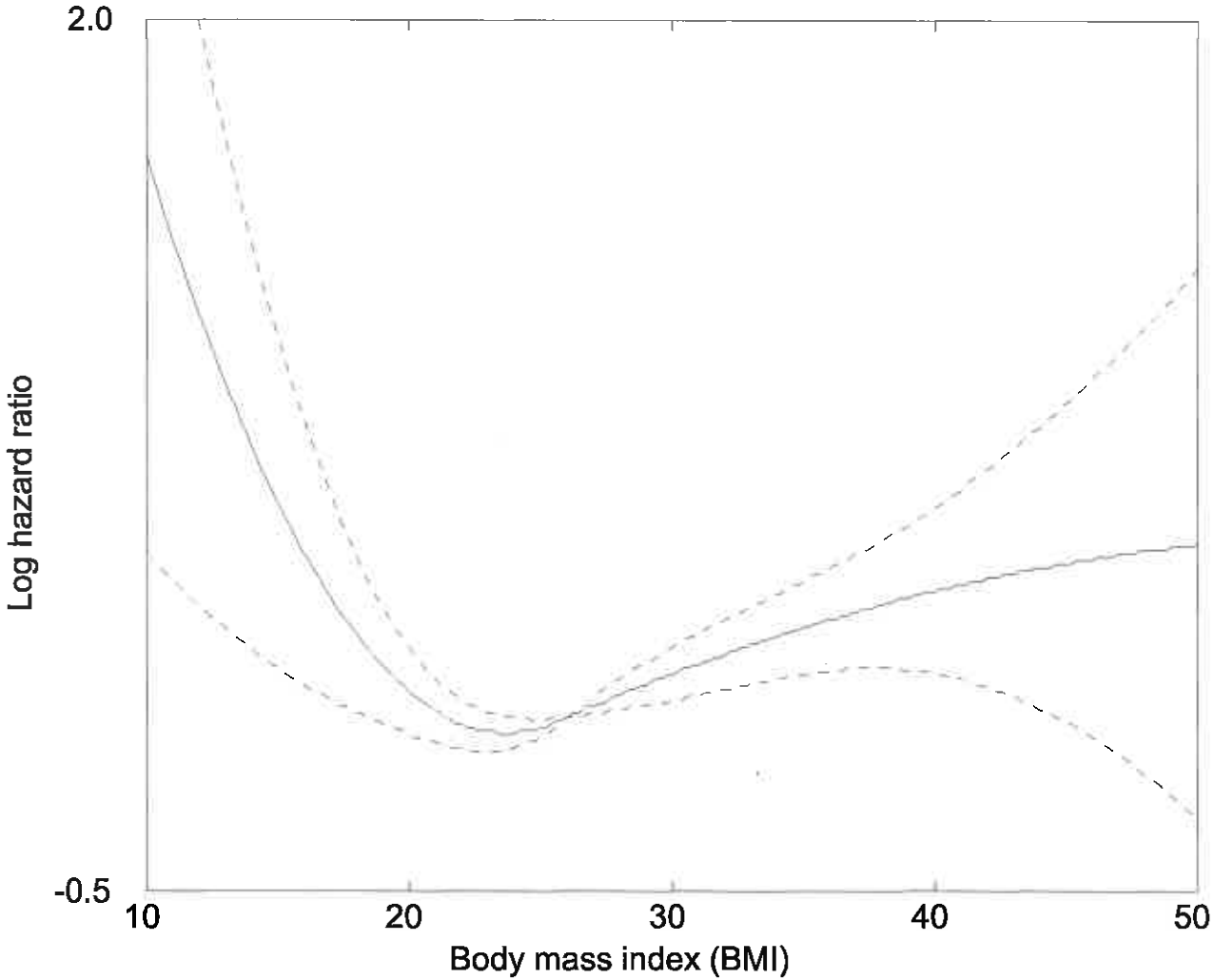

Figure C.7 Flexible quadratic spline (3-df) estimate of the non-linear effect of increasing Body Mass Index (BMI) on log hazard of mortality in a subset of the Harvard Six-cities Study. The horizontal axis represents the BMI, and the vertical axis corresponds to the log hazard ratio (HR), with respect to the mean BMI value (reference value). The solid curve represents the point estimate of the log HR and the dashed curves the pointwise 95% confidence intervals.

