

Interpretation of disease time trends: is cancer on the increase? A simple cohort technique and its relationship to more advanced models

M J GARDNER AND C OSMOND

From the Medical Research Council's Environmental Epidemiology Unit (University of Southampton), Southampton General Hospital, Southampton SO9 4XY, UK

SUMMARY Analysis of trends in mortality from respiratory cancer among women shows that, contrary to previous interpretation, there is no suggestion of any increase over time which might be due to recent increases in exposures to carcinogens. Although there are upward trends in the number of deaths, the crude death rate, and the age standardised rate, these are shown to be related to the aging of earlier cohorts of women who have experienced high mortality rates. More recent cohorts, born since the middle 1920s, show a decline. A simple technique to identify trends in different cohorts is described, and it is shown to be linked to the age-period-cohort modelling approach to investigating time trends. Dangers inherent in ignoring either period or cohort effects when describing one of these factors are discussed.

Epstein and Swartz contend that time trends from cancer in both England and Wales and the United States lead to the conclusion that rates are increasing, and suggest that the contribution of occupational exposures towards these rising trends has been underestimated.¹ Their interpretation is based on a study of changes in mortality between two years using crude death rates in England and Wales (1971 and 1977) and age-adjusted death rates in the United States (1969 and 1976). They show that in each country the annual average changes in rates are upwards for cancer considered overall, and that the same is true for some—but not all—individual sites.

In both countries the rise in death rates (and incidence rates in the United States) is shown to be highest from respiratory cancer in women, with annual reported percentage increases of from 5.3% to 8.6% between the years examined. By taking cancer of this site and looking at the data in more detail, it is shown that these increases cannot be ascribed to recent changes in the environment, but rather are associated with generation effects that now appear to be on the wane.

Material and methods

The table shows the basic data on mortality from cancer of the lung (ICD Nos 162, 163) for women in

England and Wales by age during the two years 1971 and 1976.^{2,3} For a reason discussed below we choose to use 1976 in place of 1977, but this change would not affect the thesis of Epstein and Swartz.¹ The crude death rates were 226 per million in 1971 and 278 per million in 1976—an increase of 23% over five years, or an annual average of 4.6% (similar to the figure of 5.3% quoted by Epstein and Swartz).

If age standardisation is incorporated into the calculations, to adjust the comparison for the effect of the older age of the 1976 population, then the standardised death rates (standardised to the 1976 population age structure) are 233 per million for 1971 and 278 per million for 1976. These figures would imply a slightly smaller increase of 19% over the five years or a 3.9% annual average.

Each of these two indices is an "average" rate over the complete age range, and it is pertinent to study carefully what is happening at separate ages. It will be seen clearly from the table that the ratios of the age specific death rates for 1976 to those for 1971 vary considerably in different five year age groups (from 0.74 to 1.26), and show an inconsistent decrease from older to younger ages. Thus the 1976/71 ratios of the crude death rates (1.23) and the age standardised death rates (1.19) are not representative of the ratios over the whole age specific range but are similar to those of the older age

Mortality from lung cancer in women during 1971 and 1976 in England and Wales by age and cohort of birth

Age group	1971			1976			Ratio of rates 1976/1971	Central years of birth cohort		Ratio of 1976 birth cohort to cohort of	
	Population (thousands)	No of deaths	Rate per million	Population (thousands)	No of deaths	Rate per million		1971	1976	1893/4*	1933/4†
0-24	9 180.8	4	0.4	8 863.6	5	0.6	1.29	—	—	—	—
25-29	1 615.8	8	5	1 843.0	9	5	0.99	1943/4	1948/9	2.23	0.66
30-34	1 413.6	17	12	1 568.9	17	11	0.90	1938/9	1943/4	2.26	0.67
35-39	1 379.9	44	32	1 394.1	33	24	0.74	1933/4	1938/9	2.51	0.74
40-44	1 464.3	103	70	1 360.3	105	77	1.10	1928/9	1933/4	3.38	1
45-49	1 569.9	288	183	1 439.6	279	194	1.06	1923/4	1928/9	3.08	0.91
50-54	1 486.2	464	312	1 539.5	531	345	1.10	1918/9	1923/4	2.92	0.86
55-59	1 534.2	675	440	1 446.4	729	504	1.15	1913/4	1918/9	2.64	0.78
60-64	1 499.3	930	620	1 461.3	1157	792	1.28	1908/9	1913/4	2.30	0.68
65-59	1 332.1	992	745	1 392.4	1274	915	1.23	1903/4	1908/9	1.81	0.53
70-74	1 085.1	916	844	1 174.0	1153	982	1.16	1898/9	1903/4	1.47	0.43
75-79	744.6	643	864	867.3	946	1091	1.26	1893/4	1898/9	1.26	0.37
≥80	789.6	587	743	880.7	771	875	1.18	—	—	—	—
All ages	25 095.4	5671	226	25 231.1	7009	278	1.23	—	—	—	—

