The influence of study characteristics on the healthy worker effect: A multiple regression analysis

R. Chen and A. Seaton
Department of Environmental and Occupational Medicine, Medical School, University of Aberdeen, Foresterhill, Aberdeen AB9 2ZD, UK

The so-called 'healthy worker effect' (HWE) describes a reduced mortality rate in occupational populations. From 85 occupational cohorts of workers exposed to organic solvents, we have previously found a low weighted standardized mortality ratio (SMR) for all causes (SMR = 89.5, 95 per cent confidence interval [CI] = 89.0-90.2). Characteristics of the cohorts were examined for association with the HWE. A multiple regression analysis revealed that, for overall deaths, the closer the comparison group, the smaller the HWE (p = 0.001); the more hazardous the potential exposures, the higher the SMR (p = 0.02); the higher the social class, the greater the HWE (p = 0.02); and the higher the rates of loss to follow up, the greater the HWE (p = 0.04). These results have shown that, in addition to time-related modifiers, the comparison group, type of occupational exposure, social class and rates of loss to follow-up significantly influence the size of the so-called HWE. Differences in the HWE related to gender and race were probably attributable to different rates of loss to follow-up.

Key words: Healthy worker effect; mortality; organic solvents.

INTRODUCTION

In the past two decades, many epidemiologists have studied a phenomenon occurring in occupational cohort studies referred to as the 'healthy worker effect' (HWE). The term was first used by McMichael et al. in 1974, in a paper in which they found that death rates from all causes within a cohort of rubber-workers were lower than those of the USA male population. They concluded that there was a strong selection process at play, wherein, to be employable in an industrial workforce, an individual must be relatively healthy and active, and that this selection factor acted to produce a relatively low mortality in industrial cohorts. The phenomenon has subsequently been recorded and discussed in many other reports of occupational cohort studies.

At present, the most generally accepted definition of the healthy worker effect is an observed decrease in mortality ratios among workers when compared with the general population, the likely explanation being self-selection by the employee and selection by the employer. Healthier people are better able to seek and compete for work, especially in industry, while employers prefer to select healthier candidates by such processes as pre-employment medical examinations, questionnaire or interview. Furthermore, after getting the job, employees with ill health may leave the cohort early. An additional factor is that workers in the cohort benefit from employment and regular life routine, income and medical care/insurance, which may directly or indirectly affect their health. In contrast, the general population, which includes the sick and unemployed, would be expected to have a worse experience of mortality.

Many factors have been considered to contribute to the HWE, including age, gender, race, length of follow-up, duration of exposure, time from exposure to disease recognition, socio-economic status, particular causes of deaths and occupations. However most observations on the HWE have been based on a single cohort or relatively few separate cohorts, and a comprehensive investigation of the overall effect of
methodological variables on the HWE in occupational cohort studies has rarely been reported. One recent cross-sectional study has reported such an analysis, finding that chemical industries, more person-years of follow-up, shorter time of follow-up and negative study outcome increased the so-called HWE significantly. We have carried out a meta-analysis of mortality among workers exposed to organic solvents, in which we found that many of the studies attributed their low mortality rates simply to the HWE. In order to examine further some of the study characteristics associated with the HWE, we have analyzed 85 occupational cohorts from that database.

METHODS

The methods of selecting the published papers to set up the database have been described in an accompanying paper. In this study we only used the papers which provided overall mortality with the standardized mortality ratios (SMR) or relative risk (RR) as a risk index. In total, 52 studies were included and three other studies, which did not give the SMRs for all causes, were excluded from this analysis. The analysis is based on cohorts, rather than individual papers. Therefore, when a paper had two or more different cohorts separated by sex, race, social class or occupational exposures, these results were analyzed independently in our study. Also when a study had two or three results based on different comparison populations (e.g., national/regional) or for different restricted periods (e.g. > 15 years since first time exposure), we only chose the result from the most closely compared group or the entire follow-up period for analysis. As a result, a total of 85 cohorts were available for the analysis. From them, the following information was identified and quantified:

X1. Year of publication (e.g., 1988 = 88)
X2. Year of start of follow-up (e.g., 1940 = 40)
X3. Selection criterion for subjects, i.e., minimum months of work/exposure (e.g., 1 year = 12)
X4. Gender (male = 1, male + female = 2, female = 3)
X5. Race (white = 1, white + black = 2, black/non-white = 3)
X6. Size of cohort (numbers of subjects/1,000)
X7. Cohort definition (entire cohort = 1, period cohort = 2)
X8. Method of follow-up (during follow-up, no subjects enter = 1, some time-limited subjects enter = 1.5, allowing subjects to enter = 2)
X9. Length of follow-up (years)
X10. Rate of loss to follow up (e.g., 1% = 1)
X11. Social class (average of cohort: professional = 1, managerial = 2, other white collar = 3, skilled blue collar = 4, unskilled blue collar = 5)
X12. Occupational exposure or hazard level (non-exposure = 1, petrol industry = 2, painting or dry cleaning = 3, pigments/chromates involved = 4, asbestos/radiation/benzene involved = 5)
X13. Comparison group (national = 1, province/state = 2, regional/city/county = 3, controlled cohort/internal cohort = 4)
X14. The observed numbers of all causes of death/1,000
X15. 'Crude death rate' (surrogate for mean age of cohort: = X14 x 1000 / (X6 x X9)

The six outcomes of SMR (Y1–6) as dependent variables were copied only after all 15 possibly relevant factors (X1–15) as independent variables had been identified.

A simple regression analysis was first carried out to study the relations between the relevant variables (X) and the SMRs (Y) among all these cohorts, calculating the regression coefficient (b), correlation coefficient (r) and statistical significance (p value). Then, a multiple regression was used to consider 'confounding' effects among some relevant variables, calculating the coefficient of multiple correlation (R), partial regression coefficient (b), partial correlation coefficient (P) and statistical significance (P or p value). Of the 15 variables, further multiple regression was carried out on eight because of their possible epidemiological and statistical significance and the availability of appropriate data. All analyses were completed using a standard statistical package (SPSS).

RESULTS

Out of 85 cohorts, 65 showed a low mortality for all causes (the SMR being less than 100) called an estimated HWE. On calculating the weighted SMRs (Poisson model), we found that there was an obviously low mortality for all causes (SMR = 89.5, 95% confidence interval [CI] = 89.0–90.2), circulatory system diseases (SMR = 87.2, CI = 74.2–90.2), respiratory system diseases (SMR = 87.1, CI = 76.7–97.2), and accidents (SMR = 71.7, CI = 68.8–74.7) and a slightly reduced mortality for all cancers (SMR = 95.5, CI = 94.2–96.9), but not for lung cancer (SMR = 98.4, CI = 96.0–100.7). Also, we found that there was a lower combined mortality of all causes for the female workers (SMR = 87.1, 83.0–91.5) than for the males (SMR = 91.3, CI = 90.7–92.0), and that non-white workers had...
a higher HWE (SMR = 73.7, CI = 69.8–77.8) than white workers (SMR = 91.4, CI = 90.7–92.0).

The simple regression analysis showed that there was no relationship between the SMR and year of start of follow-up (X2) \((b = 0.20, r = 0.10, p = 0.35)\), size of cohort (X6) \((b = 0.10, r = 0.05, p = 0.64)\), cohort definition (X7) \((b = -3.10, r = -0.05, p = 0.63)\), or the observed number of deaths (X14) \((b = 0.106, r = 0.01, p = 0.92)\), while other factors were possibly related to the SMRs for overall deaths, all cancer, lung cancer, circulatory diseases, respiratory diseases or accidents. More detailed results are shown in Table 1. It was found that the closer the comparison group (X13) was to the cohort, the higher the SMRs for overall death, circulatory system diseases and accidents. The higher the social class (X11), the lower the SMRs (save for accidents). The more potentially harmful the exposure (X12), the higher the mortality for all cancer, lung cancer, circulatory and respiratory system diseases. Non-white workers (X5) showed a lower SMR than whites. Higher rates of loss to follow-up (X10) reduced the SMR for all causes, in other words increasing the HWE.

