ESTIMATING THE EFFECT OF FORMALDEHYDE EXPOSURE ON LUNG CANCER AND NON-MALIGNANT RESPIRATORY DISEASE (NMRD) MORTALITY USING A NEW METHOD TO CONTROL FOR THE HEALTHY WORKER SURVIVOR EFFECT.

JAMES M. ROBINS, MONIQUE PAMBRUN, CHRIS CHUTE, DONALD BLEVINS

Occupational Health Program and Department of Biostatistics, Harvard School of Public Health, 665 Huntington Avenue, Boston, Mass. 02115 U.S.A.

SUMMARY

Blair et al. conducted a historical cohort study of 26, 561 chemical workers. They reported (1) an elevated SMR (130) for lung cancer in workers exposed to formaldehyde but no dose-response relationship, and (2) an SMR of 96 for emphysema. We report a reanalysis of Blair et al.’s data using newly developed methods (2-3) that control for the tendency of workers at increased risk of death (e.g., cigarette smokers, workers with emphysema, or lung cancer) to preferentially terminate employment. This reanalysis confirmed Blair et al.’s findings for lung cancer. In contrast, the new methods demonstrated a statistically significant positive association between recent formaldehyde exposure and NMRD mortality, even though the association of NMRD mortality with cumulative exposure was significant and negative! The artificial negative association between cumulative exposure and NMRD mortality was due to the fact that the empirical rate of NMRD mortality in workers terminating employment was nearly twice that of workers continuing at work over the subsequent 15 years.

INTRODUCTION

Blair et al. report an elevated SMR (130) for lung cancer in workers exposed to formaldehyde but no dose-response relationship. They argue that the absence of a dose-response relationship is evidence against a causal effect. In contrast, critics attributed the absence of a dose-response relationship to two biases: the healthy worker survivor effect bias and exposure misclassification. Specifically, if workers at increased risk of lung cancer death leave employment early we say that the healthy worker survivor effect is operating. If so, workers who leave employment will tend to have less cumulative exposure to formaldehyde than other workers (since they cease to accumulate exposure) and therefore, an artificial negative association between cumulative exposure and lung cancer mortality may result. This bias can be controlled as described below.

METHODS

Blair et al.’s data was kindly provided by Dr. Blair. For each lung cancer death (ICD codes 162-163) and NMRD death (ICD codes 480-497 and 510-520) we collected up to 50 controls, alive and at risk at the death time of the case (with time as "years since hire") matched to the case on race, sex, salary status,
and to within two years on age at hire and within five years on calendar
of hire. We obtained 266 lung cancer and 189 NMIRD matched sets with at
one control. For each case and control we developed an exposure vector by
tracting, for November 15th of each year, data on whether the subject was at
or off work and, if at work, whether the subjects’ formaldehyde exposure
was "zero" (<.01 ppm TWA), low (.1-.5 ppm), medium (.5-2.0 ppm), or high
(.0 ppm)

Association with Cumulative Exposure

We fit a conditional logistic regression model

$$\text{logit} \ p(\text{case} | \text{control}) = \beta_{0i} + \beta_1 \ \text{cum}(0,t_i - x)$$

(1)

where indexes a matched set, \(t_i\) is years from hire for the case at time of
and \(\text{cum}(0,t_i - x)\) is cumulative exposure (estimated as in Blair et al.) up \(t_i - x\) years after hire. To see whether recent exposure to formaldehyde was
important we also fit Eq. 1 using [in place of \(\text{cum}(0,t_i - x)\)] the covariate
\(\text{cum}(t_i - x, t_i)\), i.e., cumulative exposure in the last \(x\) years before \(t_i\). The re-

TABLE 1

<table>
<thead>
<tr>
<th>Lag x</th>
<th>0</th>
<th>5</th>
<th>9</th>
<th>15</th>
<th>20</th>
<th>-5</th>
<th>-10</th>
<th>-15</th>
<th>-20</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-.016*</td>
<td>-.016*</td>
<td>-.019*</td>
<td>-.019</td>
<td>-.011</td>
<td>-.08*</td>
<td>-.029*</td>
<td>-.028*</td>
<td>-.025*</td>
</tr>
<tr>
<td>CA</td>
<td>-.003</td>
<td>-.0015</td>
<td>-.0006</td>
<td>.0009</td>
<td>.005</td>
<td>-.0553*</td>
<td>-.0189</td>
<td>-.0099</td>
<td>-.0071</td>
</tr>
</tbody>
</table>

is model with \(\text{cum}(0,t_i - x)\); Lag (-x) has covariate \(\text{cum}(t_i - x, t_i)\); *-p<.03

Table 1 demonstrates statistically significant inverse associations (1) be-
NMIRD mortality and (a) cumulative formaldehyde exposure in the last 20
years and (b) cumulative exposure lagged 0, 5 and 10 years and (2) between lung
mortality and cumulative exposure in the last five years. To determine
whether these negative associations were artifactual consequences of the healthy
worker survivor effect, we fit Eq. 1 with the covariate \(\text{cum}(0,t_i - x)\) replaced by
the covariate \(I_{ow}(t_i - x)\), an indicator variable that takes the value 1 if the
subject was off work at time \(t_i - x\) and a value 0 otherwise. Results are in
Table 2. It follows from Table 2 that, among subjects at risk at \(t\), subjects who
were off work five years previously had an NMIRD mortality rate 1.69 times that of
subjects still at work five years previously. From Table 2 we see that the
healthy worker survivor effect is greater for NMIRD than for lung cancer.

