STATISTICS IN MEDICINE
Statist. Med. **17**, 1305–1323 (1998)

AGE–PERIOD–COHORT ANALYSIS OF CHRONIC DISEASE RATES. I: MODELLING APPROACH

CHRIS ROBERTSON* AND PETER BOYLE

Division of Epidemiology and Biostatistics, European Institute of Oncology, via Ripamonti 435, Milano 20141, Italy

SUMMARY

Age–Period–cohort models are widely used by epidemiologists to analyse trends in disease incidence and mortality. The interpretation of such models is fraught with difficulty in view of the exact linear dependency between the three variables. It is the purpose of this paper to review, compare and contrast some of the more common approaches to this problem based on Poisson regression and a linear model for the log rates. Also the results of using the different approaches on a single series of data on breast cancer incidence among females in Scotland from 1960–1989 are presented for comparison. Recommendations as to the merits and drawbacks of the approaches are also given in the conclusions. Models which are based upon the estimable contrasts such as local curvatures and deviations from linearity are most suitable.

1. INTRODUCTION

In Table I the age specific incidence rates per 100,000 of breast cancer among women, aged 30–84, in Scotland between 1960 and 1989 are presented. The rates show an increasing incidence of breast cancer with increasing age at all six time periods. Among older age groups there is a 50 per cent increase in incidence rates from 1960–1964 to 1985–1989 but in the age groups under 40 the increase is not as great. Such a pattern is common among many cancers in many countries[1] and points to an interaction between age group and time period. This may be of significance for the aetiology or may reflect an increase in the completeness of coverage for registration of incident cases, or components of both. Age–period–cohort models have been proposed as an attempt to summarize such data succinctly and have been used recently to analyse and summarize trends in cancer incidence and mortality in all major cancer sites.[2] Similar two-way tables are also common in sociological research.[3] This paper will review and compare a number of different modelling approaches to age–period–cohort investigation.

One aim of fitting age–period–cohort models is to estimate the effects of each of these three factors on the rates. Birth cohort is a restricted version of a more general age by period interaction and age–period–cohort models are a way of looking at interaction in two-way tables with one observation per cell.[4] In this model all the factors have clear interpretations and implications. Reviews of certain aspects of age–period–cohort models are given in Kupper *et al.*,[5] Clayton and

* Correspondence to: Dr. C. Robertson, Division of Epidemiology and Biostatistics, European Institute of Oncology, via Ripamonti 435, Milano 20141, Italy

CCC 0277–6715/98/121305–19$17.50

Received April 1996
Accepted October 1997

Table I. Age specific incidence rates per 100,000 of breast cancer for women in Scotland

Age group	Time period 1960–1964	1965–1969	1970–1974	1975–1979	1980–1984	1985–1989
20–24	1·61	0·79	1·93	1·16	0·77	0·84
25–29	4·94	7·89	6·68	7·51	5·99	5·79
30–34	18·58	20·34	26·33	24·27	25·38	24·50
35–39	45·38	46·36	58·59	56·15	58·53	54·75
40–44	73·38	92·21	99·71	106·88	110·87	108·16
45–49	104·94	121·22	135·80	159·34	167·52	158·83
50–54	104·58	119·70	125·05	148·32	161·96	171·84
55–59	120·07	128·56	148·01	161·67	187·28	194·02
60–64	124·08	151·38	168·94	185·43	204·75	206·62
65–69	144·78	165·45	183·48	202·17	213·48	217·77
70–74	165·23	173·66	185·44	221·57	239·28	248·57
75–79	176·65	200·90	208·54	228·95	261·12	258·98
80–84	189·73	188·78	215·92	277·97	283·71	281·10

Schifflers[6] and Holford.[7] This paper differs from these in that direct comparisons of a variety of methods are made using a single series of data. Also some more recent approaches to the problem are also considered.

Most of the models fall into the class of generalized linear models and the assumptions usually made are:

(i) The number of cases in age group i at time period j, denoted y_{ij}, is Poisson,[8] mean θ_{ij}, where $i = 1, \ldots, m$ and $j = 1, \ldots, n$.
(ii) The number of persons at risk (N_{ij}) is a fixed known value.
(iii) The random variables, y_{ij}, are jointly independent.
(iv) The logarithm of the expected rate is a linear function:

$$\ln(E[r_{ij}]) = \ln\left(\frac{\theta_{ij}}{N_{ij}}\right) = \mu + \alpha_i + \beta_j + \gamma_k \tag{1}$$

where μ represents the mean effect, α_i represents the effect of age group i, β_j the effect of time period j, and γ_k the effect of birth cohort k.

The person-years-at-risk, N_{ij}, is not, strictly speaking, a fixed quantity. It is a population estimate based upon census figures and updated each year from birth and death records, estimated immigration and emigration rates. No random variation in N_{ij} is considered in view of it being many times larger than y_{ij}. Brillinger[8] discussed this aspect of event modelling in detail and concluded that the Poisson approximation was valid over a wide variety of circumstances.

The parameters of the models may be estimated quite easily using GLIM,[9] or any other statistical package which has a generalized linear modelling procedure. The populations are required, see Table II. The cases, y_{ij}, are specified as the y-variate, Poisson errors with log link and $\ln(N_{ij})$ as an offset and then fit the factors age, period and cohort.

There are two major problems: first, there is an exact linear dependency among the three factors; secondly, the birth cohorts form a sequence of overlapping intervals. The cell representing

Table II. Person years at risk (in hundreds) for Scottish women in five year age groups and five year time periods

Age group	Time period					
	1960–1964	1965–1969	1970–1974	1975–1979	1980–1984	1985–1989
30–34	8395	7868	7595	8240	8788	8775
35–39	8660	8089	7612	7498	8167	8676
40–44	8654	8372	7863	7466	7477	8099
45–49	8510	8366	8122	7694	7414	7379
50–54	8864	8179	8061	7895	7570	7274
55–59	8345	8455	7790	7732	7609	7319
60–64	7439	7828	7932	7302	7277	7192
65–69	6161	6697	7107	7197	6647	6695
70–74	4757	5148	5700	6102	6202	5781
75–79	3323	3570	3913	4442	4810	4985
80–84	1850	2103	2311	2565	2989	3376

age group 30–34 in the period 1960–1964 is associated only with the 1925–1934 cohort. Similarly age group 35–39 in period 1960–1964 is associated only with the 1920–1929 cohort. Thus the cohorts form a sequence of overlapping intervals with the consequent difficulties of interpretation. This problem has usually been ignored[10,11] and the cohorts are taken to be centred on the mid-years …, 1925, 1930, 1935, … .

The exact linear dependency among the factors arises because only one cohort is associated with each cell of the two way table; thus $k = j - i + m$, $k = 1, 2, \ldots, m + n - 1$. Consequently a further constraint (over and above the location constraint for each of the three factors) is required to estimate the parameters.[4,12,13,14] Holford[14] showed that it is impossible to separate the linear effect of one factor from the linear effects of the other two. Thus it is not possible to identify any linear increase in $\ln(r_{ij})$ with any one factor.

Kupper *et al.*[4] showed that to ensure identifiability of the parameters it is necessary to impose only one further linear constraint, which is written as $\mathbf{c}^{\mathrm{T}}\theta = 0$ where $\theta^{\mathrm{T}} = (\mu, \alpha^{\mathrm{T}}, \beta^{\mathrm{T}}, \gamma^{\mathrm{T}})$ is the vector of parameters. This constraint is in addition to the location constraints which in GLIM[9] are $\alpha_1 = \beta_1 = \gamma_1 = 0$. Any constraint $\mathbf{c}$ will do to ensure identifiability but if the parameter estimates are to have any validity then $\mathbf{c}$ must be based on defensible prior information.[12] The easiest form for $\mathbf{c}$ is to constrain two parameters to have the same value, for example, $\alpha_1 = \alpha_2$, $\alpha_1 = \alpha_n$ or $\gamma_1 = \gamma_{m+n-1}$.

The above procedure is fine provided the choice of constraint is made on sound theoretical grounds. This is unlikely to be the case.[7] Initial analyses of tables of rates are usually exploratory which suggests an absence of exact prior knowledge about any possible constraints. The effect of using different constraints is extremely marked.[6] Consequently the imposition of arbitrary constraints without prior knowledge cannot be recommended.

