The Sciences 1 (2): 48-51
March-April, 2001

Age-Period-Cohort Modelling of the Incidence of Breast and Cervical Cancer In New South Wales

[1]Halimah Awang, [2]Don McNeil, [2]Marilyn Chilvers,
[2]Jennifer Middeldorp, [3]Richard Taylor

The incidence of breast and cervical cancer in New South Wales women, diagnosed between 1972 and 1991, is modelled using an age-period-cohort poisson regression model. The age-period-cohort models are fitted to the data without the age constraint, using 5-year age and cohort effects and annual period effects. These models show significant age, period and cohort effects for breast cancer and significant age and cohort effects for cervical cancer. Both models confirm expected aetiological patterns. A sharply increasing age effect is exhibited until age 45 with a change in slope thereafter, the effect at age group of 80-84 is approximately 11% higher than at age group of 60-64.

Key words: Poisson regression, age-period-cohort models, cancer incidence

Halimah Awang
Department of Applied Statistics, Faculty of Economics & Administration, University of Malaya, 50603 Kuala Lumpur, Malaysia

[1]Department of Applied Statistics, Faculty of Economics and Administration, University of Malaya, 50603 Kuala Lumpur, Malaysia
[2]Department of Statistics, Macquarie University, NSW 2109, Australia, [3]Department of Public Health, University of Sydney, NSW 2006, Australia

Introduction

Cancer incidence is defined as the number of new cases of cancer reported in a specific population over a particular time period. In this study the population comprises women in the state of New South Wales (NSW), Australia. Data were obtained from the NSW Central Cancer Registry which was established in 1971 as a population-based registry, with data collection commencing in January, 1972. Cases notified to the Registry are classified according to the International Classification of Disease 9th Revision (ICD-9), and are accepted by IARC (International Agency for Research in Cancer) for publication in Cancer Incidence in Five Continents (Parkin *et al.*, 1992).

Breast cancer is the most frequent cancer among women worldwide, followed by cervical cancer. An increase in the incidence of breast cancer over the last century has commonly been reported and is greatest in affluent societies. An increased risk of breast cancer is associated with a positive family history, early menarche and late menopause, while early age at first pregnancy has a protective effect (Anderson *et al.*, 1990).

In recent years, it has been noted that the over all incidence rate of cervical cancer has fallen as a result of the introduction of screening in the early nineteen sixties. However, the incidence rate has risen among young people, while for older women there has been no change. In NSW it has fallen from the fourth most common cancer in the population in 1973 to eighth in 1991 (NSW Central Cancer Registry, 1994). The most important risk factors for cervical cancer appear to be early age at first intercourse and multiple sex partners (Kjaer *et al.*, 1989). The over all trend has been an increase in the incidence rate up to about 40 years with an approximately constant rate thereafter.

Materials and Methods

Modelling cancer incidence: The model demonstrating the incidence of breast cancer in NSW should reflect similar trends observed in other countries. An increasing age effect and perhaps some period effects reflecting changes in community activities such as education and large scale mammographic screening projects would be expected. In particular, we might expect to detect a significant period effect in 1991 when mammographic screening was introduced on a large scale in NSW. Similarly, one might expect a model of cervical cancer to demonstrate a decreasing incidence, coinciding with the introduction of the screening program in the early 1960s.

In looking at trends of the incidence of cancer over time, therefore, three factors need to be considered. These are the age at which the subject is diagnosed with cancer (age effect), the year of diagnosis (period effect) and the year of birth of the subject (cohort effect). The problem with including all three factors in a model to describe the incidence of cancer has been that of non-identifiability where identical descriptions of data may be obtained from different sets of parameter values leading to different, sometimes contradictory, interpretations (Clayton and Schifflers, 1987).

It is clear that any model describing the incidence of cancer must include the effect of age, so the problem remains of separating the period and cohort effects to address the non-identifiability problem. Clayton and Schifflers (1987) found that the age-period or the age-cohort models fit some cancer data well. For example, they found that either model fits data describing mortality rates for lung cancer in Belgium equally well. They also fit an age-period-cohort model to age specific mortality rates of breast cancer in Japan.

The present paper aims to model the incidence of both breast and cervical cancer in NSW, Australia using the age-period-cohort model, and to address the non-identifiability problem associated with it.