Controlling for the Healthy Worker Survivor Effect With the G-Null Test

When the healthy worker survivor effect is operating, if, among subjects at
at time \(t\) with identical past exposure and employment history, exposure
<table>
<thead>
<tr>
<th>x Years</th>
<th>0</th>
<th>5</th>
<th>10</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>NMRE</td>
<td>4.95</td>
<td>1.69</td>
<td>1.56</td>
<td>1.30</td>
</tr>
<tr>
<td></td>
<td>(.0001)</td>
<td>(.02)</td>
<td>(.03)</td>
<td>(.19)</td>
</tr>
<tr>
<td>Lung CA</td>
<td>3.78</td>
<td>1.47</td>
<td>1.33</td>
<td>1.16</td>
</tr>
<tr>
<td></td>
<td>(.0001)</td>
<td>(.01)</td>
<td>(.07)</td>
<td>(.36)</td>
</tr>
</tbody>
</table>

2-sided p-values in parenthesis

level (i.e., job assignment) at t is unrelated to unmeasured risk factors, a valid test of the null hypothesis of no exposure effect can be based on the following:

G-null test algorithm:

(1) For each matched set, at each tk at which the case was at work, determine the subset of the case's matched controls who were at work at tk and whose exposure and employment history up to tk are identical to that of the case; (2) for each case and this subset of matched controls into a 2XK table of case-control status versus the K possible (recorded) exposure levels that a subject could receive at time tk; (3) select a score for each exposure level and, for each table, construct a log rank (Mantel) trend statistic numerator (i.e., observed - expected) that compares the exposure received by the case at tk to the exposures received by the subset of matched controls; (4) sum these numerators over all tables possibly after multiplication by a table specific weight chosen so as to increase power against alternatives felt to be a priori likely; (5) multiply the null variance associated with each table by the square of the table weight and sum over tables; (6) divide the square of the numerator sum by the variance and compare to a χ² distribution.

If, among subjects at work at any tk, the probability of receiving any of exposure levels at tk depends only on employment and exposure history in the m years, then a valid, and often much more powerful, test of the null hypothesis of no exposure effect can be obtained based on the following (3).

m-Modified G-null test algorithm. Remembering that each table constructed in step 2 of the G-null test algorithm is determined by a death time ts, an exposure time tk, and a unique exposure history through tk-1 and employment history to tk, we add a step 2a to the algorithm in which we combine into a single 2XK table of exposure at tk versus case-control status at ts, those tables that had the same exposure history in the interval [tk-m,tk-1] and employment history in the interval [tk-m,tk].
In Table 3 we report the results of the G-null test and the m-G-null test with m = 5 years. To detect an effect of an exposure with a biological latent period of x years, we gave weight 1 to tables associated with exposures more than x years before the death of the case and weight 0 to other tables (positive lag x tests). We gave adequate power to detect an effect of recent formaldehyde exposure, we gave table weight functions that gave weight 1 to exposures received within x years of the death of the case and weight 0 to other tables (negative lag tests). In Table 3 there is a significant association between exposure to formaldehyde in the last 15 years and NMRD mortality. Nonetheless, in Table 3 there remains suggestion of an adverse effect of formaldehyde exposure on lung cancer mortality. Further G-null analyses also failed to find an exposure effect on lung cancer mortality when we determined exposure level either by peak exposure or by exposure to formaldehyde in particulates.

| TABLE 3 |
|-----------------|---|---|---|---|---|---|---|---|
| **Effect Estimates and p-values for G-null and m-G-null Tests with m=5 Years** |
| **Lag x** |
| 0   | 5  | 9  | 15 | 20 | -5 | -10 | -15 | -20 |
| null | .024 | .001 | .03 | .03 | .01 | .40 | .34 | .10 | .07 |
| (.62) | (.99) | (.58) | (.55) | (.86) | (.04) | (.004) | (.04) | (.34) |
| 11  | .027 | .001 | .02 | .02 | .02 | .44 | .32 | .18 | .07 |
| (.60) | (.99) | (.66) | (.74) | (.98) | (.03) | (.01) | (.01) | (.36) |
| null | -.01 | -.013 | -.01 | -.02 | -.03 | .03 | -.05 | .02 | .01 |
| (.75) | (.70) | (.84) | (.48) | (.48) | (.78) | (.46) | (.66) | (.80) |
| 11  | .001 | -.001 | .001 | .001 | -.02 | .12 | -.05 | .001 | .03 |
| (.99) | (.89) | (.94) | (.99) | (.68) | (.44) | (.52) | (.98) | (.60) |

Effect estimate = (O-E)/Var(O-E) (2) 2-sided p-values in parenthesis

The effect of recent exposure on NMRD mortality was detected only by the G-analysis. [As an example, in a logistic regression analysis no significant association was found between average exposure over the previous 15 years and mortality (among those employed 15 years previously).] Finally, we do not have an adequate biological explanation for our finding that only exposure in the 15 years influences NMRD mortality.

REFERENCES: