STATISTICS IN MEDICINE, VOL. 6, 449–467 (1987)

MODELS FOR TEMPORAL VARIATION IN CANCER RATES. I: AGE–PERIOD AND AGE–COHORT MODELS

D. CLAYTON

Department of Community Health, Clinical Sciences Building, Leicester Royal Infirmary, PO Box 65, Leicester LE2 7LX, U.K.

AND

E. SCHIFFLERS

Facultés Universitaires N.-D. de la Paix, Rue de Bruxelles, 61, B5000, Namur, Belgium

SUMMARY

A main concern of descriptive epidemiologists is the presentation and interpretation of temporal variations in cancer rates. In its simplest form, this problem is that of the analysis of a set of rates arranged in a two-way table by age group and calendar period. We review the modern approach to the analysis of such data which justifies traditional methods of age standardization in terms of the multiplicative risk model. We discuss the use of this model when the temporal variations are due to purely secular (period) influences and when they are attributable to generational (cohort) influences. Finally we demonstrate the serious difficulties which attend the interpretation of regular trends. The methods described are illustrated by examples for incidence rates of bladder cancer in Birmingham, U.K., mortality from bladder cancer in Italy, and mortality from lung cancer in Belgium.

KEY WORDS Cohort analysis Cancer trends Age–period–cohort models Age standardization

INTRODUCTION

In recent years, advances in statistical theory (particularly in the fields of log-linear models and survival analysis) have led to a re-evaluation of traditional methods of analysis of vital rates. Such methods as direct and indirect standardization have been based upon the definition of summary indices with desirable properties. For example, age standardized rates are indices of mortality which, for fixed age specific rates, remain constant under changes to the age structure of the population or cohort under study. A more modern approach, however, views such indices as estimates of parameters of a probabilistic model for mortality.

This approach has brought great benefits in the shape of a unification of the methodologies for the analysis of vital rates in descriptive epidemiology,[1] for the regression analysis of individual records in cohort studies[2] and for the analysis of matched and unmatched case-control studies.[3] However, this advance has not been achieved without cost. The purposes and methods of probabilistic modelling are still not as widely understood as statisticians tend to assume, and the correct interpretation of such analyses depends upon a level of understanding of applied mathematics beyond that demanded of previous generations of medically qualified epidemiologists. This may lead to incomprehension, or, perhaps worse, to serious over-interpretation.

0277–6715/87/040449–19$09.50

Received 6 March 1986
Revised 5 September 1986

The advantages and difficulties accompanying the model-based approach are illustrated in recent approaches to an old and fundamental methodological problem in epidemiology: the analysis of temporal variation in disease incidence or mortality. The new methods lead naturally to generalization of the method of indirect standardization to, eventually, estimation of parameters of the age–period–cohort model.[1,4] We believe that these methods are a useful advance, particularly for the purposes of comparison of temporal variation of disease in different populations, but they have brought serious difficulties and dangers. The models are beset by problems of identifiability, by which we mean that identical descriptions of data may be obtained from different sets of parameter values. Also, two such indistinguishable sets of parameter values may lead to quite different interpretations. Therefore, it is essential that the epidemiologist working with these models should fully understand the strengths and weaknesses of the approach and should be aware of the limits to inference. Alas, this process has not been aided by several recent papers which apparently have resolved the identifiability problems.[5-7] Unfortunately, all such attempts depend upon mathematical assumptions which have no biological basis.

This first paper deals with the use of log-linear models to describe variations in rates simply in terms either of the calendar period of observation or of the cohort or generation to whom the rates apply. Since no analysis in cancer epidemiology can ignore age, we are led to the age–period and age–cohort models, respectively. In the section on regular trends we encounter the problem of 'drift', a type of variation described equally well by either model. This introduces the problems of identifiability we tackle in the second paper[8] when we discuss the full age–period–cohort model.

ESTIMATION OF PERIOD EFFECTS: THE AGE–PERIOD MODEL

Table I displays incidence rates of bladder cancer together with the corresponding cases (expressed in 100,000 person-years observation) for males in the Birmingham (U.K.) cancer registry during the period 1960–1976 as published in *Cancer Incidence in Five Continents*.[9-12] These data are laid out in the table in a manner which reflects the method of collection; that is, with columns defining P calendar periods of observation and rows defining A age groups.

Naively, the most natural examination of the data is to plot, for each age group, the age-specific incidence rates against the central date of each period of observation. Here, as in all cancers, the age-specific rates vary over several orders of magnitude. However, by using a logarithmic scale for incidence rate, the trends for all age groups may be plotted on the same graph to facilitate comparison. The data from Table I are plotted in Figure 1.

The dominant impression given by this figure is one of parallelism of the curves. There was a sharp increase in incidence between the second and third observation periods in just about every age group. This parallelism implies two important characteristics of the causal influence which gave rise to this observation:

(i) it has either an immediate or a fixed delayed effect upon incidence, and
(ii) it is constant across all age-groups; that is, the logarithmically transformed incidence rates are increased (or decreased) by the same quantity regardless of age. Such an effect is termed a period effect.

When the observed variation in age-specific rates is entirely consistent with such influences, the curves of age-specific incidence rates against the date of observation (equivalent to Figure 1) would be parallel. If we denote by Y_{ap} the logarithm of the age-specific incidence rate for the ath age group measured during the pth observation period, then this parallelism may be expressed

Table I. Age-specific incidence rates (per 100,000 person-years observation) of bladder cancer for males in the region of Birmingham (U.K.), during the period 1960–1976. Numbers of cases on which rates are based are in parentheses. (Source: *Cancer Incidence in Five Continents*,[9–12] Vol. 1–Vol. 4)

Age/period	1960–1962	1963–1966	1968–1972	1973–1976
25–29	0·42 (2)	0·31 (2)	0·55 (5)	1·10 (9)
30–34	0·00 (0)	0·65 (4)	1.73 (14)	1·15 (8)
35–39	2·06 (11)	1·21 (8)	4·02 (31)	2·49 (16)
40–44	1·62 (8)	4·03 (28)	6·74 (55)	5·29 (33)
45–49	9·40 (48)	7·02 (45)	14·95 (126)	16·80 (107)
50–54	13·90 (67)	16·65 (108)	25.73 (199)	24·41 (164)
55–59	24·25 (102)	29·15 (171)	41·06 (309)	44·81 (245)
60–64	44·50 (141)	50·51 (253)	71·39 (469)	70·25 (372)
65–69	60·47 (135)	66·97 (226)	100·69 (514)	101·97 (440)
70–74	94·84 (150)	95·73 (210)	141·96 (450)	142·70 (420)
75–79	116·08 (116)	118·16 (159)	154·19 (276)	174·42 (270)

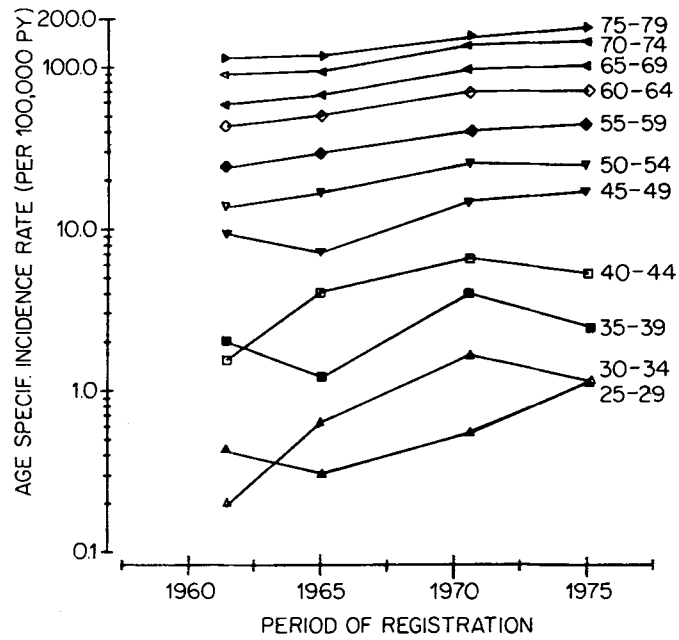

Figure 1. Incidence of bladder cancer in males in the region of Birmingham (U.K.) against central date of each period of observation. Rates are plotted using a logarithmic scale. The quasi parallelism of the age-specific curves constitutes the empirical basis for the models discussed in this paper (source: see Table I)

mathematically by the relationship

$$Y_{ap} = \alpha_a + \beta_p. \tag{1}$$

Thus, rather then displaying the $A \times P$ logarithmic-rates, Y_{ap}, the data may be reported in terms of the A parameters, α_a, which describe the relationship between age and incidence, and the P parameters, β_p, which describe the temporal relationship. It may be easier to think about the

relationship set out in equation (1) in terms of the incidence rates themselves rather than their logarithms. We denote these by ρ_{ap} and use a prime notation to denote antilogarithms, so that $\rho_{ap} = Y'_{ap}$. Then equation (1) can be re-expressed as

$$\rho_{ap} = \alpha'_a \beta'_p. \tag{2}$$

That is, the age-specific rates can be represented by products of A parameters describing the dependence upon age, α'_a, and P parameters describing their dependence upon period of observation, β'_p.

Earlier we introduced the term period effect in a descriptive sense. However, the term *effect* is widely used by statisticians in a quantitative sense to describe parameters of models such as (1) and (2). Thus, α_a and β_p would be termed (additive) age effects and period effects, respectively. It may be difficult to think in terms of an additive model for the logarithms of rates, and in general we prefer the form (2) in which the parameters should be termed multiplicative effects. Unless otherwise specified, we shall use the terms age effects and period effects to represent the multiplicative parameters, α'_a and β'_p, respectively.

Of course, the relationship (1) will never hold exactly, but it may hold sufficiently closely to provide a useful description of the data. In these circumstances, equation (1) (or, equivalently (2)) provides a statistical model for expected rates; the discrepancies between the rates observed and their expected values being regarded as random fluctuations.

With a model such as (1) or (2), a statistical analysis proceeds in two stages. First, by a process of estimation of the values of the parameters of the model, α_a and β_p here, which give rise to expected incidence rates which are as close as possible in some sense to the observed rates, and secondly by a process of criticism in which the discrepancies between observed and expected rates are examined to determine whether the model describes the data adequately. In the remainder of this section, we consider the estimation of the parameters α_a and β_p of the age–period model. The next section is dedicated to model criticism.

The detailed theory of estimation of parameters for best fit of the model need not concern us here. It falls within the theory of generalized linear models as described by Nelder and Wedderburn,[13] reviewed and extended by McCullagh and Nelder.[14] There are, however, two aspects of this problem which are of practical relevance and warrant some discussion. The first of these is the criteria we use to assess the overall goodness-of-fit of the model. One set of parameter values might give a very good fit in certain cells of the table, but may perform less well in others. On the other hand, a different set of parameter values might correct these discrepancies, but at some cost in terms of the fit to the remaining cells. The choice depends upon the predominant reasons for discrepancy between the observed rates and those predicted by the model. Often the dominant reason for fluctuation of observed rates is the natural fluctuation of numerators, which vary according to the well-known Poisson law. It is then appropriate to give less weight to poor prediction of observed rates based upon small numbers of cases. With this Poisson criterion, good fit of the rates based on larger numerators is more important.

Sometimes the Poisson criterion can be inappropriate. When considering a very common cancer, rates may be based upon such large numbers of incident cases that the variability predicted by the Poisson law becomes negligible for all practical purposes. In such situations, it is unlikely that the model may still be a useful one if the discrepancies between observed and expected rates, although larger than would be predicted by the Poisson law, are small enough to be considered of no importance or they show no systematic pattern. In these circumstances it is, perhaps, more appropriate to weight cells with few cases according to the Poisson assumption, but to restrict the weight given to cells with more substantial numerators (we shall return to this problem when discussing model criticism).

