eAppendix 1. Direct calculation of bias

Recall that:

$$BIAS_{Lung} = \sum^{t} \omega_t \frac{\{\pi_{1,2,t}\exp(\beta_2) + \pi_{1,3,t}\exp(\beta_3) + (1 - \pi_{1,2,t} - \pi_{1,3,t})\}}{\{\pi_{0,2,t}\exp(\beta_2) + \pi_{0,3,t}\exp(\beta_3) + (1 - \pi_{0,2,t} - \pi_{0,3,t})\}}$$

and,

$$BIAS_{COPD} = \sum^{t} \omega_t \frac{\{\pi_{1,2,t}exp(\theta_2) + \pi_{1,3,t}\exp(\theta_3) + (1 - \pi_{1,2,t} - \pi_{1,3,t})\}}{\{\pi_{0,2,t}\exp(\theta_2) + \pi_{0,3,t}\exp(\theta_3) + (1 - \pi_{0,2,t} - \pi_{0,3,t})\}}.$$

The difference between the true bias correction factor for the target parameter, log($BIAS_{Lung}$), and the quantity log($\mathrm{RR}_{\mathrm{COPD}}^{\mathrm{unadj}}$) indicates the magnitude of bias in the target parameter after this adjustment for confounding by smoking, where $Bias_{COPD} = RR_{COPD}^{unadj}$.

The direct calculations in Figure 2 concern hypothetical study cohorts constituted by a single stratum (t=1) in which the relative rate of lung cancer among current and former smokers was exp(β_2)=23.6 and exp(β_3)=8.7, respectively, and the relative rate of COPD among current and former smokers was exp(θ_2)=12.2 and exp(θ_3)=8.4, respectively.   The proportion of current and former smokers among the unexposed was $\pi_{0,2}$=0.35 and $\pi_{0,3}$=0.31, respectively. Values for log($BIAS_{Lung}$)- log($BIAS_{COPD}$) were calculated for various scenarios regarding the prevalence of smoking among exposed workers (i.e., for various values of $\pi_{1,2}$ and $\pi_{1,3}$).

eAppendix 2. Additive Risk Model

Suppose that an investigator posits that the joint effects of smoking and occupational exposure are additive rather than multiplicative, as implied by the proportional hazards model. Let RD_{Lung}^{unadj} be the rate difference for the association between exposure and lung cancer in the study cohort unadjusted for smoking. Now, let us assume that the lung cancer rate, R_{Lung} conforms to a linear model of the form $R_{Lung}=\gamma_0 + \gamma_1 X_1 + \gamma_2 X_2 + \gamma_3 X_3$, where the target parameter of interest, γ_1, represents the rate difference of lung cancer among smokers contrasted to non-smokers (adjusted for smoking). The bias due to confounding by smoking under a linear rate model is a function of the weighted average of the stratum-specific proportions of current and former smokers among the exposed relative to the unexposed:

$$BIAS_{Lung}^{Additive} = \sum^{t} \omega_t \{(\pi_{1,2,t} - \pi_{1,0,t})\gamma_2 + (\pi_{1,3,t} - \pi_{1,0,t})\gamma_3\}$$

where $\pi_{1,0,t}$ and $\pi_{o,0,t}$ are the proportion of never smokers among the exposed and unexposed workers in covariate stratum, t, and $\gamma_1 = RD_{Lung}^{unadj}$ -$Bias_{Lung}^{Additive}$.

Let RD_{COPD}^{unadj} be the rate difference for the association between exposure and COPD in the study cohort unadjusted for smoking and assume that the COPD rate conforms to a linear model of the form $R_{COPD}=\delta_0 + \delta_1 X_1 + \delta_2 X_2 + \delta_3 X_3$. The bias due to confounding by smoking under a linear rate model is a function of the weighted average of the stratum-specific proportions of current and former smokers among the exposed relative to the unexposed:

$$BIAS_{COPD}^{Additive} = \sum^{t} \omega_t\{(\pi_{1,2,t} - \pi_{0,2,t})\delta_2 + (\pi_{1,3,t} - \pi_{0,3,t})\delta_3\}$$

where $\delta_1 = RD_{COPD}^{unadj} - Bias_{COPD}^{Additive}$.