Poisson regression model of cancer incidence: Poisson regression is used to model incidence density data when the probability that an individual subject experiencing an outcome is small and the number of subjects at risk are large. For a given anatomic site we assume that the numbers of incident cases have independent Poisson distributions with means

$$\lambda_{ijk} = N_{ij} \exp (\alpha_i + \beta_j + \gamma_k)$$

where N_{ij} is the population at risk in age group i during period j and α_i, β_j, and γ_k are age, period and cohort effects, respectively.

As a full age-period-cohort model with no constraints cannot be fitted due to the linear dependence among the indices i, j and k in the model above an additional constraint is required to obtain a specific set of parameters (Holford, 1983; 1991). That is, the same data can lead to infinitely many different models, all with the same accuracy of fit, and with the same fitted values. The problem emerges when an interpretation of the meaning of the model co- efficients is required, especially as in some cases even the direction of the co-efficient can differ. (Holford 1983; 1991) parameterised the inter-relationship between three linear parameters for age, period and cohort (β_a, β_p and β_c respectively) with v, which is an unknown indeterminate constant. In this paper we propose a similar method of addressing the non-identifiability problem, which will restrict the possible model co-efficients using "credibility" intervals based on a process similar to sensitivity analysis.

Range of models: For both sets of cancer incidence data, the model with no additional constraints (other than equating the corner cohorts) is compared with models where one fewer constraint was placed on the age parameters,

for breast cancer $\beta_{a(80\text{-}84)} = \rho * \beta_{a(60\text{-}64)}$
for cervical cancer $\beta_{a(80\text{-}84)} = \rho * \beta_{a(50\text{-}54)}$
where ρ is a constant

The particular age groups selected for breast and cervical cancer were chosen based on observed trends with respect to age (NSW Central Cancer Registry, 1994) The range of models using the techniques described for each of the cancers should be essentially the same regardless of which age groups are selected to include in the equation with ρ. The constrained model has one more degree of freedom for error than the unconstrained model, so the change in deviance may be compared with chi-squared distribution with one degree of freedom.

Investigate models with a range of values of ρ, with the aim of determining a range of models that are statistically acceptable. Models with residual deviance and p-values greater than 0.05 will then give a credible range for co-

efficient values. Note that the value of ρ which gives the same deviance as the model with no additional age constraints will have p-value 1, since the change in deviance will be zero. This value of ρ may be estimated from the data being modelled, for example,
for breast cancer

$$\beta_a(80\text{-}84) = \rho * \beta_{a(60\text{-}64)}$$

so

$$\rho = \beta_{a(80\text{-}84)} / \beta_{a(60\text{-}64)}$$

where the β_a's are estimated from the model with no additional constraints (other than equating the corner cohorts). This value of ρ then becomes the 'centre' of the estimated credibility interval of models which do not differ significantly from one with no additional constraint.
The value of ρ itself is not of particular interest, as this will depend on which age groups were selected to be parameterised in this fashion. However, the ranges on the β_a, β_p and β_c, resulting from the extremes of the credibility interval should not depend on the selection of the age groups. Hence a 'credibility interval' can be obtained for these estimates, which are the ones of most interest. If the credibility intervals are small and do not contain zero, a meaningful interpretation of the co-efficient value can be made.

Results

Poisson regression results: The analysis is based on annual incidences in the 20 years from 1972 to 1991 and the assumption that the first four cohort effects are equal and last four. Since very few cases of either disease were diagnosed before age 25 years, only the age groups commencing with 25-29 years and ending with 80-84 years were used. The age-period-cohort model is fitted to the data using the 5 year age group effects, annual period effects and cohort effect centred at 5 year epochs covering the year of birth from 1892 to 1962. The three incompletely observed groups of cohorts centred at 1892, 1897 and 1902 are given in the same parameter estimates as the one centred at 1907, and similarly, those centered at 1952, 1957 and 1962, are given in the same value as that centred at 1947.
Four models are fitted to the data: the Age model, the Age-Period model, the Age-Cohort model and finally the full Age-Period-Cohort (APC) model. There are 12 age groups, 20 periods, and nine cohorts. The results are summarised in Table 1. The first four rows in the table compare the stated model with fully parameterised model, while the tests for period and cohort effects refer to the significance of these effects within the APC model.

