American Journal of Epidemiology
Vol. 143, No. 2
Printed in U.S.A.

Negative Bias in Exposure-Response Trends in Occupational Studies: Modeling the Healthy Worker Survivor Effect

Kyle Steenland,[1] James Deddens,[1] Alberto Salvan,[1,2] and Leslie Stayner[1]

Many occupational studies analyze trends between cumulative exposure and mortality. The authors show that such trends are, in general, negatively confounded by employment status. Mortality rates for workers who leave work ("inactive" workers) are higher than for active workers because some workers leave because they are ill. The percentage of inactive relative to active person-time is higher in low categories of cumulative exposure, causing employment status to act as a negative confounder of exposure-response trends (the opposite occurs for time-since-hire). We illustrate these phenomena using 10 "negative" mortality studies, in which adjustment for employment status removes false trends. However, adjustment for employment status will lead to biased estimates when it acts as an intermediate variable between cumulative exposure and death, as occurs directly when exposure causes a disabling disease that, in turn, causes death or indirectly when exposure causes workers to leave work. The authors illustrate this problem using simulated follow-up data for leaving, disease incidence, and mortality. In the null case in which cumulative exposure affects neither disease incidence (or mortality) nor leaving rates, employment status indeed acts as a negative confounder of exposure-response trends, and traditional adjustment eliminates this confounding. However, when cumulative exposure affects disease incidence or rates of leaving, adjustment for employment status will not be adequate. Employment status falls under the general rubric of variables that are simultaneously confounders and intermediate variables. *Am J Epidemiol* 1996;143:202–10.

bias; cohort studies; healthy worker effect; occupations

Observed mortality rates for workers after leaving employment at a work site under study are approximately double the rates when workers are actively employed (1–3). This phenomenon is known as the healthy worker survivor effect, one component of the healthy worker effect (the other being the selection of healthy workers into the workforce). The increase in mortality for those who become "inactive" is presumably due to some proportion of those who are inactive leaving work because they are ill, although others may go on to work elsewhere. That some workers leave because they are ill is suggested by empirical data showing that mortality is particularly elevated in the first year or two after leaving (1).

In analyzing cumulative exposure-response relations in cohort studies, employment status (active/inactive status) will generally act as a negative confounder because those with little cumulative exposure also tend to have a higher percentage of inactive person-time, which has higher mortality.

Traditional methods of controlling for a confounder (e.g., stratification) may effectively remove these false-negative trends with cumulative exposure. However, such methods may result in biased estimates of the exposure effect if employment status acts either as a true intermediate variable (exposure causes disabling disease, which then results in mortality) or only as an indirect intermediate variable (cumulative exposure causes workers to leave work, workers who leave work have higher mortality, although exposure itself does not cause higher mortality). Robins et al. (4, 5) and Weinberg (6) have discussed this problem of variables acting as both confounders and intermediate variables, and Robins et al. (5) has suggested some methods to adjust for this phenomenon.

Here we present both empirical and simulated data that illustrate these points. The simulated data consider disease incidence as well as mortality and allow for some workers to leave employment because they are ill, some to leave because of death, some to leave for reasons unrelated to health, and others to stay on the job.

Received for publication August 31, 1994, and in final form October 20, 1995.

[1] National Institute for Occupational Safety and Health, 4676 Columbia Parkway, Cincinnati, OH.
[2] LADSEB-CNR, National Research Council, Corso Stati Uniti 4, Padova, Italy I-35127.
Reprint requests to Dr. Kyle Steenland, National Institute for Occupational Safety and Health (NIOSH), R-13, 4676 Columbia Parkway, Cincinnati, OH 45226.

MATERIALS AND METHODS

For our empirical data, we have used the same data from our earlier work (1). Briefly, the data are taken from 10 published mortality studies in which no occupational risks were observed by the authors. The studies included 89,000 workers and 1,980,000 person-years. Person-years were divided between active and inactive, with inactive being defined as beginning 1 week after the end of employment (1 week was used to allow for inaccuracies in personnel records for the actual day last employed). In these studies, employment and exposure were the same since all employed workers were considered exposed. Poisson regression with the EGRET (7) package was used to analyze the data.