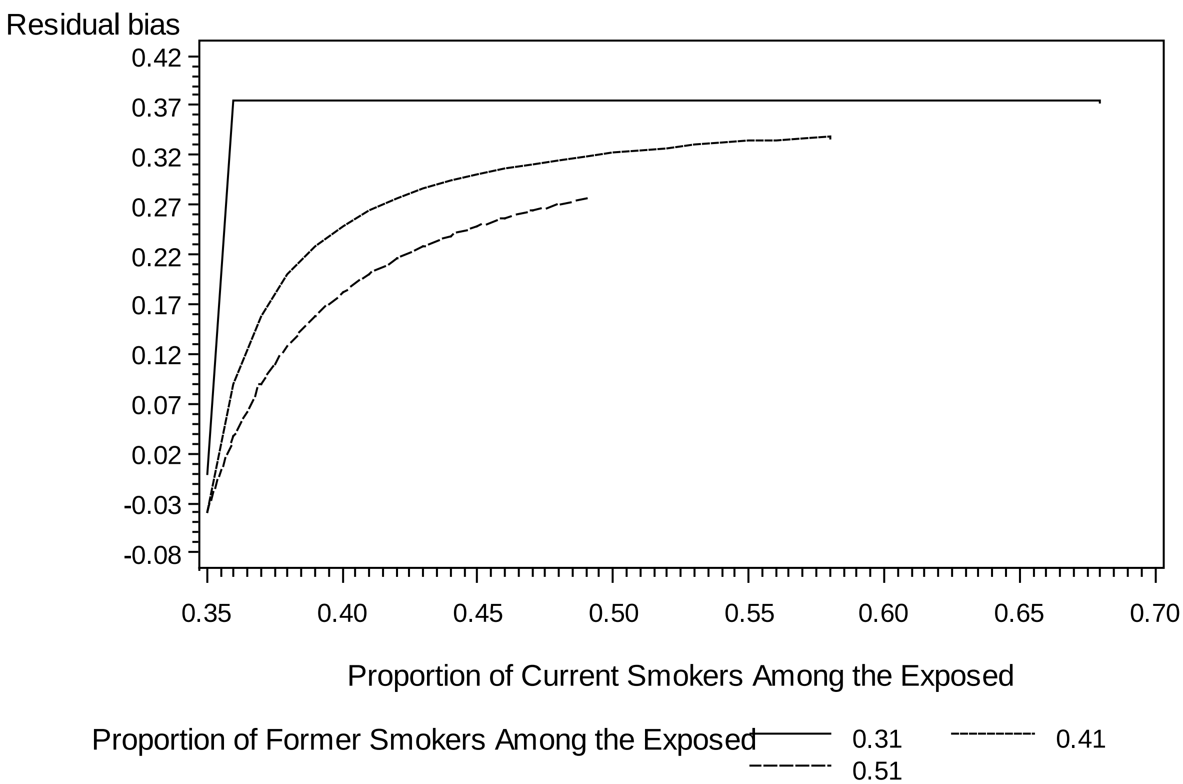
If there is no true causal association between exposure and *COPD* (i.e., $\delta_1 = 0$) then $Bias_{COPD}^{Additive} = RD_{COPD}^{unadj}$. If we use $Bias_{COPD}^{Additive}$ in place of $Bias_{Lung}^{Additive}$ as the correction factor to apply to the crude estimate of the exposure-lung cancer association we will obtain an indirectly adjusted estimate of the association between exposure and lung cancer. The indirectly adjusted estimate of the target parameter under the linear rate model is equal to $\gamma_1 = RD_{Lung}^{unadj} - RD_{COPD}^{unadj}$.

The performance of this indirect approach to adjustment for bias due to confounding can be assessed via direct calculations (eFigure). Illustrative calculations were derived for a hypothetical study in which the rate of lung cancer among current, former, and never smokers was 2.49, 0.68, and 0.17 per 1000 person-years respectively, the rate of COPD among current, former, and never smokers was 1.56, 0.64, and 0.11, respectively per 1000 person-years (values obtained from the UK study of British doctors [8], and the proportion of current and former smokers among the unexposed was 35% and 31%, respectively. This indirect approach results in adjustment for at least 70% of confounding due to smoking under the scenarios considered.

Example

The hypothetical data in Table 3 conform to a multiplicative model. If confounding by smoking were assessed by fitting of a linear rate model, this would incorrectly specify the true model for joint action of these exposures. The crude lung cancer rate difference when contrasting exposed to unexposed was 0.0448 (Table 1) and the crude COPD rate difference was 0.0037 (Table 2). Therefore, the adjusted estimate of the associations between exposure and lung cancer is equal to (0.0448 - 0.0037)=0.0411. Given that the underlying model is multiplicative, calculation of a common rate difference (i.e., common across stratum defined by smoking status) masks the true differences in this quantity between current smokers, former smokers, and never smokers (Table 3).

eFigure. Residual bias* after indirect adjustment for confounding by smoking. Hypothetical cohort study in which the rate of lung cancer among current, former, and never smokers was 2.49, 0.68, and 0.17 per 1000 person-years respectively, and the rate of COPD among current, former, and never smokers was 1.56, 0.64, and 0.11, respectively per 1000 person-years. The prevalence of current and former smokers among the unexposed is 35% and 31%, respectively.



*Residual bias= $(Bias^{Additive}_{Lung} - Bias^{Additive}_{COPD}) / Bias^{Additive}_{Lung}$