Table 1: Summary of Model Deviances

		Breast Cancer		Cervical Cancer	
Model	DF	Deviance	P-value	Deviance	P-value
Age only	228	562.12	0.000	291.78	0.003
Age-cohort	220	281.52	0.003	220.23	0.483
Age-Period	209	194.14	0.762	223.66	0.232
Age-Per-Cohort	201	176.52	0.893	198.13	0.544
Period effects	19	105.00	0.000	22.1	0.279
Cohort effects	8	17.62	0.024	25.53	0.001

Reference groups: Age 25-29, Period 1981, Cohort 1922

The results in Table 1 were obtained by fitting the data to Poisson regression models using the SPIDA statistical package (Statistical Package for Interactive Data Analysis) Lunn *et al.*, 1992. The full model provides a good fit for breast cancer and both period and the cohort effects are statistically significant after adjusting for each other (p-value = 0 and 0.024 respectively). For cervical cancer, the full model provides a good fit, but the period effects are not statistically significant (p-value = 0.279) after adjusting for the cohort effects.

Model range calculation: For both cancers we then obtained the value of ρ and fitted the constrained age-period-cohort model to obtain a range of acceptable models (those with a non-significant change in deviance).
The fitted age effects and 95% confidence intervals, based on the age-period-cohort model with no constraint for breast or cervix, are plotted in Fig. 1 (the middle curve for each cancer) and shows that the sharp increase in the age effect for breast cancer (upper curve) up to age group 45-49 years with a steady increase from thereon, whereas the age effects for cervical cancer are approximately constant

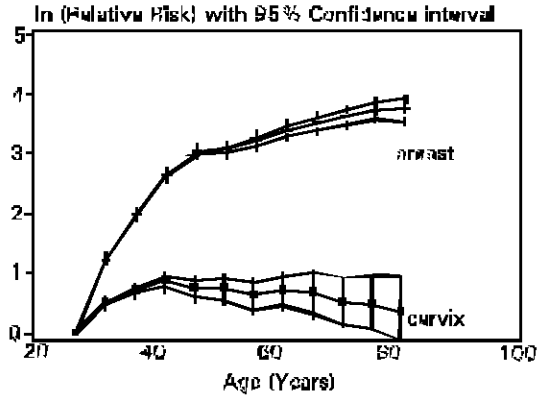

Fig. 1: Breast and Cervical Cancer in NSW: 1972-1991. Age Effect Controlling for Period and Cohort

after age group 40-45 years. The upper and lower curves indicate fitted age effects at both extreme points of acceptable ρ-values using the constrained age-period-cohort model.
Fig. 2a and 3a graph the period effects and Fig 2b and 3b graph the cohort effects, also based on the full model. For cervical cancer the significance of the period effects is due to a lower incidence in 1979 compared to the referent group 1981, while for breast cancer, there is a distinction between 1991 and the referent year 1981. The cohort effects for the two cancers have different trends, with breast cancer showing an almost perfect linearly increasing trend and cervical cancer showing an irregular trend. For cervical cancer, the significance of cohort effects is contributed by cohorts 1927, 1932, 1937 and 1942.

Discussion

The results of our analysis clearly shows that the need for incorporating all three factors when modelling cancer incidences and that an acceptable range of models can be obtained should additional constraints be placed on either age, period or cohort. It is desirable that such constraints be made on some biological or aetiological grounds. For both breast and cervical cancer, that the age effects confirm previously reported trends. The significance of year 1991 for breast cancer indicates that the significant period

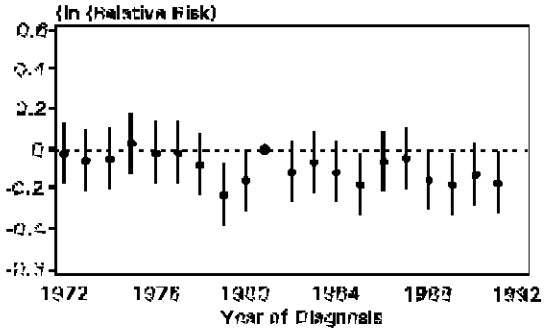

Fig. 2a: Cervical Cancer in NSW: 1972-1991 Period Effect Controlling for Age and Cohort