The choice of constraint to achieve identifiability requires careful consideration and many authors have proposed 'solutions'. The more popular of these approaches are reviewed in Section 2. The results of applying these 'solutions' to the data in Tables I and II are also presented in Section 2. The conclusions are discussed in Section 3. There are other ways of proceeding to take into account the linear dependency. These include the use of polynomials, excluding one of the two linear effects of period or cohort,[2] the use of non-linear functional forms based on the multi-stage model of Armitage and Doll,[7,15] and the use of covariates in the place of period or cohort, such as replacing period by the mean exposure to tar in the population for that year in a study of lung cancer mortality.[16] Such extensions have not been considered here as for breast cancer incidence, in common with the great majority of common cancers, polynomials are not appropriate. Also, there is no suitable covariate as in the lung cancer example.

2. IDENTIFIABILITY AND 'SOLUTIONS'

There are four broad classes of 'solutions' other than those which are based on arbitrary linear constraints. The first are based on the use of a penalty function which is minimized to derive the necessary extra linear constraint. Secondly there are the methods which rely on having individual records of cases so that a three way age–period–cohort table can be constructed. Thirdly, there are methods which impose a time series structure on the time effects. Finally there are the methods which concentrate solely on the estimable functions. These approaches are all discussed and illustrated in this section.

2.1. Penalty Function Approach

Osmond and Gardner[13] suggested that a constraint may be imposed by ensuring that a penalty function is minimized. This function measures the distance, in the parameter space, between each of the three two-factor models (age–period, age–cohort, period–cohort) and the three-factor model. All models were fitted by weighted least squares. The identifiability constraint is obtained by choosing as the three-factor solution the set of estimates which minimize this distance. The period–cohort model has the age effects constrained to be the average age specific rates over the whole period in view of the importance of the age effect.

Osmond and Gardner[13] wrote

$$\mu' = \mu$$
$$\alpha'_i = \alpha_i + \lambda(m - i) \quad i = 1, \ldots, m$$
$$\beta'_j = \beta_j + \lambda j \quad j = 1, \ldots, n$$
$$\gamma'_k = \gamma_k - \lambda k \quad k = 1, \ldots, m + n - 1$$

where $k = j - i + m$, and showed that the parameters μ', α'_i, β'_j, γ'_k yield exactly the same fitted values as μ, α_i, β_j, γ_k – a consequence of the linear relationship. Thus there are infinitely many solutions to the three-factor model, each one indexed by λ. Let $\theta^{\mathrm{T}}(\lambda) = (\mu, \alpha^{\mathrm{T}}, \beta^{\mathrm{T}}, \gamma^{\mathrm{T}})$, be the parameter vector of length $2(m+n) - 3$ of the full age–period–cohort model which depends upon λ. Denote the parameter vector for the age–period model by θ_{C} (that is, $\theta(\lambda)$ with $\gamma = 0$); θ_{P} is similarly defined but θ_{A} has age effects based on the average age specific rates. Note that θ_{C}, θ_{A} and θ_{P} are all of length $2(m+n) - 3$. Let R_{C}, R_{P} and R_{A} be the mean residual sum of squares of the two factor models. The penalty function[13] is

$$g(\lambda) = \frac{\|\theta_{\mathrm{A}} - \theta(\lambda)\|}{R_{\mathrm{A}}} + \frac{\|\theta_{\mathrm{C}} - \theta(\lambda)\|}{R_{\mathrm{C}}} + \frac{\|\theta_{\mathrm{P}} - \theta(\lambda)\|}{R_{\mathrm{P}}}$$

where $\|\theta_1 - \theta_2\| = [(\theta_1 - \theta_2)^{\mathrm{T}}(\theta_1 - \theta_2)]$ is the squared Euclidean distance between the two sets of parameter estimates. λ represents the unidentifiable linear relationship among the three factors and is estimated by minimizing $g(\lambda)$. Holford[7] suggests that it may be possible to specify bounds for λ to give a range of estimates.

Decarli and La Vecchia[17] proposed a similar solution, using Poisson regression, and provided GLIM macros. There are only slight differences between the two approaches and only the results of fitting the Decarli–La Vecchia version are presented in Table III. In view of the estimation of the parameters via the minimization of a penalty function, no standard errors are directly available.

The standard errors presented in this paper were calculated using a parametric bootstrap method.[18] Using the observed counts as the means of independent Poisson random variables the rates were simulated and the age–period–cohort model fitted to them. The standard deviation of the parameter estimates over the 100 bootstrap simulations are the standard errors in Table III. This reveals the sampling distribution of the estimate of λ and so emphasizes the random nature of the linear constraint.

The visual impression obtained from the Decarli–La Vecchia estimates is that the age and cohort effects are important but that the time period effect, while still significant, is of much less importance, with a slightly increasing trend. The cohort effect is curved with incidence rates increasing steadily from the earliest cohorts to a peak in 1925–1934 followed by a fairly flat plateau from 1930 to 1959. The standard errors of the estimates from the younger cohorts are large and there is no evidence to support a difference in risk form the cohort centred on 1940. The estimates are similar to the age and cohort model estimates, which is the two-factor model with the most parameters, despite there being clear evidence that a time period effect is necessary. The deviance of the three-factor age–period–cohort model is 57·8 on 44 d.f. and the log-likelihood G^2 statistic testing for the inclusion of the period effect, controlling for the age and cohort effects, is 53·65 and 4 d.f.

2.2. Individual Records Approach

Robertson and Boyle[10] attempted to overcome the non-identifiability problem by using the individual records of cases and forming a three-way table of age group, time period and birth cohort (see, also, Tango[19]). All three variables are recorded in the Scottish incidence data. Effectively this procedure splits each cell of Table I into two with the different cohorts (known as the older and younger cohort) in each part. It also has birth cohorts with minimal overlap. The population figures are published in 5 year age groups each year and the person years at risk for the two cohorts are calculated as weighted averages of the population sizes in the single years assuming a uniform distribution of births. The fitted model is

$$\ln(E[r_{ijk}]) = \mu + \alpha_i + \beta_j + \gamma_k \tag{2}$$

where $i = 1, \dots, m, j = 1, \dots, n, k = j - i + m$, for the older cohort and $k = j - i + m + 1$ for the younger cohort. As there are two cohorts associated with each (i, j) combination, the exact linear dependency between the parameter indices is broken. Within the same assumptions as discussed in Section 1 this model is identifiable.

The parameter estimates for model (2), Table III, contrast with those obtained from Decarli and La Vecchia[17] in that they suggest an increasing trend in the rates with time. The age curve is not as steep and the cohort curve appears to have a more shallow rise in the older cohorts to a peak at 1935–1939 followed by a decrease among the younger cohorts. There were no recorded

