Evaluating Bias From Birth-cohort Effects in the Age-based Cox Proportional Hazards Model

Misty J. Hein,[a] *Mary K. Schubauer-Berigan,*[a] *and James A. Deddens*[a,b]

Abstract: The nested case–control design is frequently used to evaluate exposures and health outcomes within the confines of a cohort study. When incidence-density sampling is used to identify controls, the resulting data can be analyzed using conditional logistic regression (equivalent to stratified Cox proportional hazards regression). In these studies, exposure lagging is often used to account for disease latency. In light of recent criticism of incidence-density sampling, we used simulated occupational cohorts to evaluate age-based incidence-density sampling for lagged exposures in the presence of birth-cohort effects and associations among time-related variables. Effect estimates were unbiased when adjusted for birth cohort; however, unadjusted effect estimates were biased, particularly when age at hire and year of hire were correlated. When the analysis included an adjustment for birth cohort, the inclusion of lagged-out cases and controls (assigned a lagged exposure of zero) did not introduce bias.

(*Epidemiology* 2011;22: 249–256)

S ince its use was first illustrated in 1977,[1] the nested case–control study design has been frequently used to evaluate exposures and health outcomes within occupational cohort studies (eg, lung cancer mortality among silica exposed workers[2]) and population-based cohort studies (eg, incident maternal breast cancer and insulin-like growth factor during early pregnancy[3]). Cases are cohort members who develop the outcome of interest, and incidence density or "risk-set" sampling is used to identify eligible controls. In addition to the case, each risk set includes all cohort members who are under observation and outcome-free at the time of the case's failure. The nested case–control study design has been shown to produce unbiased risk estimates under incidence-density sampling, in which controls are sampled at random (without replacement) from each risk set.[4–5] Variants of the nested case–control study design include randomly selecting m controls from the risk set for each case (simple 1:m design) or matching on additional factors such as sex and race (matched 1:m design).

Potential loss of information compared with the full cohort analysis is offset by gains in computational efficiency and reduced study costs. The nested case–control study design requires fewer evaluations of time-dependent covariates (eg, cumulative exposure or cumulative dose), which must be evaluated within risk sets at the case's failure time. Additional cost savings can result if further data collection is required, such as additional questionnaires or analysis of stored biologic samples.

When constructing risk sets, candidates for the time scale include calendar year, age, and time since cohort entry. Age is frequently used, because controlling for confounding from age is a primary concern.[6] Analysis of the age-based risk sets via conditional logistic regression is asymptotically equivalent to a Cox proportional hazards regression analysis with age as the time scale.[7] Use of age as the time scale precisely controls for age-related effects, but not necessarily for other time-related effects, which can be controlled in the analysis by including the appropriate terms in the regression model. This technique is often preferred over further matching of controls to cases.[8]

In studies of chronic diseases including cancer, exposure lagging (in which exposures occurring in the most recent L years are discounted) is often used to account for disease latency.[9] For example, a nested case–control study of incident breast cancer among Finnish flight attendants (1:4 matched on year of birth [± 1 year]) incorporated a 10-year exposure lag period for cumulative occupational radiation dose "to allow for the induction period of at least 10 years for radiation-induced solid tumors."[10] Another nested case–control study (1:5 matching on race and attained age) of male beryllium-exposed workers at a Reading, PA facility (hereafter, "the Reading cohort") considered several metrics of beryllium exposure (duration, cumulative, average, and maximum) and 3 exposure lag periods (0, 10, and 20 years).[11]

A subsequent reanalysis[12] of the nested case–control data identified confounding from birth cohort. Although this confounding was not considered in the original analysis, most positive associations between lagged beryllium exposure and lung cancer remained after adjusting for birth cohort.[12]

Critics of the study design have claimed that the results were biased, in part, due to the use of risk-set sampling in conjunction with exposure lagging.[13,14] Two simulation studies to evaluate the nested case–control study design in conjunction with exposure lagging did not observe bias.[15,16] Langholz and Richardson[15] further described the theoretical basis for nested case–control studies both with and without exposure lagging. These studies have nonetheless been criticized because they did "not fully model important associations" in the Reading cohort.[17]

Here we report the results of a third simulation study that evaluates incidence-density sampling in conjunction with exposure lagging. The main study objective was to determine whether there is bias in cumulative exposure effect estimates from the Cox proportional hazards regression model in simulated cohorts in which risk was related to age, birth cohort, and lagged cumulative exposure. The simulated cohorts included associations among age at hire, year of hire, and exposure intensity. We also compared methods of adjusting for confounding by birth cohort and methods of treating lagged-out cases and controls, and we assessed the effect of incomplete follow-up.