When Poisson errors are believed to predominate, the most common statistical method for fitting a model is that of maximum likelihood which provides a theoretical justification for the traditional method of indirect standardization. This is demonstrated by considering the multiplicative form (2) and pretending first that the α'_a are known. These known α'_a may be regarded as a standard set of age-specific rates and the parameters β'_p as relative risks expressing the ratio of incidence rates at period p to the standard set. In these circumstances, it can be shown that the method of maximum likelihood leads to the choice of the Standardized Mortality Ratios (SMRs) as the best estimate of the relative risks, β'_p.[15] However, this argument holds equally well if we regard the period effects, β'_p, as known and estimate the age effects, α'_a; we would again use SMR calculations. When neither set of parameters is known, a convenient method simply alternates between these computations until a stable solution is obtained. This was first proposed in this context by Mantel and Starke,[16] who termed it 'internal indirect standardization', but the method is a special case of a general algorithm for maximum likelihood estimation of log-linear models and is usually termed 'iterative proportional fitting' (IPF) (see for instance Bishop, Fienberg and Holland,[17] Section 3·5).

An alternative method which has been used in the past (for example, Barrett[18, 19]) is based upon the fact that the Poisson law predicts that the variance of a logarithmically transformed rate is given (approximately) by the inverse of its numerator. This suggests choosing α_a and β_p to minimize a weighted sum of squared deviations of logarithmic rates in which each cell of the table is weighted by the number of observed cases upon which it is based. This leads to nearly the same estimates as the method of maximum likelihood and avoids iterative computation.

Another technical problem in fitting the age–period model arises because it does not have a single set of best parameter values, α_a and β_p; they are therefore termed unidentifiable. Each set of expected rates, ρ_{ap}, which obey the model may indeed be obtained from more than one ensemble of parameter values, since if we add some constant, to all the α'_as in (1), but subtract the same constant from all the β'_ps, then the fitted logarithmic rates, Y_{ap} remain unchanged.

A model whose parameters are unidentifiable might seem of little practical utility. However, if we consider two periods, p and q, the model predicts that, for all age groups, the difference in logarithmetic incidence rates is constant, that is

$$(Y_{ap} - Y_{aq}) = \beta_p - \beta_q,$$

for all age groups. In terms of the incidence rates themselves this implies that the ratios of age-specific rates are also stable across age groups, that is

$$\rho_{ap}/\rho_{aq} = \beta'_p/\beta'_q$$

for all a. Thus, the difference between β_p and β_q is interpreted as the logarithm of the relative risk of period p relative to period q. All parameterizations which lead to the same fitted rates have the same differences, $(\beta_p - \beta_q)$. Similarly the difference $(\alpha_a - \alpha_b)$ is interpretable as the log relative risk of age group a relative to age group b. Such functions of the parameters are termed identifiable. Note that interpretations based solely on identifiable functions do not depend on any arbitrary selection of a specific parameterization. The model may be communicated in terms of any set of such comparisons. A natural choice is the P-1 first differences, $(\beta_2 - \beta_1)$, $(\beta_3 - \beta_2)$, whose antilogs represent the relative risk of the second period relative to the first, the third period relative to the second and so on. This set of comparisons focuses our attention upon regular trend models in which the first differences are approximately constant and upon deviations from such models. The same technique may be used to describe the age effects.

It is usual to choose some parameterization which leaves the αs looking rather like logarithms of age-specific incidence rates, and the βs looking like log-SMRs (relative risks). In this way, by taking

Table II. Bladder cancer incidence in the region of Birmingham. Age (α_a) and period (β_p) parameters estimated from the rates of Table I using the iterative proportional fitting (IPF) procedure. The scales for the multiplicative effects were chosen to facilitate interpretation

Age	Additive effects α_a	Multiplicative effects $\alpha'_a \times 100{,}000$
25–29	−12·31	0.45
30–34	−11·58	0·71
35–39	−10.90	1·85
40–44	−10·28	3·43
45–49	−9·32	8·97
50–54	−8·80	15·12
55–59	−8·26	25·94
60–64	−7·73	43·99
65–69	−7·39	61·62
70–74	−7·04	87·88
75–79	−6·87	103·43
Period	β_p	$\beta'_p \times 100(\%)$
1960–1962	0·00	100·0
1963–1966	0·09	109·2
1968–1972	0·49	162·4
1973–1976	0·50	165·4

antilogs, the estimates of α'_a and β'_p look like numbers we are well accustomed to in epidemiology. The simplest way in which this can be done is to adopt a parameterization in which one of the βs, usually β_1, is taken as zero; this is equivalent to displaying the results in terms of the differences $\beta_p - \beta_1$. With this convention, α'_a are the (fitted) age-specific rates for period 1, and β'_p are the (fitted) relative risks of each period relative to period 1. This method leads to the simplest interpretation. The main rival method centres the period effects, β_p, around zero (so that their antilogs, β'_p, are arranged around 1). Centring can be achieved by adding a constant to all the βs such that their mean becomes zero and subtracting the same constant from all the αs. Again this has the effect of making, the α'_a look like age-specific rates, and the β'_p look like SMRs, but they are not simply interpretable as such. Except for considerations of mathematical symmetry, the only motivation for this latter procedure is to preserve some analogy with the traditional technique of calculating SMRs using the marginal age-specific rates as standard (that is as estimates of α'_a). However, Mantel and Starke[16] showed that this procedure is not to be recommended since the marginal age-specific rates may badly misrepresent the time age gradient and do not, in general, form an appropriate base for comparisons.

The estimated parameter values for the data of Table I were calculated using the IPF method and are shown in Table II. It can be seen that the period effects, β'_p, confirm and quantify the impression of Figure 1 that there is a rather sharp increase in incidence between the second and third period.

It must be stressed that however we choose to communicate the model, the degree to which the analysis reproduces an accurate representation of the observed data depends upon whether the model does indeed fit our data. If not, then any parameterization is worthless. The next section is dedicated to the process of model criticism – the examination of the fit of the model.

ASSESSING THE FIT OF THE MODEL

The process of model criticism involves an examination of the discrepancies between the observed rates and the rates predicted by the best fit model. The differences between the observed log-rates, Y_{ap}, and the fitted log-rates $\hat{Y}_{ap}$, should be standardized to take account of the Poisson variability of the numerators. These quantities are called standardized residuals. Writing D_{ap}, for the observed number of cases in age group a at period p, and $\hat{D}_{ap}$ for the corresponding expected number obtained by multiplying the corresponding person-years observation by the fitted rate, three alternative definitions for the standardized residuals are widely used:

(i) a log-residual:

$$S_{ap}^{(1)} = (Y_{ap} - \hat{Y}_{ap})\sqrt{D_{ap}} \tag{3a}$$

(ii) a chi-residual:

$$S_{ap}^{(2)} = (D_{ap} - \hat{D}_{ap})/\sqrt{\hat{D}_{ap}} \tag{3b}$$

(iii) the deviance residual:

$$S_{ap}^{(3)} = \{2[D_{ap}\log(D_{ap}/\hat{D}_{ap}) - D_{ap} + \hat{D}_{ap}]\}^{1/2} \tag{3c}$$

with the sign of $(Y_{ap} - \hat{Y}_{ap})$.

For most purposes, these three definitions are equivalent. If any of the cells have no observed cases, the observed rate is zero, so that we cannot calculate its logarithm, Y_{ap}. In these circumstances, (3*a*) and (3*c*) break down and for this reason one might prefer to use (3*b*). These definitions of standardized residuals are closely related to our discussion, in the second section, of the choice of criterion for best model used for the purpose of estimation of parameters. The method of maximum likelihood minimizes the total sum of the squared deviance residuals, $S_{ap}^{(3)}$, over all cells of the tables (this is termed the deviance). The method of weighted least-squares minimizes the sum of squares of $S_{ap}^{(1)}$.

The decision as to the acceptability of the model depends first upon whether the residual variability is small enough to be of little practical importance. There can be no general rules on this point. If the residual variability is not negligible, it may be for any of three reasons:

(i) there may be widespread deviations from model assumptions which exhibit no discernable pattern,
(ii) there may be a few isolated cells with very large residuals, or
(iii) the residuals may exhibit a systematic pattern, for example, consistent underestimation of rates in one corner of the table.

If the first reason is the cause of our problems we have no further use of models – there is no option, but to present all the cells of the table in as clear a manner as is possible. Of course, it may be that the reason for the difficulty is simply poor quality of the basic data either in respect of numerators or denominators of rates. The second type of important deviation, if not attributable to simple transcribing errors, would indicate careful enquiry as to possible causes. For the presentation of data, it will often be acceptable to refit the model omitting aberrant cells, and report the observed (and expected) rates for them separately. If the residuals exhibit some clear pattern, however, this should indicate that some alternative model may provide a better data description. We shall discuss one special case in more detail in the next section and shall return to the general issue of interaction between age and period effects later. If, after examining the residuals, we are

Table III. Standardized residuals (chi-residuals) from the age–period model for the data of Table I

Age/period	1960–1962	1963–1966	1968–1972	1973–1976
25–29	−0·09	−0·65	−0·63	1·18
30–34	—	−0·37	1·50	−0·06
35–39	0·36	−1·46	1·63	−0·82
40–44	−2·18	0·38	1·41	−0·41
45–49	0·32	−2·25	0·28	1·29
50–54	−0·69	0·08	0·66	−0·31
55–59	−0·69	0·37	−0·45	0·68
60–64	0·13	0·79	−0·02	−0·68
65–69	−0·22	−0·07	0·14	0·18
70–74	0·93	−0·04	−0·12	−0·38
75–79	1·24	0·56	−1·43	0·32

reassured that the age–period model gives a good description of the data, then it is reasonable to proceed to test the statistical significance of the period effects.

Table III shows the standardized residuals (method (3*b*)). They are not larger than expected; the deviance is 41·17, and the corresponding degrees of freedom are $10 \times 3 = 30$. This value corresponds to a value of P, about 0·10 on the chi-squared distribution and hence is not unduly large.

When examining the standardized residuals for isolated aberrant cells we may assume that, if the model were true, the standardized residuals would be approximately normally distributed with zero mean and standard deviation, $\sqrt{\{(A-1)(P-1)/(AP)\}}$, in this case 0·91. Probability plotting methods can be used, but usually tables of the expected extreme range are sufficient. For example, the expected extreme values of 44 residuals with zero mean and standard deviation 0·91 are $\pm 2{\cdot}1 \times 0{\cdot}91 = \pm 1{\cdot}9$. Table IV reveals no systematic pattern nor does it contain aberrant values. In fact all but two residuals lay within the expected range.

If we conclude by this process of criticism that the age–period model with Poisson errors is an adequate description of our data, then it is appropriate to test the statistical significance of the period effect. To do this, we fit the model which omits period effects: the null hypothesis states that the same age-specific rates apply at all periods. This may be carried out by simply pooling the data over periods and calculating the marginal age-specific rates; these estimate the age effects, α'_a, under the null hypothesis. We then calculate the deviance test of the overall fit of this age-only model; the difference between this test statistic and that for the age-period model provides a test for the significance of the period effects. If there were no period effect, then this difference would be distributed as chi-squared with $(P-1)$ degrees of freedom. For the data of Table I, we have:

Model	Chi-squared	d.f.
Age	309·69	33
Age–period	36·27	30
Difference	273·42	3

Clearly, in this case, there is no doubt as to the significance of this period effect! Note, however, that this test, on $(P-1)$ degrees of freedom, is a test for any difference between the P periods and is

not especially sensitive to smoothly increasing or decreasing trends. We shall discuss this problem in the section on regular trends.