Table III. Parameter estimates for models

	Clayton–Schifflers Curvature estimate	Clayton–Schifflers Standard error	Holford Deviation estimate	Holford Standard error	Decarli–La Vecchia Effect estimate	Decarli–La Vecchia Standard error	Lee and Lin Effect estimate	Lee and Lin Standard error		Robertson–Boyle Effect estimate	Robertson–Boyle Standard error
Age											
20–24	–	–	−1·841	0·098	0·00	–	0·0	–		0·0	–
25–29	−0·382	0·174	−0·602	0·055	1·65	0·144	1·74	0·097		1·53	0·135
30–34	−0·406	0·086	0·255	0·040	2·92	0·136	3·00	0·091		2·73	0·130
35–39	−0·194	0·053	0·706	0·032	3·78	0·139	3·82	0·086		3·50	0·131
40–44	−0·247	0·038	0·963	0·027	4·45	0·135	4·43	0·082		4·10	0·132
45–49	−0·366	0·032	0·973	0·022	4·87	0·137	4·80	0·074		4·47	0·134
50–54	0·153	0·030	0·617	0·018	4·93	0·138	4·80	0·074		4·46	0·137
55–59	−0·013	0·029	0·415	0·014	5·14	0·137	4·95	0·071		4·61	0·140
60–64	−0·030	0·028	0·199	0·012	5·33	0·137	5·09	0·068		4·74	0·143
65–69	−0·001	0·029	−0·046	0·012	5·50	0·137	5·20	0·066		4·85	0·147
70–74	−0·006	0·030	−0·292	0·014	5·66	0·138	5·31	0·067		4·96	0·151
75–79	−0·004	0·034	−0·545	0·018	5·82	0·139	5·41	0·065		5·05	0·155
80–84	–	–	−0·801	0·023	5·98	0·137	5·51	0·069		5·16	0·160
Time											
1960–1964	–	–	−0·028	0·008	0·0	–	0·0	–		0·0	–
1965–1969	−0·024	0·026	0·003	0·009	0·041	0·015	0·100	0·017		0·095	0·018
1970–1974	0·019	0·025	0·010	0·009	0·058	0·015	0·176	0·021		0·171	0·022
1975–1979	−0·034	0·024	0·035	0·009	0·094	0·014	0·267	0·027		0·267	0·029
1980–1984	−0·067	0·023	0·027	0·008	0·097	0·018	0·335	0·032		0·330	0·036
1985–1989	–	–	−0·048	0·007	0·032	0·014	0·314	0·036		0·331	0·043
Cohort											
1875–1884	–	–	−0·111	0·059	−0·910	0·081	−0·240	0·075	1875–1879	−0·320	0·125
1880–1889	0·089	0·087	−0·134	0·042	−0·881	0·058	−0·235	0·081	1880–1884	−0·323	0·098
1885–1894	−0·051	0·058	−0·069	0·032	−0·763	0·048	−0·206	0·084	1885–1889	−0·218	0·086
1890–1899	−0·020	0·044	−0·054	0·025	−0·695	0·044	−0·192	0·088	1890–1894	−0·248	0·077
1895–1904	0·070	0·036	−0·060	0·018	−0·648	0·039	−0·190	0·081	1895–1899	−0·177	0·068
1900–1909	−0·049	0·030	0·005	0·014	−0·532	0·039	−0·142	0·087	1900–1904	−0·170	0·060
1905–1914	0·016	0·029	0·021	0·015	−0·464	0·032	−0·126	0·087	1905–1909	−0·128	0·053
1910–1919	−0·017	0·029	0·052	0·019	−0·380	0·031	−0·101	0·092	1910–1914	−0·102	0·046
1915–1924	0·059	0·030	0·067	0·024	−0·312	0·027	−0·083	0·089	1915–1919	−0·169	0·039
1920–1929	−0·035	0·031	0·141	0·030	−0·186	0·030	−0·023	0·095	1920–1924	−0·036	0·033
1925–1934	−0·013	0·035	0·180	0·037	0·095	0·025	0·013	0·100	1925–1929	−0·022	0·027
1930–1939	−0·062	0·040	0·206	0·043	0·016	0·027	0·029	0·098	1930–1934	−0·007	0·024
1935–1944	−0·001	0·049	0·170	0·051	0·0	–	0·0	–	1935–1939	0·0	–
1940–1949	0·004	0·066	0·133	0·058	0·016	0·032	−0·032	0·094	1940–1944	−0·078	0·029
1945–1954	−0·035	0·100	0·101	0·068	0·036	0·041	−0·060	0·110	1945–1949	−0·080	0·038
1950–1959	−0·171	0·182	0·033	0·084	0·021	0·078	−0·083	0·116	1950–1954	−0·278	0·057
1955–1964	−0·030	0·437	−0·207	0·127	−0·166	0·128	−0·109	0·109	1955–1959	−0·399	0·093
1960–1969	–	–	−0·476	0·285	−0·383	0·969	−0·108	0·089	1960–1964	−0·600	0·183

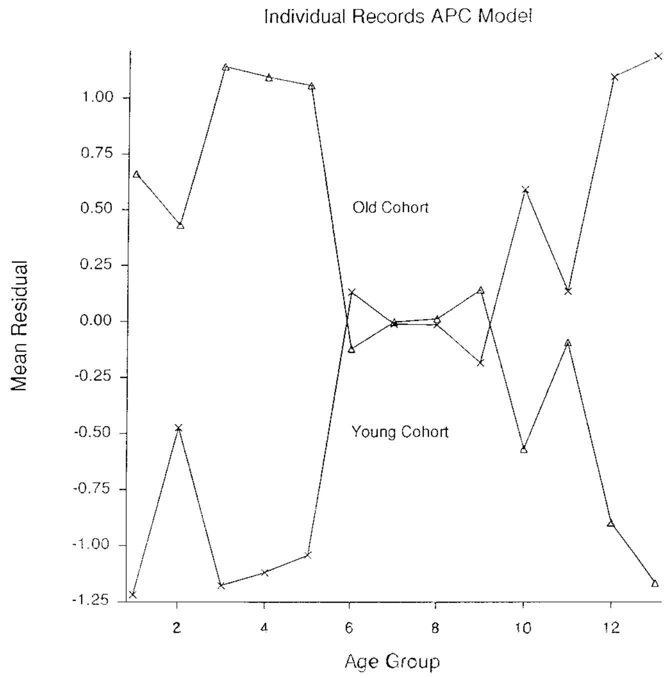

Figure 1. Residual plots for the Robertson and Boyle[10] solution (model 2)

cases for the 1965–1969 cohort and as only one observation is associated with this cohort the estimated effect would be large and negative indicating that the estimated effect is tending towards negative infinity with an infinite standard error. This was dealt with by weighting out this point, see also Section 4, but prior information[20] could have been used.

The age–period–cohort model does not provide a very good fit to the data yielding a deviance of 239 on 120 d.f. Initially Robertson and Boyle[10] pursued an approach based upon extra Poisson variation[21] but this is not the correct procedure in this case as inspection of the residuals suggests that there are systematic departures from the model assumptions, Figure 1. These show that, at the younger age groups, model (2) overestimates the numbers of cases when cohort k forms the younger cohort in the (i, j)th cell, and underestimation among the older cohort within the age group. The reverse is true in the older age groups.

The assumption of a common age effect within the two cohorts associated with each cell of the two way table of rates (Table I) is invalid.[6,19,22] With reference those in age group j the average age of the older cohort is $15 + 5j + 3\frac{1}{3}$ years whereas the average age in the younger cohort is $15 + 5j + 1\frac{2}{3}$ years. These values are obtained by integrating age over the older and younger cohorts within a cell and are based on a uniform distribution of age within a time period. So for those aged 60–64 the average ages are $63\frac{1}{3}$ years and $61\frac{2}{3}$ years. Similarly the ‘average’ years of incidence and ‘average’ years of birth are not equal. As the age curve tends to be very steep, particularly for certain forms of cancer, the assumption of equal age effects is not valid and is most crucial.

Table IV. Deviances and degrees of freedom for the models fitted with an individual records approach

Model			Deviance	d.f.
Age +	Period +	Cohort	239·0	120
Age +	Period +	Cohort + Old	239·0	120
Age × Old +	Period +	Cohort	150·3	108
Age +	Period × Old +	Cohort	226·4	115
Age +	Period +	Cohort × Old	149·2	104
Age × Old +	Period × Old +	Cohort	135·6	103
Age × Old +	Period +	Cohort × Old	111·0	92
Age +	Period × Old +	Cohort × Old	124·9	99
Age × Old +	Period × Old +	Cohort × Old	99·3	88

By introducing a factor which picks out the old–young cohort within each cell the above criticisms can partially be taken into account.[23] The old–young factor is completely confounded within cohort so that adding in the main effect produces no change in the deviance. However, interactions with age, time and cohort are estimable and can be interpreted. They adjust the age effects to allow for differences between the old and young cohorts within each age group. The deviances and degrees of freedom are presented in Table IV. Only 16 degrees of freedom are lost when fitting the old–young by cohort interaction because the oldest and youngest cohorts in the series (cohort 1: 1875–1879, and cohort 19: 1965–1969) are only observed once each and cohort 19 has been weighted out as there were no observed cases in the data. Consequently there are no degrees of freedom for the interaction with the old–young factor for these two cohorts.

This provides a test of the validity of the assumption of the common age, period and cohort effects within a cell of the table. The addition of the old–young cohort by both age and cohort interactions yields a final model with an acceptable deviance in relation to its degrees of freedom. The reduction in deviance associated with the interaction with time is 11·7 on 4 d.f. which suggests that the interaction with period is less important than the interactions with age and cohort. However this extension does confirm that the individual records approach is not particularly useful in this case. There is clear evidence of differential effects among the old and young cohorts within a cell. With lung cancer incidence in Scotland some progress can be made using interactions with the old–young cohort indicator variable.[23,24]

A model with all two-way interactions between old–young factor and age, period and cohort will have the identifiability problem. The two cohort layers of the three-way table are modelled separately and within each layer, which is effectively a two-way table of age group by time period, only one cohort is associated with each cell. The use of this interaction term is only feasible if at least one of the factors, but preferably two, are aggregated on a coarser level from the others, that is, if the effect of the period and cohort factors is assumed to be the same over the old and young cohorts within a cell. While such extensions to the simple age–period–cohort model are possible they cannot be recommended other than for investigating the adequacy of the assumption of common effects across the old and young cohort.

The individual records procedure is asymmetric in the factors and it is this lack of symmetry which is utilized to break the linear relationship, which is exact in continuous time.[25] Although identifiability results there is an induced bias as a result of forcing the age effects to be common across two adjacent birth cohorts. For increasing age effects this yields a bias towards decreasing cohort effects and increasing period effects. With the addition of interaction terms the bias is not as severe but there is much more collinearity leading to less stable estimates.

2.3. Autoregressive Models

Recently a first-order autoregressive time series model for the cohort effects was proposed by Lee and Lin.[26] This approach considers the effects of cohort to be stochastic rather than deterministic. Also, the researchers argue that there is usually some dependency between cohorts which can be modelled by an autoregressive process. Furthermore, an identifiable set of parameters can be obtained within this model.

The model proposed by Lee and Lin[26] takes the following form:

$$\begin{aligned} \ln(E[r_{ij}]) &= \mu + \alpha_i + \beta_j + \gamma_k \\ \gamma_k &= \phi\gamma_{k-1} + \delta_k \end{aligned} \tag{4}$$

where δ_k are independent and normally distributed with mean zero and variance σ^2. The δ_k represent the random white noise shocks to the birth cohort process. The stationary restrictions applied to the autoregressive process imply that $-1 < \phi < 1$ and in view of the anticipated dependency of one cohort on another, negative values are not anticipated. The cohort effects are stochastic and only the two parameters, ϕ and σ^2, are used to represent them within the autoregressive process with white noise. There are a further $m + n - 1$ parameters for the intercept and the age and period effects in model (4) above. This is a big reduction in the number of parameters compared to the models in Sections 2.1 and 2.2.

The parameters are obtained by maximum likelihood estimation and the likelihood has two portions. The first is the Poisson likelihood for μ, α_i, β_j, γ_k and the second is the likelihood for the parameters of the autoregressive process, ϕ and σ^2. The unconditional likelihood function for the $m + n + 1$ parameters is obtained by integrating over the unobserved values of $\gamma_1, \ldots, \gamma_{m+n-1}$. However, Lee and Lin[26] do not choose to maximize the unconditional likelihood as it may not have a closed form solution, and, more importantly, there would be no estimates of the cohort effects. Instead, they suggest that the conditional likelihood, which is just the product of the Poisson and autoregressive portions, be maximized. The log-likelihood reduces to

$$\begin{aligned} L(\mu, \alpha_i, \beta_j, \gamma_k, \phi, \sigma^2) = C &+ \frac{1}{2}\ln(1-\phi^2) - \frac{(m+n-1)}{2}\ln(\sigma^2) \\ &- \frac{1}{2\sigma^2}\left[\gamma_1^2(1-\phi^2) + \sum_{k=2}^{m+n-1}(\gamma_k - \phi\gamma_{k-1})^2\right] \\ &+ \sum_i \sum_j (y_{ij}\ln(E[r_{ij}]) - E[r_{ij}]) \end{aligned}$$

where C is a constant.

The parameter estimates presented in Table III were obtained by maximizing this function numerically using the non-linear function minimization routines in S-plus. A number of different starting values were used and all converged to the same point. The starting values requiring the fewest iterations were obtained using the Decarli and La Vecchia[17] estimates supplemented by the autocorrelation and variance of the cohort parameters. Although there are range constraints on two of the parameters, ϕ and σ^2, no boundary problems were experienced in the maximization. Standard errors were obtained from a parametric bootstrap. Lee and Lin[26] used bootstrap confidence intervals but this has not been done here to maintain comparability with the other methods. In fact the univariate distributions of the estimates from the bootstrap samples were all normally distributed apart from those for the two parameters with range constraints, where the distributions were both slightly skew.

The age effects from the Lee and Lin[26] solution are similar to those produced by the Decarli and La Vecchia approach though the bootstrap standard errors are much smaller. There is a strong indication of an increasing risk with time period which has levelled off in the last three periods. In this respect the estimates are similar to the Robertson and Boyle[10] estimates. The cohort effects are curved with a peak at 1930–1939 and while the estimates among the recent cohorts are declining the standard errors are large. The rise and fall of the cohort effects is much less than the corresponding rise and fall observed in the Decarli and La Vecchia estimates. The autocorrelation parameter is estimated as $\hat{\phi} = 0{\cdot}946(0{\cdot}016)$, indicating a strong relationship between the cohorts, and the variance of the white noise process is estimated as $\hat{\sigma}^2 = 1{\cdot}13\times 10^{-2}(0{\cdot}086\times 10^{-2})$.

Although the model is specified as one with stochastic cohort components, the estimation process makes it clear that the 'solution' has much in common with penalized likelihood and the constraint used by Decarli and La Vecchia.[17] In the function which is maximized the term

$$-\frac{1}{2\sigma^2}\left[\gamma_1^2(1-\phi^2)+\sum_{k=2}^{m+n-1}(\gamma_k-\phi\gamma_{k-1})^2\right]$$

attempts to keep γ_1 close to zero and to constrain $\gamma_k = \phi\gamma_{k-1}$. Thus, it is evident that a location constraint such as $\gamma_1 = 0$ will not be useful here and in the example the constraint used was $\sum\gamma_k = 0$. The estimates were subsequently transformed for comparability with the other 'solutions'. The parameter, σ^2, can be thought of as a smoothing parameter. Large values of σ^2 mean that little emphasis is placed on the autoregressive constraint whereas small values of σ^2 emphasize the importance of the constraint. Both of the terms $\ln(1-\phi^2)$ and $\ln(\sigma^2)$ can be interpreted as penalizing the log-likelihood to ensure that both ϕ and σ^2 stay away from their respective boundaries.

Viewed in this way it is clear that this method is constraining the cohort effects to follow a smooth curve. It is entirely possible to use different constraints. If the first differences were constrained to be equal then this is equivalent to constraining the estimates to lie as close as possible to a straight line. Also, there is nothing to prevent constraints being applied to the age effects and the period effects.

Berzuini and Clayton[27] have developed a Bayesian framework for the analysis of temporal trends. This is very similar to the work of Lee and Lin,[26] particularly in the use of the autoregressive process. The main developments of Berzuini and Clayton[27] focused on the estimation of future rates, based on the autoregressive process in both period and cohort, and not on the estimation of any age, period and cohort effects. Consequently this method is not applied to the data here. Essentially they acknowledged the existence of the linear dependency between the three components, age, period and cohort, but because the predictions are the same irrespective of which two components are used in the model to obtain the parameter estimates it is not necessary to focus on the estimation of the trends.

2.4. Curvature and Drift

Holford[14] concentrated only on the estimable functions and did not make any attempt to impose any constraints to ensure identifiability. Instead, there are two estimates which summarize the linear effects of age, period and cohort, and parameters associated with the deviations of the effects of each of these factors from linearity. If α_L, β_L and γ_L denote the linear components of age, period and cohort, respectively then any linear combination of these of the form $d_1\alpha_L + d_2\beta_L + (d_2 - d_1)\gamma_L$, where d_1 and d_2 are constants, is estimable. For example, $\alpha_L + \beta_L$ and $\beta_L + \gamma_L$ are both estimable functions but say nothing about which of the two factors are associated with any linear change in $\ln(r_{ij})$. Holford[14] then goes on to estimate the deviations from linearity, which are measures of curvature and are identifiable.

Details of how to fit this model are presented in Holford.[14] Genstat[28] was used as matrix calculations are required. The deviations in Table III are linear combinations of the parameter estimates based upon a design matrix which is orthogonal to the linear trend and so it is possible to find their standard errors. The estimated linear components are (standard errors, in brackets) $\alpha_L + \beta_L = 0{\cdot}422(0{\cdot}006)$ and $\beta_L + \gamma_L = 0{\cdot}073(0{\cdot}004)$. There are clear and systematic deviations from linearity in the age and cohort effects. At the older ages and younger ages the effects are below the linear increase but in the middle age groups they are above. Similar curves can be seen among the period and cohort deviations.