METHODS

Cohorts were simulated based on methods developed by Richardson and Loomis,[18] using SAS Software (version 9.2, SAS Institute Inc., Cary, NC). Each worker was randomly assigned a year of hire and age at hire (described below), a maximum exposure duration (exponential distribution [mean, 5 years]), and, to introduce censoring due to loss to follow-up, a maximum follow-up length (100 years minus an exponential random variable [mean, 15 years]). Three methods of assigning year of hire and age at hire were evaluated:

1. Year of hire was a uniform random variable (1935–1974) and age at hire was 18 years plus a gamma random variable ($\alpha = 1.6$, $\beta = 7.5$).
2. Year of hire and age at hire as in (1) if hire year $\geq$1945; otherwise, age at hire was 18 years plus 8 times an exponential random variable (mean, 1 year) with probability 0.4 or uniform (18–65 years) with probability 0.6.
3. Year of hire and age at hire taken directly from the 3569 workers in the Reading cohort.[11]

The second method was designed to produce an association between year of hire and age at hire that was similar to the Reading cohort.[18] The third method was added because hire year and hire age distributions and associations would exactly mimic the Reading cohort. For each worker, birth year was calculated from the assigned hire age and hire year.

Exposure intensities were randomly assigned to each worker-year, using a lognormal distribution with a geometric mean of 2 units for workers hired in 1935, with a 2.5% decrease in each subsequent year, and a geometric standard deviation of 2.2. Supplemental analyses considered different methods of assigning the geometric mean: constant (one unit) for all calendar years and increasing (one-half unit in 1935 with a 2.5% increase in each subsequent year).

In the Cox proportional hazards model, the effect (ϕ) for a continuous exposure variable is the log hazard ratio for a 1-unit exposure increase. Four risk scenarios defined by the true exposure effect and lag period were considered:

1. $\phi = \ln(1)$.
2. $\phi = \ln(2)$ and lag $= 0$ years.
3. $\phi = \ln(2)$ and lag $= 10$ years.
4. $\phi = \ln(2)$ and lag $= 20$ years.

Each worker was followed from his hire year through his randomly assigned maximum follow-up. At each follow-up year, cumulative exposure under the assigned exposure lag period (CE_L, unit-years) was evaluated as:

$$CE_L(t) = \Sigma \; intensity(t) \text{ for } t \leq \text{year} - L \qquad (1)$$

where, $intensity(t)$ was the randomly assigned exposure intensity for year t, and L was the exposure lag period. Each worker-year was assigned a probability of mortality from lung cancer $h(t)$ and a probability of mortality from all other causes, $c(t)$, that were based on published age- and year-specific death rates for US white males.[19] Probability $h(t)$ was a function of $CE_L(t)$, ϕ, age(t), and birth year, conditional on survival to that age:

$$h(t) = exp[-8.03 + 0.139(\text{age}(t) - 50)$$

$$+ \; -0.00175(\text{age}(t) - 50)^2 + 0.0347(\text{birth year} - 1900)$$

$$+ \; -0.000639(\text{birth year} - 1900)^2 + \phi CE_L(t)/10] \quad (2)$$

In equation (2), cumulative exposure was divided by 10 so that the exposure effect was the log hazard ratio for a 10 unit-year increase in cumulative exposure. Probability $c(t)$ was a function of age(t) and birth year, conditional on survival to that age:

$$c(t) = exp[-4.72 + 0.0603(\text{age}(t) - 50)$$

$$+ \; 0.000423(\text{age}(t) - 50)^2 + \; -0.0113(\text{birth year} - 1900)$$

$$+ \; 0.0000103(\text{birth year} - 1900)^2] \quad (3)$$

Models (2) (without the CE_L term) and (3) each accounted for 98.8% of the variability in the natural-log-trans-

formed US white male death rates due to lung cancer and all other causes, respectively. These probabilities were used to assign Bernoulli random variables indicating the occurrence of death (lung cancer or all other causes combined) in a given follow-up year. Workers not assigned to die were censored at the end of their maximum follow-up. Workers assigned to die were cases if their earliest death was from lung cancer; otherwise, workers were censored at their earliest death.

Complete follow-up included all outcomes and person-time through death or loss to follow-up. Approximately 4% of the subjects were lost to follow-up under the null risk scenario. In addition, a study end year of 1990 was implemented to evaluate the effect of incomplete follow-up. In these analyses, workers who had not died or been lost to follow-up by 1990 were administratively censored at the study end date (31 December 1990). Approximately one-third of deaths were expected to occur through 1990 under the null risk scenario.