Before leaving the topic of examination of residuals, we should consider what might be done if the standardized residuals are larger than expected but do not exhibit any systematic pattern. As we stated earlier, this can occur when some rates are based on large numbers of cases such that even small and unimportant residuals are large in comparison with Poisson variability and in such situations we use unweighted least-squares backed up by analysis of variance. The model is fitted by minimizing a weighted sum of squared residuals, that is

$$\Sigma_{a,p} w_{ap} (Y_{ap} - \hat{Y}_{ap})^2 D_{ap} \tag{4}$$

where the weights, w_{ap}, are given by

$$w_{ap} = (1/\hat{D}_{ap} + \sigma^2)^{-1}. \tag{5}$$

Note that σ^2 is an unknown constant representing the squared coefficient of variation of the rates over and above their Poisson variability. This constant must be estimated from the table, and this is achieved by an iterative method which adjusts σ^2 at each step until the residual weighted sum of squares is equal to its degrees of freedom. Details of the iterative method and of the modifications necessary for testing hypotheses are given by Breslow.[20] However, such procedures are largely untried in practice, and in our experience are seldom necessary. Further discussion of the extra-Poisson variability is given in McCullagh and Nelder.[14]

Often the examination of residuals shows clearly that the model is systematically misleading. Then, the analysis of residuals can be instructive in suggesting alternative models. In the next section, we consider the most important simple alternative model for the age–period model.

ESTIMATION OF COHORT EFFECTS: THE AGE–COHORT MODEL

Table IV shows deaths from bladder cancer and corresponding mortality rates for Italian males in the period 1955 to 1979. By contrast to the Birmingham bladder cancer incidence data, an attempt to fit the age–period models is not very successful. The global deviance chi-squared test of fit of the age–period model yields deviance = 455 on 40 degrees of freedom; highly significant indeed.

Table V examines the residuals for this model in more detail. Looking carefully at this table, it can be seen that there is a systematic pattern in the residuals; there is a tendency for ratios of observed to expected mortality rates to decrease regularly along diagonals running downwards from left to right across the table. The explanation for this becomes apparent when one asks what cells along a diagonal have in common; the answer is a high proportion of the same people! Since the periods are spaced by five years, and the age groups also spaced by five years, on average the people studied in age group a at period p will be in age group $(a+1)$ when their mortality is studied at period $(p+1)$. Note that the identification of the diagonals of a table with birth cohorts is only possible for tables in which the grouping interval is equal on both sides (5 years in our example). Note also that the identification is only approximate. We shall discuss problems related to this approximation in our second paper.[8]

This pattern in the residuals therefore indicates that a rather different sort of time effect has been observed. Rather than rates being affected equally across all age groups at a specified period, we may consider influences which affect rates in a specified generation or birth cohort equally throughout life. Such effects are known as cohort effects, and the model which describes time trends in these terms is known as the age–cohort model. The model is easier to understand if we rewrite the table of rates corresponding to Table IV with each diagonal as a column. This is shown in Table VI;

Table IV. Age-specific mortality rates (per 100,000 person-years observation) of bladder cancer in Italian males during the period 1955–1979. Numerators are in parentheses. (Source of data: WHO mortality database)

Age/period	1955–1959	1960–1964	1965–1969	1970–1974	1975–1979
25–29	0·03 (3)	0·03 (3)	0·01 (1)	0·04 (4)	0·12 (12)
30–34	0·17 (16)	0·18 (17)	0·12 (11)	0·08 (8)	0·09 (8)
35–39	0·32 (24)	0·31 (29)	0·35 (33)	0·42 (39)	0·32 (30)
40–44	1·04 (79)	1·05 (76)	0·91 (82)	1·04 (95)	1·27 (115)
45–49	2·86 (234)	2·52 (185)	2·61 (183)	3·04 (267)	3·16 (285)
50–54	6·64 (458)	7·03 (552)	6·43 (450)	6·46 (431)	8·47 (723)
55–59	12·71 (720)	13·39 (867)	14·59 (1069)	14·64 (974)	16·38 (1004)
60–64	20·11 (890)	23·98 (1230)	26·69 (1550)	27·55 (1840)	28·53 (1811)
65–69	24·40 (891)	33·16 (1266)	42·12 (1829)	47·77 (2395)	50·37 (3028)
70–74	32·81 (920)	42·31 (1243)	52·87 (1584)	66·01 (2292)	74·64 (3176)
75–79	45·54 (831)	47·94 (937)	62·05 (1285)	84·65 (1787)	104·21 (2659)

Table V. Chi-residuals from the age–period model for the mortality rates of Table IV. Most residuals are more extreme than expected and decrease regularly along diagonals running downwards from left to right

Age/period	1955–1959	1960–1964	1965–1969	1970–1974	1975–1979
25–29	−0·17	−0·40	−1·67	−0·48	2·31
30–34	2·44	2·10	−0·26	−1·65	−1·78
35–39	1·31	0·42	0·24	0·47	−1·85
40–44	2·87	1·46	−1·21	−1·40	−0·64
45–49	5·17	0·70	−1·00	−1·04	−2·45
50–54	5·77	4·00	−1·71	−4·56	−1·82
55–59	5·52	2·99	0·88	−3·53	−3·83
60–64	3·05	4·26	2·69	−1·86	−5·46
65–69	−4·87	−0·46	2·83	2·37	−1·02
70–74	−5·47	−2·91	−0·79	2·81	3·47
75–79	−2·96	−6·38	−3·95	2·38	7·49

columns represent cohorts of individuals born within a period surrounding some central date and rows again refer to age groups.

Table VI represents the data as a series of longitudinal studies rather than, as in Table IV, a series of cross-sectional studies. This longitudinal view of the data is also emphasized in Figure 2, which plots, again on a logarithmic scale, mortality rates against age for each birth cohort.

We are struck by the parallelism of these curves, indicating that, on the logarithmic scale, the differences in age-specific mortality between any pair of birth cohorts is approximately constant throughout life. Again we could not expect real data to yield exactly parallel curves but nevertheless the age cohort model can provide a useful description of the data. If we denote the logarithm of the incidence or mortality rate for cohort c at age a by Y_{ac}, then the age cohort model implies that

$$Y_{ac} = \alpha_a + \gamma_c. \tag{6}$$

As before, α_a measure age effects but now γ_c measure cohort effects. By taking antilogs, this can

Table VI. Bladder cancer in Italian males. The age-specific rates of Table IV are rearranged here by central date of birth. Each column contains a longitudinal series of rates

Age	Date of birth														
	1880	1885	1890	1895	1900	1905	1910	1915	1920	1925	1930	1935	1940	1945	1950
25–29	—	—	—	—	—	—	—	—	—	—	0·0	0·0	0·0	0·0	0·1
30–34	—	—	—	—	—	—	—	—	—	0·2	0·2	0·1	0·1	0·1	—
35–39	—	—	—	—	—	—	—	—	0·3	0·3	0·4	0·4	0·3	—	—
40–44	—	—	—	—	—	—	—	1·0	1·1	0·9	1·0	1·3	—	—	—
45–49	—	—	—	—	—	—	2·9	2·5	2·6	3·0	3·2	—	—	—	—
50–54	—	—	—	—	—	6·6	7·0	6·4	6·5	8·5	—	—	—	—	—
55–59	—	—	—	—	12·7	13·4	14·6	14·6	16·4	—	—	—	—	—	—
60–64	—	—	—	20·1	24·0	26·7	27·6	28·5	—	—	—	—	—	—	—
65–69	—	—	24·4	33·2	42·1	47·8	50·4	—	—	—	—	—	—	—	—
70–74	—	32·8	42·3	52·9	66·0	74·6	—	—	—	—	—	—	—	—	—
75–79	45·5	47·9	62·1	84·7	104·2	—	—	—	—	—	—	—	—	—	—

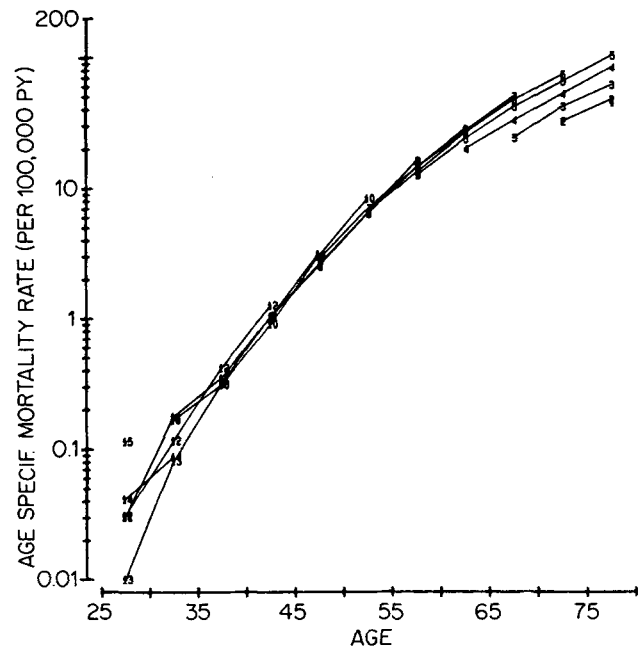

Figure 2. Mortality rates of bladder cancer in Italian males. Each curve in this graph represents a longitudinal series of rates; it depicts the evolution of mortality within a birth cohort. Rates are plotted using a logarithmic scale. Note again the quasi parallelism of the cohort curves (source: see Table IV)

also be written as a multiplicative model for the rates, ρ_{ac}, that is

$$\rho_{ac} = \alpha'_a \gamma'_c. \tag{7}$$

Once we have rearranged the data of Table IV into the form of Table VI, there is no essential difference between fitting the age-cohort model and the procedures discussed in the second section. When errors are assumed to obey the Poisson law, the iterative SMR calculations, or equivalently the IPF, again yield the best fitting model.

The same problems of parameterization apply here as in the age–period model and it is again conventional to adopt a representation in which the parameters α'_a look like age-specific rates and the parameter γ'_c look like relative risks. Again there are three main approaches. If we fix one of the γ_c at zero then the α'_a are indeed fitted age-specific rates for this reference cohort and the remaining γ'_c are relative risks of each cohort relative to the reference cohort. However, in this case it is often not very satisfactory to choose the first or last cohort as reference, since, these are represented by only one cell each and risk is not estimated as reliably as for the central cohorts. One of the cohorts with most complete data should be used. The second strategy of choosing the γ'_cs to have zero mean might be adopted. Just as before, however, they are then not directly interpretable as relative risks although the ratio of γ_c/γ_d still gives the relative risk of cohort c relative to cohort d. The third approach is again to focus attention on regular trend by reporting the first differences $(\gamma_2 - \gamma_1)$, $(\gamma_3 - \gamma_2)$ whose antilogs give the relative risks between adjacent cohorts.

For the data of Table IV, the age–cohort model is a much better fit than the age–period model: the global deviance chi-squared test gives 36·3 (on 30 degrees of freedom; not significant, $P \simeq 0{\cdot}20$). Notice, however, that the considerable improvement in fit is accompanied by a loss of 10 degrees of freedom. This reflects the fact that the Table VI has 15 columns rather than the 5 columns of

Table VII. Age (α_a) and cohort (γ_c) parameters estimated from the rates of Table VI using the IPF procedure. The scales for the multiplicative effects were chosen to facilitate interpretation. The estimates shown in parentheses are based on considerably less data than the other estimates

Age	Additive effects α_a	Multiplicative effects $\alpha'_a \times 100{,}000$
25–29	−15·36	0·02
30–34	−13·88	0·09
35–39	−12·89	0·25
40–44	−11·79	0·76
45–49	−10·73	2·19
50–54	−9·77	5·70
55–59	−8·99	12·49
60–64	−8·34	23·93
65–69	−7·77	42·05
70–74	−7·31	66·95
75–79	−6·87	103·64
Cohort	γ_C	$\gamma'_C \times 100(\%)$
1880	−0·822	44·0
1885	−0·743	47·6
1890	−0·502	60·5
1895	−0·216	80·6
1900	0·000	100·0
1905	0·112	111·9
1910	0·171	118·6
1915	0·165	117·9
1920	0·224	125·1
1925	0·359	143·2
1930	0·362	143·6
1935	0·492	163·6
1940	0·104	111·0
1945	(0·132)	(114·1)
1950	(1·693)	(543·6)

Table IV, so that the age–cohort model requires 10 more parameters than the age–period model. This fact is often forgotten; in comparing the fit of age–period and age–cohort models, we are not comparing models of equal complexity and it is perhaps not surprising that the age–cohort model is often a better fit. For example, since the external cohorts are observed only at one age group each, the model must fit these cells perfectly. Likewise, the adjacent cohorts include data on only two age groups each, and the fit will usually be very good. For these data, however, the improvement in fit of the age–cohort model is very considerable indeed (deviance = 36·3 rather than 455) and there is no question that it justifies the extra parameters.

