Age, time and cohort factors in mortality from cancer of the cervix

By J. C. Barrett

London School of Hygiene and Tropical Medicine, London, W.C.1;
and Office of Population Censuses and Surveys, London, W.C.2

(Received 8 August 1972)

SUMMARY

Data for mortality from cancer of the cervix in England and Wales by 5-year age groups and four quinquennia (1951–70) are analysed. The logarithms of the mortality rates are regressed on age group, epoch of death and epoch of birth. The factors obtained are considered in relation to particular features of the mortality pattern, such as the reversal of trend in certain age groups.

INTRODUCTION

Patterns of mortality from cancer of the cervix present a number of peculiar features. Primarily a disease of the human species, its frequency in relation to all cancer in females ranges from 5 to 60% in different population groups, and when one reflects on the lower frequencies found, inter alia, in developed countries, rural populations, high socio-economic levels, Jewish women, the Amish women, and single women, as compared to higher frequencies in developing countries, urban populations, low socio-economic levels, negroes, married women and other associations (Dunham, 1968; Voors, 1967), one is led to suspect a local cause in women who later develop the disease.

Hill & Adelstein (1967), presenting data for mortality from cervical carcinoma, 1951–65, England and Wales, showed that, although a downward trend existed in the standardized mortality rates, rises in age-specific death-rates had occurred in certain age groups, beginning with those aged 30–34 and later changing to older groups up to ages 50–54. Similar intrusions in the generally falling rates were reported by Adelstein, Hill & Maung (1971) for carcinoma of the uterus in several countries, especially England and Wales, Scotland, the Netherlands, Denmark, Norway, Sweden, New Zealand and Chile; but the intrusion was absent in the United States and Japan. They considered that ‘the intrusion is related to change in the customary sexual relationships brought about by two world wars and their aftermath and, taken with other evidence, the pattern is best explained by an infective cause of the cancer’.

Higgins (1971) has compared recent mortality from cancer of the cervix uteri in the United States and the United Kingdom, pointing out that the increase in each 10-year age group did not occur after intervals of 10 years, and suggesting
Table 1. Cancer of cervix deaths, England and Wales, 1951–70 (e_{ij})

(From the Registrar General's Statistical Review of England and Wales, Part I, 1951–70, Table 17.)

<table>
<thead>
<tr>
<th></th>
<th>25-79</th>
<th>62</th>
<th>38</th>
<th>59</th>
<th>49</th>
<th>65-74</th>
</tr>
</thead>
<tbody>
<tr>
<td>1951-55</td>
<td>79</td>
<td>251</td>
<td>450</td>
<td>789</td>
<td>1123</td>
<td>1553</td>
</tr>
<tr>
<td>1956-60</td>
<td>62</td>
<td>283</td>
<td>615</td>
<td>920</td>
<td>1284</td>
<td>1467</td>
</tr>
<tr>
<td>1961-65</td>
<td>38</td>
<td>132</td>
<td>514</td>
<td>1097</td>
<td>1380</td>
<td>1530</td>
</tr>
<tr>
<td>1966-70</td>
<td>49</td>
<td>108</td>
<td>313</td>
<td>809</td>
<td>1424</td>
<td>1519</td>
</tr>
</tbody>
</table>

Table 2. Female populations (thousands person-years) (P_{ij})

(From the Registrar General's Statistical Review of England and Wales, Part II, 1951–70, Table A 2, home populations.)

<table>
<thead>
<tr>
<th></th>
<th>25-79</th>
<th>62</th>
<th>38</th>
<th>59</th>
<th>49</th>
<th>65-74</th>
</tr>
</thead>
<tbody>
<tr>
<td>1951-55</td>
<td>77,740</td>
<td>82,220</td>
<td>78,240</td>
<td>84,850</td>
<td>82,670</td>
<td>76,620</td>
</tr>
<tr>
<td>1956-60</td>
<td>72,330</td>
<td>77,440</td>
<td>82,890</td>
<td>77,400</td>
<td>83,280</td>
<td>80,880</td>
</tr>
</tbody>
</table>

that the intrusions may be associated with changes in sexual behaviour which, however, were considered to affect different cohorts at different times. The difficulty of interpreting the patterns is increased by the possible effects of the introduction of screening programmes for cancer of the cervix, immigration, rising trends in hysterectomy, and changes in death certification.

METHOD AND RESULTS

In the present work the data have been analysed by a regression method similar to that used by Sacher (1960) in an application to mortality from tuberculosis, which in turn bears some similarity to earlier work (Kermack, McKendrick & McKinlay, 1934; Greenberg, Wright & Sheps, 1950; Spicer, 1954; Case, 1956).