For our simulated data, we considered 10,000 workers followed for 40 years. All workers entered the study at the same time, the time when they began employment. Everyone was exposed when first employed. The focus of analysis was the effect of cumulative exposure, a continuous variable. Employment time and cumulative exposure were identical. After entering employment, workers could leave work due to incident disease, leave work without incident disease, die while employed, or stay at work. After leaving work, they could die subsequent to incident disease, develop the disease but survive, die not subsequent to incident disease, or survive disease-free until the end of the study. For each worker, we generated a time until stopping work (t_s), a time until getting the disease (t_d), a time until death after prior incident disease (t_{d1}), and a time until death irrespective of incident disease (t_{d2}). Workers with incident disease stopped work. The study period was 40 years. The unit of time was years. The focus of the analysis was on total mortality.

Each worker worked a minimum of t_s, t_d, t_{d2}, or 40 years. If t_{d2} or $t_d + t_{d1}$ was less than 40, then the worker died under study. Person-time was divided into active and inactive categories. Person-time data were generated by SAS (8), and Poisson regression was conducted using SAS PROC GENMOD (8). Cumulative exposure was divided into yearly intervals and treated as a continuous variable in Poisson regression. While we present standard errors for results from simulated data, the reader should be aware that they are arbitrary and a function of the sample size used in the simulation.

Three cases were examined in the simulations (figure 1). In the first case, called the "null-null" case, incident disease, mortality, and leaving rates were not a direct function of cumulative exposure (duration of exposure). Times t_s, t_d, t_{d1}, and t_{d2} were exponentially distributed (SAS RANEXP)(8), with means of 25,

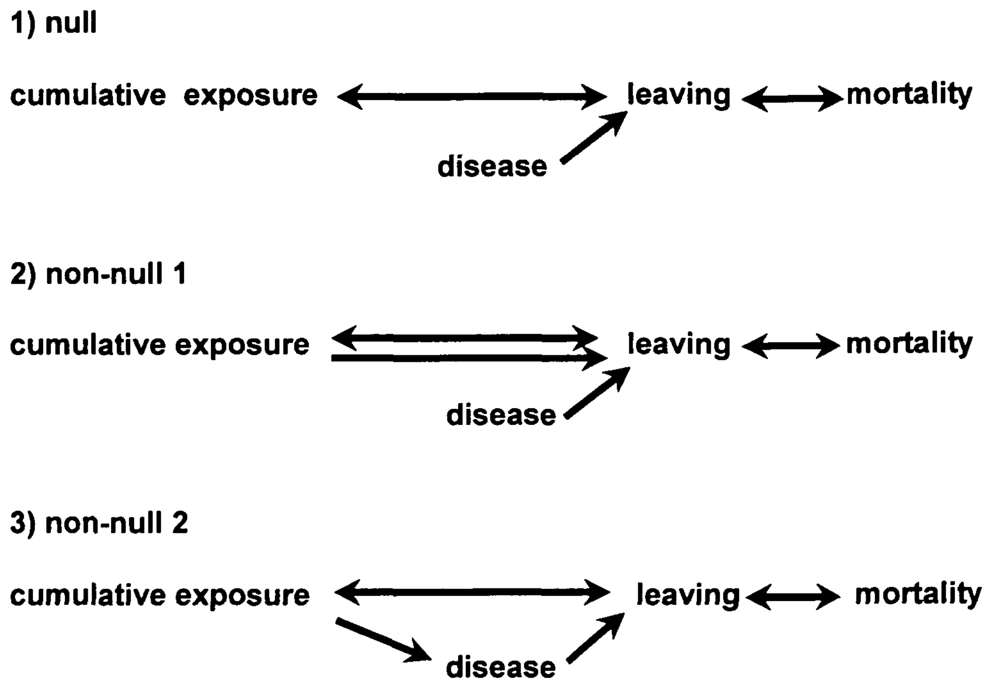

FIGURE 1. Three scenarios for relations between cumulative exposure, disease incidence, leaving work (employment status), and mortality. One-directional arrow indicates causality, two-directional arrow indicates association. Cumulative exposure is associated with leaving because leaving ends exposure. Disease incidence causes leaving. Leaving is associated with mortality because some percentage of those who leave are ill. In the null case, employment status is a confounder. In non-null 1 and non-null 2 cases, employment status is both a confounder and an intermediate variable.