Five hundred cohorts were generated for each of the 36 scenarios (3 hire year/age methods, 3 exposure-intensity methods, and 4 risk scenarios). Each cohort, consisting of 4000 (hire year/age methods 1 and 2) or 3569 (method 3) workers, was analyzed using Cox proportional hazards regression (SAS PHREG procedure). Each analysis included all eligible controls (identified using incidence-density sampling by matching on attained age) rather than sample from the risk sets. Initial effect estimates were unadjusted for birth cohort. Three methods of adjusting for birth cohort were then compared:

1. Birth category (<1900 (reference), 1900–1919, 1920–1929, and 1930+), with cutpoints selected based on expected birth-cohort effects.
2. Birth year as a continuous variable.
3. Restricted cubic spline terms (ie, natural splines) for birth year (k = 4 knots at the 5th, 35th, 65th, and 95th percentiles of birth year for the cases).[20]

Cumulative exposure, truncated at the case's age at death, was lagged by zero, 10, or 20 years for cohorts generated with exposure lag periods of zero, 10, or 20 years, respectively. In lagged analyses, 2 methods of treating lagged-out workers were compared:

1. Excluding cases who died in the first 10 (or 20) years of follow-up and, when risk sets were determined, excluding controls with fewer than 10 (or 20) years of follow-up at the case's age at death.
2. Including lagged-out cases and controls (assigned a cumulative exposure of zero unit-years).

Model fit was measured by Akaike's Information Criterion and Schwarz's Bayesian Criterion.[20]

Exposure effect estimates were summarized for each simulation scenario. Summary statistics included the mean of the estimated hazard ratios and the empirical 5% type I error rates (ie, the fraction of cohorts for which the 95% confidence interval for the log hazard ratio did not contain the true value). We used a 2-sided *z*-test for proportions to determine if the empirical 5% type I error rate was significantly different from the nominal 0.05.

RESULTS

As expected, year of hire and age at hire were not correlated in cohorts simulated under hire year/age method 1 (mean correlation: 0.00), whereas hire year and hire age were negatively correlated in cohorts simulated under hire year/age method 2 (mean correlation, −0.18) and method 3 (correlation, −0.13) (Table 1). Also as expected, birth year and unlagged cumulative exposure were not correlated in cohorts simulated with constant exposure intensity (mean correlations, 0.01 to 0.03). In contrast, birth year and unlagged cumulative exposure were negatively correlated in cohorts simulated with decreasing exposure intensities (mean correlations, −0.23 to −0.10) and positively correlated in cohorts with increasing exposure intensities (mean correlations, 0.14 to 0.20). For scenarios with a positive lag, the correlation between birth year and lagged cumulative exposure increased with the lag period. For example, for simulations based on the Reading cohort and a decreasing exposure intensity, mean correlations between birth year and cumulative exposure were −0.10, −0.02, and 0.09 for lag periods of 0, 10, and 20 years, respectively.

Cox proportional hazards regression analyses of cohort data simulated under the decreasing exposure intensity method are summarized in Table 2 for complete follow-up. For hire year/age method 1, there was no confounding by birth cohort and the empirical 5% type I error rates were not inflated. For hire year/age method 2, some confounding was observed for unadjusted analyses, but only for the positive, unlagged scenario (mean hazard ratio 1.95 compared with 2.0; 10.8% empirical type I error rate). For birth-cohort-adjusted analyses, the empirical 5% type I error rates were preserved for category of birth and restricted cubic spline adjustments, but with some minor residual confounding for continuous year-of-birth adjustment. For hire year/age method 3 (Reading cohort), results were similar to method 2, but minor bias was also observed for unadjusted analyses in the positive, 20-year lagged scenario when lagged-out workers were included in the analysis (mean hazard ratio 2.06 compared with 2.0) with an empirical 5% type I error rate of 10.2%. No bias was observed in any analysis that included an adjustment for birth-year category or included restricted cubic spline terms for birth year, regardless of lagging or the treatment of lagged-out workers; some bias was observed for continuous year-of-birth adjustment under a 20-year lag period.

Similar results were observed when outcomes and person-time were administratively censored in 1990 (Table 3).

TABLE 1. Mean Correlation Coefficients for Hire Year and Hire Age $\bar{\rho}_1$ and for Birth Year and Cumulative Exposure $\bar{\rho}_2$, by Hire Year/Age Method, Exposure Intensity, Hazard Ratio, and Lag Period[a]