The relevant data for mortality from cancer of the cervix in England and Wales, now including the quinquennium 1966–70, are shown in Tables 1 and 2.

Kermack et al. (1934) represented age-specific mortality rates from all causes as the product of two factors – one a function of age (x) alone, and the other a function of the year of birth alone:

\[ q(x, t + x) = f(t)g(x), \]

where \( f(t) \) is a function acting at the date of birth and \( g(x) \) is the age function. They found that in England and Wales, and also in Scotland and to a lesser extent in Sweden, the relative mortality (i.e. the age-specific mortality as a percentage of the specific mortality for the same age group in a standard period such as 1845–55) depended primarily on the date of birth of the individuals concerned and only indirectly on the particular year under consideration.

Following Sacher, in the present analysis the rates of mortality from the
particular cause of death are first transformed logarithmically to provide
Gompertzians,

\[ G_{ij} = \log (e_{ij}/P_{ij}) \]  

(2)

where \( i = 1, 2, \ldots, 10 \) are the 5-year age classes, \( j = 1, 2, 3, 4 \) are the epochs of
death (that is, the quinquennia for which the current data were obtained), and
\( k = 1, 2, \ldots, 13 \) are the epochs of birth (from 1881 to 1941). In this case,

\[ k = j - i + 10. \]  

(3)

It seems best to work with the logarithms of the rates since the rates vary by
several orders of magnitude between age groups. Accordingly, the following
relation is postulated for the expectation of the Gompertzian:

\[ E(G_{ij}) = \alpha_i + \beta_j + \gamma_k \]  

(4)

where \( \alpha \) is the age factor, \( \gamma \) is the factor for influences operative in relation to the
epoch of birth (called here the cohort factor), and \( \beta \) is the secular factor which
operates in relation to the epoch of death (called here the time factor).

The following normal equations are derived from a 4 x 10 table of mortality
rates, for least squared deviations of Gompertzians \( G_{ij} \):

\[ \alpha_i \sum_j w_{ij} + \beta_j \sum_i w_{ij} + \gamma_k \sum_i w_{ij} - \sum_j w_{ij} G_{ij} = 0 \quad (i = 1, 2, \ldots, 10), \]  

(5)

\[ \beta_j \sum_i w_{ij} + \alpha_i \sum_j w_{ij} + \gamma_k \sum_i w_{ij} - \sum_j w_{ij} G_{ij} = 0 \quad (j = 1, 2, 3, 4), \]  

(6)

\[ \gamma_k \sum_i w_{ij} + \alpha_i \sum_j w_{ij} + \beta_j \sum_i w_{ij} - \sum_j w_{ij} G_{ij} = 0 \quad (k = 1, 2, \ldots, 13), \]  

(7)

where \( L \) is a variable orthogonal to \( k \), so that \( L = i + j - 1 \). Thus in any one of the
13 equations (7), summation is for the fixed \( k \), i.e. for all the range values of \( i \) and
\( j \) such that \( j - i = k - 10 \).
The weights \( w_{ij} \) have been chosen to be inversely proportional to the sampling variances of the Gompertzians \( G_{ij} \): that is,

\[
\frac{1}{w_{ij}} = \left( \frac{dG_{ij}}{de_{ij}} \right)^2 \text{var} (e_{ij}) \approx \frac{1}{e_{ij}}, \tag{8}
\]

assuming that the deaths in each group follow a Neyman Type A generalization of the Poisson distribution (with the extra parameter \( m \) common to all groups); and that the variances \( e_{ij}(1+m)/P_{ij}^2 \) of the rates are entirely due to the comparatively small number \( e_{ij} \) of deaths.

The program then computes the combinations of weights to be attached to each of the 27 factors according to the 27 normal equations. There are 10 factors for age, 4 for time and 13 for cohort, giving a 27 \( \times \) 27 matrix of coefficients. The solution to the normal equations has 3 arbitrary constants, since the matrix of coefficients of the factors, though symmetrical and positive semi-definite, is singular and of nullity 3. A generalized inverse is obtained by means of a subroutine based on the method of Healy (1968). The program sets \( \beta_4 = 0 \), \( \gamma_{13} = 0 \) and \( \gamma_{12} = 0 \) in the particular solution provided. The conditions \( \beta_4 = 0 \) and \( \gamma_{13} = 0 \) present no difficulty as they correspond to datum levels relative to the most recent quinquennium of deaths (in 1966–70) and to the most recent cohort of women respectively. If desired, a constant \( k_1 \) can be added to the four time factors, and another constant \( k_2 \) to the 13 cohort factors, provided that \( k_1 + k_2 \) is subtracted from each age factor. This would not affect the overall form of the factors. The third condition arises as follows: an arbitrary linear trend can be added to the time factors, and the same linear trend to the age factors (beginning at ages 25–29, 1966–70), provided that the same trend is subtracted from the cohort factors. This has been done so as to transform \( \gamma_{12} = 0 \) into \( \beta_3 = 0 \), leaving \( \beta_4 = 0 \) and \( \gamma_{13} = 0 \) as before. The particular and transformed solutions are shown in Table 3.