Parameter estimates using the method of maximum likelihood based upon Poisson deviations are shown in Table VII. The reference cohort is taken as the 1900 birth cohort – the first cohort to be observed throughout the full 25 years of study. Note also that the estimates of the relative risks for the two last cohorts are shown in parentheses, reflecting the fact (which we have already discussed) that these are based on considerably less data and are, therefore, less reliably estimated. This is particularly the case for the last cohorts which are only observed during the youngest age

groups so that γ_{14} and γ_{15} are both estimated from only 12 deaths each. This fact is also frequently lost sight of when considering the age–cohort model and it is frequently found that plots of the cohort effects, γ_c, show extremely large fluctuations for the latest cohorts, and to a lesser extent, for the earliest cohorts. It is easy to be impressed by such, often artifactual, suggestions. We can guard against this danger by also estimating the standard errors of the parameter estimates as measures of the reliability of estimation. Note that this problem does not arise in the interpretation of period effects, which are all estimated with approximately equal reliability.

REGULAR TRENDS: THE LOG-LINEAR DRIFT MODEL

In the previous section we showed a data set which was well fitted by the age–cohort model but not by the age–period model. We shall discuss the analysis of data sets which are not well described by either model in paper II.[8] In this section we consider the interesting phenomenon of a data set described equally well by both models.

Table VIII shows mortality from lung cancer in females in Belgium during the period 1955 to 1978, and Table IX shows deviance chi-squared tests for the age only, the age–period and the age–cohort models.

At first sight, the results of Table IX seem paradoxical. Comparison of the age only and age–period models indicates a highly significant period effect, while comparison of the age only and age–cohort models indicates a highly significant cohort effect. Yet, both age–period and age–cohort models fit the data very well, with chi-squares very close to their expected values (the corresponding degrees of freedom). The only possible resolution of this paradox is that there must be some temporal variation of rates which does not distinguish between period and cohort influences, that is, a variation over time which could be predicted either by the age–period model or by the age–cohort model. This is indeed the case, and we introduce the term 'drift' to describe such variation.

In Table X we examine the estimated additive period effects, β_p, of the age–period model and the estimated cohort effects, γ_c, of the age–cohort model and these indicate empirically the nature of drift; both models show mortality increasing almost monotonically at an average rate of about 10 per cent per five-year period (or cohort). This suggests a log-linear trend model; for the age–period model it predicts that the logarithmic age-specific rates, Y_{ap}, may be represented by

$$Y_{ap} = \alpha_a + \delta_p(p - p_0) \tag{8}$$

where p_0 is the reference period and δ_p is the (constant) change in log-rates from one period to the next. Similarly, a log-linear version of the age–cohort model is

$$Y_{ac} = \alpha_a + \delta_c(c - c_0) \tag{9}$$

where c_0 is the reference cohort. In (8) α_a are the fitted age-specific rates in the reference period, while in (9), α_a are the fitted age-specific rates for the reference cohort.

We shall not discuss here the technicalities of fitting either of these models. It is sufficient for our purpose to note that we can do so by means of a suitable computer program, using either least-squares or maximum likelihood. Here the latter is more appropriate, and when we fit these models we obtain an interesting result; the likelihood ratio chi-squares assessing the fit of each of these models are equal; 42·06 on 43 degrees of freedom. This reflects the fact that the fitted rates are identical: the two models give identical predictions. Also, in both cases the estimate of the linear trend coefficient, δ, is the same, 0·1025, so that $\delta' = 1{\cdot}11$.

Table VIII. Age-specific mortality rates (per 100,000 person-years observation) of lung cancer in Belgian females during the period 1955–1978. Numerators are shown in parentheses. (Source of data: WHO mortality database)

Age/period	1955–1959	1960–1964	1965–1969	1970–1974	1975–1978
25–29	0·19 (3)	0·13 (2)	0·50 (7)	0·19 (3)	0·70 (10)
30–34	0·66 (11)	0·98 (16)	0·72 (11)	0·71 (10)	0·57 (7)
35–39	0·78 (11)	1·32 (22)	1·47 (24)	1·64 (25)	1·32 (15)
40–44	2·67 (36)	3·16 (44)	2·53 (42)	3·38 (53)	3·93 (48)
45–49	4·84 (77)	5·60 (74)	4·93 (68)	6·05 (99)	6·83 (88)
50–54	6·60 (106)	8·50 (131)	7·65 (99)	10·59 (142)	10·42 (134)
55–59	10·36 (157)	12·00 (184)	12·68 (189)	14·34 (180)	17·95 (177)
60–64	14·76 (193)	16·37 (232)	18·00 (262)	17·60 (249)	23·91 (239)
65–69	20·53 (219)	22·60 (267)	24·90 (323)	24·33 (325)	32·70 (343)
70–74	26·24 (223)	27·70 (250)	30·47 (308)	36·94 (412)	38·47 (358)
75–79	33·47 (198)	33·61 (214)	36·77 (253)	43·69 (338)	45·20 (312)

Table IX. Lung cancer mortality in Belgian females (data of Table VIII): goodness of fit of various log-linear models

Model	Deviance	Degrees of freedom	p-value
Age	196·3	44	—
Age + period	38·2	40	0·5
Age + cohort	29·5	30	0·5
Age + drift	42·1	43	0·5

Table X. Lung cancer mortality in Belgian females. Additive effects, deviances (dev) and degrees of freedom (d.f.) for the age–period and age–cohort models estimated using the IPF procedure

	Age + period (deviance = 38·2, d.f. = 40)			Age + cohort (deviance = 29·5, d.f. = 30)			
Age	α_a	period	β_p	Age	α_a	Cohort	γ_c
25–29	−12·82	1955–1959	0·000	25–29	−13·54	1880	−0·331
30–34	−12·01	1960–1964	0·107	30–34	−12·42	1885	−0·318
35–39	−11·43	1965–1969	0·162	35–39	−11·85	1890	−0·231
40–44	−10·58	1970–1974	0·278	40–44	−10·92	1895	−0·105
45–49	−9·99	1975–1978	0·423	45–49	−10·22	1900	0·000
50–54	−9·55			50–54	−9·66	1905	0·055
55–59	−9·12			55–59	−9·14	1910	0·203
60–64	−8·83			60–64	−8·73	1915	0·331
65–69	−8·50			65–69	−8·29	1920	0·470
70–74	−8·25			70–74	−7·94	1925	0·484
75–79	−8·07			75–79	−7·67	1930	0·655
						1935	0·740
						1940	0·717
						1945	0·361
						1950	1·664

Taking antilogs of (8) and (9) gives the multiplicative models for the ratios themselves. That is, for the age–period model

$$\rho_{ap} = \alpha'_a \, (\delta'_P)^{(p-p_0)} \tag{10}$$

and, for the age–cohort model

$$\rho'_{ac} = \alpha'_a \, (\delta'_C)^{(c-c_0)}. \tag{11}$$

Thus, δ'_P is the relative risk between adjacent periods and δ'_C is the relative risk between adjacent cohorts.

Does this mean, then, that the two models (8) and (9) are the same? Unfortunately it does not; they merely make the same predictions for rates. This may be demonstrated by first considering the case where the log-linear age–period model (8) is known to hold. That is

$$Y_{ap} = \alpha_a + \beta_P(p - p_0).$$

However, from the structure of the table, the logarithmic-rate corresponding to age group a and period p may also be indexed by a and cohort c in the longitudinal table, where $c = A - a + p$. This in turn means that $p = c + a - A$, and we may substitute this value for p in the age–period model and obtain

$$Y_{ac} = \alpha_a + \delta_P(c + a - A - p_0).$$

The term in parentheses is linear in c so that the model may also be written as an age–cohort model, but, and this is important, the term also includes a. Further rearrangement of the expression gives

$$Y_{ac} = [\alpha_a + \delta_P(a - a_0)] + \delta_P(c - c_0)$$

where $c_0 = A - a_0 + p_0$. Thus, although we started from a firm assumption of the age–period model, we find that it may be written as an age–cohort model, but the age–cohort model shows a different (incorrect) age relationship for the rates. The age gradient is enhanced by δ'_P per age interval.

We can also apply the same argument in reverse; starting from a firm assumption that the log-linear age–cohort model, (9), holds we find it equivalent to an age–period model with an identical linear trend parameter, δ_C, but with an incorrect age relationship. In this case, the age gradient is attenuated by δ_C per age interval.

Table XI shows the estimated age relationship according to which of the log-linear models, (8) or (9), we fit. The reference period for the age–period model is period 1 (1955–1959), while the reference cohort for the age–cohort model is cohort 5 (1900 birth cohort). We are not surprised by now that the values of α_a differ between the age–period and age–cohort models – these will depend on the choice of reference period or cohort as explained in earlier sections. However, now even the difference between adjacent parameters depend on which model is adopted. This demonstrates the algebra above; the fitted age gradient depends upon the model assumed, but we have seen that there is no information within the data to allow us to discriminate between the models; we do not know which model is true. Indeed, the position is even more difficult since both types of influence may operate simultaneously, and this problem is the subject of our second paper.[8]

We suggest, therefore, the term 'drift parameter' for the coefficient, δ, of the log-linear trend model, since it is free of any connection with either the age–period or the age–cohort model specifically. When α_a are estimated from the age–period model (8) we suggest they should be reported as an estimate of the 'cross-sectional' age curve, while if (9) is used, α_a would be referred to

Table XI. Lung cancer mortality in Belgian females. Additive age effects for the two age-drift models. Both models yield a deviance of 42·1 on 43 degrees of freedom and an estimated drift δ of 0·103

Age class	Period drift model	Cohort drift model
25–29	−12·83	−13·45
30–34	−12·02	−12·53
35–39	−11·44	−11·85
40–44	−10·59	−10·90
45–49	−10·00	−10·20
50–54	−9·56	−9·66
55–59	−9·14	−9·14
60–64	−8·84	−9·74
65–69	−8·52	−8·31
70–74	−8·26	−7·96
75–79	−8·08	−7·67

as the 'longitudinal' age curve, although strictly it only approximates the curve which would be obtained from longitudinal studies.

DISCUSSION

In this paper we have introduced the age–period and the age–cohort models. We have discussed the problem of identifiability arising because the same fitted rates are predicted by many different sets of parameter values. In the third and fourth sections we showed that these problems may be resolved fairly easily by choosing a conventional way of writing the model (a parameterization) in which the parameters have simple interpretations in terms of relative risks, taking one period or cohort as reference. In the last section we encountered the phenomenon of 'drift', which leads to a much more difficult identifiability problem. We shall deal with this in our second paper.[8]

It is, perhaps, surprising that the age–drift model has received little or no attention in the literature. Frequently analyses start with either an age–period or an age–cohort model. We recommend that this model should always be the next possibility considered after the model of no temporal variation (the age-only model). As we have seen, drift is regular trend which cannot be ascribed to either period or cohort influences, and it is only when we observe irregular or sudden changes that we must consider age–period or age–cohort models.