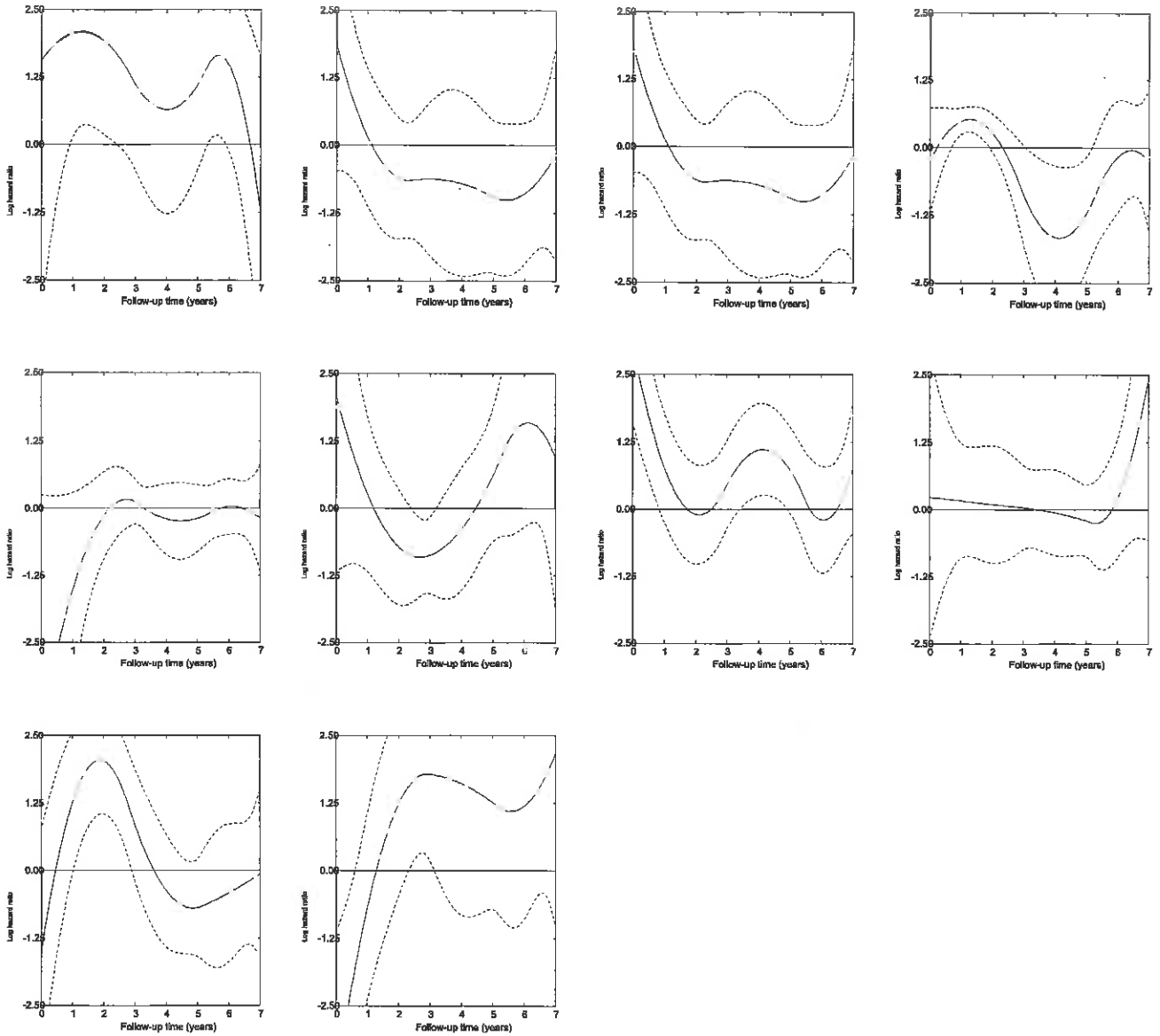

Figure C.8. Flexible quadratic spline estimates (5-df) of the time-dependent effect of Fine Particles on the log hazard of mortality of ten disjoint subsets of the American Cancer Society Study. The horizontal axes represent the follow-up time, and the vertical axes correspond to the log hazard (HR) ratio associated with increasing the Fine Particles level by 24.5 micrograms/m3, which corresponds to the difference between the highest and the lowest of the city-specific levels. The solid curves represent the point estimates of the log HR and the dashed curves represent the pointwise 95% confidence intervals.

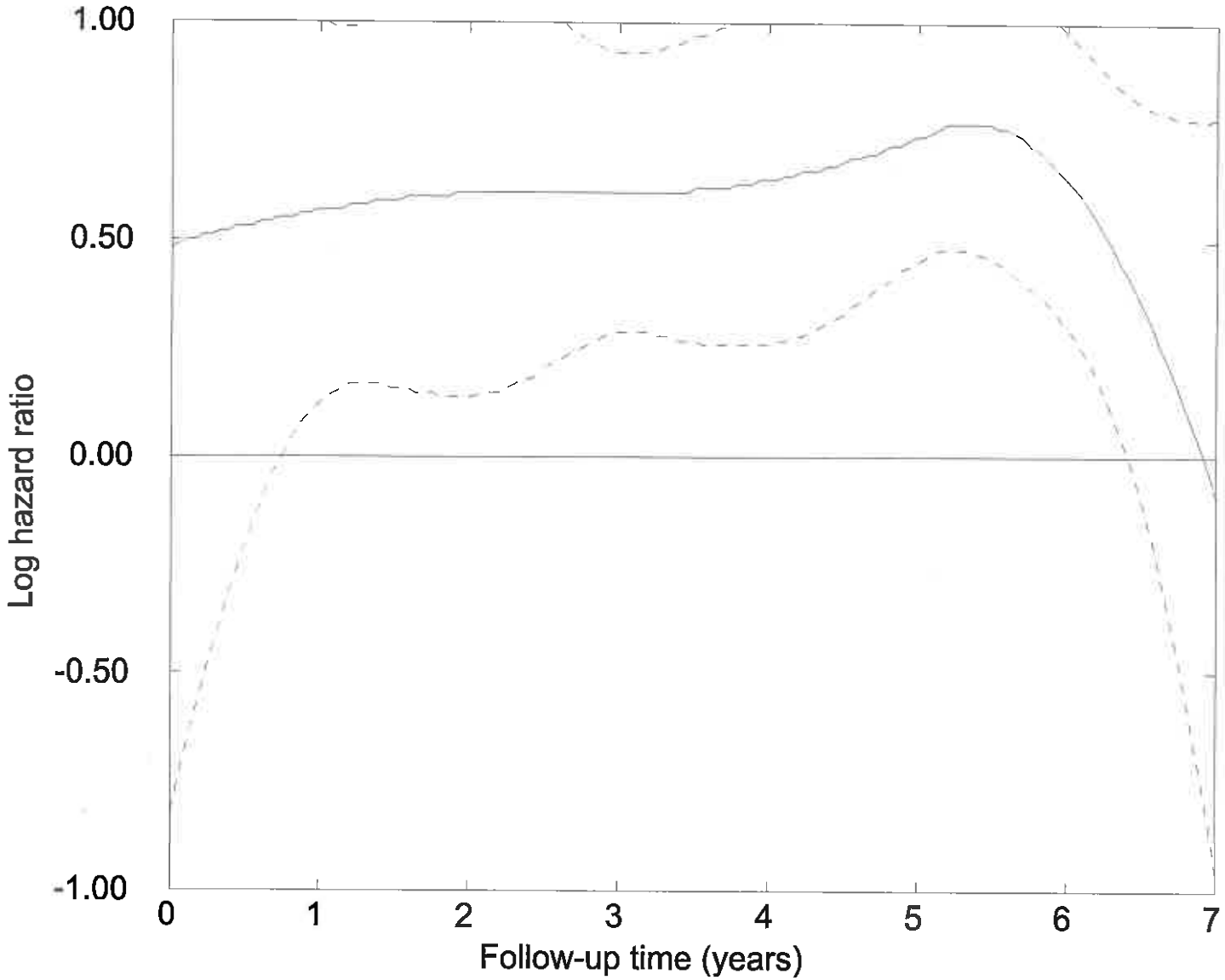

Figure C.9. Flexible quadratic spline estimate (5-df) of the time-dependent effect of Fine Particles on the log hazard of mortality of a "case-cohort" type subset of the American Cancer Society Study. The horizontal axis represents the follow-up time, and the vertical axis corresponds to the log hazard (HR) ratio associated with increasing the Fine Particles level by 24.5 micrograms/m3, which corresponds to the difference between the highest and the lowest of the city-specific levels. The solid curves represent the point estimates of the log HR and the dashed curves represent the pointwise 95% confidence intervals.