120, 20, and 150 years. Under these conditions, the average length of employment (active time) was 16.0 years, and the average time after employment (inactive time) was 16.9 years. Thirty-six percent of the workers died, 11 percent while active and 25 percent while inactive. Thirteen percent of the workers left work because of disease, 11 percent left because of death, 65 percent left alive without disease, and 11 percent remained employed.

In the second, or "non-null-1" case, rates of leaving were made to increase with cumulative exposure, but incident disease and mortality were again not directly a function of cumulative exposure. The exponential distribution of t_s used under the null-null case was modified for the non-null-1 case so that the hazard rate for leaving was no longer constant but was a function of cumulative exposure, or $h(t_s) = a + bt_s$, where $a = 0.033$ and $b = 0.003$ (see Appendix 1 for details). Under these conditions, the average active time (length of employment) was 12.7 years, while the average inactive time (after employment) was 20.2 years. Thirty-six percent of the workers died during the study, 9 percent while active and 28 percent while inactive. Ten percent of the workers left work because of disease, 9 percent left because of death, 80 percent left alive without disease, and 1 percent remained employed.

In the third case, or "non-null-2" case, increasing exposure increased the probability of disease incidence (and therefore, indirectly, of mortality). Rates of leaving were not related to cumulative exposure directly; however, since workers with disease leave work, rates of leaving would increase with increasing cumulative exposure. The exponential distribution of t_d was modified so that the disease hazard rate was $h(t_d) = a + bt_d$, where $a = 0.007$ and $b = 0.008$ (see Appendix 1 for details). Once exposure/employment ceased, the hazard rate was kept constant. Under these conditions, the average active time was 15.7 years, and the average inactive time was 15.7 years. Forty-six percent of the workers died, 13 percent while active and 33 percent while inactive.

The parameters for the exponential distributions for the null-null case were chosen so that conditions would be typical of many occupational cohort mortality studies in which approximately one third of the workers die during the study, only a small minority (here about 10 percent) of the workers never leave work during the study period, and there is an approximately equal division between active and inactive person-times. We also chose initial parameters so that only a minority of workers who left work would do so because they were ill (approximately 13 percent under the null-null situation). Additional parameters for the two non-null cases were chosen to have a significant observable effect on the distributions of leaving times and disease incidence, but without so drastically altering the scenario that the resulting study clearly would be unlike any actual study. By way of sensitivity analysis, we varied the parameter "b" for the two non-null cases and present some of these results as well.

The focus of the analyses of the simulated data was the effect of cumulative exposure on mortality, while controlling or not controlling for employment status by inclusion of a dichotomous variable for employment status in Poisson regression. Mortality is generally the only observable event in occupational cohort studies, in which data on disease incidence are unavailable. We also present mortality rates stratified by employment status and cumulative exposure categories for descriptive purposes. We believe the simulations reflect the actual situation in occupational cohorts, where most individuals leave work for reasons unrelated to health, but a minority leave because they are sick, and the increase in mortality for inactive workers is due to the increase in mortality for the minority of inactive workers who leave due to illness.

Although not presented here, we also conducted a dichotomous exposure analysis rather than one focused on cumulative exposure. We conducted simulations in which everyone was initially nonexposed, and we then created an additional variable for "time-until-first-exposure." Approximately one third of workers remained nonexposed throughout their employment in these simulations. The overall percentage of active and inactive person-years over the follow-up period differed only slightly for exposed and nonexposed workers in these simulations, and hence no substantive confounding by employment status occurred.