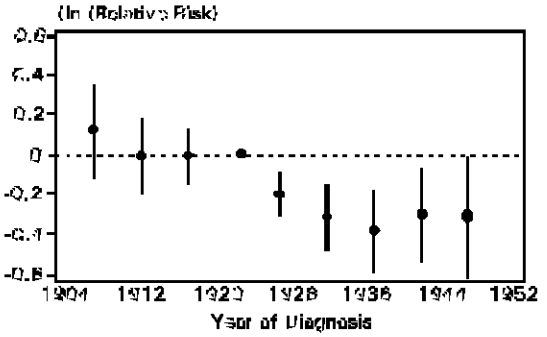

Fig. 2b: Cervical Cancer in NSW: 1972-1991 Cohort Effect Controlling for Age and period

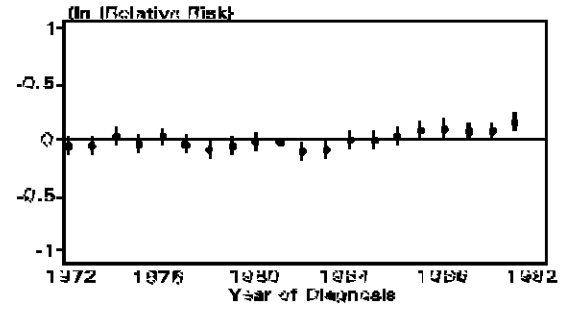

Fig. 3a: Breast Cancer in NSW: 1972-1991. Period Effect Controlling for Age and Cohort

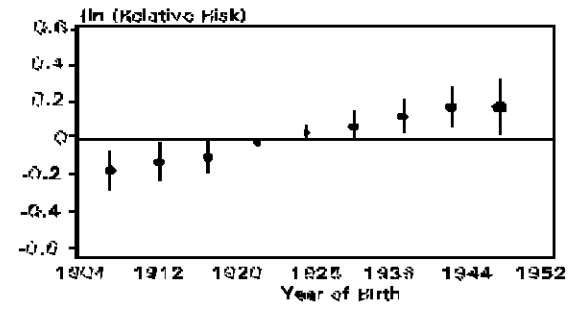

Fig. 3b: Breast Cancer in NSW: 1972-1991. Cohort Effect Controlling for Age and period

effect may have been due to the large scale mamographic screening in that year.

It was expected that a model demonstrating the incidence of breast cancer in NSW should reflect similar trends noticed in other developed countries. Such trends include an increasing age effects, especially to age 45 years, a strong increasing birth cohort effect, and perhaps some period effects reflecting changes in community activities such as education and large scale mammographic screening projects.In fact, the proposed model for breast cancer satisfies these expectations.

Fig. 3a shows that there is apronounced period effect for breast cancer incidence in 1991 and Fig. 3b shows an increasing cohort effect with perhaps some levelling off in the later cohort. The age effects demonstrated by our final model show a sharply increasing trend until about age 45 years, with a change in slope thereafter and specifically the effect at age group 80-84 about 11% higher than at age group 60-64 years. Hence, the age period cohort model with a set of realistic constraints has a set of identifiable parameters and fits the data most satisfactorily.

References

Anderson, B. and E. Smith, 1990, Epidemiology of Breast Cancer In: Mitchell GW, Bassett LW, eds. The Female Breast and its Disorders. Maryland: Williams and Wilkins, 218-226.

Clayton, D.G. and E. Schifflers, 1987, Models for temporal variation in cancer rates. I: Age-period-cohort models, Statistics in Medicine, 6: 449-467.

Clayton, D.G. and E. Schifflers, 1987. Models for temporal variation in cancer rates. II: Age-period and age-cohort models, Statistics in Medicine, 6: 469-481.

Holford, T.R., 1983, The estimation of age, period and cohort effect for vital rates, Biometrics, 39: 311-324.

Holford, T.R., 1991. Understanding the effects of age, period and cohort on incidence and mortality rates. Annual Review of Public Health, 12: 425-437.

Kjaer, S.K., C. Teisen and B.J. Haugaard, 1989. Risk factors for cervical cancer in Greenland and Denmark: a population-based cross-sectional study, Int. J. Cancer, 44: 40-47.

Lunn, A.D.,V. Gebski, O. Leung and D. McNeil,1992. SPIDA User Manual, version 6 Statistical Computing Laboratory. Macquarie University.

NSW Central Cancer Registry,1994. Cancer in New South Wales: Incidence and Mortality, 1991, 44-47.

Parkin, D.M., C.S. Muir, S.L. Whelan, Y.T. Gao, J. Ferlay and J. Powell (eds). 1992. Cancer Incidence in Five Continents, Volume 17, IARC Scientific Publication No. 120.