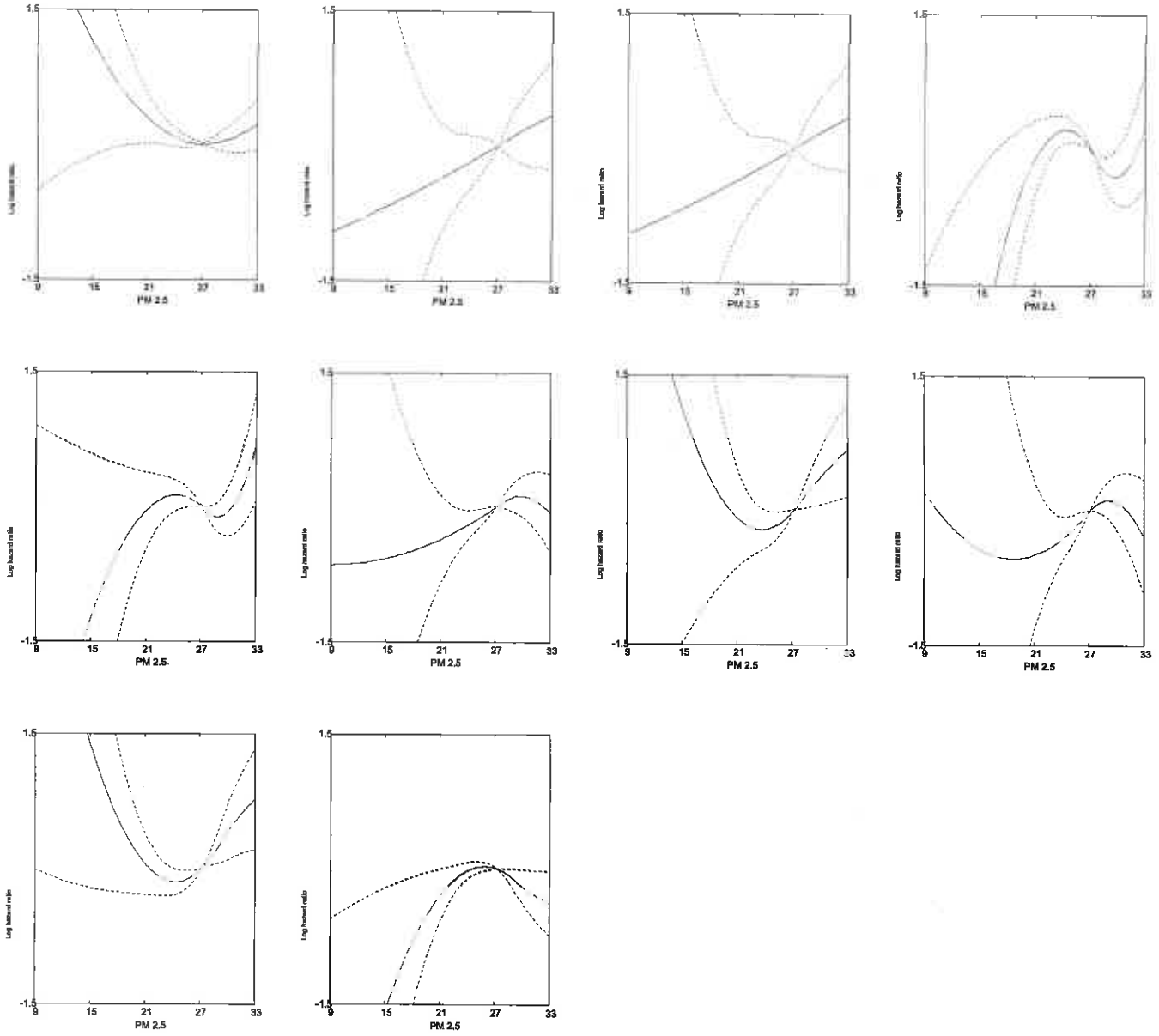

Figure C.10. Flexible quadratic spline (3-df) estimates of the non-linear effect of increasing the level of exposure to Fine Particles on the log hazard of mortality of ten disjoint subsets of the American Cancer Society Study. The horizontal axes represent the cumulative exposure to Fine Particles, and the vertical axes correspond to the log hazard (HR) ratio associated with increasing the Fine Particles level by 24.5 micrograms/m3, which corresponds to the difference between the highest and the lowest of the city-specific levels. The solid curves represent the point estimates of the log HR and the dashed curves represent the pointwise 95% confidence intervals.

