Journal of Epidemiology and Community Health, 1978, **32**, 314-316

The redundant factor method and bladder cancer mortality

JOHN C. BARRETT

From the Department of Medical Statistics and Epidemiology, London School of Hygiene and Tropical Medicine

SUMMARY Of the three factors, age at death, epoch of death, and epoch of birth, one seems almost superfluous. It may nevertheless be worthwhile to include all three in a mortality analysis, allowing for the constraints that the redundancy imposes. This procedure is applied to data for England and Wales on bladder cancer mortality from 1951 to 1970.

Sex and age have long been recognised as among the most important factors affecting mortality from various causes. There is less agreement, however, about whether the deceased should be classified further according to the secular years in which they died, or preferably according to their birth cohort, or both, for the purposes of analysis. If we know the cohort (by birth or marriage, etc.) to which a woman belongs, and her age (at death, or at the diagnosis of an illness, or at the initiation of some disease process), then we also know the time period in which this latter event occurred. At first sight, this last piece of information may seem redundant, since it can be calculated from the other two; thus, if a woman is aged 71 and was born in 1906, we must be dealing with the year 1977 or 1978. Nevertheless, it sometimes appears advantageous to include in a mortality analysis all the three factors of age, cohort (of birth), and time (year of death), recognising that one of them is redundant.

Writing on 'generation' mortality, Kermack *et al.* (1934) showed that specific death rates can be usefully represented as the product of two factors, one of which is a function of age alone, and the other a function of the year of birth alone. They inferred that the important factor from the point of view of the health of the individual during his whole life was his environment up to the age of say 15 years. Their results stemmed from the comparative constancy of the terms in any diagonal table of mortalities (relative to a base year) for Sweden, Scotland, and the United Kingdom.

The use of cohorts is essential for diseases with long periods of induction, such as cancers, where a decline in cohort factors may reflect a progressive environmental diminution of a carcinogenic agent, so that at certain ages men or women are exposed to less hazard, or fewer are exposed, than their forerunners at the same ages. On the other hand, a change in the accuracy, or extent of coverage, of death certification, occurring in a particular year, and affecting all age groups for that cause of death, may be indicated by the year of death, and so also may the sudden introduction of a more effective treatment at all ages, or possibly a screening programme. Conceivably, both kinds of influences may be acting at once, making it desirable to consider all three factors of age, cohort, and time. Because of the aforementioned redundancy, therefore, a linear trend in the cohort factors, if the age groupings and time periods are all equal, can always be represented instead by a corresponding equal and opposite trend in the age and time factors. Thus, the product of a doubling of mortality in each five-year age group from 25 upwards, and a halving of mortality in each successive quinquennium, can just as well be regarded instead as a halving of mortality in successive five-year birth cohorts. The solution proposed here is to confine the interpretation to those patterns in the resulting factors that are free of linear trend, when the cohort and time factors retain their separate utility as regards sudden changes of slope and other features. A similar problem occurs in analyses of employment (Price, 1976), in educational achievement, and, more acutely, in fertility.

Data and methods

The method is applied to data for mortality from cancer of the bladder, using data for England and Wales published by the Office of Population Censuses and Surveys (1975), shown in Tables 1 and 2. It has previously been used in application to mortality from cancer of the cervix (Barrett, 1973), and it is a development of earlier work by Sacher (1960) in application to tuberculosis.

The mortality rates are transformed logarithmically:

$$G_{ij}=\log (m_{ij}/P_{ij})$$

where m_{ij} are the deaths and P_{ij} the person-years at risk for the five-year age classes $i=1, 2, \ldots 10$ and the epochs of death $j=1, 2, 3, 4$ (in this case the quinquennia for the current data 1951–1970). The birth groups from 1876 to 1936 when centred on means are denoted by $k=1, 2, \ldots 13$. In this case:

$$k=j-i+10.$$

The model proposed is:

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

where α is the age factor, β the secular or time factor, and γ the cohort factor.

Weights for the data are chosen to be inversely proportional to the sampling variances of the G_{ij}: that is:

$$\frac{1}{W_{ij}}=\left(\frac{dG_{ij}}{dm_{ij}}\right)^2 \text{Var } m_{ij}=\frac{1}{m_{ij}}$$

assuming that the deaths in each group follow a Poisson distribution and that the variances m_{ij}/P_{ij}^2 of the rates are entirely due to the comparatively small number m_{ij} of deaths.

The programme computes the combinations of weights to be attached to each of the 27 factors in the 27 normal equations (10 factors for age, 13 for cohort, and 4 for time). The solution has 3 arbitrary constants, since the singular matrix of coefficients is of nullity 3. A generalised inverse matrix is obtained by the routine of Healy (1968). Alternatively, the GLIM package may be used to obtain the same results.

Because of the redundancy mentioned earlier, only one solution among many equivalent ones is provided, in this case with $\beta_4=0$, $\gamma_{13}=0$ and $\gamma_{12}=0$. The first two conditions correspond to datum levels for time and cohort factors: a constant k_1 can be added to the four time factors and another constant k_2 to the 13 cohort factors, if k_1+k_2 is subtracted from each age factor. As regards the third condition, an arbitrary linear trend can be added to the time factors, and the same linear trend added to the age factors (beginning at ages 30–34, in the quinquennium 1966–70), if the same trend is subtracted from the cohorts. Here the condition $\gamma_{12}=0$ has been transformed into $\beta_1=\beta_4$, leaving $\beta_4=0$ and $\gamma_{13}=0$ as in the particular solution.

Results

The fitted factors are shown in Table 3. The values G_{ij} were converted to expected numbers of deaths to compare with observed numbers. This indicated $\chi^2_{16}=21{\cdot}7$ for females and $\chi^2_{16}=15{\cdot}8$ for males, so the model appears to be acceptable ($P>0{\cdot}1$). It turns out in this case that the time factors are all close to zero. This may be because treatment and death certification have not changed significantly in the period under consideration. Substantially the same results would be obtained if the time factor were omitted from this analysis. It has been included, however, since the procedure illustrated should in some applications indicate sharp changes in time factors.

Table 1 *Cancer of the bladder deaths* (m_{ij})

	30–	35–	40–	45–	50–	55–	60–	65–	70–	75–79
Females										
1951–55	14	17	46	113	194	293	401	592	725	783
1956–60	3	21	43	93	166	286	444	665	794	875
1961–65	4	18	58	96	187	322	464	646	829	938
1966–70	0	12	36	96	155	328	500	732	926	1043
Males										
1951–55	12	36	112	311	600	845	1268	1580	1742	1541
1956–60	9	38	122	285	530	1001	1340	1724	1863	1715
1961–65	10	44	116	247	551	1049	1692	1899	2102	1834
1966–70	8	33	93	268	505	1083	1801	2515	2321	2043

Table 2 *Populations (thousands person-years)* (P_{ij})

	30–	35–	40–	45–	50–	55–	60–	65–	70–	75–79
Females										
1951–55	8330	7815	8472	8279	7664	6880	6148	5328	4300	2939
1956–60	7747	8289	7740	8328	8088	7400	6525	5600	4553	3272
1961–65	7243	7726	8213	7663	8104	7830	7017	5946	4800	3552
1966–70	7047	7189	7663	8101	7462	7918	7435	6454	5141	3746
Males										
1951–55	8161	7553	8245	8013	7001	5657	4776	3915	2967	1905
1956–60	7696	8066	7467	8097	7689	6551	5061	4005	2987	1956
1961–65	7407	7752	8082	7385	7797	7245	5948	4233	3059	1982
1966–70	7314	7316	7631	7882	7086	7344	6512	4949	3201	2017

Table 3 *Derived factors*

Age factors	30–	35–	40–	45–	50–	55–	60–	65–	70–	75–79
Males	−9·1	−7·8	−6·8	−6·0	−5·3	−4·6	−4·0	−3·4	−3·0	−2·6
Females	−9·8	−9·1	−8·2	−7·5	−6·9	−6·3	−5·8	−5·3	−4·9	−4·4

Time factors	1951–55	1956–60	1961–65	1966–70
Males	0·00	−0·03	−0·03	0·00
Females	0·00	−0·01	−0·02	0·00

Cohort factors	1876	1881	1886	1891	1896	1901	1906	1911	1916	1921	1926	1931	1936
Males	0·09	0·18	0·24	0·33	0·37	0·45	0·37	0·35	0·32	0·28	0·20	0·16	0·00
Females	0·80	0·80	0·82	0·84	0·84	0·83	0·83	0·81	0·80	0·86	0·56	0·28	0·00

Discussion

The problem of the arbitrary linear trends contained in the solution, mentioned above, can be circumvented by restricting the interpretation to those features which are trend-free, for example, by considering the points at which relatively large changes of slope occur. The interpretation is not then affected by the three constraints, such as $\gamma_{13}=0$, $\beta_1=0$, $\beta_4=0$. The chief features that persist, notwithstanding the standard errors, are the downturn in the cohort factors for males born after 1901 and, similarly, the downturn for females born after 1921. For males, the mortality change may be attributable to smoking, as other studies have indicated (Cole *et al.*, 1971; Armstrong and Doll, 1974), since a related pattern is found in cigarette consumption. At the same time, it is noteworthy that the male cohort with the peak rate was one of those heavily involved in the armed services in the first world war, and since nitrites are a known precursor of certain nitroso-compounds that have been implicated (Hicks *et al.*, 1977), the larger quantities of tinned meat which they consumed may also deserve consideration. It is difficult to account for the peak among women born about 1920, that is, those who became adult in the second world war, with a subsequent decline in younger women. Perhaps the use of cosmetics at that time (European Committee, 1962) as well as employment in explosives, paint, or relevant chemical industries could warrant further investigation.

Professor D. D. Reid showed considerable interest in this technique, and suggested that it should be applied to other types of cancer mortality.

Reprints from John Barrett, Department of Medical Statistics and Epidemiology, Keppel Street, London WC1E 7HT.

References

Armstrong, B., and Doll, R. (1974). Bladder cancer mortality in England and Wales in relation to cigarette smoking and saccharin consumption. *British Journal of Preventive and Social Medicine,* **28,** 233–240.

Barrett, J. C. (1973). Age, time and cohort factors in mortality from cancer of the cervix. *Journal of Hygiene* (Cambridge), **71,** 253–259.

Cole, P., Monson, R. R., Haning, H., and Friedell, G. H. (1971). Smoking and cancer of the lower urinary tract. *New England Journal of Medicine,* **284,** 129.

European Committee on Chronic Toxicity Hazards (1962). The prevention of hazards from chronic toxicity of cosmetics and toilet preparations. *Journal of the Society of Cosmetic Chemists,* **13,** 322–347.

Healy, M. J. R. (1968). Multiple regression with a singular matrix. *Applied Statistics,* **17,** 110–117 and 195–199.

Hicks, R. M., Walters, C. L., Elsebai, I., El Aasser, A., El Merzebani, M., and Gough, T. A. (1977). Demonstration of nitrosamines in human urine: preliminary observations on a possible aetiology for bladder cancer in association with chronic urinary tract infections. *Proceedings of the Royal Society of Medicine,* **70,** 413–417.

Kermack, W. O., McKendrick, A. G., and McKinlay, P. L. (1934). Death rates in Great Britain and Sweden: some general regularities and their significance. *Lancet,* **1,** 698–703.

Office of Population Censuses and Surveys (1975). *Cancer mortality, England and Wales 1911–1970.* Studies in Medical and Population Subjects No. 29. HMSO: London.

Price, D. O. (1976). A respecification of variables in cohort analysis. American Statistical Association Meetings, August 23–26.

Sacher, G. A. (1960). Analysis of life tables with secular terms. In *The Biology of Aging. American Institute of Biological Sciences Symposium,* **6,** 253.