Modelling the Age-Period-Cohort Trend Surface

WEN C. LEE and RUEY S. LIN
Graduate Institute of Epidemiology,
College of Public Health,
National Taiwan University

Summary

Age-period-cohort analysis of incidence and/or mortality data has received much attention in the literature. Though several authors seem to offer solutions to the non-identifiability problem inherent in this kind of analysis, the constraints imposed lack sound biological bases and the interpretability of the parameters is in doubt. By introducing the axes of 'adjusted' age, period, and cohort variables in the age-period-cohort trend surface, the authors find that separate effects from these variables can be estimated and can be clearly interpreted. The constraint behind the procedure is also found to be mathematically simple and elegant. Since the method still lacks a biological foundation, it is best viewed as an adjunct to the graphical trend-surface analysis rather than as a solution to the non-identifiability problem. Previously published mortality data of prostate cancer is used to illustrate the methodology.

Keywords: Age-period-cohort models, Trend surface analysis, Epidemiologic methods, Statistics, Prostate cancer, Mortality

1. Introduction

Birth cohort analysis for disease incidence and mortality has a long history in epidemiology (FROST, 1939). In its original form, the technique adopted a graphical approach and the contributions of various time factors were determined visually. In the past decade, the statistical age-period-cohort (APC) analysis, which attempts to quantify the separate effects due to age, period, and cohort variables by introducing a three-factor, analysis of variance-type model, has received much attention in the literature (KUPPER, 1985; HOLFORD, 1991). Although the technique has been applied to the study of secular trends of various diseases, it has been well known that the exact linear relation of the three factors (cohort = period - age) creates a non-identifiability problem. For determining an identifiable set of parameters, one has to add one or more constraints or assumptions into the APC model. For example, OSMOND and GARDNER (1982) introduced the constraint of minimal weighted distances between the APC model and the three 2-factor models (AP, AC, and PC). Whereas ROBERTSON and BOYLE (1986) assumed a constant effect within each age, period, and cohort cross-classified cell. It should be noted that, without adding extra constraints, the unidentifiable APC model can still be interpreted in terms of the "curvature components"
(Holford, 1983) and/or of the "drift parameter" (Clayton and Schifflers, 1987). However, in instances when knowing the separate trends of age, period, as well as cohort variables is important for understanding the etiology of the disease under concern, imposing extra constraints or assumptions seems unavoidable.

In this paper, we first show that by introducing the proper axes of the 'adjusted' age, period, and cohort variables in the APC trend surface (Weinkam and Sterling, 1991; Jolley and Giles, 1992), separate effects from these variables can be estimated and clearly interpreted. The particular constraints behind such a procedure are then discussed. We use data on prostate cancer mortality among nonwhites in the U.S. from 1935 to 1969, which was previously analyzed by Holford (1983), to illustrate this methodology.

2. Methods

We begin with a brief review of the graphical APC trend surface analysis. To graph the age-period-cohort trend surface, it is sufficient to choose any two of the three indices as the coordinates with respect to which specific disease rates can be plotted. This is because specifying any two of the indices determines the third. The same is true for modelling purposes. However, we found that the coordinate system introduced by Weinkam and Sterling (1991) is appealing since any age, period, or cohort effects, if they exist, are equally represented and distortion can be avoided. This property is desirable particularly when we have no a priori information about the relative importance of the three variables. Figure 1 depicts this coordinate system. Note that we have made slight modifications to the original coordinate system for ease of presentation. Lines of constant age are horizontal and increase from the top of the figure to the bottom. Lines of constant period slope downward at the same angle toward the right. These two indices increase from the left side to the right. Two additional axes are introduced, the horizontal axis (labelled 'X'), and the vertical one (labelled 'Z'). Let the unit of measurement on all these axes be one year. Then the coordinates of \((a, p, c)\) in the age-period-cohort system are \((p - a/2, \sqrt{3} \cdot a/2)\) in the \(X-Z\) system. With the coordinate system determined, the relation between the disease rates and the three time variables can be presented in a three dimensional trend surface or in a two dimensional contour plot. Therefore, any non-trivial age, period, or cohort effects are easily recognized from particular types of features in the graph.