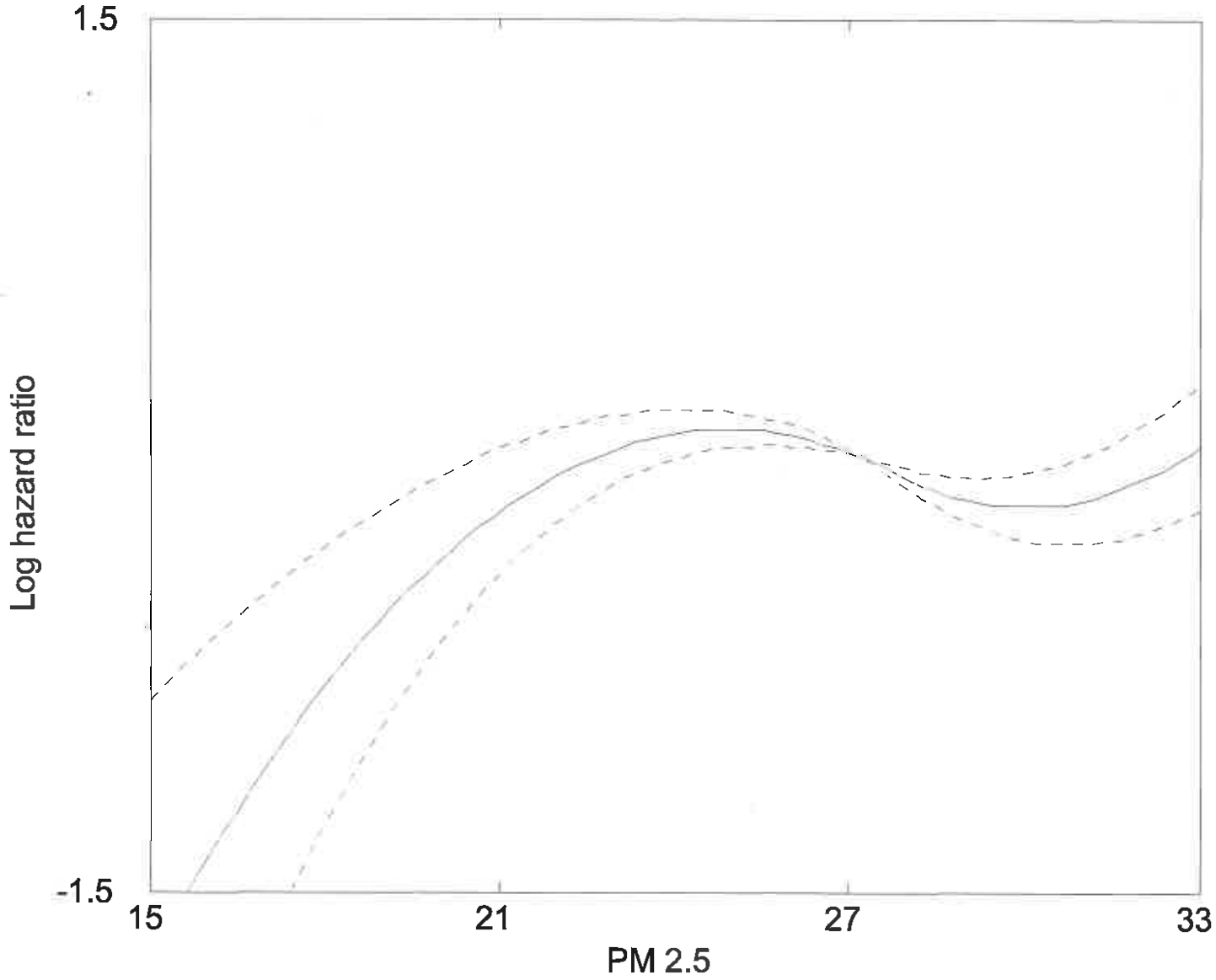

Figure C.11. Flexible quadratic spline (3-df) estimate of the non-linear effect of increasing the level of exposure to Fine Particles on the log hazard of mortality of a "case-cohort" type subset of the American Cancer Society Study. The horizontal axis represents the cumulative exposure to exposure to Fine Particles, and the vertical axis corresponds to the log hazard (HR) ratio associated with increasing the Fine Particles level by 24.5 micrograms/m3, which corresponds to the difference between the highest and the lowest of the city-specific levels. The solid curves represent the point estimates of the log HR and the dashed curves represent the pointwise 95% confidence intervals.

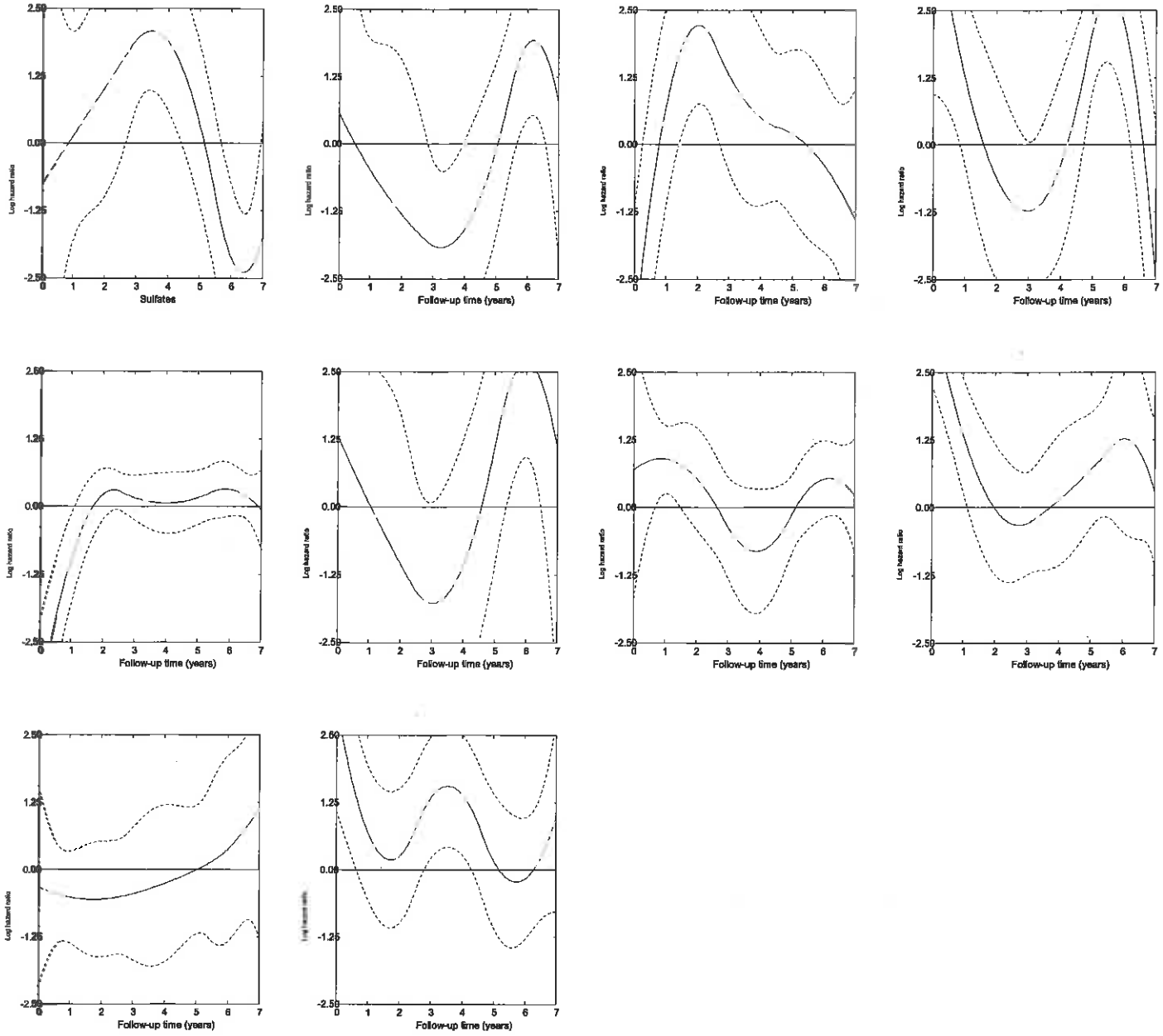

Figure C.12. Flexible quadratic spline estimates (5-df) of the time-dependent effect of Sulfates on the log hazard of mortality of ten disjoint subsets of the American Cancer Society Study. The horizontal axes represent the follow-up time, and the vertical axes correspond to the log hazard (HR) ratio associated with increasing the Sulfates level by 19.9 micrograms/m3, which corresponds to the difference between the highest and the lowest of the city-specific levels. The solid curves represent the point estimates of the log HR and the dashed curves represent the pointwise 95% confidence intervals.