It has been suggested[7,29] that restricted ranges for the time trends can be obtained by consideration of the aliasing parameter, which is effectively λ in the Osmond and Gardner[13] approach. The estimable linear combination of the linear trends $d_1\alpha_L + d_2\beta_L + (d_2 - d_1)\gamma_L$, where d_1 and d_2 are constants, can also be written as $(d_2 - a)\alpha_L + (d_1 + a)\beta_L + a\gamma_L$, where $a = d_2 - d_1$ is an alias parameter. Varying a and adding the deviations in Table III to the age, period and cohort trends (which are functions of a) gives the graphs in Figure 2. The deviations from Table III are plotted on the lower graphs and the trend plus deviations on the upper ones. They vary according to the value of the alias parameter and three scenarios are presented.

The shape of the estimated age incidence curve is relatively unaffected by the small changes in the alias parameter illustrated. For example the change in the age slope at age group 6 (45–49) is visible in all three displays. However, the interpretation of the period and cohort trends are affected by small changes to the alias parameter. However, if one was to postulate that, adjusting for age and cohort, it is unlikely that the rates would decrease noticeably with period, then the solid line in Figure 2, in particular, would be untenable. This would lead to the conclusion that displays such as the dashed and dotted line would be more acceptable with an increase in rates associated with period, virtually no cohort effect until about cohort 13 (1940) whereupon the rates begin to decline.

Tango and Kurashina[30] adopt substantially the same position as Holford.[14] Clayton and Shifflers[6] suggest that the two degrees of freedom associated with the linear trends be partitioned into two components. One is the linear age effect and the other is the 'drift'. The rationale for this approach is that age is assumed to be the dominant factor and it is natural to assign one of the linear components to this factor. Having done this the time and cohort linear effects are linked together and are known as the linear drift. They suggest that only the curvatures of the effects of the three factors be presented when dealing with the age–period–cohort model. The curvatures of the estimates are the second differences of the parameter estimates, for example, $\alpha_1 - 2\alpha_2 + \alpha_3$, and do not depend upon the linear constraint used.

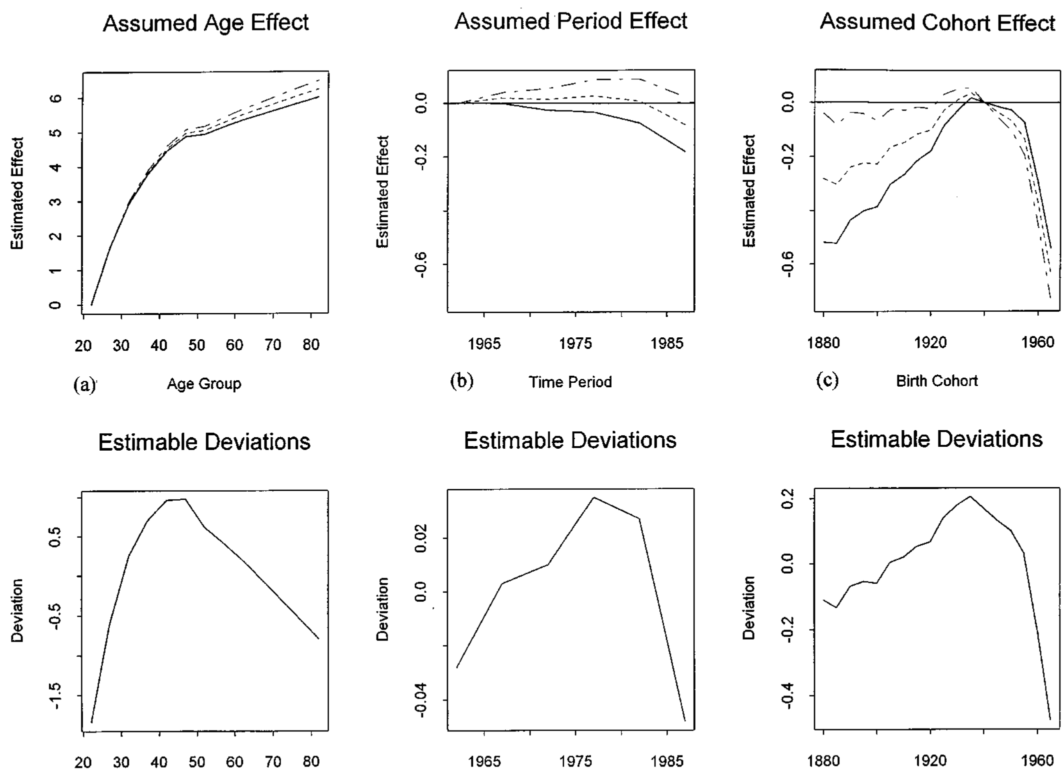

Figure 2. Estimated trends in breast cancer incidence. These were obtained by superimposing the deviations from Holford's[14] approach on a linear trend under three different values of the alias parameter, see Section 2.4. Upper plots are estimated trends, lower plots are deviations: (*a*) alias = – 0·02 (solid line); (*b*) alias = 0·0 (dashed line); (*c*) alias = 0·02 (dashed and dotted line)

The estimates can be obtained by fitting the linear age and linear period (drift) terms followed by the age, period and cohort factors. There is no need to redefine the levels of these factors to ensure identifiability and all that will happen is that the first and last levels of each factor will be aliased to zero. The curvatures can then be calculated directly from the estimates of the effects of the different levels of the factors. However, if different location constraints are applied to the factors to replace the usual ones of $\alpha_1 = 0$, $\beta_1 = 0$, $\gamma_1 = 0$, say $\gamma_{11} = 0$, then the estimates of the intercept, linear age and linear period (drift) change although the deviance remains unchanged.

As the curvatures are just linear combinations of the estimated parameters the standard errors are simple, in principle, to calculate from the variance-covariance matrix of the parameter estimates. An alternative approach is to transform the design matrix so that the curvatures of age, period and cohort are estimated directly along with the two linear components. This has two advantages: (i) the standard errors of all parameters are directly available; (ii) there is no need to specify any location constraint for each factor so the estimates of the linear age, linear drift and intercept term are also identifiable. This transformation is simply achieved by considering the age group effects $\alpha^{\mathrm{T}} = (\alpha_1, \ldots, \alpha_m)^{\mathrm{T}}$, and the associated curvatures $\alpha_c^{\mathrm{T}} = (\alpha_{c1}, \ldots, \alpha_{c(m-2)})^{\mathrm{T}}$, where

$\alpha_c = \mathbf{A}\alpha$. $\mathbf{A}$ is an $(m-2)\times m$ matrix with zero elements except for $a_{ii} = 1$, $a_{i,i+1} = -2$, $a_{i,i+2} = 1$, $i = 1, \ldots, (m-2)$. If $\mathbf{A}^*$ is the matrix derived from $\mathbf{A}$ with the first and last columns deleted, corresponding to the constraint $\alpha_1 = \alpha_m = 0$, then $\mathbf{A}^*$ is non-singular. If $\mathbf{X}_\mathrm{T}$ is the design matrix associated with the parameters $\alpha_2, \ldots, \alpha_{m-1}$ then the time period curvatures and standard errors are obtained by using the columns of the matrix $\mathbf{X}_\mathrm{T}\mathbf{A}^{*-1}$ as explanatory variables in the model.

Clayton and Schifflers[6] suggest looking at intervals where there is a sharp change in the curvature estimates. There would appear to be little curvature associated with time period apart from a change in trend at 1980–1984 (Table III). Individually none of the cohort curvatures are significantly different from zero though they are large and positive for 1895–1904 and 1915–1924. Apart from these two estimates the others tend to be negative, which is consistent with a slowing down of the linear increase in drift. There is also considerable variability in the estimates of the curvatures of the oldest and youngest cohorts. The age curvatures are consistent with an age effect which is predominantly concave, with the exception of the positive curvature at age 52. The age and cohort patterns are consistent with patterns obtained using the other methods. The cross-sectional linear age effect is 0·467 (0·020) and the longitudinal age effect is 0·505 (0·012) and the drift is 0·037 (0·022).