RESULTS FOR EMPIRICAL DATA

Table 1 shows results from the empirical data (10 cohorts) for the percentage of active and inactive person-years by category of duration of exposure. The increasing percentage of active person-years with duration of exposure (employment) accounts for the observed negative trend in mortality with duration (table 2). We also show the same data for time-since-hire categories. Note that for time-since-hire the opposite effect occurs: Active person-years decrease with time-since-hire, which results in an artificial positive trend in mortality with time-since-hire (table 2, bottom). The increasing proportion of inactive person-time with increasing time-since-hire for an occupational cohort is the basis for the well-known "wearing off" of the healthy worker effect with increasing time-since-hire.

TABLE 1. Active/inactive person-years by duration of exposure and time-since-hire, empirical data

	Person-years	% active person-years
Duration (years)		
<1	418,114	14
1–2	208,422	27
2–5	319,913	42
5–10	316,913	59
10–20	402,182	68
≥20	319,147	63
Time-since-hire (years)		
<10	738,274	59
10–20	614,918	44
≥20	631,313	32

TABLE 2. Rate ratios from Poisson regression for mortality (all causes), empirical data, with and without control for employment status

Duration of exposure (years)	Rate ratio, model 1* (no control)	Rate ratio, model 1a† (with control)
<1	1.00	1.00
1–2	0.85	0.91
2–5	0:82	0.93
5–10	0 77	1.00
10–20	0.73	1.06
≥20	0.69	1.07
Time-since-hire (years)	**Model 2* (no control)**	**Model 2a† (with control)**
<10	1.00	1 00
10–20	1 11	0.95
≥20	1.30	0.99

* Models 1 and 2 were controlled for age (5-year categories) and calendar time (10-year categories). Model 1 included a categorical variable for duration, while model 2 included the same categorical duration variable and also a categorical time-since-hire variable. The negative trend for duration in model 1 and the positive trend for time-since-hire in model 2 were highly significant.

† Models 1a and 2a added a dichotomous variable for employment status (inactive/active) to models 1 and 2. The rate ratio in model 1 for employment status (inactive vs. active) was 3.32

When the cohort is restricted to active workers, the healthy worker effect does not wear off over time (9).

Table 2 illustrates Poisson regression results for the empirical data for mortality from all causes. The first section shows the negative trend in mortality with duration, which is eliminated after addition of employment status to the model. The second section of the table shows that there is a positive trend in mortality with increasing time-since-hire that is eliminated when employment status is added to the model. While we have shown all-cause mortality here, results were substantially the same for ischemic heart disease and all cancers.

RESULTS FOR SIMULATED DATA

Tables 3–8 present results the simulated data results. Table 3 presents descriptive results from the null-null case in which cumulative exposure is independent of disease incidence (or mortality) and of leaving rates. Note that there is an increasing percentage of active person-years across categories of cumulative exposure, as in the empirical data in table 1. Death rates are approximately constant across cumulative exposure categories when stratified by employment status, but there is a downward trend in rates without such stratification.

Table 4 shows the results of Poisson regression analyses on the data in table 3. As expected, there is a significant downward trend with cumulative exposure. Also as expected, employment status itself (active = 1, inactive = 0) is a strong predictor of mortality. The false-negative trend in mortality with increasing cumulative exposure is eliminated when employment status is entered into the model. There is no significant interaction between exposure and employment status.

Table 5 presents descriptive results from the non-null-1 case in which cumulative exposure is independent of disease incidence (or mortality), but now leaving rates are a function of cumulative exposure. Now the percentage of active person-years does not monotonically increase with cumulative exposure. While a similar downward trend with unstratified mortality

TABLE 3. Simulated data for the null-null case in which exposure affects neither mortality nor leaving rates

Years of exposure	Person-years			Death rate × 10^2		
	Active	Inactive	% active	All	Active	Inactive
0–5	43,579	61,320	42	1.17	0.65	1.53
5–10	33,287	39,114	46	1.13	0.63	1.55
10–15	25,420	26,950	49	1.18	0.66	1.67
15–20	19,238	18,159	51	1.09	0.69	1.52
20–30	25,679	17,051	60	0.99	0.64	1.52
≥30	14,845	3,563	81	0.90	0.67	1.85
Total	162,048	166,157	49	1.11	0.65	1.56