To present separate age, period, and cohort linear trends (slopes) in the trend surface, a direction each time must be chosen along which one of the variables will change while the other two are held constant. The change in disease rates per unit change in the particular direction can thus be interpreted as an 'adjusted' linear trend effect for this variable of concern. Evidently, the exact relation of 'cohort = period - age' forbids us from doing this. However, after inspecting Figure 1, it appears sensible to choose the directions labelled 'A', 'P', and 'C' for the evaluation of the linear trend effects of age,
period, and cohort indices, respectively. Further examining direction A, we find that for every unit increment in this direction, age will increase $2/\sqrt{3}$ years, period increases $1/\sqrt{3}$ years, and cohort decreases $1/\sqrt{3}$ years. The changes in the period and cohort variables are the same and both are half the change of the age variable. Appendix 1 shows that this is also the direction which represents the maximal square age increment relative to the sum of the squared period and cohort increments. Therefore, with respect to the requirement that period and cohort variables should be held constant, direction A can be considered as a compromised direction which is not biased toward the period or the cohort variable. Appendix 2 further clarifies that for a one unit increment in this direction, the "adjusted" age variable increases $\sqrt{3}$ units rather than $2/\sqrt{3}$ units. With the direction and the scale being determined, the axis of the adjusted age is thus defined. The axes of the adjusted period and cohort variables can be similarly determined.

It then becomes possible to express the linear trend effects of the adjusted age, period, and cohort variables (denoted $\alpha_t$, $\beta_t$, and $\gamma_t$, respectively) in terms of the linear trend effects of the X and Z variables (denoted $X_t$ and $Z_t$) by the following transformation equations (see appendix 2):

Fig. 1: The coordinate system of Weinkam and Sterling; for details, see text.
Thus we see that if $X_t$ and $Z_t$ can be uniquely determined, so can $\alpha_t$, $\beta_t$ and $\gamma_t$. A unique determination of $X_t$ and $Z_t$ is possible since they are the linear trend effects of the $X$ and $Z$ variables in the APC trend surface, and it is clear that these two variables do not have a linear relation such as that which exists among the age, period, and cohort variables. The detailed estimation procedure is outlined in appendix 3.

Our 'adjusted' trends are, in fact, linear combinations of the true unknown linear trends. If we denote the true linear trends of age, period, and cohort as $\alpha_L$, $\beta_L$, $\gamma_L$, then the adjusted trends can be expressed as $\alpha_t = 2\alpha_L/3 + \beta_L/3 - \gamma_L/3$, $\beta_t = 2\beta_L/3 + +\alpha_L/3 + \gamma_L/3$, and $\gamma_t = 2\gamma_L/3 - \alpha_L/3 + \beta_L/3$. Since Holford (1983) has proved that $d\alpha_t + d\beta_t + (d_2 - d_1)\gamma_t$ with arbitrary $d_1$ and $d_2$ can be estimated, the estimability of the adjusted trends is again shown to be possible. From the above equations or from the transformation equations, it can be seen that $\alpha_t - \beta_t + \gamma_t = 0$. This relation is the constraint we imposed on the parameters. To understand what the constraint means, one can consider a three-dimensional space with rectangular axes of age, period, and cohort. The exact linear relation of the three variables implies that the available data of disease rates will lie on the plane defined by age - period + cohort = 0. It then becomes clear that the constraint we used amounts to an additional requirement that the gradient vector of the disease rates lie on the same plane.

3. Example

To illustrate the calculations, we consider data on prostate cancer mortality among nonwhites in the U.S. from 1935 to 1969, taken from table 2 in Holford’s paper (1983). Using the coordinate system of Weinkam and Sterling (1991), isopleths of logarithm of the mortality rates (level curves) are constructed (Figure 2) to visualize the age-period-cohort trend surface. The contouring algorithm is done by procedures G3GRID and GCONTOUR of SAS/GRAPH package (SAS Institute Inc., 1988a) which are based on a bivariate spline interpolation (Meinguet, 1979). It can be seen from Figure 2 that for cohorts born before 1870 the contour lines appear to be inconsistent with the rest of the graph. This is possibly due to sampling variability, since between 1935 and 1950 only data for ages which were older than 70 are available. Otherwise, the level curves run almost in parallel curvilinear lines. The spacings between them are also about equal. It should also be noted that there is an abrupt change in directions at the line of the 1890 birth cohort (dotted line in Figure 2), and that the level curves become almost horizontal afterwards.

A log-linear model for the mortality rates with a Poisson error distribution is employed (Frome, 1983). Using the method described above, we obtain the same fit of

\[
\alpha_t = \sqrt{3}/3 \ Z_t \\
\beta_t = X_t/2 + \sqrt{3}/6 \ Z_t \\
\gamma_t = X_t/2 - \sqrt{3}/6 \ Z_t
\]
Fig. 2: Age-period-cohort trend surface (contour plot) of prostate cancer morality among nonwhites in the U.S., 1935–1969.

The model as HOLFORD (1983) did, i.e. $G^2 = 98.91$ with 25 degrees of freedom. However, the linear trend effects of the $X$ and $Z$ variables as well as those of the adjusted age, period, and cohort can be estimated (Table 1). The linear trend effects of the $X$ and $Z$ variables are also plotted as an arrow in Figure 2. For graphing purposes, the arrow is not to scale. It can be seen that it points in the direction of the steepest increment in the age-period-cohort trend surface as shown by the level curves.