*1893/4 ratio = 1 †1893/4 ratio = 0.30

groups. Moreover, the large overall increases in the crude and age standardised rates mask the fact that for women under the age of 40 the death rates from lung cancer actually decreased, although the numbers of deaths were relatively small.

The table also gives the central years of births for women who were in each of the five year age groups during 1971 and 1976. For example, to be aged 45-49 during 1971 (and at risk for that age specific death rate) a woman would have had to be born during the years 1921 to 1926 inclusive. We have labelled this birth cohort by the central years included—that is, 1923/4. These same women would have aged to be 50-54 years during 1976, and at risk in this higher age group. Extending the argument, it is seen that the cohorts in the table are five years apart by central date of birth, and (except for the extremes) are represented by two successive age specific rates. Deaths under the age of 25 and at age 80 or over are excluded from the main analyses. From this starting point the trends in age specific rates between 1971 and 1976 can thus be related to different birth cohorts.

Results

The ratio of the lung cancer rates in one cohort to the preceding cohort may be obtained by dividing the 1976 age specific rate by that for 1971. For example, the ratio of the 1898/9 to 1893/4 cohorts is given by $1091/864 = 1.26$, exactly the age specific rate ratio mentioned before. If we take 1893/4 as the "base" cohort it is possible to relate all other cohorts to this

base by multiplying ratios as we move upwards through the table. Thus the estimated ratio for the 1903/4 cohort relative to that of 1893/4 would be $(1091/864) \times (982/844) = 1.26 \times 1.16 = 1.47$. The results of these calculations are shown in the penultimate column of the table, and it can be seen that there is a rise until the 1933/4 cohort followed by a decline. The peak reached is over three times higher than the mortality for the base cohort of 1893/4.

It is possible, very simply, to use any other birth cohort as the base since the multiplicative relations between the rows of the table will hold wherever the starting point. As an alternative presentation, the final column of the table shows the results using the cohort with the highest ratio relative to 1893/4 as the base. This enables other cohorts to be seen easily as proportions of the maximum—for example, the 1908/9 cohort had an estimated 53% of the mortality of the peak 1933/4 cohort.

Figure 1 shows the results for successive birth cohorts graphically related to the 1893/4 cohort as unity. The results for men, obtained in a similar manner to those for women but not given in tabular detail, are plotted alongside for comparison. The much earlier peak cohort of 1898/9 for men, with a continuing decline thereafter, is noticeable.

Discussion

The apparent period increase in lung cancer mortality among women between 1971 and 1976 has thus been shown to be related to the aging of the cohorts with highest death rates. Since the death rate

from lung cancer also rises with increasing age, this results in both higher crude and age standardised death rates in 1976 than in 1971. Nevertheless, the limitation of the overall 1976/71 increase to ages over 40—corresponding to cohorts born before the middle 1930s—is clear, with decreases taking place at younger ages. There is seen to be a remarkable increase from the cohorts born at the end of the nineteenth century, reaching a peak (with over three times the mortality) for the cohort born around 1933/4. Since then the mortality among future cohorts of women has so far shown a decline, although the last two points are based on small numbers of deaths.