There are two fewer parameters with Clayton and Schifflers[6] for each factor compared to Holford.[14] This represents the maximum number of parameters which can be estimated free from drift effects. The Holford[14] parameters represent deviations from linearity and consequently for each factor the parameter estimates sum to zero and have a zero slope when regressed against age, period or cohort, as appropriate. Thus the deviations are not independent of each other. Both approaches have the same number of unique estimates for each factor. The curvatures are local curvatures and are not unduly influenced by the curvatures at other intervals other than those close to it. For example, the curvature at the 1925 cohort is correlated with the curvatures at 1930 and 1935 as they have 2 and 1 coefficients in common, respectively. The curvatures at 1925 and 1940 have no coefficients in common and are much less correlated. Thus the curvatures of Clayton and Schifflers[6] are possibly easier to use. In fact the curvatures and the deviations are linearly related. For example, with cohort 1880–1889 the second differences of the cohort deviations are $-0{\cdot}1166 - 2(-0{\cdot}1407) - 0{\cdot}0759 = 0{\cdot}0889$, which is the curvature estimate within rounding error.

Tarone and Chu[31] published an analysis of birth cohort patterns in breast cancer mortality rates in the United States and Japan. They also extended the estimable curvatures of Clayton and Schifflers[6] and developed a methodology for testing changes in the trends associated principally with birth cohort, but which are also applicable to trends in age and period. In line with their previous paper,[11] Tarone and Chu also advocated the use of two year age groups and time periods to reduce the overlap in the birth cohorts. This method uses functions which are estimable.

Using the individual records of all cancer incidence in Scotland from 1960 until 1989, a two-way table of age group by time period was constructed using intervals of width two years. All ages from 20 and 83 were used giving $n = 32$ age groups, $m = 15$ time periods and 46 birth cohorts. The first cohort corresponds to those who were 82–83 in 1960–1962 and so were born in 1877–1880, the second cohort were 80–81 in 1960–1962 and so were born in 1879–1882. Thus there is some overlap of cohorts and the convention used here is to take the central two years 1878–1979, 1880–1901 etc. and to refer to the cohort by the first of these two years.

As the population data are only available by five year age groups for each year from 1950 the population by single years of age was estimated using smoothed two-dimensional cubic interpolation.[32] This is not as sophisticated as Beer's method[31,33] as it does not guarantee that the five year totals are preserved. However, the differences are not likely to have a considerable effect as the number of incidences are small relative to the population sizes.

From Section 2.1, $\hat{\theta}(\lambda)^{\mathrm{T}} = (\mu, \alpha^{\mathrm{T}}, \beta^{\mathrm{T}}, \gamma^{\mathrm{T}})$, denotes the estimated parameter vector of length $2(m+n)-3$ of the full age–period–cohort model which depends upon λ. A linear function of these parameters is estimable if $\mathbf{c}^{\mathrm{T}}\theta(\lambda_1) = \mathbf{c}^{\mathrm{T}}\theta(\lambda_2)$, where $\mathbf{c}^{\mathrm{T}} = (0, \mathbf{c}_\alpha, \mathbf{c}_\beta, \mathbf{c}_\gamma)^{\mathrm{T}}$ is a contrast vector, and, λ_1 and λ_2 index two sets of parameters using different constraints. Tarone and Chu[31] were particularly interested in evaluating if there were changes in the slopes of the cohort trends when comparing groups of cohorts, which are specified prior to fitting the models.

Here, the investigation of a decrease in risk associated with recent cohorts is of particular importance and one way to achieve this is to compare the trend in the cohort estimates for a group of 8 two year cohorts born before the Second World War (1924–1938) and a group born after (1946–1960). Groups of 8 cohorts are used in line with Tarone and Chu.[31] The linear trend in the cohort effects for the first group of cohorts beginning at cohort h can be estimated by regressing $\gamma_h, \ldots, \gamma_{h+7}$ on the integers $h, \ldots, h+7$, and similarly for the cohort group beginning at cohort k, $h+7 \leqslant k$. The contrast of the two slopes can be estimated by

$$\begin{aligned} C_{hk}^{(8)} = {} & 7\gamma_{k+7} + 5\gamma_{k+6} + 3\gamma_{k+5} + 1\gamma_{k+4} - 1\gamma_{k+3} - 3\gamma_{k+2} - 5\gamma_{k+1} - 7\gamma_k \\ & - (7\gamma_{h+7} + 5\gamma_{h+6} + 3\gamma_{h+5} + 1\gamma_{h+4} - 1\gamma_{h+3} - 3\gamma_{h+2} - 5\gamma_{h+1} - 7\gamma_h). \end{aligned}$$

A contrast of the cohort parameters will be identifiable if $\mathbf{c}_\gamma^{\mathrm{T}}\gamma = \Sigma_k c_{\gamma k}\gamma_k$ takes the same value for all parameterizations of γ. From Section 2.1, $\gamma'_k = \gamma_k - \lambda k$, and so it is sufficient to have $\Sigma_k c_{\gamma k} k = 0$, over and above the contrast requirement $\Sigma_k c_{\gamma k} = 0$, to ensure that the contrast is identifiable. It is readily verified that C_{hk}, above, is identifiable, as are other similar contrasts.

The estimated value of $C_{1946,1960}^{(8)} = -3{\cdot}81(1{\cdot}32)$ which is consistent with a major change in cohort based trends in breast cancer incidence. For the period trends the important comparison in the Scottish incidence data is the first half compared to the second. Using a similar contrast to $C_{hk}^{(8)}$ the estimated change in period slopes over 12 year periods from 1962–1972 compared to 1978–1988 is $P_{1962,1978}^{(6)} = -0{\cdot}97(0{\cdot}19)$. This estimate is consistent with a decrease in the trend in the later period compared to the earlier one. Among other possible explanations this is consistent with a registration effect which has levelled off.

The contrasts and curvatures[6] are related as the comparison of the slopes in two successive pairs of cohorts is the curvature:

$$C_{h,h+1} = (\gamma_{h+2} - \gamma_{h+1}) - (\gamma_{h+1} - \gamma_h).$$

The Tarone and Chu[31] approach has more scope when there are prior hypotheses to be investigated. Indeed, it does not appear to have much use if there are no prior hypotheses as there are many possible identifiable contrasts. If the contrasts are chosen on the basis of an inspection of the estimates there are severe problems with the interpretation of the statistical significance of any tests. There is nothing to prevent the use of the method with data arranged in 5 year age groups but then there will be much less scope to look at independent groups of cohorts or periods.

3. DISCUSSION

Age–period–cohort models would not be contemplated in continuous time in view of the linear relationship among the three variables and consequently by analogy should not be used in discrete time. The separate estimation of the linear effects of age, period and cohort only becomes apparently possible by grouping the data and assuming constant effects within each interval. The approaches of Holford,[14,34] Tango and Kurashina,[30] Clayton and Schifflers[6] and Tarone and Chu[31] come closest to the approach that one would adopt in a continuous time model. They assign two degrees of freedom to the linear effects and then estimate the higher order curvatures or deviations from linearity.

Such approaches have no failings in terms of the assumptions made. They do not provide any information on the assignment of the linear trend to the three factors without some extra assumptions being made. Thus it is clear that, for example, a conclusion that the increase in incidence rates is due to an increasing time effect is based upon extra assumptions made by the researcher. The two-way table of rates by itself will not support such a conclusion.

The estimates from the Osmond and Gardner[13] and the Decarli and La Vecchia[17] approach are modified by the estimates from all three two-factor models through the minimization of a penalty function. This is always likely to move the estimates towards the best fitting two factor model or the one with most parameters (age–cohort). The linearity constraint depends on the data and is random. This means that if this model is used in two countries a different identifiability constraint is used in each. The standard errors of the parameter estimates are not easily available though they can be obtained by simulation. In view of these comments and also those of Clayton and Schifflers[6] it would appear that such an approach should not be considered.

Incidentally, if one calculates the curvatures of the age, period and cohort estimates for both the Decarli and La Vecchia and Osmond and Gardner approaches the same values are obtained as for the Clayton and Schifflers approach. This illustrates that the curvatures are estimable but not the linear components. The linear components associated with the Decarli and La Vecchia[17] approach are age–0·412, period–0·010 and cohort–0·053.