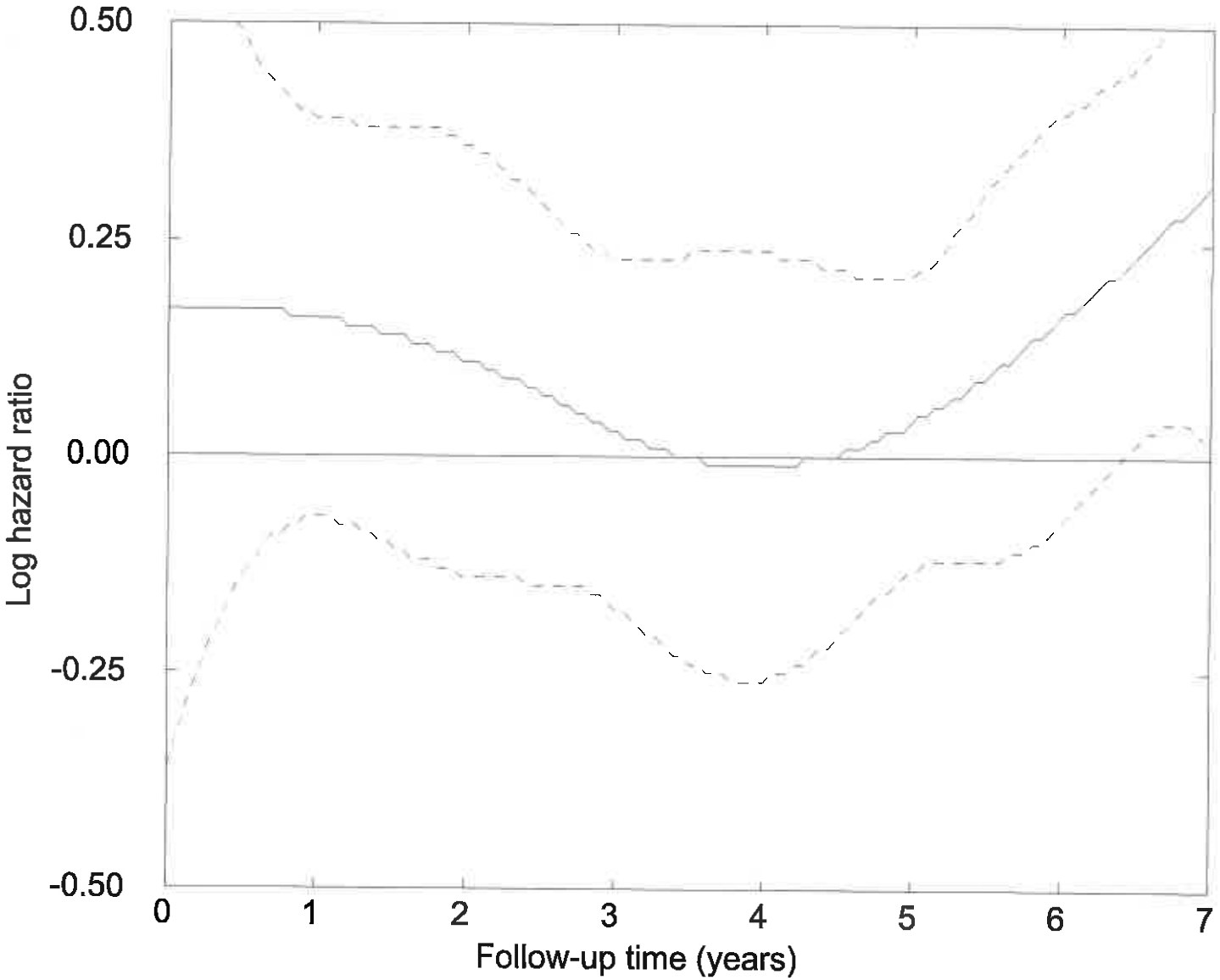

Figure C.13. Flexible quadratic spline estimate (5-df) of the time-dependent effect of Sulfates on the log hazard of mortality of a "case-cohort" type subset of the American Cancer Society Study. The horizontal axis represents the follow-up time, and the vertical axis corresponds to the log hazard (HR) ratio associated with increasing the Sulfates level by 19.9 micrograms/m3, which corresponds to the difference between the highest and the lowest of the city-specific levels. The solid curves represent the point estimates of the log HR and the dashed curves represent the pointwise 95% confidence intervals.