TABLE 4. Poisson regression results for simulated data, null-null case

Model	Coefficient (± SE)*	Deviance
Cumulative exposure	−0.0066 (± 0.0018)	664.03
Cumulative exposure	0.0021 (± 0.0019)	65.13
Employment status	−0.8807 (± 0.0371)	
Cumulative exposure	−0.0028 (± 0.0024)	64.90
Employment status	−0.8588 (± 0.0589)	
Interaction	−0.0018 (± 0.0038)	

* SE, standard error.

rates occurs as in table 3, control for employment status no longer eliminates this trend. Table 6 shows the Poisson regression results for the non-null-1 case. Again, we see a significant negative trend in mortality with increasing cumulative exposure, and again employment status itself is a strong predictor of mortality, but this time the false-negative trend for cumulative exposure is not eliminated with inclusion of employment status in the model. This situation illustrates the failure of traditional methods of adjustment for employment status (failure of the adjustment to eliminate the false-negative exposure-mortality trend) when employment status is simultaneously a confounder and an intermediate variable. There is a significant interaction between employment status and cumulative exposure, with a negative exposure-response trend seen for inactive person-time, but little trend seen for active person-time. The false-negative trend for inactive person-time results from the fact that the proportion of workers who leave because of disease is higher for those with less cumulative exposure. For those who leave with more cumulative exposure, a higher proportion leave without being ill because of the higher rate of leaving with increased exposure, so their mortality is relatively lower.

The footnote to table 6 reports the effects of varying the parameter relating cumulative exposure to leaving rates on the observed parameter relating cumulative exposure to mortality. When leaving rates are only weakly a function of cumulative exposure, the confounding effects of employment status predominate, and the unadjusted exposure-mortality trend is negative, while the adjusted trend is flat. However, as rates for leaving increase more markedly with cumulative exposure, adjustment for employment status no longer eliminates false-negative trends.

Table 7 illustrates the data for the non-null-2 case, in which increasing exposure increases disease incidence. Here the mortality rate exhibits neither a monotonic increase nor decrease with cumulative exposure, but neither is it constant. Stratification of the data by employment status shows a relatively constant mortality rate across increasing cumulative exposure during active person-time. For inactive person-time, there is a monotonically increasing mortality rate with increasing cumulative exposure, as might be expected since the simulation called for increasing exposure to increase disease incidence, which would be reflected in increasing mortality only after workers had left work. Table 8 shows Poisson regression results for these data. The unadjusted results indicate a positive trend for cumulative exposure, which doubled in magnitude after inclusion of employment status in the model. However, again there is a significant interaction in the model, as one might expect given the difference between trends for active and inactive person-time seen in table 7. While it appears that inclusion of employment status in this model corrects a presumed underestimation of the trend for cumulative exposure and mortality to a more positive trend, we do not know the expected true value of this positive trend, and it may be the coefficient for cumulative exposure after adjustment for employment status remains biased. The true expectation for this coefficient is difficult to calculate, given the complicated mixture of survival distributions used to simulate our data.

The footnote to table 8 reports the effects of varying the exposure-disease parameter in the simulation upon the estimated coefficient for the exposure-mortality trend. When the exposure-disease association is weak, confounding effects produce false-negative trends that are converted to positive trends by adjustment for employment status. As the exposure-disease parameter is increased, the unadjusted trend becomes positive but is still negatively confounded, so that the adjusted trend is considerably higher.

DISCUSSION

We have shown in empirical and simulated data that employment status can act as a confounder, producing false-negative trends between cumulative exposure and mortality. In our simulated data, our model allows for a minority of workers to leave work because of disease (which can be exposure related or not), and because disease incidence increases mortality, these workers experience higher overall mortality. Other workers leave without disease. Under these conditions, in our simulated data mortality after leaving work is higher than that during work, as is true empirically in observational studies. We believe that our model is a reasonable reflection of what actually occurs in working cohorts. We have shown earlier (1) that most excess mortality among those who leave work occurs in the first few years after leaving, suggesting that the subset of workers who leave because of disease is