Multiple regression analysis of the six SMR variables on all 15 factors indicated that there were several significant coefficients, such as overall death \((R = 0.68, P = 0.003)\), lung cancer \((R = 0.73, P = 0.01)\), circulatory system disease \((R = 0.73, P = 0.02)\), accidents \((R = 0.79, P = 0.04)\) and cancer \((R = 0.61, P = 0.04)\), but respiratory disease did not achieve statistical significance \((R = 0.68, P = 0.28)\). The HWE for overall death was significantly associated with comparison group (X13) \((b = 8.85, \beta = 0.52, p = 0.0004)\) and social class (X11) \((b = 5.30, \beta = 0.28, p = 0.03)\). The HWE for cancer was associated with average social class of the cohort (X11) \((b = 6.51, \beta = 0.26, p = 0.05)\) and relative hazard of exposure (X12) \((b = 4.27, \beta = 0.24, p = 0.065)\). The HWE in circulatory system diseases was significantly associated only with the type of comparison group (X13) \((b = 13.42, \beta = 0.69, p = 0.0008)\). A higher SMR from lung cancer was associated with the ‘crude death rate’, (X15) \((b = -5.59, \beta = -0.68, p = 0.002)\), lower rate of loss to follow up (X10) \((b = -4.20, \beta = -0.34, p = 0.02)\), lower social class \((b = 16.24, \beta = 0.32, p = 0.03)\) and more harmful exposure (X12) \((b = 11.79, \beta = 0.32, p = 0.02)\). The comparison group (X3) \((b = 33.89, \beta = 0.81, p = 0.002)\), the observed numbers of deaths (X14) \((b = 20.57, \beta = 0.97, p = 0.01)\) and size of cohort (X6) \((b = -3.53, \beta = -0.78, p = 0.03)\) affected the SMR for accidents.

Some of these 15 factors were interrelated. For example, female workers had a higher rate of loss to follow-up than males \((b = 2.49, r = 0.39, p = 0.001)\) and blacks had a higher rate of loss to follow-up than whites \((b = 3.08, r = 0.36, p = 0.002)\). In view of these relations and in order to increase statistical power, we chose eight factors for further multiple regression analysis. The result is shown in Table 2. It indicates that contributions to the HWE were still made by different comparison groups, occupational exposures, social class and rate of loss to follow-up. It was also found that SMRs for accidents, all cancer and lung cancer, but not for overall deaths, circulatory and respiratory deaths, were decreased with length of follow-up, and that the shorter the time required for selection into studies, the higher the SMR for cancer.

### Table 1. Simple regression and correlation of SMRs and characteristics of studies (X)

<table>
<thead>
<tr>
<th>Cause of deaths</th>
<th>X1</th>
<th>X3</th>
<th>X4</th>
<th>X5</th>
<th>X8</th>
<th>X9</th>
<th>X10</th>
<th>X11</th>
<th>X12</th>
<th>X13</th>
<th>X15</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall death</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>((n = 85))</td>
<td>(b) = -0.26, (r) = -0.05, (p) = 0.30</td>
<td>(b) = -0.11, (r) = -0.09, (p) = 0.17</td>
<td>(b) = -0.02, (r) = 0.04, (p) = 0.12</td>
<td>(b) = -0.01, (r) = -0.05, (p) = 0.09</td>
<td>(b) = 0.07, (r) = 0.14, (p) = 0.40</td>
<td>(b) = 0.02, (r) = 0.02, (p) = 0.02</td>
<td>(b) = 0.01, (r) = 0.01, (p) = 0.01</td>
<td>(b) = 0.00, (r) = 0.00, (p) = 0.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cancer ((n = 82))</td>
<td>(b) = -1.36, (r) = -0.35, (p) = 0.30</td>
<td>(b) = -0.16, (r) = 0.05, (p) = 0.06</td>
<td>(b) = 0.02, (r) = 0.14, (p) = 0.68</td>
<td>(b) = 0.02, (r) = 0.62, (p) = 0.94</td>
<td>(b) = 0.07, (r) = 0.49, (p) = 0.99</td>
<td>(b) = 0.01, (r) = 0.01, (p) = 0.01</td>
<td>(b) = 0.00, (r) = 0.00, (p) = 0.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lung cancer ((n = 59))</td>
<td>(b) = -2.10, (r) = -0.18, (p) = 0.40</td>
<td>(b) = -0.12, (r) = 0.09, (p) = 0.12</td>
<td>(b) = 0.15, (r) = 0.12, (p) = 0.40</td>
<td>(b) = 0.49, (r) = 0.35, (p) = 0.04</td>
<td>(b) = 0.25, (r) = 0.02, (p) = 0.02</td>
<td>(b) = 0.04, (r) = 0.77, (p) = 0.92</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Circulatory ((n = 62))</td>
<td>(b) = -0.56, (r) = 0.23, (p) = 0.00</td>
<td>(b) = -0.14, (r) = 0.00, (p) = -0.14</td>
<td>(b) = 0.40, (r) = 0.33, (p) = 0.99</td>
<td>(b) = 0.29, (r) = 0.20, (p) = 0.05</td>
<td>(b) = 0.68, (r) = 0.50, (p) = 0.02</td>
<td>(b) = 0.02, (r) = 0.02, (p) = 0.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Respiratory ((n = 48))</td>
<td>(b) = -0.42, (r) = 0.34, (p) = 16.91</td>
<td>(b) = -2.71, (r) = 0.01, (p) = 0.00</td>
<td>(b) = 0.30, (r) = 0.00, (p) = 0.30</td>
<td>(b) = 0.82, (r) = 0.27, (p) = 0.97</td>
<td>(b) = 0.80, (r) = 0.01, (p) = 0.00</td>
<td>(b) = 0.79, (r) = 0.05</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Accidents ((n = 47))</td>
<td>(b) = 1.40, (r) = 0.26, (p) = 0.00</td>
<td>(b) = -0.16, (r) = 0.02, (p) = -0.22</td>
<td>(b) = 0.14, (r) = 0.01, (p) = 0.31</td>
<td>(b) = 0.29, (r) = 0.00, (p) = 0.46</td>
<td>(b) = 0.52, (r) = 0.00, (p) = 0.96</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(b\) = Regression coefficient;  
\(r\) = Correlation coefficient;  
\(p\) = Statistical probability;  
Bold means a statistical significance.
Table 2. Multiple regression and correlation of SMRs and eight relevant factors (X)