If there is a sudden change in all age groups simultaneously, then the age–period model will describe data well. Such changes might occur in mortality rates, for example as a result of an advance in treatment which benefits all age groups equally, or as a result of the introduction of a screening programme which is equally applied and equally effective over all age groups. For incidence rates, the same pattern might occur if there is some change in population exposure to a late-stage carcinogen, again affecting all age groups equally, or (more likely) as a result of changes in the completeness of registration. The sudden increase of incidence of bladder cancer in the region of Birmingham (Table I) is a consequence of an artifact of registration. In this instance, so-called benign papillomas of the bladder were not registered as bladder cancer cases until the start of the third period of observation (see *Cancer Incidence in Five Continents*[10,11]). However, changes in

disease classification will result in a time effect only if the induced percentage variation of rates does not depend upon age.

Most causes of cancer require prolonged exposure, determined by an aspect of life-style, such as occupation or smoking habits, which is fixed very early in adult life. In these cases, a change in population exposure is more likely to manifest many years subsequently and will not occur simultaneously in all age groups; certain generations or cohorts will have greater exposure than others and the age–cohort model will provide a better description of the data.

It is important to recognize, however, that both models represent rather simplistic modes of action of risk factors on disease. The age–period model will only fit the data if the external influence changes all age-specific mortality rates by the same multiplicative factor. If a factor operating at or near the time of death operates differentially in different age groups (that is, there is an age–period interaction) then the age–cohort model may give a better fit. Thus, we should be careful to avoid over-interpretation of the better fit of the age–cohort model, which often arises out of its greater complexity; it has more parameters simply because there are more diagonals in the table than there are columns. There is a further reason why age–cohort models should be interpreted with great care. Strictly, cohort parameters describe relative risks for the diagonals of a table. The extent to which cohort parameters measure cohort relative risks depends on how closely diagonal rates reflect the actual cohort rates. There is at least one instance in which rates on a diagonal may differ appreciably from the true cohort rates. We show in our second paper[8] that a marked dip in the birth rates (like those due in some countries to the World Wars) can produce on its own a cohort effect.

In our second paper[8] we discuss analyses in which neither age–period nor age–cohort models provide an adequate fit to the data. Often such tables are analysed using the full age–period–cohort model and in this more complex model it is doubly important to be wary against fallacious interpretations.

Finally, we shall make a few observations concerning computer software for fitting the models we describe in these papers. For the method of weighted least-squares, any general linear model program may be used, for example those in BMDP or SAS packages. The most convenient general purpose program for all methods, including maximum likelihood under Poisson assumption, is GLIM.[13] The experienced user may readily carry out the residual analyses we have suggested with this program, and Breslow[20] has shown it may be used when the residual errors are more dispersed than would be expected from the Poisson assumption.

ACKNOWLEDGEMENTS

This collaborative work was supported by I.A.R.C. (Lyon). We should like to thank Dr. C. S. Muir in particular for his support and encouragement. We also thank Mr. A. De Coninck for computing assistance and Miss F. De Carli for typing the manuscript.

REFERENCES

1. Breslow, N. E. and Day, N. E. 'Indirect standardization and multiplicative models for rates, with reference to the age adjustment for cancer incidence and relative frequency data', *Journal of Chronic Diseases*, **28**, 289–303 (1975).
2. Breslow, N. E., Lubin, J. H., Marek, P. and Langholz, B. 'Multiplicative models and cohort analysis', *Journal of the American Statistical Association*, **78**, 381, 1–12 (1983).
3. Breslow, N. E. and Day, N. E. 'Statistical method in cancer research, Vol. I—The analysis of case-control studies', International Agency for Research on Cancer, Lyon, (IARC Scientific Publications No. 32) 1980.
4. Holford, T. R. 'The estimation of age, period and cohort effects for vital rates', *Biometrics*, **39**, 311–324 (1983).
5. Boyle, P., Day, N. E. and Magnus, K. 'Mathematical modelling of malignant melanoma trends in Norway 1953–1978', *American Journal of Epidemiology*, **118**, 887–896 (1982).

6. Osmond, C. and Gardner, M. J. 'Age, period and cohort models applied to cancer mortality rates', *Statistics in Medicine*, **1**, 245–259 (1982).
7. Robertson, C. and Boyle, P. 'Age, period and cohort models: the use of individual records', *Statistics in Medicine*, **5**, 527–538 (1986).
8. Clayton, D. and Schifflers, E. 'Models for temporal variation in cancer rates. II: Age–period–cohort models', *Statistics in Medicine*, **6**, 469–481 (1987).
9. Doll, R., Payne, P. and Waterhouse, J. A. H. *Cancer Incidence in Five Continents – A Technical Report*, Springer-Verlag, New York, 1966.
10. Doll, R., Muir, C. S. and Waterhouse, J. A. H. *Cancer Incidence in Five Continents, Vol II*, Springer-Verlag, New York, 1970.
11. Waterhouse, J. A. H., Muir, C. S., Correa, P. and Powell, J. *Cancer Incidence in Five Continents, Vol III*, International Agency for Research on Cancer, Lyon, (IARC Scientific Publications No. 15), 1976.
12. Waterhouse, J. A. H., Muir, C. S., Shanmugaratnam, K. and Powell, J. *Cancer Incidence in Five Continents, Vol IV*, International Agency for Research on Cancer, Lyon, (IARC Scientific Publications No. 42), 1982.
13. Nelder, J. A. and Wedderburn, R. W. M. 'Generalized linear models', *Journal of the Royal Statistical Society*, **135**, 3, 370–384 (1972).
14. McCullagh, P. and Nelder, J. A. *Generalized Linear Models*, Chapman and Hall, London, 1983.
15. Kilpatrick, S. J. 'Occupational mortality indices', *Population Studies*, **16**, 175–187 (1962).
16. Mantel, N. and Starke, C. R. 'Computation of indirect-adjusted rates in the presence of confounding', *Biometrics*, 997–1005 (1968).
17. Bishop, Y. M. M., Fienberg, S. E. and Holland, P. W. *Discrete Multivariate Analysis – Theory and Practice*, The MIT Press, Massachusetts, 1975.
18. Barrett, J. C. 'The redundant factor method and bladder cancer mortality', *Journal of Epidemiology and Community Health*, **32**, 314–316 (1978).
19. Barrett, J. C. 'Cohort mortality and prostate cancer', *Journal of Biosociological Sciences*, **12**, 341–344 (1980).
20. Breslow, N. E. 'Extra-Poisson variation in log-linear models', *Applied Statistics*, **33**, 1, 38–44 (1984).

STATISTICS IN MEDICINE, VOL. 6, 469–481 (1987)

MODELS FOR TEMPORAL VARIATION IN CANCER RATES. II: AGE–PERIOD–COHORT MODELS

D. CLAYTON

Department of Community Health, Clinical Sciences Building, Leicester Royal Infirmary, PO Box 65, Leicester LE2 7LX, U.K.

AND

E. SCHIFFLERS

Facultés Universitaires N.-D. de la Paix, Rue de Bruxelles, 61, B5000, Namur, Belgium

SUMMARY

Our first paper reviewed methods for modelling variation in cancer incidence and mortality rates in terms of either period effects or cohort effects in the general multiplicative risk model. There we drew attention to the difficulty of attributing regular trends to either period or cohort influences. In this paper we turn to the more realistic problem in which neither period nor cohort effects alone lead to an adequate description of the data. We describe the age–period–cohort model and show how its ambiguities surrounding regular trends 'intensify'. We recommend methods for presenting the results of analyses based upon this model which minimize the serious risk of misleading implications and critically review previous suggestions. The discussion is illustrated by an analysis of breast cancer mortality in Japan with special reference to the phenomenon of 'Clemmesen's hook'.

KEY WORDS Cohort analysis Cancer trends Age–period–cohort models Standardized rates

INTRODUCTION

In our first paper[1] we described an approach to the analysis of data on the variation of cancer incidence or mortality with time. This approach is in the main stream of modern methodology in chronic disease epidemiology, being based upon the proportional hazards model. This is an empirically based general model which, in its simplest form, holds that the ratios of age-specific rates between two groups of individuals with different exposures to carcinogenic influences are constant for all age groups. This general model underpins traditional methods of age standardization[2] as well as relative risk analyses in age-matched case–control studies.[3]

We described two different models for variation over time, the age–period model and the age–cohort model. These models predict constant ratios of age-specific rates

(i) between different periods, that is the calendar periods during which the incidence (or mortality) rates were observed, or
(ii) between different cohorts, that is longitudinally observed groups of people born within specific periods.

Cross-sectional tables of rates by age and calendar period allow us to fit the former (age–period) model precisely and the latter (age–cohort) model to a close approximation, at least when the data are grouped with almost equal time intervals on both age and calendar period axes (say 5 years).

0277–6715/87/040469–13$06.50

Received 6 March 1986
Revised 5 September 1986

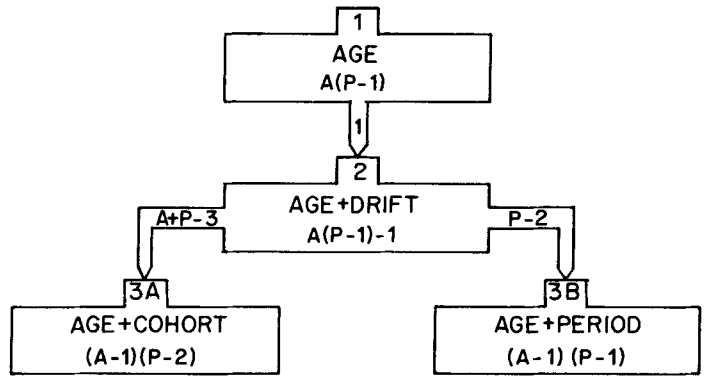

Figure 1. Logical order in which to consider models with at most one type of temporal variation. The formula on the second line of a box gives the number of degrees of freedom for a table with A age classes and P time periods

Finally, we considered the problem posed by a specific type of regular trend in which the ratio of age-specific rates between two adjacent time periods is not only constant across age groups, but is constant for any pair of adjacent time periods. We showed this model to be indistinguishable from the equivalent regular age–cohort trend model in which the relative risk between adjacent birth cohorts is constant. The models are indistinguishable, in the sense that either may generate identical predictions for the data using the same trend parameter – the relative risk between adjacent periods or cohorts. However, in so doing, the models must adopt different age curves so that if the relationship between age and incidence or mortality rates were known then the models would be distinguishable. In the absence of such knowledge, however, the models are equivalent and we introduced the term drift to describe this regular trend, unattributable to specifically period or cohort influences.

We also introduced the terms cross-sectional and longitudinal age curves for the age relationships estimated by the age–period and age–cohort versions, respectively, of the regular drift model. It follows directly from our remarks of the previous paragraph that, in the absence of extra information allowing us to determine which model is true, we are likewise unable to decide which of these represents the true age curve. This fundamental difficulty is central to understanding the problems of interpretation with the more complex models described in this paper.

Figure 1 illustrates the logical order in which to consider the models we have encountered so far. The first model is the null hypothesis of no temporal variation, while the second model is the model of regular drift unattributable to period or cohort influences. Only if this model does not adequately describe the data are we justified in considering either specifically age–period or age–cohort models.