TABLE 5. Simulated data for the null-null-1 case in which exposure does not affect mortality, but does increase leaving rates

Years of exposure	Person-years			Death rate × 10²		
	Active	Inactive	% active	All	Active	Inactive
0–5	43,596	62,994	41	1.14	0.66	1.48
5–10	31,830	53,703	37	1.10	0.62	1.38
10–15	21,974	37,304	37	1.11	0.77	1.31
15–20	13,845	25,091	36	1.16	0.71	1.41
20–30	12,669	19,953	39	0.98	0.79	1.10
≥30	3,352	2,707	55	0.76	0.75	0.78
Total	127,626	201,752	39	1.11	0.69	1.37

TABLE 6. Simulated data, Poisson regression, non-null-1 case in which exposure increases leaving rates*

Model	Coefficient (± SE)	Deviance
Cumulative exposure	−0.0055 (± 0.0022)	453.59
Cumulative exposure	−0.0059 (± 0.0022)	104.38
Employment status	−0.6869 (± 0.0371)	
Cumulative exposure	−0.0119 (± 0.0027)	86.14
Employment status	−0.8927 (± 0.0589)	
Interaction	*0.0208 (± 0.0048)*	

* $h(t_s) = a + b$ * cumexp = 0.033 + 0.003 * cumexp. When b = 0, the unadjusted coefficient for cumulative exposure is −0.0070 (standard error (SE), 0.0017), reflecting the same negative confounding as the null case, while the adjusted coefficient is 0.0028 (SE, 0.0018). When b = 0.001, the unadjusted coefficient is −0.0066 (SE, 0.0019), while the adjusted coefficient is −0.0020 (SE, 0.0019). As b increases, both adjusted and unadjusted coefficients become more strongly negative.

accounting for the excess mortality observed for those who leave.

In both our empirical and simulated data, when the focus of analysis is on trends in mortality with cumulative exposure, adjustment for employment status via traditional methods (stratification, inclusion of employment status in an exposure-response model) will suffice to eliminate the confounding under the null-null case in which exposure is not related either to the probability of leaving work or to disease incidence (and, hence, to mortality). However, our simulated data indicate that traditional adjustment techniques will not suffice to eliminate false-negative trends when increasing exposure increases the probability of leaving work, even when cumulative exposure is not related to disease incidence. In this case, employment status is de facto an intermediate variable between cumulative exposure and disease (and, hence, mortality), because increasing exposure increases the probability of leaving, and leaving in turn increases the risk of disease. This scenario has been discussed by Robins et al. (4, 5) and Weinberg (6). These authors point out that the problem of variables that are simultaneously confounder and intermediate variables may be quite common in epidemiologic analyses. While epidemiologists recognize that one should not control for variables that are clearly intermediate on a causal pathway between exposure and disease, it is not so clearly recognized that variables may act partly as confounders but partly as intermediate variables if exposure increases the probability of their occurrence and that their intermediate role can occur even without any true underlying causal pathway. Employment status is an example. Cumulative exposure can increase the probability of leaving work, and those who leave have higher mortality independent of cumulative exposure (some are ill), so that there is no true causal pathway between exposure, leaving, and mortality. Still, employment status acts as an intermediate variable, and traditional methods of control do not yield the expected null relation between cumulative exposure and mortality.

A third case considered here is the one in which cumulative exposure does increase disease incidence and, hence, mortality. Since in our simulation disease causes workers to leave work, in this case employment status is partly a true intermediate variable. Without adjustment for employment status, the trend between cumulative exposure and mortality is positive, but is still depressed downward due to negative confounding. Adjustment for employment status doubles the positive trend with cumulative exposure. However, because employment status is acting as an intermediate variable, the observed positive trend after adjustment may be biased. To date, there are no methods to derive an unbiased estimate of the effect of cumulative exposure for this situation. The methods suggested by Robins et al. (5) cannot be employed when all workers are exposed at the beginning of follow-up, a situation that is common in occupational cohort mortality studies. Our simulation, like all such simulations, is dependent on key assumptions. For example, one key assumption is that a minority of workers leave work because of illness and that it is this minority that is the cause of the increased mortality rates observed in