For men the peak mortality was in a much earlier cohort 1898/9 with a decline in following cohorts. The graphs in figure 1 are similar to those produced for lung cancer in England and Wales from mortality data over a much longer period (1951–80 inclusive), and using more advanced statistical modelling techniques.⁴⁻⁶ The intuitive approach adopted in this paper, however, has a close relationship with these age-period-cohort models. It is shown, in the appendix, to be equivalent to an age-cohort model—that is, a model without a term representing period of death. In this sense the method would be misleading if there were substantial period effects taking place—but for lung cancer these are not evident.⁶

The results of a complete analysis for the 1951–80 lung cancer mortality data, including age, period, and cohort effects, are shown diagrammatically in fig 2. The age values obtained from the model outlined in the appendix are not depicted for brevity, but are given elsewhere.⁶ For both men and women the

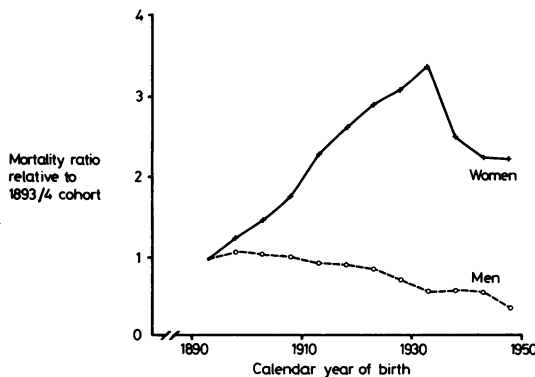


Fig 1 Mortality from lung cancer in England and Wales by year of birth relative to the cohort born around 1893/4, based on deaths during 1971 and 1976.

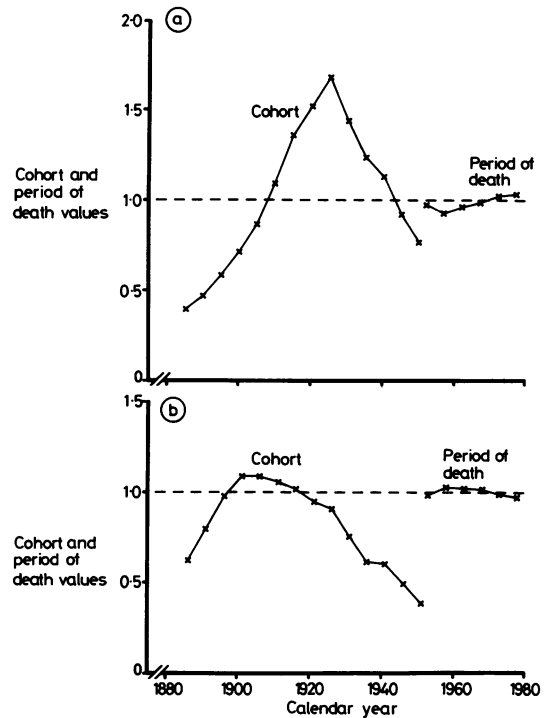


Fig 2 Mortality from lung cancer in England and Wales by calendar year of birth and death based on deaths during 1951–80. Cohort and period values, each with weighted averages of unity,⁵ are from the age-period-cohort model (see appendix). (a) women, (b) men.

dominant effects are those of the birth cohorts, with little variation being indicated between the different periods of death. Using the larger set of data gives a similar overall picture to that from the two separate years, with increasing cohort values being followed by declines. The dates of the peaks are slightly different in fig 2—and presumably more precise than in fig 1—being 1925/6 for women and 1900/1 for men.

The difficulty described in the last but one paragraph also makes the use of the standardised cohort mortality ratio, described by Beral,⁷ potentially misleading. The use of these techniques in the presence of any period effects, consistent in each age group, would produce apparent rather than real differences between cohorts. Similarly, the presence of pronounced cohort effects makes the use of simple period indices—such as the SMR—equally prone to misinterpretation. This has happened in the present example.¹ A more detailed description of these problems is given by Gardner and Osmond (submitted for publication).

It was mentioned above that we had chosen to use 1971 and 1976 (rather than 1971 and 1977) to describe the simple cohort approach adopted in our analysis. The reason for this is that it enables the same birth cohorts to be represented in the table by the use of years that are five years apart in time. To use the original years included by Epstein and Swartz,¹ it would be necessary to make some assumption about the relative mortality rates of successive cohorts—such as in the following. The 1977 lung cancer death rate at ages 75–79 in women was 1084 per million, and related to births around 1899/1900 (rather than around 1898/9 for 1976). This rate has a ratio of 1.25 to that of the 1893/4 cohort and, if a constant increase between successive annual birth cohorts is assumed, an estimate for 1898/9 compared with 1893/4 would be $1 + 0.25 \times 5/6 = 1.21$. The remainder of the analysis would follow as before.