Exposure Intensity[b]	Hazard Ratio	Lag Period (Years)	Hire Year/Age Method					
			1: Independent		2: Negative		3: Reading Cohort	
			$\bar{\rho}_1$	$\bar{\rho}_2$	$\bar{\rho}_1$	$\bar{\rho}_2$	$\bar{\rho}_1$	$\bar{\rho}_2$
Constant	1	0	0.0	0.01	−0.18	0.02	−0.13	0.03
	2	0	0.0	0.01	−0.18	0.02	−0.13	0.03
		10	0.0	0.04	−0.18	0.08	−0.13	0.10
		20	0.0	0.09	−0.18	0.15	−0.13	0.18
Decreasing	1	0	0.0	−0.23	−0.18	−0.21	−0.13	−0.10
	2	0	0.0	−0.22	−0.18	−0.21	−0.13	−0.10
		10	0.0	−0.19	−0.18	−0.14	−0.13	−0.02
		20	0.0	−0.13	−0.18	−0.04	−0.13	0.09
Increasing	1	0	0.0	0.20	−0.18	0.20	−0.13	0.14
	2	0	0.0	0.20	−0.18	0.21	−0.13	0.15
		10	0.0	0.22	−0.18	0.24	−0.13	0.20
		20	0.0	0.25	−0.18	0.28	−0.13	0.26

[a]Five hundred cohorts were generated for each exposure intensity method, hazard ratio, lag period, and hire year/age method. Each cohort contained 4000 (hire year/age methods 1 and 2) or 3569 (hire year/age method 3) workers.

[b]Exposure intensity was assigned to each worker-year of exposure using a lognormal distribution with geometric standard deviation of 2.2 and a geometric mean that was constant, increasing, or decreasing with calendar year.

However, greater bias was observed in unadjusted analyses in the positive, 20-year lagged scenario for analyses that included lagged-out workers (mean hazard ratio 2.13 compared with 2.0; 21.4% empirical 5% type I error rate) or excluded lagged-out workers (mean hazard ratio 2.07 compared with 2.0; 9.6% empirical 5% type I error rate). For these analyses, including restricted cubic spline terms for birth year produced unbiased estimates.

Results for constant and increasing exposure intensity scenarios (eAppendix, eTables 1–4 [http://links.lww.com/EDE/A452]) were similar, with inflated empirical type I error rates and some bias for unadjusted analyses. The maximum observed bias for unadjusted analyses was approximately 15% (mean hazard ratio 2.31 compared with 2.0) for the scenario based on the Reading cohort, increasing exposure intensity, 20-year lag period (including lagged-out workers), with follow-up through 1990.

Based on Akaike's Information Criterion, analyses that included a birth-cohort adjustment (any of the methods evaluated) generally fit better than unadjusted analyses (results not shown). This was particularly true for cohorts simulated under hire year/age methods 2 and 3 and for complete follow-up. The best-fitting method varied among the adjustment methods evaluated, but the restricted cubic spline method was generally superior to the others. According to Schwarz's Bayesian Criterion, which incurs a greater penalty for increasing the number of model parameters, unadjusted analyses were generally better fitting for cohorts simulated under hire year/age method 1. For cohorts simulated under hire year/age methods 2 and 3, the best fitting method varied,

but continuous metrics (birth year, restricted cubic splines) were generally better fitting than the categorical metrics (category of birth year).

DISCUSSION

Risk-set sampling in conjunction with exposure lagging was evaluated by Langholz and Richardson[15] in a simulation study based on the Colorado Plateau uranium miners cohort. In this study, no bias was observed when a simple random sample of controls was selected from the risk set for each case identified using attained age alone.[15] Hein et al[16] simulated occupational cohorts with time-dependent exposures to evaluate incidence-density sampling in the presence of exposure lagging. In these simulations, exposure effect estimates from analyses that matched on attained age alone were unbiased, and, in lagged analyses, including lagged-out cases and controls (assigned zero lagged exposure) did not introduce bias.[16] Yet, in a rejoinder to these 2 studies, Deubner and Roth[17] suggested that the Reading cohort data "display a complex web of associations" and expressed a reservation "that these simulations do not fully model important associations" in the data. Although the Hein et al[16] simulated cohorts were fairly realistic, with loss-to-follow-up censoring and with mechanisms for assigning dates and causes of death based on age and cumulative exposure, they did not include associations among hire year, hire age, and exposure intensity that have been observed in the Reading cohort.[11]

In our simulations, birth-cohort effects were integrated into the risk model and associations among other time-related variables were introduced by the distributions of hire year,

TABLE 2. Mean Hazard Ratios and Empirical Type I Error Rates ($\alpha = 0.05$) for Cohorts Simulated Under a Decreasing Exposure Intensity for Complete Follow-up (Through Death or Loss to Follow-up), by Hire Year/Age Method, Hazard Ratio, and Lag Period[a]