The approach by Lee and Lin[26] does not yield the same curvature estimates as those of the other methods in Table III. This is as a result of the constraint imposed to get the parameter estimates. Although the method is couched in terms of an autoregressive model for the cohort parameters it is really just an autoregressive constraint to ensure identifiability. The validity of this constraint needs to be established as Lee and Lin[26] acknowledge. Whether or not this penalized autoregressive constraint on the cohort parameters is applicable to breast cancer incidence in Scotland is open to question. There is not going to be much help from the data in terms of the usual means of investigating a time series for autocorrelation through the partial and autocorrelation functions as the series of cohort effects is very short in traditional times series terms.

Robertson and Boyle[10] ensured that the cohorts formed a series of non-overlapping intervals by constructing a three-way table using individual records. From the point of view of interpretation it is important that a person remains in the same cohort throughout the period of study as she ages. This is not the case when dealing with a two-way table only. This point of detail is of little importance to epidemiologists and is not a stumbling block to the analysis of two-way tables in five or ten year groups, though there have been other recent attempts to minimize the effect of the overlap.[10,31]

The use of a three-way table also leads to the identifiability of the parameters. However, there is an induced bias as a result of the different average ages, years and birth years of the two cohorts in each age group, time period cell. If there is only an increasing age effect then forcing the data into a three-way table and assuming a common age effect over the two cohorts in each age group by time period cell will induce an increasing time effect and a decreasing cohort effect.[22,25]

The introduction of a factor to compensate for the differences between the two cohorts within a cell does enable the development of models which may be better descriptions of the data.[23,24] The cost of this is in complexity in terms of the numbers of parameters required and also in terms of interpretation of the estimates. This 'solution' is only a superficial one. The hidden price is multi-collinearity which is severe in the individual records approach. For the problem in hand the condition index[35] is 25·5 for the age–period–cohort model, 112 for the old–young interaction with age group, and 190 for the model with the two interactions between old–young and age and cohort. This is a reflection of the futility of such modelling in continuous time and that age–period–cohort models only have any scope as a result of aggregation. The effects of multi-collinearity on the parameter estimates and their standard errors are well known[35,36] and these are noticeable in any models with an interaction with the old–young factor. The old–young interaction models have similarities to the interaction model proposed by James and Segal.[37]

The age and period factors exhibit similar curvature estimates to those obtained using the Clayton and Schifflers[6] approach. The cohorts have a different grouping and so are not expected to have similar curvatures. The estimates of the linear components of the age, period and cohort effects from the individual records' approach are 0·430(0·013), 0·066(0·009) and $-0{\cdot}120(0{\cdot}037)$, respectively. These contrast with the corresponding estimates from Osmond and Gardner's approach which had a broadly increasing linear cohort effect and virtually no linear time effect, albeit without any standard errors.

In all approaches to age–period–cohort models the cells at the top right, $(1, n)$, and bottom left, $(m, 1)$ of the table have fitted values which are equal to the observed values. This occurs because the cohort parameters γ_1 and γ_{m+n-1} occur only in these cells. Thus these two corner cells have a leverage of one and have one cohort parameter associated with each. Consequently they do not contribute to the residual deviance of the model. These two cells could be weighted out of the analysis with little loss. This would serve to reduce the emphasis placed upon any trend among younger cohorts with a view to getting some indication of the likely rates in the future. The curvature estimates are not greatly influenced by the values in these two corner cells, apart from the curvatures for cohorts 2 and $(m + n - 2)$, though the two linear terms can be.

The approaches which focus on the estimable functions should be adopted. While the estimates are more difficult to interpret than the trends there is not the danger of concluding spurious trends which are based on unknown constraints and assumptions. The Decarli–La Vecchia[17] and the Osmond and Gardner[13] 'solutions' both introduce a bias towards cohort based changes in rates, while the Robertson and Boyle[10] proposals also have a bias. In a recent review, Holford[7] considered the use of models in which a parametric functional form is used for the age, period or cohort effects. Within these assumptions the model is fully specified and no extra constraints are needed to ensure identifiability. Holford also reviewed models in which covariates, such as exposure of a population to smoking, are used. These models are useful in the correct setting and are more specific in their application than the methods discussed in this paper.

Some recent publications on age–period–cohort analysis have adopted an approach advocated by Holford[7] and illustrated in Figure 2 whereby a range of curves are presented under various assumptions about the aliasing parameter. These assumptions are considered to be reasonable by the investigators.[38,39] There are, however, other approaches. Lee *et al.*[40] use the Osmond and Gardner approach, Xu *et al.*[41] used two separate models age–period and age–cohort to estimate the effects of age, period and cohort, while Wilcox *et al.*[42] had a hypothesis to test about one specific cohort and so included a dummy variable for this cohort as the only cohort effect.

The most appropriate uses of such models are as a means of describing the patterns of rates within the two- or three-way tables. The most important analytic use is likely to be in the comparison of the pattern of rates over regions, sites and gender.[43] In this respect the interaction of the stratifying factor, say gender, with the age, period and cohort curvatures will provide the appropriate estimates and will permit tests of the similarity of the patterns of rates to be made. The stratifying factor by linear components interactions can also be interpreted but not the main effects of these linear components.

With regard to the interpretation of the trends in the breast cancer incidence rates only the estimates based on the models discussed in Section 2.4 are without untested assumptions. The age–period–cohort model provides a reasonable description of the data with a deviance of 57·8 on 44 degrees of freedom. None of the factors can be omitted from this model so there is evidence of significant curvatures in each of these factors. The significant age curvatures all take place before age 50–54. There are negative curvatures prior to 50–54 implying that the rate of increase is declining and a positive one at 50–54 indicating an abrupt change of slope. After 50–54 the changes in the rates with age are not significantly different from linear trends though they are all negative. These curvature estimates are consistent with 'Clemmesen's Hook'.[44]

A large negative period curvature of $-0{\cdot}067$ (0·023) occurs in 1980–1984. This means that the risk of breast cancer in 1985–1989 relative to 1980–1984 is 94 per cent of the risk of breast cancer in 1980–1984 relative to 1975–1979. This is the only major period curvature and is consistent with rates which had increased with period and are now levelling off.

The three most important cohort based curvatures take place at 1900, 1920 and 1935. The first one is associated with a sharp increase in risk associated with the 1905 cohort compared to the 1900 one relative to the increase in risk of the 1900 cohort compared to the 1895 one. Similarly in 1920. The relative risk among the 1940 cohort compared to the 1935 cohort is less than that of the 1935 cohort compared to the 1930 one. There is no evidence that there has been any significant change in cohort trend after 1945. The standard errors for the curvatures in the later cohorts are large as there are few cases. All but one of the curvatures after 1925 are negative and this is consistent with rates which are decreasing faster than linear with cohort.

This information is available from Clayton and Schiffler's[6] curvatures and Holford's[14] deviations in Table III, and also the contrasts of Tarone and Chu[31] in Section 2.4. These patterns can also be seen in the other estimates in Table III provided only changes in comparable groups of estimates are considered and not the magnitude of the estimates. For example, in the simple individual records age–period–cohort model in Table III, to say that the 1955–1959 cohort has a significantly smaller risk than that of the 1935–1939 cohort is based on the individual records assumption and is biased. However, a comparison of the trend of the estimates in the 1925–1929, 1930–1934 and 1935–1939 cohorts with the trend in the 1945–1949, 1950–1954 and 1955–1959 is just one of the contrasts of Tarone and Chu[31] and is valid.

This analysis cannot say whether breast cancer incidence is increasing or not without some extra assumptions such as those used to construct Figure 2. There, it appears as if breast cancer incidence is decreasing in the most recent cohorts but the standard errors of the estimates are very large and so there is little confidence in the estimated effects. The period effects are certainly of a much smaller magnitude than the cohort ones and this is consistent with the trends in breast cancer incidence being driven by cohort effects rather than period ones.