TABLE 7. Simulated data for the non-null-2 case in which exposure increases disease incidence, but does not affect leaving rates

Years of exposure	Person-years			Death rate × 10²		
	Active	Inactive	% active	All	Active	Inactive
0–5	43,534	57,890	43	1.31	0.79	1.71
5–10	33,392	35,915	48	1.35	0.81	1.86
10–15	25,390	25,379	50	1.59	0.89	2.28
15–20	18,828	17,666	52	1.68	0.82	2.59
20–30	23,662	16,399	59	1.79	0.88	3.10
≥30	12,161	3,487	78	1.48	0.77	3.96
Total	156,967	156,736	50	1.48	0.83	2.13

TABLE 8. Poisson regression, simulated data, non-null-2 case in which exposure increases disease incidence*

Model	Coefficient (± SE)	Deviance
Cumulative exposure	0.0111 (± 0.0016)	1,365.63
Cumulative exposure	0.0217 (± 0.0016)	129.08
Employment status	−1.1822 (± 0.0363)	
Cumulative exposure	0.0311 (± 0.0019)	54.26
Employment status	−0.7680 (± 0.0584)	
Interaction	−0.0311 (± 0.0037)	

* $h(t_d) = a + b * \text{cumexp} = 0.007 \pm 0.0008 * \text{cumexp}$. When $b = 0$ (the null case), the unadjusted coefficient for cumulative exposure is −0.0064 (standard error (SE), 0.0018), reflecting negative confounding, while the adjusted coefficient is 0.0018 (SE, 0.0019). When $b = 0.0001$, the respective results are −0.0006 (SE, 0.0018) and 0.0079 (SE, 0.0018), so that the adjustment causes a flat trend to become correctly positive. When $b = 0.0002$, the respective results are 0.0023 (SE, 0.0017) and 0.0112 (SE, 0.0018). Increasing levels of b result in increasing coefficients, with the unadjusted coefficient always being less than the adjusted.

empirical data during inactive person-time. Another assumption is that exponential or modified exponential distributions should be used to generate our times until various events and that choosing the minimum among several times until event is an adequate method to mimic empirical data.

Our simulations are also limited because, in the case of most interest in which cumulative exposure results in increased disease (non-null-2), we cannot determine the true value of the cumulative exposure parameter we are trying to estimate. Therefore, we are unable to determine to what degree control over employment status fails to yield the desired parameter, and we cannot determine the relative importance of the negative confounding by employment status versus the importance of its role as an intermediate variable.

We have considered exposure as a continuous variable, and the focus of analysis has been on evaluating exposure response trends within an exposed cohort. In this situation, confounding by employment status will inevitably occur when cumulative exposure is the focus of the analysis, whether the measure of exposure is simple cumulative duration or whether exposure intensity is also taken into account when calculating cumulative exposure. Such confounding will also occur when a nonexposed group is used as the referent rather than a low-exposed group. The problem resides in using cumulative exposure as the measure of interest, since inevitably those with less cumulative exposure will have less active employment time and, hence, higher mortality.

Many studies, however, consider exposure dichotomously, comparing the mortality of an exposed group with that of a nonexposed group. If both populations are workers and if the overall employment time is approximately equal between the two groups, then no confounding by employment status would be expected since the proportions of active and inactive person-time should be similar in both groups. A special case arises, however, when a working population is compared with the general population, which is a common practice. In this case, one observes the typical healthy worker effect, which results because the general population includes a higher proportion of unemployed (and unhealthy) individuals. This again is a form of confounding by employment status, but it cannot be controlled analytically because no information is available on the employment status of individuals in the general population. As is well recognized, this healthy worker effect will diminish with increasing follow-up time as the employed worker cohort increasingly becomes unemployed and resembles the general population.