Hire Year/ Age Method	Hazard Ratio	Lag (Years)	Includes Lagged Out	Mean Hazard Ratio				Empirical Type I Error Rate ($\alpha = 0.05$)[b]			
				Unadjusted	Category-of-birth Adjusted	Year-of-birth Adjusted	Restricted-cubic-spline Adjusted	Unadjusted	Category-of-birth Adjusted	Year-of-birth Adjusted	Restricted-cubic-spline Adjusted
1: Independent	1	0	—	0.99	1.00	1.01	1.00	0.028	0.038	0.038	0.030
	2	0	—	1.97	2.01	2.01	2.01	0.060	0.036	0.034	0.040
	2	10	Yes	1.99	2.01	2.02	2.01	0.064	0.058	0.058	0.050
			No	1.98	2.01	2.02	2.01	0.062	0.056	0.058	0.048
	2	20	Yes	2.00	2.02	2.03	2.01	0.054	0.054	0.056	0.048
			No	2.00	2.02	2.02	2.02	0.044	0.048	0.052	0.040
2: Negative	1	0	—	0.98	1.00	1.01	0.99	0.044	0.034	0.042	0.042
	2	0	—	**1.95**	2.01	2.03	2.00	**0.108**[c]	0.040	0.060	0.034
	2	10	Yes	1.98	2.01	2.04	2.00	0.034	0.044	0.062	0.040
			No	1.97	2.01	2.04	2.01	0.046	0.044	0.062	0.038
	2	20	Yes	2.02	2.02	**2.06**	2.00	0.046	0.054	**0.100**[c]	0.044
			No	2.00	2.02	2.04	2.01	0.032	0.048	0.048	0.044
3: Reading	1	0	—	0.99	1.00	1.02	1.00	0.050	0.056	0.066	0.056
	2	0	—	**1.97**	2.00	2.03	2.00	**0.088**[c]	0.042	0.058	0.048
	2	10	Yes	2.00	2.01	2.04	2.00	0.016	0.020	0.052	0.016
			No	1.99	2.01	2.04	2.01	0.026	0.020	0.044	0.022
	2	20	Yes	**2.06**	2.02	**2.06**	2.01	**0.102**[c]	0.052	**0.104**[c]	0.052
			No	2.02	2.02	**2.05**	2.02	0.062	0.062	**0.090**[c]	0.058

[a]Five hundred cohorts were generated for each combination of hire year/age method and risk scenario. Each cohort contained 4000 (hire year/age methods 1 and 2) or 3569 (hire year/age method 3) workers. Exposure intensity was assigned to each worker-year of exposure using a lognormal distribution with a geometric standard deviation of 2.2 and a geometric mean that decreased 2.5% for each calendar year.

[b]The empirical type I error rate ($\alpha = 0.05$) is the fraction of cohorts for which the 95% confidence interval for the log hazard ratio does not contain the true value.

[c]Significantly elevated (based on the z-test for proportions) empirical type I error rates (ie, type I error rate greater than 5% with a 2-sided $P < 0.05$) are indicated in bold-face type. Actual *P* values for the indicated results are <0.0001, 0.0002, 0.0027, 0.0001, 0.0001, and 0.0018 for type I error rates of 0.108, 0.100, 0.088, 0.102, 0.104, and 0.090, respectively.

TABLE 3. Mean Hazard Ratios and Empirical Type I Error Rates (α = 0.05) for Cohorts Simulated Under a Decreasing Exposure Intensity for Incomplete Follow-up (Through the Earlier of Death, Loss to Follow-up, or Administrative Censoring in 1990) by Hire Year/Age Method, Hazard Ratio, and Lag Period[a]

Hire Year/ Age Method	Hazard Ratio	Lag (Years)	Includes Lagged Out	Mean Hazard Ratio				Empirical Type I Error Rate (α = 0.05)[b]			
				Unadjusted	Category-of-birth Adjusted	Year-of-birth Adjusted	Restricted-cubic-spline Adjusted	Unadjusted	Category-of-birth Adjusted	Year-of-birth Adjusted	Restricted-cubic-spline Adjusted
1: Independent	1	0	—	0.99	1.00	1.01	1.00	0.038	0.036	0.038	0.032
	2	0	—	1.98	2.01	2.03	2.01	0.040	0.038	0.044	0.034
	2	10	Yes	1.99	2.00	2.02	2.00	0.044	0.046	0.046	0.042
			No	1.99	2.01	2.02	2.01	0.036	0.048	0.046	0.040
	2	20	Yes	2.04	2.02	2.03	2.01	0.052	0.054	0.056	0.052
			No	2.03	2.03	2.04	2.03	0.048	0.052	0.054	0.050
2: Negative	1	0	—	0.99	1.00	1.01	1.00	0.040	0.038	0.040	0.034
	2	0	—	1.97	2.00	2.03	2.00	0.036	0.020	0.040	0.024
	2	10	Yes	2.03	2.02	2.03	2.00	0.046	0.044	0.052	0.040
			No	2.01	2.02	2.03	2.01	0.038	0.044	0.050	0.036
	2	20	Yes	**2.09**	2.03	2.03	2.00	**0.102**	0.062	0.066	0.054
			No	2.03	2.03	2.04	2.02	0.056	0.060	0.058	0.050
3: Reading	1	0	—	1.00	1.00	1.00	1.00	0.062	0.060	0.062	0.056
	2	0	—	2.00	2.00	2.01	2.00	0.056	0.072	0.062	0.060
	2	10	Yes	2.06	2.01	2.01	2.00	0.072	0.034	0.036	0.034
			No	2.04	2.01	2.02	2.01	0.046	0.036	0.032	0.034
	2	20	Yes	**2.13**[c]	2.04	**2.02**	2.00	**0.214**[c]	0.068	**0.076**[c]	0.058
			No	**2.07**	**2.04**	2.04	2.02	**0.096**[c]	**0.080**[c]	0.070	0.072