REFERENCES

1. Waterhouse, J. A. H., Muir, C. S., Shanmugaratnam, K. and Powell, J. *Cancer Incidence in Five Continents*, Vol IV, IARC, Lyon, 1982.
2. Coleman, M., Esteve, J., Damiecki, P., Arslan, A. and Renard, H. *Trends in Cancer Incidence and Mortality*, IARC, Lyon, 1993.
3. Mason, W. M. and Fienberg, S. E. (eds). *Cohort Analysis in Sociological Research*, Springer-Verlag, New York, 1985.
4. Kupper, L. L., Janis, J. M., Salama, I. A., Yoshizawa, C. N. and Greenberg, B. G. 'Age–period–cohort analysis: an illustration of the problems in assessing interaction in one observation per cell data', *Communications in Statistics – Theory and Methods*, **12**, (23), 2279–2807 (1981).
5. Kupper, L. L., Janis, J. M., Karmous, A. and Greenberg, B. G. 'Statistical age–period–cohort analysis: a review and critique', *Journal of Chronic Diseases*, **38**, 811–830 (1985).
6. Clayton, D. and Schifflers, E. 'Models for temporal variation in cancer rates II: Age–period–cohort models', *Statistics in Medicine*, **6**, 469–481 (1987).
7. Holford, T. R. 'Analysing the temporal effects of age, period and cohort', *Statistical Methods in Medical Research*, **1**, 317–337 (1992).
8. Brillinger, D. F. 'The natural variability of vital rates and associated statistics', *Biometrics*, **42**, 693–734 (1986).
9. Francis, F. F., Green, M. and Payne, R. D. (ed.). *GLIM 4 The Statistical System for Generalized Linear Interactive Modelling*, Clarendon Press, Oxford, 1993.
10. Robertson, C. and Boyle, P. 'Age, period and cohort models: The use of individual records', *Statistics in Medicine*, **5**, 527–538 (1986).
11. Tarone, R. E. and Chu, K. C. 'Implications of birth cohort patterns in interpreting trends in breast cancer rates', *Journal of the National Cancer Institute*, **84**, 1402–1410 (1992).
12. Fienberg, S. E. and Mason, W. M. 'Identification and estimation of age–period–cohort models in the analysis of discrete archival data', *in* Schuessler, K. F. (ed.), *Sociological Methodology*, Josey-Bass, San Francisco, 1978, pp. 1–67.
13. Osmond, C. and Gardner, M. J. 'Age, period and cohort models applied to cancer mortality', *Statistics in Medicine*, **1**, 245–259 (1982).
14. Holford, T. R. 'The estimation of age, period and cohort effects for vital rates', *Biometrics*, **39**, 311–324 (1983).
15. Armitage, P. and Doll, R. 'The age distribution of cancer and a multi stage theory of carcinogenesis', *British Journal of Cancer*, **8**, 1–12 (1954).
16. Brown, C. C. and Kessler, L. G. 'Projections of lung cancer mortality in the United States: 1985–2025', *Journal of the National Cancer Institute*, **80**, 43–51 (1988).
17. Decarli, A. and La Vecchia, C. 'Age, period and cohort models: A review of knowledge and implementation in GLIM', *Rivisita di Statistica Applicata*, **20**, 397–410 (1987).
18. Efron, B. and Tibshirani, R. A. *Bootstrap Methods*, Chapman and Hall, London, 1994.
19. Tango, T. 'Letter to the editor, re Statistical modelling of lung cancer and laryngeal cancer in Scotland, 1960–1979', *American Journal of Epidemiology*, **127**, 677–678 (1988).
20. Firth, D. 'Bias reduction, the Jeffrey's prior and GLIM', *in* Fahrmeir, L., Francis, B., Glichrist, R. and Tutz, G. (eds), *Advances in GLIM and Statistical Modelling*, Springer-Verlag, New York, 1992.
21. Breslow, N. 'Extra Poisson variation in log linear models', *Applied Statistics*, **39**, 38–44 (1984).
22. Osmond, C. and Gardner, M. J. 'Letter to the editor, re Statistical modelling of lung cancer and laryngeal cancer in Scotland, 1960–1979', *American Journal of Epidemiology*, **129**, 31–35 (1988).
23. Robertson, C. and Boyle, P. 'Age, period and cohort models: A comparison of methods', *in* Fahrmeir, L., Francis, B., Glichrist, R. and Tutz, G. (eds), *Advances in GLIM and Statistical Modelling*, Springer Verlag, New York, 1992.
24. Robertson, C. and Boyle, P. 'A review of age, period and cohort models in chronic diseases', World Health Organisation Conference on Modelling Chronic Diseases, Heidelberg, 1994.
25. Boyle, P. and Robertson, C. 'Letter to the editor, re Statistical modelling of lung cancer and laryngeal cancer in Scotland, 1960–1979', *American Journal of Epidemiology*, **129**, 225–226 (1988).
26. Lee, W. C. and Lin, R. C. 'Autoregressive age period cohort models', *Statistics in Medicine*, **15**, 273–281 (1996).

27. Berzuini, C. and Clayton, D. 'Bayesian analysis of survival on multiple time scales', *Statistics in Medicine*, **13**, 823–838 (1994).
28. Payne, R. D. (ed.). *Genstat 5 Reference Manual*, Clarendon Press, Oxford, 1987.
29. Wickramaratne, P. J., Weissman, M. M., Leaf, P. J. and Holford, T. J. 'Age period and cohort effects on the risk of major depression: Results from 5 United States communities', *Journal of Clinical Epidemiology*, **42**, 333–343 (1989).
30. Tango, T. and Kurashina, S. 'Age, period and cohort analysis of trends in mortality from major disease in Japan, 1955 to 1979: peculiarity of the cohort born in the early Showa Era', *Statistics in Medicine*, **6**, 709–726 (1987).
31. Tarone, R. E. and Chu, K. C. 'Evaluation of birth cohort patterns in population disease rates', *American Journal of Epidemiology*, **143**, 85–91 (1996).
32. Akima, H. 'A method of bivariate interpolation and smooth surface fitting for irregularly distributed data points', *ACM Transactions on Mathematical Software*, **4**, 148–164 (1978).
33. Shyrock, H. S., Siegel, J. S. and Stockwell, E. G. *The Methods and Materials of Demography*, (*Condensed Edition*), Academic Press, New York, 1976.
34. Holford, T. R. 'An alternative approach to statistical age–period–cohort analysis', *Journal of Chronic Diseases*, **38**, 831–836 (1985).
35. Belsley, D. A. *Conditioning Diagnostics*: *Collinearity and Weak Data in Regression*, Wiley, New York, 1991.
36. Besley, D. A., Kuh, E. and Welsch, R. E. *Regression Diagnostics*: *Identifying Influential Data and Sources of Collinearity*, Wiley, New York, 1980.
37. James, I. R. and Segal, M. R. 'On a method of mortality analysis incorporating age-year interaction, with application to prostrate cancer mortality', *Biometrics*, **39**, 311–324 (1982).
38. Chen, Y. T., Zheng, T., Holford, T. R., Berwick, M. and Dubrow, R. 'Malignant melanoma incidence in Connecticut (United States): time trends and age period cohort modelling by anatomical state', *Cancer Causes and Control*, **5**, 341–350 (1994).
39. Zheng, T., Holford, T. R., Boyle, P., Chen, Y., Ward, B. A., Flannery, J. and Mayne, S. T. 'Time trend and the age period cohort effect on the incidence of histological types of lung cancer in Connecticut, 1960–1989', *Cancer*, **74**, 1556–1567 (1994).
40. Lee, L. T., Lee, W. C., Lin, R. S., Chen, S. C., Chen, C. Y., Luh, K. T. and Chen, C. J. 'Age period cohort analysis of lung cancer mortality in Taiwan 1966–1990', *Anticancer Research*, **14**, 673–676 (1994).
41. Xu, X., Laird, N., Dockery, D. W., Schouten, J. P., Rijcken, B. and Weiss, S. T. 'Age period and cohort effects on pulmonary function', *American Journal of Epidemiology*, **141**, 554–586 (1995).
42. Wilcox, A. J., Skaerven, R. and Irgens, L. M. 'Harsh social conditions and perinatal survival: an age period cohort analysis of the World War II occupation of Norway', *American Journal of Public Health*, **84**, 1463–1467 (1994).
43. Dubrow, R., Johansen, C., Skov, T. and Holford, T. R. 'Age period cohort modelling of large bowel cancer incidence by anatomical subsite and sex in Denmark', *International Journal of Cancer*, **58**, 324–329 (1994).
44. Clemmesen, J. *Statistical Studies on Malignant Disease*, Munksgaard, Copenhagen, 1994.