<table>
<thead>
<tr>
<th>Cause of deaths</th>
<th>X3</th>
<th>X6</th>
<th>X8</th>
<th>X9</th>
<th>X10</th>
<th>X11</th>
<th>X12</th>
<th>X13</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall death</td>
<td>b</td>
<td>0.12</td>
<td>0.14</td>
<td>0.07</td>
<td>0.05</td>
<td>0.65</td>
<td>5.07</td>
<td>3.47</td>
</tr>
<tr>
<td>R = 0.64; P = 0.0001;</td>
<td>β</td>
<td>0.07</td>
<td>0.07</td>
<td>0.09</td>
<td>0.05</td>
<td>0.22</td>
<td>0.26</td>
<td>0.25</td>
</tr>
<tr>
<td>b0 = 58.01; P = 0.0001</td>
<td>ρ</td>
<td>0.50</td>
<td>0.49</td>
<td>0.39</td>
<td>0.64</td>
<td>0.04</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>Cancer</td>
<td>b</td>
<td>-0.53</td>
<td>0.038</td>
<td>5.12</td>
<td>-0.52</td>
<td>-0.93</td>
<td>7.69</td>
<td>4.55</td>
</tr>
<tr>
<td>R = 0.56; P = 0.005;</td>
<td>β</td>
<td>-0.26</td>
<td>0.02</td>
<td>0.09</td>
<td>-0.26</td>
<td>-0.15</td>
<td>0.31</td>
<td>0.26</td>
</tr>
<tr>
<td>b0 = 58.42; P = 0.004</td>
<td>ρ</td>
<td>0.03</td>
<td>0.89</td>
<td>0.54</td>
<td>0.03</td>
<td>0.21</td>
<td>0.01</td>
<td>0.03</td>
</tr>
<tr>
<td>Lung cancer</td>
<td>b</td>
<td>-1.30</td>
<td>0.05</td>
<td>10.16</td>
<td>-1.28</td>
<td>-2.09</td>
<td>12.34</td>
<td>10.41</td>
</tr>
<tr>
<td>R = 0.59; P = 0.018;</td>
<td>β</td>
<td>-0.25</td>
<td>0.01</td>
<td>0.09</td>
<td>-0.33</td>
<td>-0.17</td>
<td>0.25</td>
<td>0.28</td>
</tr>
<tr>
<td>b0 = 49.72; P = 0.292</td>
<td>ρ</td>
<td>0.07</td>
<td>0.94</td>
<td>0.54</td>
<td>0.02</td>
<td>0.22</td>
<td>0.08</td>
<td>0.05</td>
</tr>
<tr>
<td>Circulatory</td>
<td>b</td>
<td>0.28</td>
<td>0.13</td>
<td>-5.64</td>
<td>0.15</td>
<td>-0.46</td>
<td>3.71</td>
<td>4.23</td>
</tr>
<tr>
<td>R = 0.59; P = 0.024</td>
<td>β</td>
<td>0.17</td>
<td>0.05</td>
<td>-0.11</td>
<td>0.08</td>
<td>-0.09</td>
<td>0.17</td>
<td>0.26</td>
</tr>
<tr>
<td>b0 = 49.76; P = 0.014</td>
<td>ρ</td>
<td>0.24</td>
<td>0.69</td>
<td>0.47</td>
<td>0.61</td>
<td>0.56</td>
<td>0.25</td>
<td>0.08</td>
</tr>
<tr>
<td>Respiratory</td>
<td>b</td>
<td>0.52</td>
<td>0.076</td>
<td>-16.94</td>
<td>0.41</td>
<td>-0.52</td>
<td>5.65</td>
<td>9.02</td>
</tr>
<tr>
<td>R = 0.52; P = 0.222</td>
<td>β</td>
<td>0.18</td>
<td>0.02</td>
<td>-0.22</td>
<td>0.16</td>
<td>-0.06</td>
<td>0.18</td>
<td>0.38</td>
</tr>
<tr>
<td>b0 = 42.08; P = 0.232</td>
<td>ρ</td>
<td>0.29</td>
<td>0.89</td>
<td>0.23</td>
<td>0.40</td>
<td>0.71</td>
<td>0.32</td>
<td>0.03</td>
</tr>
<tr>
<td>Accidents</td>
<td>b</td>
<td>-1.23</td>
<td>-0.20</td>
<td>23.50</td>
<td>-1.90</td>
<td>0.11</td>
<td>-6.21</td>
<td>-3.17</td>
</tr>
<tr>
<td>R = 0.64; P = 0.045</td>
<td>β</td>
<td>-0.21</td>
<td>-0.07</td>
<td>0.23</td>
<td>-0.48</td>
<td>0.01</td>
<td>-0.15</td>
<td>-0.10</td>
</tr>
<tr>
<td>b0 = 117.34; P = 0.013</td>
<td>ρ</td>
<td>0.22</td>
<td>0.68</td>
<td>0.23</td>
<td>0.01</td>
<td>0.95</td>
<td>0.38</td>
<td>0.57</td>
</tr>
</tbody>
</table>