[a]Five hundred cohorts were generated for each combination of hire year/age method and risk scenario. Each cohort contained 4000 (hire year/age methods 1 and 2) or 3569 (hire year/age method 3) workers. Exposure intensity was assigned to each worker-year of exposure using a lognormal distribution with a geometric standard deviation of 2.2 and a geometric mean that decreased 2.5% for each calendar year.
[b]The empirical type I error rate (α = 0.05) is the fraction of cohorts for which the 95% confidence interval for the log hazard ratio does not contain the true value.
[c]Significantly elevated (based on the z-test for proportions) empirical type I error rates (ie, type I error rate greater than 5% with a 2-sided P < 0.05) are indicated in bold-face type. Actual P values for the indicated results are 0.0001, <0.0001, 0.028, 0.0005, and 0.013 for type I error rates of 0.102, 0.214, 0.076, 0.096, and 0.080, respectively.

hire age, and exposure intensity. That is, there was potential for bias from confounding factors; however, the presence of birth-cohort effects did not necessarily require adjustment in the analysis to obtain unbiased exposure effect estimates. Indeed, for simulated cohorts with age at hire independent of year of hire, unadjusted analyses fit the best and resulted in little bias, regardless of the method of assigning exposure intensities. For simulated cohorts with hire age negatively correlated with hire year, adjustment for birth cohort was sometimes necessary to obtain unbiased estimates. In these simulations, however, failure to control for birth-cohort effects did not result in substantial bias. The maximum bias observed for any scenario was only 15% in the hazard ratio. Furthermore, the need for adjustment was not limited to scenarios for which the exposure-response relation was lagged. Incorporating an exposure lag period requires consideration of the treatment of lagged-out cases and controls. In these simulations, the inclusion of lagged-out cases and controls did not introduce bias as long as the analysis included an adjustment for birth cohort. Excluding lagged-out cases and controls reduced, but did not necessarily eliminate, bias associated with unadjusted analyses.

In these simulations, controls were not "matched" to cases on birth year; rather, birth-cohort confounding was controlled by including terms in the regression model because adjusting for differences in potential confounders is generally recommended over matching.[8] Options for adjusting for continuous confounders include categorization, the inclusion of higher-order terms, and the use of regression splines. In these simulations, the type I error rate was always preserved and results were unbiased when restricted cubic splines were used to adjust for birth cohort. Generally, including terms for category of birth year was unbiased, but including birth year as a linear term in the model occasionally resulted in bias (likely from residual confounding). This was not unexpected because birth-cohort effects are highly nonlinear. Here, the risk model (based on actual death rates for US white men) included first- and second-order terms for birth year. Restricted cubic splines have been promoted as one method to minimize residual confounding when adjusting for a continuous covariate.[21–23]

Compared with Hein et al,[16] the cohorts simulated here were more realistic because the risk model incorporated calendar-year effects and because exposure intensity was varied over time and within-worker. In addition, relations among year of hire and age at hire approximated (method 2) or represented exactly (method 3) relations in the Reading cohort. Greater degrees of bias were observed for simulated cohorts based on the Reading cohort compared with method 2, likely due to differences in the distribution of hire year (uniform in the simulated cohorts, but highly right skewed in the Reading cohort).