In summary, there is no support in the available data for a specific period effect on lung cancer mortality, such as might be expected if carcinogenic substances had been recently introduced into the environment. In particular, lung cancer death rates among younger generations of women who started work recently show a decline. Equally, this is true for men. The observed trends are more suggestive of relationships with the increasing cigarette smoking habits of earlier generations, with the increasing use of filter tips and decreasing tar content of later cigarettes, and with the reduction in levels of atmospheric pollution after the Clean Air Act, all of which have been discussed by several authors.^{8–11} Lung cancer mortality remains high relative to other cancers, however, and further falls in both sexes are needed to reach the low rates of cohorts born earlier in the nineteenth century.

In this analysis cancer of the lung in women has been looked at in detail because it was the site suggested by Epstein and Swartz as having the fastest rising time trend. Examination of the trends in mortality from cancer overall for men and for women likewise show no indication of an increase in recent years or in recent generations.⁶ Several other individual cancers, such as those of the testis and cervix, have pronounced cohort effects, and these must always be considered when drawing conclusions about time trends.

Appendix

We use the following notation to describe the age-period-cohort modelling approach and its relationship to the technique described in the paper. In a collection of data, such as in the table, let:

- I = number of consecutive five year age groups
- J = number of five year spaced periods
- K = number of cohorts ($K = I + J - 1$)
- R = (r_{ij}) matrix of death rates
- D = (d_{ij}) matrix of numbers of deaths.

The age-period-cohort model may then be written as:

$$\log r_{ij} = \log a_i + \log p_j + \log c_{I-i+j}$$

where a_i , p_j , and c_{I-i+j} are the age, period, and cohort effects respectively. The method of estimation of the parameters in the model is described by Osmond and Gardner,⁵ based on the minimisation of

$$\sum_{i,j} d_{ij} (\log r_{ij} - \log a_i - \log p_j - \log c_{I-i+j}).^2$$

If it is assumed that period effects are absent, that is $p_j = 1$ for all $j = 1, \dots, J$, then the model reduces to an age-cohort representation. The minimisation produces $2I + J - 2$ independent normal equations for $2I + J - 1$ parameters, which requires one constraint to be put on the parameters to obtain solutions. If for $J = 2$ the constraint $c_1 = 1$ is imposed, then the solutions of the normal equations are equivalent to those from the intuitive approach described in the body of the paper—namely, for successive cohorts forward in time,

$$\hat{c}_{I-i+2} = \prod_{l=i}^{I-1} \frac{r_{l2}}{r_{l1}}$$

for $i = I, I - 1, \dots, 1$.

References

- ¹ Epstein SS, Swartz JB. Fallacies of lifestyle cancer theories. *Nature* 1981; **289**: 127–30.
- ² Registrar General. *Statistical review of England and Wales*. Part 1. *Medical tables, 1971*. London: HMSO, 1973.
- ³ Office of Population Censuses and Surveys. *Mortality statistics: cause, 1976*. Monitor series DH2, No 3. London: HMSO, 1977.
- ⁴ Osmond C, Gardner MJ, Acheson ED. An analysis of trends in cancer mortality in England and Wales during 1951–80 separating changes associated with period of birth and period of death. *Br Med J* 1982; **284**: 1005–8.
- ⁵ Osmond C, Gardner MJ. Age, period and cohort models applied to cancer mortality rates. *Statistics in Medicine* 1982; **1**: 245–59.
- ⁶ Osmond C, Gardner MJ, Acheson ED, Adelstein AM. Trends in cancer mortality separating changes associated with age, period of birth and period of death. Office of Population Censuses and Surveys publication (in press).
- ⁷ Beral V. Cancer of the cervix: a sexually transmitted infection? *Lancet* 1974; **i**: 1037–9.
- ⁸ Lee PN, ed. *Statistics of smoking in the United Kingdom*. Research paper 1. 7th ed. London: Tobacco Research Council, 1976.

- ⁹Todd GF, Lee PN, Wilson MJ. *Cohort analysis of cigarette smoking and of mortality from four associated diseases*. Occasional paper 3. London: Tobacco Research Council, 1976.
- ¹⁰Wald N, Doll R, Copeland G. Trends in tar, nicotine, and carbon monoxide yields of United Kingdom manufactured cigarettes since 1934. *Br Med J* 1981; **282**: 763-5.
- ¹¹Adelstein AM. Encouragement from recent statistics. In: Raven RW, ed. *Outlook on cancer*. New York and London: Plenum Press, 1977.