Figure 3 presents the birth cohort effects (with 95% confidence intervals) of the U.S. nonwhite population derived by our method. Between 1855 and 1890, the effects

<table>
<thead>
<tr>
<th>variables</th>
<th>estimates</th>
<th>standard errors</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X$</td>
<td>0.1151</td>
<td>0.00471</td>
</tr>
<tr>
<td>$Z$</td>
<td>0.6080</td>
<td>0.00473</td>
</tr>
<tr>
<td>Age #</td>
<td>0.3511</td>
<td>0.00273</td>
</tr>
<tr>
<td>Period #</td>
<td>0.2331</td>
<td>0.00266</td>
</tr>
<tr>
<td>Cohort #</td>
<td>-0.1180</td>
<td>0.00278</td>
</tr>
</tbody>
</table>

* for every five years in each variable
# the `adjusted` variables; see the text for explanations
due to birth cohort remain fairly constant, though slightly wider confidence intervals are observed for the pre-1870 birth cohort. A sharp downward turn emerges after the 1890 birth cohort, which is comparable to the previous analysis by level curves. However, one may argue that since the level curves in Figure 2 indicate essentially constant age-specific death rates for the post-1890 birth cohort, period or cohort effects should not be operating after then. This may in fact be the case. However, it is difficult to say with certainty which is the correct explanation because of the non-identification problem. In fact, the answer depends on the particular constraint(s) being used. Nevertheless, we believe that a zero period linear trend, which is also assumed in Holford's paper (1983), may not be a reasonable assumption in this example, since the improved accuracy of diagnosis and death certification in more recent periods, especially for the nonwhite population, should be taken into account. Furthermore, the cohort effects derived from our method are comparable with those observed in Australia (HOLMAN et al., 1981), England and Wales (HOLMAN et al., 1981; JAMES and SEGAL, 1982) as well as in the U.S. white population (HOLFORD, 1983). It has been suggested that male sexual frustration during the Great Depression years may be a possible explanatory factor for the cohort effects of this disease (HOLMAN et al., 1981). Consequently, it is not difficult to explain why people in different countries and/or of different races should display similar birth-cohort phenomena.

4. Discussion

![Graph showing cohort effects with 95% confidence intervals of prostate cancer mortality among nonwhites in the U.S., 1935-1969.](image)
Similar to the philosophy in the method of Osmond and Gardner (1982), which seeks to find the compromised solution between the AP, AC, and PC models, our methodology can be viewed as an adjunct to the graphical trend-surface analysis (Weinkam and Sterling, 1991; Jolley and Giles, 1992), which tries to minimize the biases of the three time variables in a geometrical sense. Though the constraint behind our procedure is mathematically simpler than that of Osmond and Gardner (1982), further studies using simulated and real-world datasets should be undertaken to compare the performance of bias-reduction of these two methodologies. It should be noted that since neither method is biologically based, the true linear trends of age, period, and cohort indices in some instances can become quite different from the estimates using either one. Recently, Holcomb et al. (1994) incorporated biologic information regarding functional forms of age effects as derived from multistage carcinogenesis models for cancers into the APC analysis. It was shown to be possible to determine an identifiable set of parameters this way. Yet the parameter estimates were highly unstable and among the five different functional forms for age they tried, four resulted in biologically implausible estimates. Therefore there is no definite evidence, at present, to support that the use of "biologic" constraints is superior to the use of non-biologic ones.

It is advised that, if one of the three time variables is known a priori not to operate on the disease rates, our methodology should not be employed. This is because, in this circumstance, identifiability does not pose a problem and one can, in the usual ways, allow one variable to change in the trend surface while holding the others constant. Likewise, there is no need to plot the age-period-cohort trend surface according to the coordinate system of Weinkam and Sterling (1991).

To implement the analysis, the authors have developed a program written in SAS/IML language (SAS Institute Inc, 1988b) which is available by request. However, readers may also find out that using a simple program to modify the design matrix (appendix 3) and allowing it to be analyzed by commercialized packages such as GLIM (Baker and Nelder, 1978) or EGRET (Statistics and Epidemiology Research Corporation, 1988) is more convenient. Otherwise, one can also tailor the constraint, $\alpha_t - \beta_t + \gamma_t = 0$ to his/her need, if a particular solution has been found (Holcomb, 1991) or if a constrained optimization algorithm is to be employed.