Simulation studies have been used to evaluate the nested case–control study design in a variety of settings.[24–26] In fact, Deubner et al[27] conducted their own "empirical evaluation" of incidence-density sampling in conjunction with exposure lagging using the same Reading cohort data. In their evaluation, "case" status was randomly assigned to 142 cohort subjects, all subjects were assigned an average exposure of 20 units, and incidence-density sampling was used to identify 5 controls from the risk set for each case. Because they observed that cumulative exposure was higher among "cases" compared with "controls" when a 20-year lag period was implemented, they concluded that "the study design produced a biased case–control lagged exposure difference under the null hypothesis and could not distinguish qualitatively between null and alternate [sic] hypotheses."[27]

This "evaluation," however, was fundamentally flawed.[28,29] Langholz and Richardson[28] noted that the empirical evaluation detected a "real association between length of employment and age at exit from the study" because case status was randomly assigned to workers but age at exit and duration of employment were not. Because their explanation did not consider birth cohort, we provide here an alternative. Because Deubner et al[27] randomly assigned case status to the entire cohort, workers who had not died by the study end date of 31 December 1992 were allowed to be selected as cases. Each person in the risk set for a case who was alive at the end of follow-up must have been born earlier than the case. For example, if a case was 50 years old and alive at the study end date, then every control in the risk set must have been either (i) alive as of 31 December 1992 and at least 50 years old on that date or (ii) deceased prior to this date and at least 50 years old at death. In both situations, the control would have had to have been born before 31 December 1942 (the case's birthdate). Thus, as long as a significant proportion of the cohort had not died by study end, controls will have an average date of birth earlier than cases using the methods of Deubner et al.[27] In the Reading cohort, approximately 55% of the cohort was still alive as of the study end date; consequently, the empirical evaluation of Deubner et al[27] actually introduced a birth-cohort effect, that was not considered in their analysis. We replicated their results (data not shown) and then considered an alternative analysis that randomly assigned lung cancer case status to "deceased," rather than all, cohort members. In the alternative analysis, there was no association between lagged exposure and case status. This flaw in the design of the empirical evaluation[27] negates the authors' conclusions regarding the nested case–control study design with lagged exposures.

Furthermore, the "*P* values" reported in Table 2 in Deubner et al[27] are invalid because they were obtained by performing *t* tests on the combined data set from their 10 trials. Combining data from the 10 trials artificially inflates the sample size by a factor of 10, and hence decreases the *P*

values, thus making the results appear to be more significant than they might otherwise be. It would be more correct to perform and summarize 10 *t* tests, or, even better, 10 Cox regressions. Additionally, Monte Carlo simulation studies are generally conducted with more than 10 simulations.

The nested case–control study design has been recommended as an efficient study design.[30–32] Theoretical arguments have shown that the nested case–control study design is unbiased under simple random sampling of controls (without replacement) from the risk set for each case, as long as incidence-density sampling is used.[7,33] In addition, the nested case–control study design has been shown to be unbiased for other sampling strategies that incorporate matching on potential confounders or counter-matching on an exposure surrogate (as long as the appropriate sampling weights are included).[34,35] Although these simulations analyzed the full cohort rather than simple nested case–control samples, the results lend support to the unbiased properties of exposure effect estimates obtained from incidence-density sampling.

In summary, for simulated occupational cohorts with birth-cohort effects and time-related associations among year of hire, age at hire, and exposure intensity, incidence-density sampling with matching on attained age was shown to produce unbiased exposure effect estimates as long as the analysis included an adjustment for birth cohort. The results of this study lend support to the use of incidence-density sampling and the nested case–control study design. We observed biased effect estimates when there was no adjustment for birth cohort, particularly when year of hire and age at hire were correlated. Effect estimates were unbiased when adjusted for birth-cohort effects, and, as before, including lagged-out cases and controls did not introduce bias.