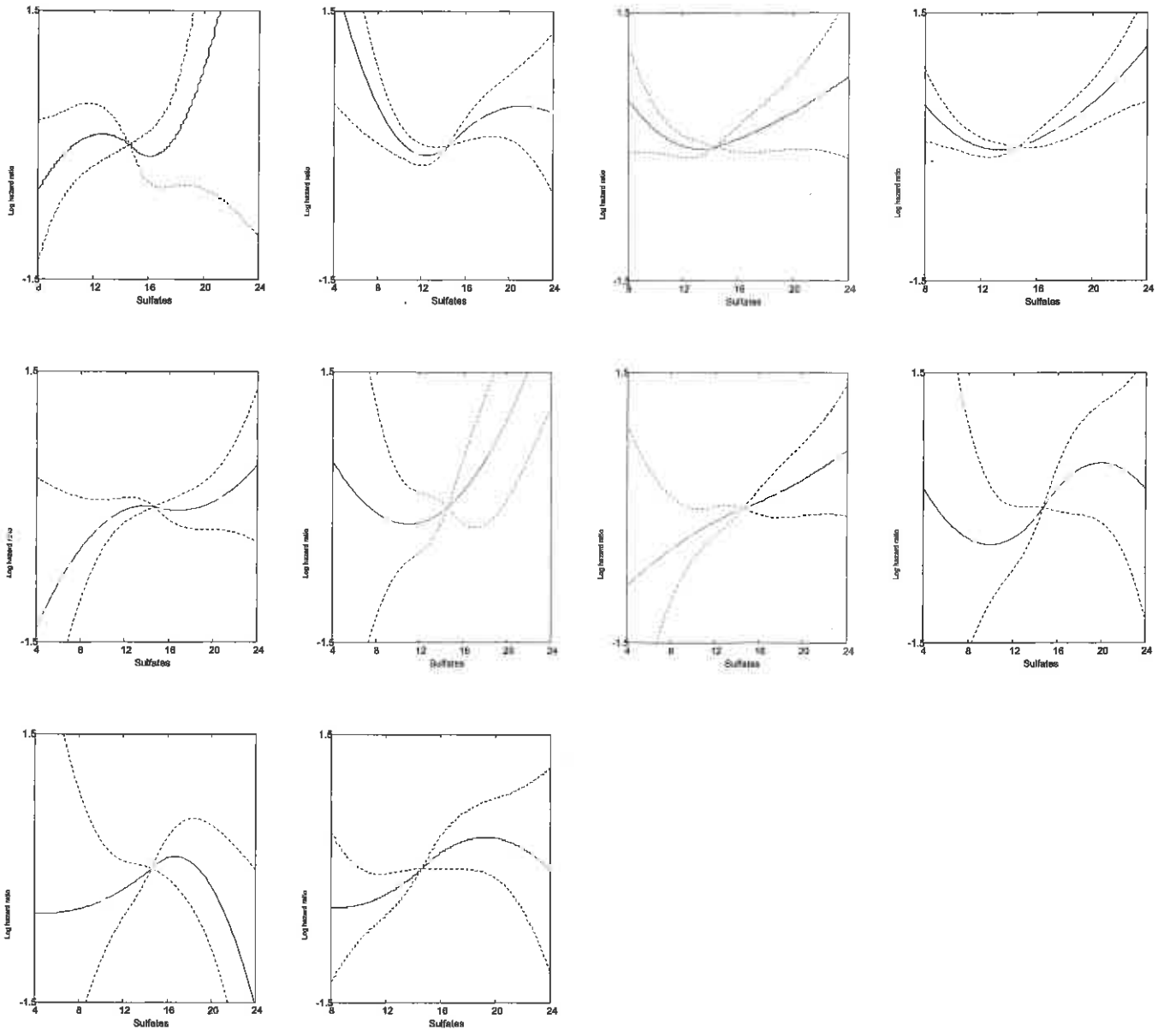

Figure C.14. Flexible quadratic spline (3-df) estimates of the non-linear effect of increasing the level of exposure to Sulfates on the log hazard of mortality of ten disjoint subsets of the American Cancer Society Study. The horizontal axes represent the cumulative exposure to exposure to Sulfates, and the vertical axes correspond to the log hazard (HR) ratio associated with increasing the Sulfates level by 19.9 micrograms/m3, which corresponds to the difference between the highest and the lowest of the city-specific levels. The solid curves represent the point estimates of the log HR and the dashed curves represent the pointwise 95% confidence intervals.

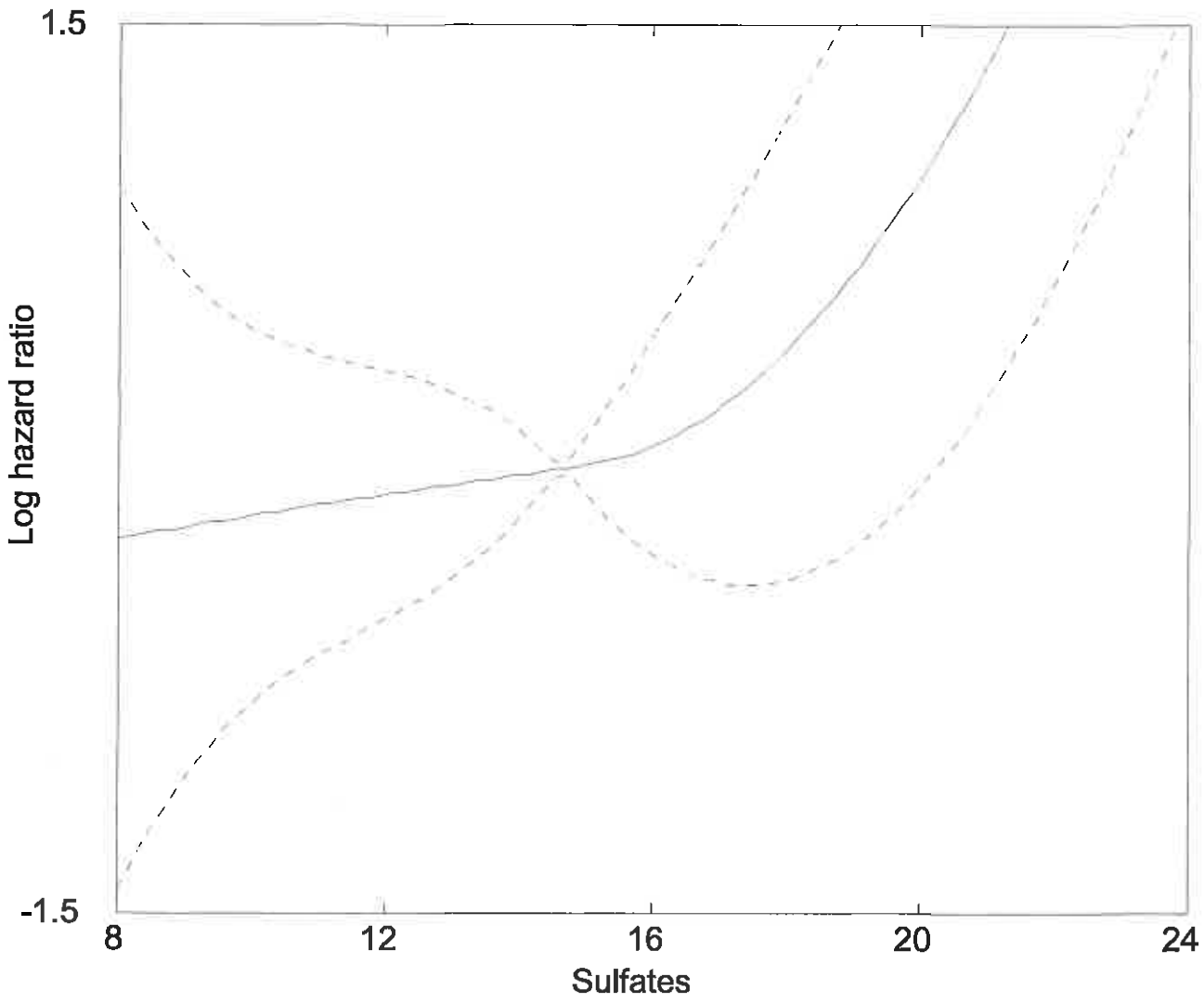

Figure C.15. Flexible quadratic spline (3-df) estimate of the non-linear effect of increasing the level of exposure to Sulfates on the log hazard of mortality of a "case-cohort" type subset of the American Cancer Society Study. The horizontal axis represents the cumulative exposure to exposure to Sulfates, and the vertical axis corresponds to the log hazard (HR) ratio associated with increasing the Sulfates level by 19.9 micrograms/m3, which correspond to the difference between the highest and the lowest of the city-specific levels. The solid curves represent the point estimates of the log HR and the dashed curves represent the pointwise 95% confidence intervals.