In our first paper[1] we also discussed the assessment of the goodness-of-fit of models and showed how the change in a global measurement of goodness-of-fit (or badness-of-fit, deviance) may be used to construct statistical significance tests. Thus, comparison of model (2) with model (1) provides a one-degree of freedom test for trend (drift). Comparison of model (3A) with model (2) provides a $(P-2)$ degree-of-freedom test for irregular trends incidence (death) attributable to period while comparison of models (3B) and (2) provides a $(C-2)$ degree-of-freedom test for irregular cohort effects. Note that models (3A) and (3B) may not be compared directly in this way and it is not possible to construct a formal test of whether the age–cohort model is significantly better than the age–period model.

Table I. Age-specific mortality rates (per 100,000 person-years observation) of breast cancer in Japan, during the period 1955–1979. Numbers of cases on which rates are based are in parentheses (source: WHO mortality data base).

Age/period	1955–1959		1960–1964		1965–1969		1970–1974		1975–1979	
25–29	0·44	(88)	0·38	(78)	0·46	(101)	0·55	(127)	0·68	(179)
30–34	1·69	(299)	1·69	(330)	1·75	(363)	2·31	(509)	2·52	(588)
35–39	4·01	(596)	3·90	(680)	4·11	(798)	4·44	(923)	4·80	1056)
40–44	6·59	(874)	6·57	(962)	6·81	(1171)	7·79	(1497)	8·27	(1716)
45–49	8·51	(1022)	9·61	(1247)	9·96	(1429)	11·68	(1987)	12·51	(2398)
50–54	10·49	(1035)	10·80	(1258)	12·36	(1560)	14·59	(2079)	16·56	(2794)
55–59	11·36	(970)	11·51	(1087)	12·98	(1446)	14·97	(1828)	17·79	(2465)
60–64	12·03	(820)	10·67	(861)	12·67	(1126)	14·46	(1549)	16·42	(1962)
65–69	12·55	(678)	12·03	(738)	12·10	(878)	13·81	(1140)	16·46	(1683)
70–74	15·81	(640)	13·87	(628)	12·65	(656)	14·00	(900)	15·60	(1162)
75–79	17·97	(497)	15·62	(463)	15·83	(536)	15·71	(644)	16·52	(865)

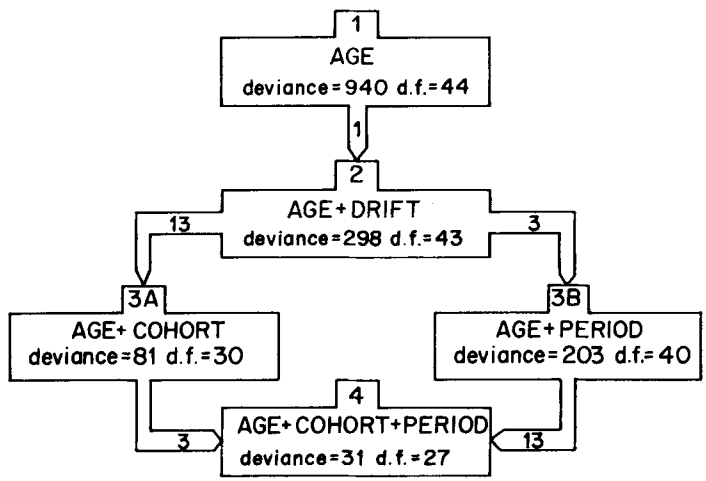

Figure 2. Deviances obtained from fitting various multiplicative models to the data of Table I. Numbers in the connecting arrows give the loss of degrees of freedom (d.f.)

In this paper we consider the case where neither the age–period nor the age–cohort model provides an adequate fit to the observed table of rates.

THE AGE–PERIOD–COHORT MODEL

Table I shows mortality from breast cancer in Japan during the period 1955–1979, and the upper section of Figure 2 shows the results of fitting the four models of Figure 1.

In this case the fit of the models is measured using a deviance or likelihood-ratio criterion which assesses the deviations between observed and fitted rates relative to Poisson variability expected on the basis of the numerators of the observed rates (see paper I, third section).[1] If any model gives a

true description of the underlying rates, the corresponding deviance should be distributed as chi-squared with the appropriate degrees of freedom.

It can be seen that although the age–cohort model is clearly the best of those considered it does not fit the data adequately; a value of chi-squared of 81 on 30 degrees of freedom is highly significant.

This suggests that the model to consider next is one in which both cohort and period effects are included; this is the age–period–cohort model which has received much attention in recent literature. As for the age–period and age–cohort models, it may be written either as a model for the log-rates in which the effects of age, period and cohort combine additively, or as a model for the rates themselves in which the effects combine multiplicatively. That is, writing Y_{ap} for the logarithms of the rates, the age–period–cohort model is

$$Y_{ap} = \alpha_a + \beta_p + \gamma_c, \tag{1}$$

$$c = A - a + p.$$

Here, as in our first paper,[1] c indexes diagonals of age × period table, which approximate to birth cohorts.

Writing, ρ_{ap} for the untransformed rates and α'_p, β'_p and γ'_c for the antilogs of the corresponding parameters, the multiplicative form of the model is

$$\rho_{ap} = \alpha'_a \beta'_p \gamma'_c \tag{2}$$

so that age-specific rates, α'_a, are multiplied by factors β'_p corresponding to the calendar period of incidence or death, and by factors γ'_c corresponding to the birth cohort of individuals affected.

This model may be fitted to the data either by weighted least-squares or by Poisson maximum likelihood. In these papers the latter method has been used throughout. When fitted to the data of Table I, model (1) gives a deviance of 30·7 on 27 degrees of freedom, which is consistent with chance (Poisson) fluctuations ($P \simeq 0{\cdot}28$). Figure 2 shows the sequence of models leading to this final, acceptable model. By comparing models 3B and 4 we conclude that, after adjusting for period effects, cohort effects are statistically significant (deviance = 172, on 13 degrees of freedom). Likewise, comparing models 3A and 4 we see that, after adjusting for cohort effects, the effect of calendar period is significant (deviance = 50, on 3 degrees of freedom). Note that the degrees of freedom for each of these tests mirrors the value for the test corresponding to the opposite corners connecting models 2, 3A, 3B and 4. These tests are the corresponding crude tests. For example, comparison of 3A and 2 tests for cohort effects but does not adjust for period effects.

Exactly as the adjusted tests, (3A − 2) and (3B − 2), test for attributable cohort and period effects, respectively (that is effects over and above regular drift), so do the corresponding tests based upon the full age–period–cohort model. The impossibility of ascribing drift to either specifically period or specifically cohort influences must and does persist. This in turn presents us with serious problems in displaying and interpreting the estimates of the model parameters. We shall discuss this issue in the next section.

MEASURING AGE, PERIOD AND COHORT EFFECTS

The fundamental problem in interpreting parameter estimates from the age–period–cohort model, (1) or (2), is that there is no single unique solution; indeed there are infinitely many.[4–7] We have already encountered this in our first paper,[1] in which we demonstrated that the period drift and the cohort drift models lead to identical fitted rates. Thus there are clearly limitations to the interpretation of such data. We must examine what the infinitely many possible solutions have in

Table II. Three sets of age, period and cohort effects that give identical best fitting expected rates for Table I

Set no.	(1)	(2)	(3)	Set no.	(1)	(2)	(3)
Age $\alpha'_a \times 100\,000$							
25–29	0·55	0·38	0·27	Cohort $\gamma'_c \times 100$			
30–34	2·14	1·58	1·17	1880	190·3	149·7	117·8
35–39	4·30	3·38	2·66	1885	162·0	135·3	113·0
40–44	6·84	5·71	4·77	1890	133·9	118·7	105·3
45–49	9·30	8·25	7·32	1895	113·6	107·0	100·8
50–54	11·00	10·37	9.76	1900	100·0	100·0	100·0
50–59	11·23	11·23	11·23	1905	96·3	102·3	108·6
60–64	10·38	11·02	11·70	1910	94·2	106·2	119·7
65–69	9·85	11·11	12·52	1915	92·7	111·0	132·9
70–74	9·46	11·33	13·56	1920	90·3	114·8	145·9
75–79	9·43	11·99	15·24	1925	84·9	114·6	154·7
Period $\beta'_p \times 100$				1930	77·4	111·0	159·1
1955–1959	100·0	100·0	100·0	1935	69·3	105·5	160·6
1960–1964	106·2	100·0	94·2	1940	67·1	108·4	175·2
1965–1969	121·3	107·5	95·4	1945	67·8	116·3	199·6
1970–1974	147·2	123·0	102·7	1950	71·4	130·1	237·1
1975–1978	173·9	136·8	107·6				

common with one another, for it is this we may interpret. Although this is obvious, it seems to have largely escaped attention in the (futile) search for a mathematical 'solution' to the 'problem' of identifiability. Such attempts can only invite conclusions unsupported by the data. The model as specified has more parameters than may be estimated from the data. We might attempt to proceed, as previously, to find a parameterization which has α'_a representing fitted age-specific rates by choosing one reference period and setting the corresponding β'_p to 1, and a reference cohort so that one γ'_c is also taken as 1. This would leave as unknown A age parameters, $(P-1)$ period parameters and $(C-1)$ cohort parameters. Unfortunately, however, this does not work; there is no unique solution. Table II displays three possible sets of parameter estimates and it may be verified easily that each set gives an identical prediction for the observed table. In this table the reference period and cohort are $p = 1$ and $c = 5$, respectively.

These solutions are chosen to illustrate how the unwary could be led to unjustified conclusions. Incidentally, they are not too different from age relationships observed for breast cancer cross-sectionally in different countries. In the first solution the period effects show a strong increase from the first period onwards, while the cohort effects show a reverse gradient. In this solution the age curve is unusual, with rates increasing until 55–59 and thereafter decreasing. In the second solution, there is an upward trend with calendar period but a U-shaped cohort curve with a minimal risk for the 1900 cohort. The age curve shows the phenomenon of Clemmesen's hook; rates increase to a maximum at 50–54 then fall back slightly before continuing their upward trend from the age of 65 onwards. Finally, the last solution yields U-shaped period and cohort curves but, compared to the previous parameterization, a more pronounced increase over successive cohorts from 1900 onwards. This solution, has no Clemmensen's hook, the age curve being uniformly increasing.

The reason for these seemingly paradoxical results lies with the problem of drift, which, as we showed earlier, is not specifically attributable either to period or cohort effects and is described by a single parameter in addition to the age parameter, α_a. Adoption of the age–period model adds $(P-2)$ extra parameters expressing irregular period effects (see Figure 2). Likewise, adoption of the age–cohort model adds $(C-2)$ parameters to the regular age–drift model. Finally, the

age–period–cohort model includes: (i) drift, (ii) non-drift period effects and (iii) non-drift cohort effects, that is $1 + (P-2) + (C-2)$ parameters in addition to the age-curve parameters. Table II, however, purports to estimate $(P-1)$ period effects and $(C-1)$ cohort effects which include two parameters for period drift and cohort drift. We already know that the data are not capable of distinguishing between these two effects and it is not surprising that we get into difficulty when trying to estimate two indistinguishable parameters!

To clarify the position, let us look more closely at such an attempt when there are no non-drift period or cohort effects. As in our first paper[1] we shall wirte δ_P and δ_C for the parameters of period–drift and cohort–drift, respectively. Thus, without non-drift effects, the age–period–cohort model for the logarithms of the rates is

$$Y_{ap} = \alpha_a + \delta_P(p - p_0) + \delta_C(c - c_0) \tag{3}$$

where p_0, c_0 are reference period and cohort, respectively. Thus, the antilogs of δ_P and δ_C, δ'_P and δ'_C are the relative risks between adjacent periods and adjacent cohorts, respectively (constant across age).