In Table 4 are shown the observed and fitted values of \( G_{ij} \) \((G_{10,1} \text{ and } G_{14,4} \text{ are necessarily fitted exactly})\). The fitted values are for both the particular and the

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1951–5</td>
<td>(a)</td>
<td>−6.89</td>
<td>−5.89</td>
<td>−5.89</td>
<td>−4.68</td>
<td>−4.30</td>
<td>−3.90</td>
<td>−3.67</td>
<td>−3.56</td>
<td>−3.49</td>
</tr>
<tr>
<td></td>
<td>(b)</td>
<td>−6.80</td>
<td>−5.69</td>
<td>−5.89</td>
<td>−4.61</td>
<td>−4.28</td>
<td>−3.92</td>
<td>−3.73</td>
<td>−3.56</td>
<td>−3.49</td>
</tr>
<tr>
<td>1956–60</td>
<td>(a)</td>
<td>−7.06</td>
<td>−6.61</td>
<td>−5.90</td>
<td>−4.43</td>
<td>−4.17</td>
<td>−4.01</td>
<td>−3.92</td>
<td>−3.70</td>
<td>−3.56</td>
</tr>
<tr>
<td></td>
<td>(b)</td>
<td>−7.19</td>
<td>−5.76</td>
<td>−4.94</td>
<td>−4.46</td>
<td>−4.18</td>
<td>−4.03</td>
<td>−3.85</td>
<td>−3.67</td>
<td>−3.57</td>
</tr>
<tr>
<td>1961–5</td>
<td>(a)</td>
<td>−7.55</td>
<td>−6.31</td>
<td>−5.01</td>
<td>−4.31</td>
<td>−4.02</td>
<td>−3.98</td>
<td>−4.03</td>
<td>−3.81</td>
<td>−3.77</td>
</tr>
<tr>
<td></td>
<td>(b)</td>
<td>−7.51</td>
<td>−6.12</td>
<td>−5.04</td>
<td>−4.38</td>
<td>−4.07</td>
<td>−3.97</td>
<td>−4.00</td>
<td>−3.82</td>
<td>−3.71</td>
</tr>
<tr>
<td>1966–70</td>
<td>(a)</td>
<td>−7.34</td>
<td>−6.49</td>
<td>−5.44</td>
<td>−4.55</td>
<td>−4.04</td>
<td>−3.90</td>
<td>−3.91</td>
<td>−3.95</td>
<td>−3.82</td>
</tr>
<tr>
<td></td>
<td>(b)</td>
<td>−7.34</td>
<td>−6.51</td>
<td>−5.47</td>
<td>−4.48</td>
<td>−3.99</td>
<td>−3.86</td>
<td>−3.94</td>
<td>−3.97</td>
<td>−3.87</td>
</tr>
</tbody>
</table>
Mortality from cervical cancer

transformed solution, as may be verified. In the observed and fitted values, the pattern falling with time is broken by the intrusion as described by Hill & Adelstein (1967), beginning with women aged 35–39 in 1956–60 and reaching women aged 55–59 in 1966–70. If, for the particular solution, corresponding values of $\alpha_{ij}$ are found before the addition of the time factors $\beta_j$, it appears that the pattern falls in every age group, whence it might seem that the intrusion in certain age groups could be regarded as due to the superimposition of the time factors on the age and cohort factors. That inference, however, would be incorrect, and stems from an unreasonably large positive trend in the time factors in the particular solution. Without further evidence, the most reasonable course seems to be instead to set this trend to zero in recent years, as in the transformed solution.

The fitted values in Table 4 were transformed to numbers of deaths in order to compare them with numbers of observed deaths. This test gave $\Sigma X = -1.3$ and, significantly, $\Sigma X^2 = 72.2$, with approximately 16 degrees of freedom, and no apparent pattern of individual deviations $X$. As the conditions strictly required for the use of a chi-square test are not present for these transformed values, the significance was checked by comparison with the theoretical distribution obtained from 12 Monte Carlo simulation runs, replacing the observed deaths in each cell by independent Poisson variables generated from them. This theoretical distribution has estimated mean values $\Sigma X^2 = 16$ and $\Sigma X = -0.4$, a small bias. The test indicates that the deaths are more variable than a Poisson distribution of deaths would imply (or that we do not have the true populations at risk), but the hypothesis of a Neyman Type A distribution, with $m = 3.5$, is not rejected.