REFERENCES

1. Liddell FD, McDonald JC, Thomas DC, Cunliffe SV. Methods of cohort analysis: appraisal by application to asbestos mining. *J R Stat Soc A*. 1977;140:469–491.
2. Steenland K, Sanderson W. Lung cancer among industrial sand workers exposed to crystalline silica. *Am J Epidemiol*. 2001;153:695–703.
3. Chen T, Lukanova A, Grankvist K, et al. IGF-I during primiparous pregnancy and maternal risk of breast cancer. *Breast Cancer Res Treat*. 2010;121:169–175.
4. Lubin JH, Gail MH. Biased selection of controls for case-control analyses of cohort studies. *Biometrics*. 1984;40:63–75.
5. Robins JM, Gail MH, Lubin JH. More on "biased selection of controls for case-control analyses of cohort studies." *Biometrics*. 1986;42:293–299.
6. Breslow NE, Lubin JH, Marek P, Langholz B. Multiplicative models and cohort analysis. *J Am Stat Assoc*. 1983;78:1–12.
7. Borgan Ø, Goldstein L, Langholz B. Methods for the analysis of sampled cohort data in the Cox proportional hazards model. *Ann Stat*. 1995;23:1749–1778.
8. Checkoway H, Demers PA. Occupational case-control studies. *Epidemiol Rev*. 1994;16:151–162.
9. Checkoway H, Pearce N, Kriebel D. *Research Methods in Occupational Epidemiology*. 2nd ed. New York: Oxford University Press; 2004.
10. Kojo K, Pukkala E, Auvinen A. Breast cancer risk among Finnish cabin attendants: a nested case-control study. *Occup Environ Med*. 2005;62:488–493.
11. Sanderson WT, Ward EM, Steenland K, Petersen MR. Lung cancer case-control study of beryllium workers. *Am J Ind Med*. 2001;39:133–144.
12. Schubauer-Berigan MK, Deddens JA, Steenland K, Sanderson WT, Petersen MR. Adjustment for temporal confounders in a reanalysis of a case-control study of beryllium and lung cancer. *Occup Environ Med*. 2008;65:379–383.
13. Deubner DC, Lockey JL, Kotin P, et al. Re: lung cancer case-control study of beryllium workers [letter]. *Am J Ind Med*. 2001;40:284–285.
14. Levy PS, Roth HD, Deubner DC. Exposure to beryllium and occurrence of lung cancer: a reexamination of findings from a nested case-control study. *J Occup Environ Med*. 2007;49:96–101.
15. Langholz B, Richardson D. Are nested case-control studies biased? *Epidemiology*. 2009;20:321–329.
16. Hein MJ, Deddens JA, Schubauer-Berigan MK. Bias from matching on age at death or censor in nested case-control studies. *Epidemiology*. 2009;20:330–338.
17. Deubner DC, Roth HD. Rejoinder: progress in understanding the relationship between beryllium exposure and lung cancer. *Epidemiology*. 2009;20:341–343.
18. Richardson DB, Loomis D. The impact of exposure categorisation for grouped analyses of cohort data. *Occup Environ Med*. 2004;61:930–935.
19. Hein MJ. *Occupational Cohort Studies and the Nested Case-Control Study Design* [doctoral dissertation]. Cincinnati, OH: Department of Mathematical Sciences, University of Cincinnati; 2009.
20. Harrell FE Jr. *Regression Modeling Strategies with Applications to Linear Models, Logistic Regression, and Survival Analysis*. New York: Springer; 2001.
21. Slama R, Werwatz A. Controlling for continuous confounding factors: non- and semiparametric approaches. *Rev Epidemiol Sante Publique*. 2005;53(Spec No 2):2s65–2s80.
22. Desquilbet L, Mariotti F. Dose-response analyses using restricted cubic spline functions in public health research. *Stat Med*. 2010;29:1037–1057.
23. Steenland K, Deddens JA. A practical guide to dose-response analyses and risk assessment in occupational epidemiology. *Epidemiology*. 2004;15:63–70.
24. Langholz B, Thomas DC. Efficiency of cohort sampling designs: some surprising results. *Biometrics*. 1991;47:1563–1571.
25. Essebag V, Platt RW, Abrahamowicz M, Pilote L. Comparison of nested case-control and survival analysis methodologies for analysis of time-dependent exposure. *BMC Med Res Methodol*. 2005;5:5. doi: 10.1186/1471–2288–5–5.
26. Wang MH, Shugart YY, Cole SR, Platz EA. A simulation study of control sampling methods for nested case-control studies of genetic and molecular biomarkers and prostate cancer progression. *Cancer Epidemiol Biomarkers Prev*. 2009;18:706–711.
27. Deubner DC, Roth HD, Levy PS. Empirical evaluation of complex epidemiologic study designs: workplace exposure and cancer. *J Occup Environ Med*. 2007;49:953–959.
28. Langholz B, Richardson D. Re: empirical evaluation of complex epidemiologic study designs: workplace exposure and cancer [letter]. *J Occup Environ Med*. 2008;50:1–2.
29. Wacholder S. Bias in full cohort and nested case-control studies? *Epidemiology*. 2009;20:339–340.
30. Wacholder S, Silverman DT, McLaughlin JK, Mandel JS. Selection of controls in case-control studies: III. Design options. *Am J Epidemiol*. 1992;135:1042–1050.
31. Ernster VL. Nested case-control studies. *Prev Med*. 1994;23:587–590.
32. Checkoway H, Pearce N, Kriebel D. Selecting appropriate study designs to address specific research questions in occupational epidemiology. *Occup Environ Med*. 2007;64:633–638.
33. Goldstein L, Langholz B. Asymptotic theory for nested case-control sampling in the Cox regression model. *Ann Stat*. 1992;20:1903–1928.
34. Langholz B, Clayton D. Sampling strategies in nested case-control studies. *Environ Health Perspect*. 1994;102(suppl 8):47–51.
35. Langholz B, Borgan Ø. Counter-matching: a stratified nested case-control sampling method. *Biometrika*. 1995;82:69–79.