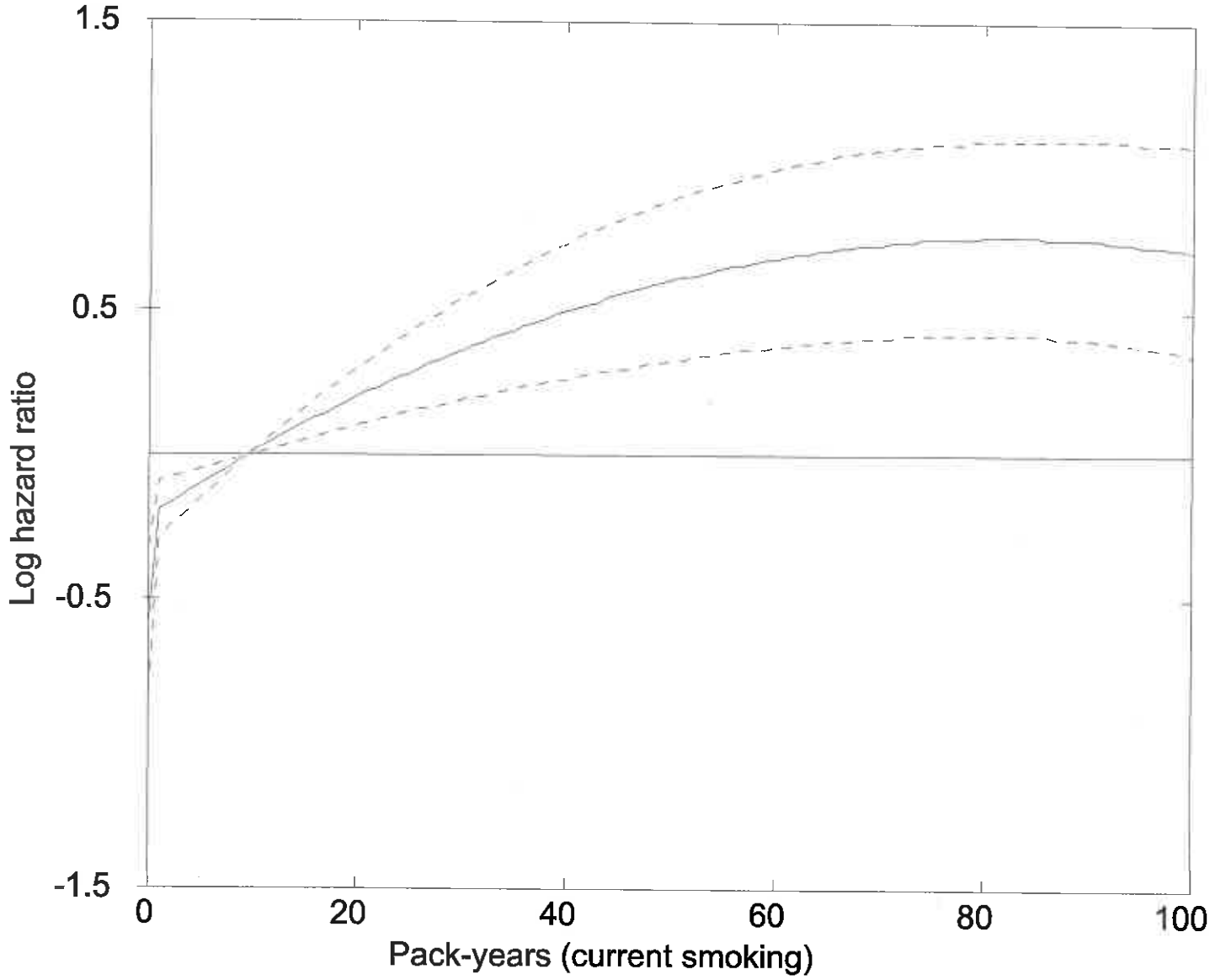

Figure C.16. Flexible quadratic spline (3-df) estimate of the non-linear effect of increasing the level of pack-years of cigarette smoking, for Current Smokers, on the log hazard of mortality of a "case-cohort" type subset of the American Cancer Society Study. The horizontal axis represents the cumulative exposure to cigarette smoke, in pack-years, and the vertical axis corresponds to the log hazard (HR) ratio with respect to the mean pack-years (reference) value. The solid curves represent the point estimates of the log HR and the dashed curves represent the pointwise 95% confidence intervals.

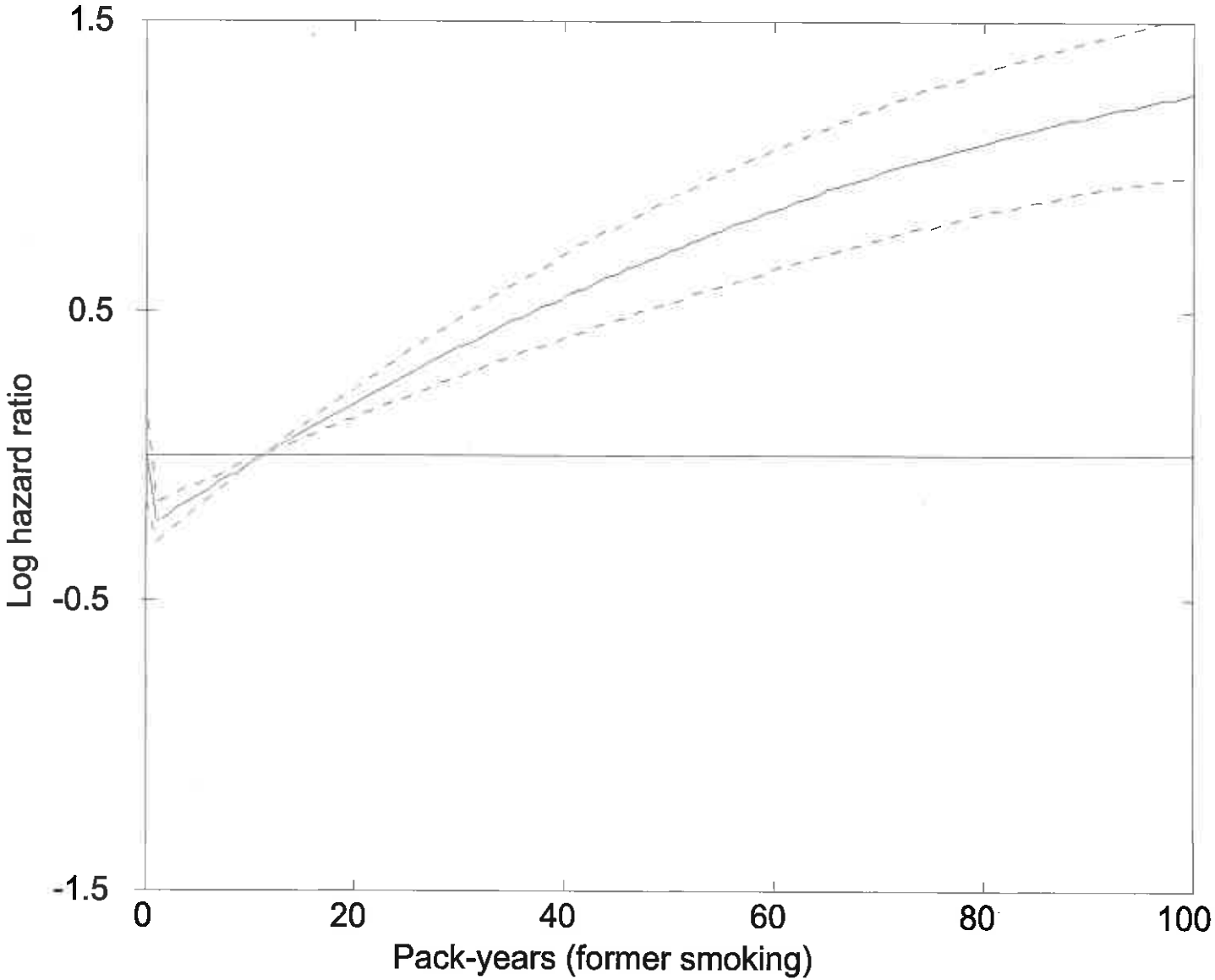

Figure C.17. Flexible quadratic spline (3-df) estimate of the non-linear effect of increasing the level of pack-years of cigarette smoking, for Former Smokers, on the log hazard of mortality of a "case-cohort" type subset of the American Cancer Society Study. The horizontal axis represents the cumulative exposure to cigarette smoke, in pack-years, and the vertical axis corresponds to the log hazard (HR) ratio with respect to the mean pack-years (reference) value. The solid curves represent the point estimates of the log HR and the dashed curves represent the pointwise 95% confidence intervals.

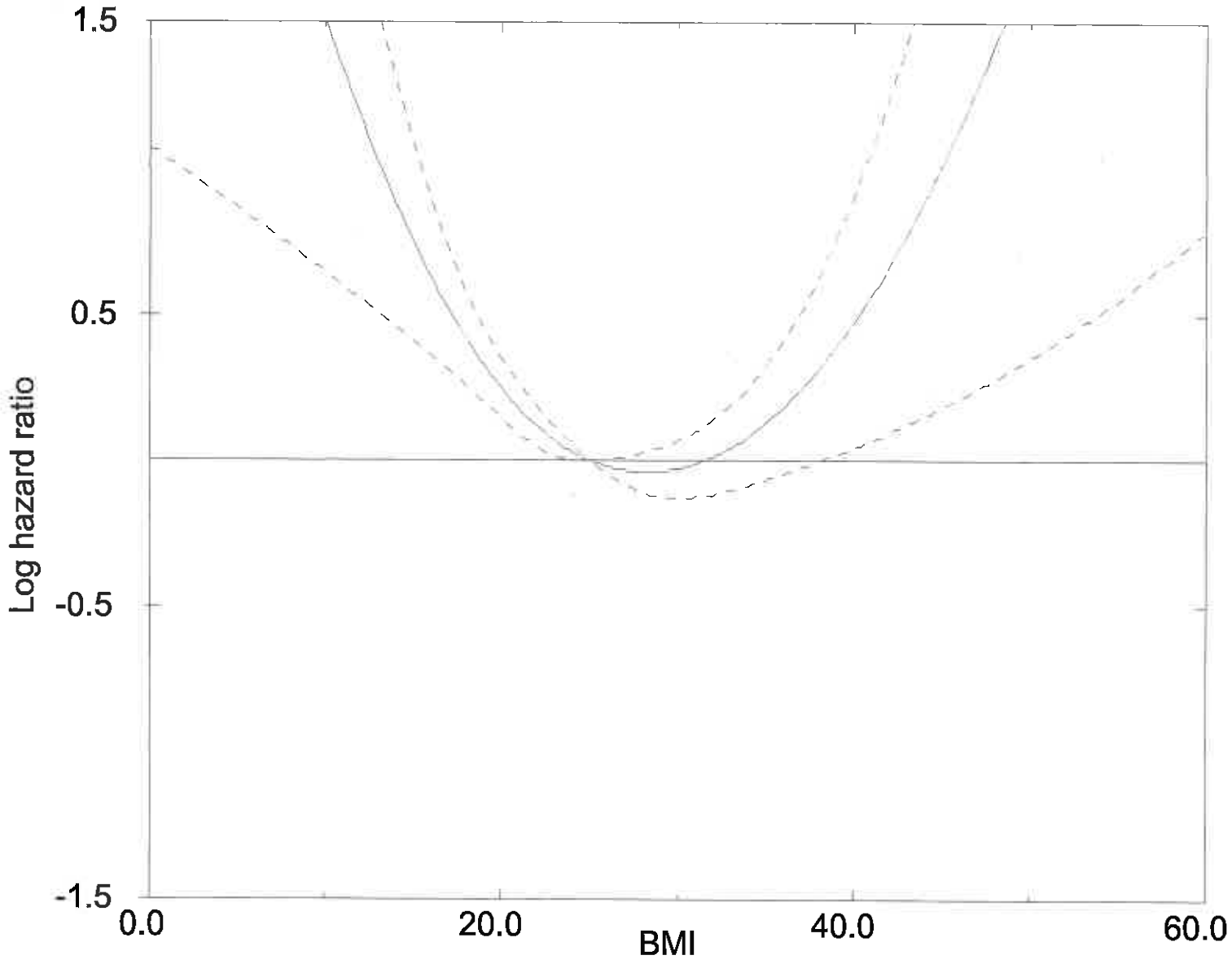

Figure C.18. Flexible quadratic spline (3-df) estimate of the non-linear effect of increasing Body Mass Index on the log hazard of mortality of a "case-cohort" type subset of the American Cancer Society Study. The horizontal axis represents BMI, and the vertical axis corresponds to the log hazard (HR) ratio with respect to the mean BMI (reference) value. The solid curves represent the point estimates of the log HR and the dashed curves represent the pointwise 95% confidence intervals.