Two further analyses were performed in order to examine the robustness of the method. The first consists in making all the weights $w_{ij}$ equal ($i = 1, \ldots, 10; j = 1, \ldots, 4$). This procedure gave factors for age ($-7.34, -6.52, -5.95, -5.55, -5.32, -5.28, -5.42, -5.56, -5.77, -5.97$), for time ($-0.57, -0.31, -0.18, 0.00$) and for cohort ($3.08, 2.83, 2.54, 2.23, 1.91, 1.59, 1.47, 1.41, 1.33, 1.09, 0.49, 0.00, 0.00$), which differ very little from the particular solution above. The second analysis consists in omitting the time variable and regressing on ages and cohorts only. Again, the resulting factors for age ($-7.34, -6.33, -6.00, -5.03, -4.63, -4.41, -4.37, -4.33, -4.36, -4.36$) and cohort ($0.90, 0.88, 0.77, 0.64, 0.50, 0.36, 0.43, 0.55, 0.65, 0.55, 0.13, -0.18, 0.00$) are fairly similar in form to the transformed solution above. The extent of agreement with results from these further analyses, to which the subsequent discussion equally applies, appears to provide some support for the present procedure, even if the method of choosing weights and populations at risk (which include single women) is not quite optimal.

**DISCUSSION**

In the transformed results the age factors rise steeply up to an age between 60 and 65 and then fall very slightly. A peak in incidence at ages 50–54 has, it is interesting to note, been found for cancer of the cervix in Norway (Breland,
1951), and peaks in morbidity in the same age group for Sweden for various periods of time between 1920 and 1949 (Lindell, 1952). Approximately 90% of deaths from cancer of the cervix occur within 5 years of the diagnosis (Bailar, 1964).

The cohort factors show decreases in successive cohorts, except for those born around 1941, 1916, 1921 and 1911, in order of magnitude of the increase from the cohort born 5 years earlier. Thus two groups of intrusions appear. One of these corresponds to that already described (Hill & Adelstein, 1967). The other refers to the most recent cohort for which data are available. But it will be necessary to look closely at data for 1971–5 to check the existence of the second intrusion, since it is based on rather few data and on an assumption about the time trend.

A general decline in cohort factors may reflect a progressive environmental diminution of a carcinogenic agent (such as a virus to which the women are exposed) so that at important ages (or even in utero) females are exposed to less hazard, or fewer are exposed, than among their forerunners at the same ages. Provided that these ages are fairly constant and that their relative importance within cohorts does not change much, one is not restricted to an influence acting only at birth, or even at a single age, in order to use cohort factors. Important ages in this connexion may be ages at marriage or at first intercourse, considering the statistical link which has been reported between these ages and the frequency of the disease (Rotkin, 1967; Terris, Wilson, Smith & Nelson, 1967); also age at menopause (Kashgarian & Dunn, 1970).

A progressive change in the importance of different ages, such as that produced perhaps by exposure to a hazard at progressively earlier ages at marriage, or by increases in proportions married, or by changes in the age at menopause, occurring during this century, might be reflected chiefly in the cohort factors. Such an effect may contribute to the intrusions seen. On the other hand, any tendency towards more accurate or greater certification of deaths from cervical cancer between 1950 and 1970 should principally be represented in time factors. Its extent is thought to be small, but a ratio of only 0.85 has been reported for clinicians’ to pathologists’ diagnoses for deaths from this cause (Heasman & Lipworth, 1966). In the opposite direction should be the effect of a rising trend in the frequency of hysterectomy (Fairbairn & Acheson, 1969), and the effect on mortality, which has been questioned (Ahluwalia & Doll, 1968), of mass screening for carcinoma in situ. Misstatements of age are thought to contribute very little to the factors, but the grouping may warrant further investigation.

The present results are consistent with those of Hill & Adelstein (1967). A more definitive interpretation than that offered here, and greater discrimination between various models, may depend on the use of further data. It remains an open question how far it is proper to represent influences that may, for all we know, span much of a lifetime, by factors depending only on age at death, epoch of birth and epoch of death.

The author is grateful to colleagues for suggestions.
Mortality from cervical cancer

REFERENCES