Appendix 1

Let $\theta$ be the angle from the $Z$ axis to the $A$ direction measured in a counterclockwise fashion. Then for a unit increment in this direction, age, period, and cohort will increase $(2 \cos \theta / \sqrt{3})$, $(\sin \theta + \cos \theta / \sqrt{3})$, and $(\sin \theta - \cos \theta / \sqrt{3})$ years, respectively. Therefore, the squared age increment divided by the sum of the squared period and cohort increments is:
\[ R = \frac{(2 \cos \theta / \sqrt{3})^2}{((\sin \theta + \cos \theta / \sqrt{3})^2 + (\sin \theta - \cos \theta / \sqrt{3})^2)} \]
\[ = 2 \cos^2 \theta / (3 - 2 \cos^2 \theta) \]

Using elementary calculus, readers can easily verify that \( R \) attains its maximum when \( \theta = 0 \) (or \( 0 = \pi \)). Also when \( \theta = 0 \), the following index achieves its maximum:

\[ D = \frac{(2 \cos \theta / \sqrt{3})^2}{((\sin \theta + \cos \theta / \sqrt{3})^2 + (\sin \theta - \cos \theta / \sqrt{3})^2)} - \]
\[ = 8 \cos^2 \theta / 3 - 2 \]

Appendix 2

Let the linear trend effects of \( X, Z \), and the adjusted variables be denoted according to the text. Then the gradient vector of the age-period-cohort trend plane is
\[ \vec{g} = \vec{i}X + \vec{j}Z, \]
where \( \vec{i} \) and \( \vec{j} \) are the unit vectors of the \( X \) and \( Z \) axes. Assume that one unit in directions \( A, P, \) and \( C \) is equal to \( k \) years for the adjusted age, period, and cohort, respectively. Then the directional derivatives in the trend plane of the three adjusted variables are:

\[ \alpha_t = (1/k) \vec{j} \cdot \vec{g} = (1/k)Z_t \]
\[ \beta_t = (1/k)(\sqrt{3} \vec{i}/2 + \vec{j}/2) \cdot \vec{g} = (\sqrt{3}X_t/2 + Z_t/2)/k \]
\[ \gamma_t = (1/k)(\sqrt{3} \vec{i}/2 - \vec{j}/2) \cdot \vec{g} = (\sqrt{3}X_t/2 - Z_t/2)/k \]

To determine \( k \), we realize that for a certain \((\Delta a, \Delta p, \Delta c)\) increment in the adjusted variables, the (logarithms of) disease rates will increase:

\[ \Delta r = \alpha_t \Delta a + \beta_t \Delta p + \gamma_t \Delta c \]
\[ = Z_t \Delta a/k + (\sqrt{3}X_t/2 + Z_t/2) \Delta p/k + (\sqrt{3}X_t/2 - Z_t/2) \Delta c/k \]
\[ = (\sqrt{3}/k) (\Delta p/2 + \Delta c/2) X_t + (\sqrt{3}/k) (\sqrt{3}\Delta a/2) Z_t \]
\[ = (\sqrt{3}/k) (X_t \Delta x + Z_t \Delta z) \]

One can see that \( k \) must be equal to \( \sqrt{3} \). Substituted back into the above equations, we observe that the transformation equations described in the text are valid.
Appendix 3

To estimate the parameters associated with age, period, and cohort effect, the design matrix: \((1 \mathbf{A}, \mathbf{P}, \mathbf{C}, \mathbf{AL}, \mathbf{PL}, \mathbf{CL})\), parametrized by Holford (1983) to represent the curvature components and linear components of age, period, and cohort variables, is to be modified as: \((1 \mathbf{A}, \mathbf{P}, \mathbf{C}, \mathbf{T})\), with \(\mathbf{T} = \mathbf{L}\mathbf{Q}\), where \(\mathbf{L} = (\mathbf{AL}, \mathbf{PL}, \mathbf{CL})\), is the matrix of the linear components, and

\[
\mathbf{Q} = \begin{bmatrix}
0 & \sqrt{3}/3 \\
1/2 & \sqrt{3}/6 \\
1/2 & -\sqrt{3}/6
\end{bmatrix}
\]

corresponds to the transformation equations listed in the text. This modified design matrix has full column rank. Therefore, the parameters associated with it, namely the grand mean, the curvature effects of age, period, and cohort variables, as well as the linear trend effects of \(X\) and \(Z\) variables, are all estimable. The linear trend effects of the adjusted age, period, and cohort can thus be obtained by applying the transformation equations again. Appropriate confidence intervals of the parameters can also be constructed under the condition of asymptotic normality of the inverse of the information matrix.

References


Received August 1994
Accepted September 1994

WEN C. LEE and RUEY S. LIN
Graduate Institute of Epidemiology
College of Public Health
National Taiwan University
No. 1, Jen – Ai Rd., 1st Sec., Taipei, Taiwan, R.O.C.