As previously, however, the cohort passing through age group a at period p is totally determined by a and p according to the relationship $c = A - a + p$, or equivalently by $p = c + a - A$. We can substitute either of these expressions for c or for p in (3) and obtain, respectively:

$$Y_{ap} = \alpha_a - \delta_C(a - a_0) + (\delta_P + \delta_C)(p - p_0) \tag{4}$$

which is the age + period–drift model, with drift parameter $\delta = (\delta_P + \delta_C)$, and:

$$Y_{ap} = \alpha_a + \delta_P(a - a_0) + (\delta_P + \delta_C)(c - c_0) \tag{5}$$

which is the age + cohort–drift model, again with drift parameter δ.

This corresponds to the problem we first discussed in the fifth section of our first paper.[1] Not only are the (age + period–drift) and (age + cohort–drift) models indistinguishable from each other, they are also indistinguishable from any (age + period–drift + cohort–drift) model in which the net drift, $\delta = (\delta_P + \delta_C)$, is held constant. It is only this net drift which can be estimated using only the data in the age × period table of rates.

Are, then, all such models identical? Again unfortunately not – they differ in the age curves which must be assumed to represent the observed data, as may be seen by comparing (4) and (5). To identify the true age curve we therefore need to partition the net drift between age and cohort influences, and this we cannot do, at least without further information or assumptions.

In our first paper[1] we suggested the term cross-sectional age curve for the age effects estimated when fitting the age–period model and longitudinal age curve for the age effects estimated when fitting the age–cohort model. Inspection of (4) shows that the cross-sectional age curve differs in gradient from the true age curve by (minus) the cohort drift, δ_C. Likewise, the longitudinal age curve differs in gradient from the true age curve by the period drift, δ_P.

We now return to the results of Table II. These represent various parameterizations of the full age–period–cohort model. This differs from the model discussed above only in that it also includes non-drift period effects and non-drift cohort effects. The difficulties concerning drift remain; the three solutions displayed all show the same net drift, but differ according to how it is partitioned between period and cohort components. Thus, the three sets of parameters all predict identical fitted rates, but suggest different age relationships. In the first parameterization there is a strong positive period drift, a strong negative cohort drift and an inverted U age curve. The other two solutions have milder period drift compensated for by an equal increase in the cohort drift. This transfer is matched by an increase in the age gradient. Note again that the transfer of drift onto the age curve causes the shape of the curve to change, in particular local extrema may be induced or erased.

How then can we present the parameters of the age–period–cohort model? It is our belief that any parameterization in a form such as we used in Table II runs the risk of over-interpretation and should be avoided, unless extra information has effectively resolved the partition of drift between period and cohort influences. However carefully one might deal with the problem in the text of a paper, the selection of one arbitrary parameterization for a table or a graphical display can be grossly misleading. It would seem wiser to report the net drift as some overall summary of the relative risks betwen adjacent intervals, and to report only non-drift period or cohort effects. This has been suggested recently by Holford.[7]

Three methods have been proposed for presentation of period or cohort effect. We shall describe these methods in relation to period effects, but the same considerations apply for cohort effects. We start from one arbitrary parameterization, say any one of the columns in Table II. If these parameters are β_p, then we may de-trend them by adding in a log-linear drift term to give new parameters

$$\beta_p^* = \beta_p + \delta(p - p_0)$$

where we choose δ so that the resultant β_p^* are free of drift. However, this raises a question as to how we define β_p^* as being free of drift. Holford[7] suggests to interpret this such that the linear regression line of the parameters β_p^* against the periods, p, has zero slope. This has the advantage that the resulting β_p^* are identifiable, that is do not depend on the repartition of drift. An even simpler alternative is based upon drift being defined as the average of the successive first differences, $(\beta_2 - \beta_1)$, $(\beta_3 - \beta_2)$. This leads to a choice of δ such that $\beta_1^* = \beta_P^*$; the period curve is restrained to return to the same level as it commenced. Since period 1 is usually taken as reference, so that $\beta_1^* = 0$, this leads also to taking $\beta_p^* = 0$ which is in computational terms very straightforward. However, it must be kept in mind that the proper interpretation of such β_p^* is not straightforward. For instance, β_p^* obtained by the latter method should be interpreted as: $\beta_p^* = (\beta_p - \beta_1) - (p-1)(\beta_P - \beta_1)/(P-1)$. The third method derives from a consideration of what it is that defines non-drift period effects. Non-drift effects operate in such a way that the relative risks between adjacent periods are not identical. Non-drift effects are, therefore, expressible as contrasts between such relative risks. Perhaps the simplest such contrast would be the ratio of two adjacent relative risks (see Figure 3):

$$\frac{\beta'_3/\beta'_2}{\beta'_2/\beta'_1}, \quad \frac{\beta'_4/\beta'_3}{\beta'_3/\beta'_2} \cdots$$

Note that in Table II these contrasts are identical in all the parameterizations. For example,

$$\frac{\beta'_3/\beta'_2}{\beta'_2/\beta'_1} = \frac{1{\cdot}213/1{\cdot}062}{1{\cdot}062/1{\cdot}00} = \frac{1{\cdot}075/1{\cdot}00}{1{\cdot}00/1{\cdot}00} = \frac{0{\cdot}954/0{\cdot}942}{0{\cdot}942/1{\cdot}00} = 1{\cdot}075$$

so that the relative risk of period 3 versus period 2 is 8 per cent higher than that of period 2 versus period 1. This may be thought of as a measure of acceleration of period trend during the time around period 2. On the logarithmic scale, these contrasts are the 'second differences': $(\beta_3 - \beta_2) - (\beta_2 - \beta_1) = \beta_3 - 2\beta_2 + \beta_1, \beta_4 - 2\beta_3 + \beta_2$, which are well-known measures of curvature. Zero value indicates that the log-risk versus calendar time curve is locally a straight line, while positive or negative values indicate convex or concave relationships, respectively.

Figure 3 shows graphically the identifiable second differences parameter estimates for our breast cancer example. While such contrasts are unfamiliar in epidemiology, they have the important property of representing characteristics specifically attributable to age, period or cohort without any arbitrary repartition of drift components. Undoubtedly there are other possibilities for presentation of the identifiable information which might be helpful. In our example the second differences show two irregularities in the birth cohort effects, indicating sudden changes in the

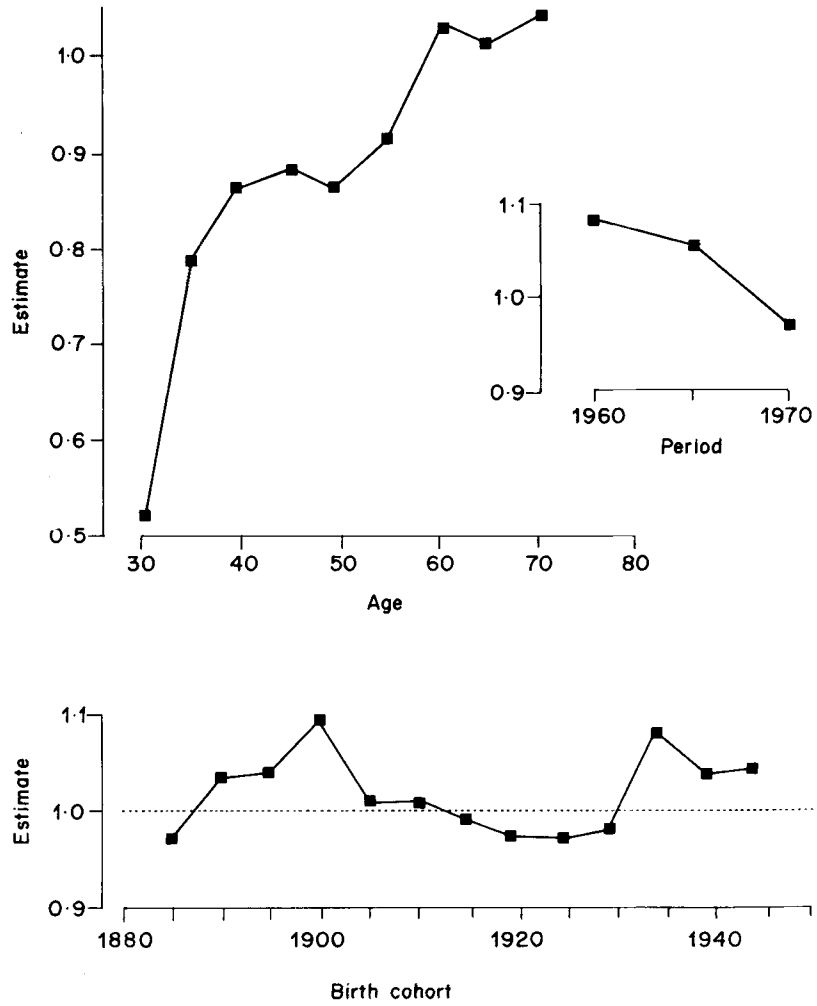

Figure 3. Mortality from breast cancer in Japan: estimates of the identifiable non-drift effects

cohort trend around 1900 and around 1935, and also the distinctive dip in the age curve around menopause, which, depending upon the partition of drift, may manifest itself as Clemmesen's hook.

These second differences have the important practical advantage that the value taken is affected by only neighbouring data. For example, the second difference around period 2 is not perceptibly influenced by trends occurring after period 3; this is not the case for the other methods we have outlined.

A NOTE ON THE EFFECT OF GROUPING

In these papers[1] we have accepted the identification of birth cohorts and diagonals of regular tables of vital rates. As stated earlier this is only approximate. Our first paper[1] draws attention to a consequence of the grouping, namely the spurious cohort effects which can result from a sudden change in birth rate. We illustrate this now.

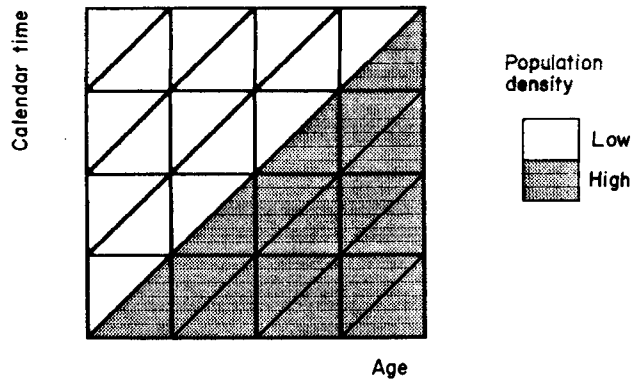

Figure 4. Lexis diagram

Figure 4 is a Lexis diagram plotting calendar time against age. The horizontal and vertical lines represent the usual age and period grouping, and birth cohorts correspond to diagonals. Suppose, for the sake of our argument, that all birth cohorts after a specified date are less numerous than before as shown by the shadings. Clearly the mean age in a square cell affected by the change in birth rates will be higher that the mid-point of the age group. Such effect will not occur elsewhere. This shift in the mean age will result in shifted (usually increased) rates. These excessive rates, aligned on a diagonal, mascarade a cohort effect.

Failure to take account of distortions due to grouping have led Boyle *et al.*[8] and more recently Boyle and Robertson[9,10] to a fallacious elimination of non-identifiability. Their argument requires access to individual records (or at least to more detailed tabulation). All arguments we presented earlier about identifiability depend in no way upon the degree of grouping, and it follows that finer grouping itself cannot resolve the non-identifiability of the model. The fallacy can be illustrated using again Figure 4. The diagonal lines delineate birth cohorts with grouping interval equal to that used for age and calendar time. Thus each age × period cell contains two triangular regions refering to adjacent birth cohorts. Consider cell (a, p), containing, say, cohorts c and $c+1$. Boyle *et al.*[8] and Boyle and Robertson[9,10] suggested modelling the log-rates in these two regions by:

$$Y_{ap}^{(1)} = \alpha_a + \beta_p + \gamma_c \tag{6}$$

$$Y_{ap}^{(2)} = \alpha_a + \beta_p + \gamma_{c+1}. \tag{7}$$

By calculating rates for the $2 \times A \times P$ triangular regions it can be shown that this model is fully identifiable. The authors claim that the fine grouping has solved the problem, but the true source of the solution is the assumption that the age and period effects are identical between the opposed triangles of each cell. Indeed the model (6) and (7) implies that the age incidence curve is a step function (similarly the period trend). If this functional form may be taken as known fact, then the model is indeed identifiable. However what if this assumption is false (as it must be)?