R = Coefficient of multiple correlation; b0 = constant; b = partial regression coefficient; β = partial coefficient of correlation; P or p = statistical probability. Bold means a statistical significance.

but not for other indices of mortality.

**DISCUSSION**

Although it has been called an unfortunate term, the expression 'healthy worker effect' is widely used to describe a cause of serious bias in occupational epidemiology. In occupational follow-up studies in which a favourable experience of mortality has been shown, most authors tend to attribute their results to this effect, even though such low SMRs may also arise from other factors such as defective study design, differing occupational exposures or negative confounders. Clearly, selection factors affecting entry to and survival in an occupational cohort, together with some benefits from employment in the cohort, could contribute to the healthy worker effect. However, the limited evidence available suggests that the effect is a complex process involving several components in addition to these selection factors. We have investigated some of them.

The usual method for analyzing SMRs is to use a Poisson regression model. But in our database, we took one SMR as one event and found that their distribution was probably symmetrical because the mean of the overall SMR was 90.8, very close to the median of 89.5. We have treated the independent variables, gender, race and exposure to harmful substances, as continuous rather than categorical data and have coded them according to the proportions of males to females, whites to blacks, or to the relative toxicities of likely exposures on an arbitrary scale. We were therefore able to use a multiple regression model to analyze our data.

Other models such as logistic and Poisson regressions have not been employed because they have some previously recognized limitations for our data, e.g., loss of information during transformation of the data to category, which leads to loss of power.

An obvious source of bias in mortality studies is the nature of the population chosen for comparison, be it a national population, a more local one or another industrial group. Availability of appropriate statistics means that the general population is usually chosen, which leads to loss of power.

Many studies have shown that social class may affect the HWE. Many studies have shown that social class may affect the HWE. Our analysis supports the suspicion that less bias occurs with more closely matched comparison populations.

The effects of occupational exposures should be considered in analysis of the HWE. Some authors have observed that job category affected the SMRs, while others have questioned the degree to which such time-related increased patterns in mortality have been confounded by the effects of exposure or accumulated exposures rather than simply the diminution of the healthy worker effect with time. Our results showed that the more hazardous the potential exposure, the higher the SMR. We have taken occupational exposure as a 'confounding' or 'adjusting' variable in
the multiple regression model. We have noticed that many of the cohorts which have been used to observe and discuss the HWE in previous studies have comprised workers in chemical plants. Such industries would be unlikely to be associated with a higher mortality than other heavier industries, showing a higher HWE.19