In summary, we have considered employment status as a confounder in studies analyzing trends with cumulative exposure and have noted that employment status can possibly also act simultaneously as an intermediate variable. This will occur any time a variable that acts as a confounder is also partly caused by exposure. However, to date we are limited by two equally unappealing choices, to control or not to control via traditional methods, neither of which is likely to provide unbiased results. To compound the problem, we often do not know a priori whether a variable is acting as an intermediate variable. That is often the case for employment status because we usually do not know if cumulative exposure increases the likelihood of leaving employment. We suspect that for many chronic diseases with multifactorial etiology the confounding effect of employment status is strong while its role as an intermediate variable may be minor. For such diseases, particularly those with long latency (e.g., cancer), cumulative exposure will often not appreciably increase, leaving rates either directly (by causing disease) or indirectly, so that traditional adjustment methods should work. However, this will vary from study to study and cannot be taken as a general recommendation.

REFERENCES

1. Steenland K, Stayner L. The importance of employment status in occupational mortality studies. Epidemiology 1991;2: 418–23.
2. Wen CP, Tsai SP, Gibson RL, et al. Long-term mortality of oil refinery workers. II. Comparison of the experience of active, terminated, and returned workers. J Occup Med 1984;26: 118–27.
3. Gilbert ES. Some confounding factors in the study of mortality and occupational exposure. Am J Epidemiol 1982;116: 177–88.
4. Robins J. A graphical approach to the identification and estimation of causal parameters in mortality studies with sustained exposure periods. J Chronic Dis 1990;40 (suppl. 2): 1091–8.
5. Robins JM, Blevins D, Ritter G, et al. G-estimation of the effect of prophylaxis therapy for *Pneumocystis carinii* pneumonia on the survival of AIDS patients. Epidemiology 1992; 3:319–35.
6. Weinberg CR. Toward a clearer definition of confounding. Am J Epidemiol 1993;137:1–8.
7. EGRET. Seattle, WA: Statistics and Epidemiology Research Corporation, 1990.
8. SAS. Version 6.08. Cary, NC: SAS Institute, Inc., 1992.
9. Howe GR, Chiarelli AM, Lindsay JP. Components and modifiers of the healthy worker effect: evidence from three occupational cohorts and implications for industrial compensation. Am J Epidemiol 1988;128:1364–75.

APPENDIX 1

Generation of survival times when the hazard is a function of time

Non-null-1 case

The hazard of leaving work (t_s) was a function of time for the non-null-1 case in which leaving time increased with cumulative exposure. Leaving times for each worker were generated as follows.

Let the hazard $h(t) = a + bt$, where a and b are constants > 0. Then the cumulative hazard

$$H(t) = \int_0^t (a + bx)dx = at + (b/2)t^2,$$

and the density function of leaving times is

$$f(t) = (a + bt)e^{-at - (b/2)t^2}$$

Let $Y = at + (b/2)t^2$. Solving, and requiring $t \geq 0$,

$$t = -a + \sqrt{a^{2 + 2bY}}/b$$

Via transformation, Y is exponential with mean 1. Then leaving times t can be generated using SAS's RANEXP (8) function to first generate Y and then derive t. The minimum of t_s, t_d, and t_{d2} would then determine when employment ceased.

Non-null-2 case

The hazard of disease incidence (t_d) was a function of time for the non-null-2 case in which disease hazard increased with cumulative exposure until employment ceased and then remained constant thereafter. Time until disease was generated as follows.

For each worker, first generate a leaving time (t_s) and a time until death not dependent on prior disease (t_{d2}) as exponentials with means 25 and 150. Choose the minimum of these and call it A. Let

$$h(t) = a + bt \quad \text{when} \quad t \leq A$$

and

$$h(t) = a + bA \quad \text{when} \quad t \geq A.$$

Then $H(t)$ is as in the non-null-1 case above when $t \leq A$

and

$$H(t) = aA + (b/2)A^2 + (a + bA)(t - A) = -(b/2)A^2 + (a + bA)t$$

when $t \geq A$.

When $t \leq A$, $Y \leq aA + (b/2)A^2$, *and then* t is as in non-null-1 case.

When $t \geq A$, then $t = (Y + (b/2)A^2)/(a + bA)$. Disease incidence times (t_d) are generated by first generating Y and then deriving t. If t_d is less than t_s or t_{d2}, then the worker gets the disease.