We have shown that the model in which period and cohort effects consist of equal and opposite drift (the EOD model, say) leads to rates which vary only with age. However, if there are no trends (the NT model) the rates also vary only with age at least when there is no constraint on the age parameters. Equations (6) and (7) imply we would be able to differentiate between these results since the EOD model predicts discrepancies between $Y_{ap}^{(1)}$ and $Y_{ap}^{(2)}$, while the NT model predicts equality. However if the age curve is not constant within cells this will also lead to discrepancies

between $Y_{ap}^{(1)}$ and $Y_{ap}^{(2)}$. Thus the truth of the conclusion to which one is led by equations (6) and (7) depends entirely upon the validity of constant incidence rates within age groups.

DISCUSSION

In this final section we address the question as to whether previous claims to have 'solved' the identifiability problem have been well founded and, if not, whether these difficulties negate the usefulness of this approach as we already stated. It is mistaken to regard the non-identifiability of drift component as a problem which needs only an advance in methodology for its solution. We have tried to demonstrate in these papers that this is a scientific problem in which our data do not discriminate between different models or explanations. A good statistical analysis will not only summarize the data in a succint and meaningful way, but will also make clear its limitations. By considering the age–period–cohort model in the context of the sequence of models set out in Figure 2, we would hope that this is achieved. Nevertheless, the view that non-identifiability is a methodological problem is prevalent as illustrated by the papers discussed in the previous paragraph as well as by the four different approaches we shall discuss next.

Recently, Osmond and Gardner[5,6] introduced a mathematical constraint in the model. Essentially, they choose one of the infinitely many possible solutions on the ground that it has certain mathematical properties. Such a strategy can only be justified if the property which identifies the unique solution has any biological basis and no such justification has been offered. Their solution, is therefore, totally arbitrary. The mathematical constraint they proposed is difficult to explain in non-technical language, but its effect is to partition the drift between period and cohort curves in a ratio which depends upon the relative magnitude of non-drift effects. Thus, if the age–cohort model is better than the age–period model then in the age–period–cohort model drift will be concentrated into the cohort effects. It is interesting to speculate on likely results of the Osmond and Gardner method when applied to data such as those for lung cancer mortality in Belgian females (see Table VIII of first paper[1]) which is well described by the age–drift model. The solution obtained would, of necessity, be determined by statistically insignificant fluctuations. There seems no scientific reason to prefer such a solution to any other.

Day and Charnay[11] considered the extra information available when analysing data from several cancer registries. They pointed out that, if the age effects may be assumed equal for different registries, then the identifiability problem is partially resolved. This assertion is undoubtedly accurate, but it is very doubtful whether one would be prepared to make such an assumption. There are different levels of carcinogenic exposure in different registry areas and it is quite conceivable that these will result in age curves of a different shape; for example, the age relationship for lung cancer differs markedly between persons with different smoking histories. A method which assumes the form of the relationship between disease rates and age to be an immutable biological constant unaffected by environmental exposure is unlikely to command widespread support.

Similarly, the identifiability problem theoretically disappears if we are prepared to assume a precise mathematical form for the age curve (an approach similar to that suggested by Boyle *et al.*,[8] and Boyle and Robertson[9,10]), provided that form does not contain a log-linear component. Now, this mathematical function must be chosen on the ground of compelling biological evidence, otherwise the whole process, even if confirmed by a good fit, amounts to a complicated but still arbitrary repartition of drift. One such curve is the Weibull law in which incidence rates are proportional to the power of age so that log-rates are linearly related to the logarithm of age rather than to age itself. This relationship is suggested by the multi-stage model for carcinogenesis[12,13] and by empirical evidence from animal carcinogenesis experiments. With the Weibull model, the

log-linear components of both cohort and period effects are identifiable. However, as already stressed, their identifiability depends upon the difference between the Weibull law and the log-linear (Gompertz) law. This difference is small and the resultant solution is unstable, the estimates of the newly identifiable linear components having very large variances and covariances. There is a further difficulty in that the empirical evidence for the Weibull law requires measurement of age not from birth but from some predefined starting point. This point may be thought of as the end of a guarantee period during which the disease may not be detected. Unfortunately the statistical information for estimation of the guarantee period is very limited since it is drawn almost entirely from the incidence rates observed at the youngest ages which are by their very nature estimated from the sparsest data. The estimates obtained for the age–period–cohort model must be expected to be heavily dependent upon the choice of guarantee period. We must therefore conclude that this approach also cannot provide a satisfactory resolution of the problem.

The last approach we consider is an extended form of age–period–cohort model, originally proposed by Moolgavkar, Stevens and Lee,[14] and discussed in detail by James and Segal.[4] In this extended model, the age effects are allowed to vary over calendar periods in such a manner that the age curve during one period, expressed as additive effects upon logarithms of rates, is a fixed multiple of the corresponding curve during another period. These multiples are an extra set of parameters over and above those required by the age–period–cohort model. Rather strangely, this extended form of the model does not suffer the same identifiability problems of the basic model. However, the model is difficult to interpret and, of course, depends upon a lack of fit of the age–period–cohort model. If the age–period–cohort model fits adequately then the extended model will degenerate to the basic form and the identifiability problem reappears. This is, therefore, not likely to prove a widely useful approach.

It is clear from the above discussion that there has been no satisfactory resolution of the problem of identifiability of log-linear trend components in age–period–cohort models. This led Kupper *et al.*[15] to conclude that, at present, such models offer little or no advantage over simple graphical methods. The same authors have recommended that future research efforts should be directed to develop and evaluate methods which bypass the identifiability problem. We would disagree with both statements. The simple facts of the information available and of the relationship between the three variables ensure that any research efforts directed at the search for this philosopher's stone of modern epidemiology is both futile and pointless. It is the purpose of statistical analysis to extract from research data the maximum information in as parsimonious and comprehensive manner as possible. No sophistication of method can create information where that information is lacking and there can surely be no other conclusion but that the observation of incidence and mortality rates in populations over time does not provide sufficient information to ascribe smooth trends to period or cohort influences with any reliability, but this is not to deny all uses for such models. In replying to the remarks of Kupper *et al.*,[15] Holford[16] pointed out that there are other aspects of such data that can be identified, and models can still provide a more parsimonious representation of the data than simply graphing the full data. While this parsimony might be considered of dubious value given the overhead of understanding necessary for interpretation of an analysis of a single table, the same cannot be said of more complex analyses over numerous registries and for many sites. It is for that purpose that we believe that the age–cohort model will continue to be of some value. Further work is necessary but should be directed at finding the most comprehensible parameterization of the model and for presenting the statistical reliability of identifiable estimates. This latter problem has been largely ignored. The problem of the analysis of tables in which the width of observation periods and of age groups is unequal (Schifflers *et al.*[17]) also requires further work, particularly as the interpretation of graphical displays is more difficult in this case.

Finally we should address a few words to the problem of forecasting future cancer rates. At first sight it might appear, since cohort risks are estimated from past observations, such forecasting is achievable without undue extrapolation. In recent years, there have been several attempts to use an age–period–cohort model fitted to past data to forecast rates. It should come as no surprise to a reader of these papers that we would in most cases doubt the wisdom of this course! Holford[16] states that, for the purpose of forecasting, the partition of drift between period and cohort components is irrelevant. Unfortunately this is only true if we are prepared to assume that the log-linear period drift which has occurred in the past will continue unchanged into the future. This is a strong assumption which will rarely be justified in practice. It is not possible to use the model to forecast under the assumption of no future period effect. It follows that forecasting is not possible without sufficient knowledge of the epidemiology of a given cancer and of the concomittant trends in population exposure to the major etiological factors to be able to resolve the underlying ambiguities. In certain situations this will be the case, for example for mesothelioma, but, when such detailed understanding is missing, we believe the place of the age–period–cohort model is in descriptive epidemiology. In this setting it has its place, provided the researcher is aware of the limits to inference from the data it is used to summarize.

ACKNOWLEDGEMENTS

This collaborative work was supported by I.A.R.C. (Lyon). We should like to thank Dr. C. S. Muir in particular for his support and encouragements. We also thank Mr. A. De Coninck for computing assistance and Miss F. De Carli for typing the manuscript.

REFERENCES

1. Clayton, D. and Schifflers, E. 'Models for temporal variation in cancer rates. I: Age–period and age–cohort models', *Statistics in Medicine*, **6**, 449–467 (1987).
2. Breslow, N. E. and Day, N. E. 'Indirect standardization and multiplicative models for rates, with reference to the age adjustment for cancer incidence and relative frequency data', *Journal of Chronic Diseases*, **28**, 289–303 (1975).
3. Breslow, N. E. and Day, N. E. *Statistical Methods in Cancer Research, Vol. I—The Analysis of Case–Control Studies*, International Agency for Research on Cancer, Lyon, (IARC Scientific Publications No. 32), 1980.
4. James, I. R. and Segal, M. R. 'On a method of mortality analysis incorporating age–year interaction, with application to prostate cancer mortality', *Biometrics*, **38**, 433–443 (1982).
5. Osmond, C. and Gardner, M. J. 'Age, period and cohort models applied to cancer mortality rates', *Statistics in Medicine*, **1**, 245–259 (1982).
6. Gardner, M. J. and Osmond, C. 'Interpretation of time trends in disease rates in the presence of generation effects', *Statistics in Medicine*, **3**, 113–130 (1984).
7. Holford, T. R. 'The estimation of age, period and cohort effects for vital rates', *Biometrics*, **39**, 311–324 (1983).
8. Boyle, P., Day, N. E. and Magnus, K. 'Mathematical modelling of malignant melanoma trends in Norway 1953–1978', *American Journal of Epidemiology*, **118**, 887–896 (1982).
9. Boyle, P. and Robertson, C. 'Statistical modelling of lung cancer and laryngeal cancer incidence in Scotland, 1960–1979', *American Journal of Epidemiology* (in press).
10. Robertson, C. and Boyle, P. 'Age, period and cohort models: the use of individual records', *Statistics in Medicine*, **5**, 527–538 (1986).
11. Day, N. E. and Charnay, B. 'Time trends, cohort effects and aging as influence on cancer incidence', in Magnus, K. (ed.) *Trends in Cancer Incidence: Causes and Practical Implications*, Hemisphere Publishing Corporation, 1982, pp. 51–65.
12. Armitage, P. and Doll, R. 'Stochastic models for carcinogenesis', *Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability*, **4**, 19–39 (1961).
13. Day, N. E. and Brown, C. C. 'Multistage models and primary prevention of cancer', *Journal of the National Cancer Institute*, **64**, (4), 977–989 (1980).

14. Moolgavkar, S. H., Stevens, R. G. and Lee, J. A. H. 'Effect of age on incidence of breast cancer in females', *Journal of the National Cancer Institute*, **62**, (3), 493–501 (1979).
15. Kupper, L. L., Janis, J. M., Karmous, A. and Greenberg, B. G. 'Age–period–cohort analysis: a review and critique', *Journal of Chronic Diseases*, **38**, 10, 811–830 (1984).
16. Holford, T. R. 'An alternative approach to statistical age–period–cohort analysis', *Journal of Chronic diseases*, **38**, 10, 831–840 (1985).
17. Schifflers, E., Smans, M. and Muir, C. S. 'Birth cohort analysis using irregular cross-sectional data: a technical note', *Statistics in Medicine*, **4**, 63–75 (1985).