Although some authors have proposed that persistent deficits in mortality might result from losses to follow-up,11,20,23 this has often been ignored in analysis of explanations for the HWE. We have shown that a higher HWE may be associated with a higher rate of loss to follow-up. It is likely that some of the lost subjects would be unemployed for a while after leaving the cohort, a recognized association with increases in mortality.28 Often these lost subjects of unknown vital status are taken as living at the end of follow-up, whereas some of them will have died and have shown a much higher relative mortality.12,20

There is disagreement as to whether or not there is a stronger HWE for females than for males.18,20 From the regression analysis, we did not find that gender played an important role, even though females had a lower SMR than males. We think that other relevant factors, such as rate of loss to follow-up, are associated with gender and the females' low SMR. This is supported by the fact that two of the studies in our database which gave the overall SMRs of the total cohort, rather than SMRs for females and males separately, also showed that the rate of loss to follow-up in the females was about five per cent higher than that of the males. It is likely that female workers who leave a cohort have reduced opportunities to find other employment, most becoming housewives;29 there is also likely to be information bias from changes in women's surnames from marriage during follow-up. It has also been observed that black workers show a higher HWE than whites in some studies,27 perhaps because census and vital statistics recording for non-whites historically has been less complete and accurate than for whites in USA24 and because non-whites in the general population have a higher overall mortality on account of multiple social disadvantages.4 However, we remain unsure of the significance of the effect of race on HWE, since blacks also had a high rate of loss to follow-up in our study.

Some authors have also indicated that the longer the time criterion required for selection into a cohort, the greater the chance of revealing a HWE, and have suggested that short-term workers, e.g., < 6 months, should be included in the cohort.20 In our analysis, this phenomenon occurred significantly only for all cancer, possibly including lung cancer, but not for death from other diseases. This is a complex issue, since short-term workers will clearly have less prolonged exposure but may, for example, be exposed to a strong carcinogen at a critical time. However, these workers are likely to include disproportionate numbers with relatively poor health and habits such as heavy smoking. The usual compromise of entering only those who have worked in the industry for a qualifying period depending on the nature of the likely exposure seems a sensible one.

Most studies have indicated that the HWE disappears with longer follow-up. However, surprisingly, we got the opposite result for accidents, all cancers and lung cancer. With respect to cancer, a possible explanation is that the majority of occupational exposures (which were to organic solvents) in these studies were of relatively low toxicity, so that the hazardous effects of cumulative exposure were not apparent, and that exposures to more highly toxic substances such as benzene and asbestos were reduced and banned in workplaces during the period of follow-up. In addition, although the prevalence of smoking among the cohorts was not available for this analysis, we think it unlikely that it was high among these workers, because the SMRs from other smoking-related conditions such as respiratory disease were low in our meta-analysis.21 Smoking would be a negative confounder among these cohorts and it might have a stronger effect with time of follow-up, compared to the general population. As for accidents, a longer period of follow-up as well as a longer time selection criterion, leading to the subjects having more working experience, is likely to have reduced the mortality.

It has also been considered that a younger age of cohort could increase the HWE.3,20 In view of the fact that few studies in the database provided the relative age of the cohort, we tried to use the 'crude death rate' (X15) to represent an approximate age of cohort initially, according to a rule that generally the overall mortality rate increases with age within a range. But we have not obtained the expected result, possibly due to its imperfect estimation of the real death rate, depending on different sizes of cohort and lengths of follow-up. Another index which indirectly reflects the age of a cohort is the method of follow-up (X8). It is likely that a dynamic cohort, allowing workers to enter during the follow-up period, could include more young people than a fixed cohort. This type of cohort showed a greater HWE for circulatory diseases in our analysis (Table 1).

Our results have further shown that the HWE varies for different specific causes of death and that it is greater for diseases which have a considerable prediagnosis symptomatology or which may appear early in life and exclude people from the labour force, e.g., respiratory diseases and circulatory diseases, than for diseases without such symptomatology or which have a high fatality rate with symptoms that appear only a few years before death occurs, e.g., lung cancer and cancers.20,23 Our multiple regression analysis has demonstrated that many factors may simultaneously affect the HWE and that this may differ for different diseases. Although various methods of reducing the bias from the HWE have been proposed,24,20 such as a constant correction factor, say 0.9, according to the magnitude of the HWE, it seems unlikely that any simple correction can take account of such complex
biases. On the other hand, a more detailed knowledge of the likely sources of bias in a mortality study should allow for better control, by taking steps to reduce them at the design stage of the study.

ACKNOWLEDGEMENTS

This research was supported by a fellowship for Dr R. Chen from the Colt Foundation. We would like to thank Ms Julie Marshall for her help in identifying references.

REFERENCES