Evaluation of Birth Cohort Patterns in Population Disease Rates

Robert E. Tarone and Kenneth C. Chu

Interpretation of trends in disease rates using conventional age-period-cohort analyses is made difficult by the lack of a unique set of parameters specifying any given model. Because of difficulties inherent in age-period-cohort models, neither the magnitude nor the direction of a linear trend in birth cohort effects or calendar period effects can be determined unambiguously. This leads to considerable uncertainty in making inferences regarding disease etiology based on birth cohort or calendar period trends. In this paper, the authors demonstrate that changes in the direction or magnitude of long term trends can be identified unequivocally in age-period-cohort analyses, and they provide parametric methods for evaluating such changes in trend within the usual Poisson regression framework. Such changes can have important implications for disease etiology. This is demonstrated in applications of the proposed methods to the investigation of birth cohort trends in female breast cancer mortality rates obtained from the National Center for Health Statistics for the United States (1970-1989) and from the World Health Organization for Japan (1955-1979).

Many investigations of incidence and mortality data involve analyses of temporal trends in rates by age and calendar year period (1, 2). Such analyses are termed “cross-sectional,” because they cut across birth cohorts at a given point in time (3). Interpretation of cross-sectional analyses can be misleading when longitudinal trends, such as trends with successive birth cohorts, influence the pattern of rates over time (4-12). Longitudinal trends can provide critical clues regarding disease etiology which would be missed by strictly cross-sectional analyses. For example, cross-sectional analyses of trends in US female breast cancer mortality rates indicate a decreasing calendar period trend in young women at the same time the calendar period trend in older women is increasing (1, 12), which has led to suggestions that there are different etiologies for pre- and postmenopausal breast cancer, differences between pre- and postmenopausal breast cancer patients in their response to chemotherapy, or different patterns of exposure to environmental carcinogens in young and old women (2, 13, 14). Analysis of birth cohort trends, on the other hand, demonstrated that the observed disparity between trends in young and old women results from a moderation of the birth cohort trend in risk, the timing of which suggests that decreasing breast cancer mortality in young women may be largely due to changes in childbearing patterns following World War II (12). This example illustrates that examination of longitudinal trends in rates can have a marked impact on inferences regarding disease etiology.

Conventional age-period-cohort analyses of variation in disease rates over time assume a log-linear relation of age, calendar period, and birth cohort effects and invoke Poisson maximum likelihood methods to estimate the corresponding parameters. In spite of the recognized importance of determining birth cohort patterns of risk, the identification of such patterns has proven difficult because of the inability to separate unequivocally linear trends in birth cohort effects from linear trends in calendar period effects (15-18). There are infinitely many sets of maximum likelihood estimates of age, calendar period, and birth cohort effects for any given data set (i.e., in statistical terms, the parameters are not identifiable). The magnitude, or even the direction, of the linear trend in birth cohort effects (or calendar period effects) cannot be determined with certainty (15, 18). For example, examination of three sets of maximum likelihood estimates resulting from age-period-cohort analyses of Japanese female breast cancer mortality rates indicates that birth cohort parameters are increasing with advancing birth cohort in the first set of estimates but
decreasing with advancing birth cohort in the third set of estimates (18).

Such uncertainty has led to an understandable reluctance on the part of researchers to make strong inferences about birth cohort trends in disease rates. In this paper, we demonstrate that, unlike the direction of a birth cohort trend in rates, a change in the magnitude or direction of the linear trend in birth cohorts (or calendar periods) can be evaluated unambiguously (i.e., such a change in magnitude is identifiable). Identifiable parameters representing local departures from linearity have been proposed (15, 18), but the etiologic interpretation of these parameters can be difficult. Although such second-order effects have traditionally been of less interest than first-order effects (e.g., the magnitude of a slope), identification of changes in the magnitude of long term trends can have important etiologic implications (12). In the current paper, identifiable parameters are derived to permit unequivocal inferences regarding such changes in the slope of birth cohort effects in different eras using standard age-period-cohort analyses, and the parametric method is applied to US white female breast cancer mortality rates to confirm the results of the previous nonparametric analysis (12). The method is also applied to Japanese female breast cancer mortality data for comparison of the patterns of birth cohort risk in US and Japanese women (18).

MATERIALS AND METHODS

Sources of mortality data

US white female breast cancer mortality data for the years 1970 through 1989 were obtained from the Division of Vital Statistics of the National Center for Health Statistics (Hyattsville, Maryland). The breast cancer mortality rates analyzed in the current study differ somewhat from those used in the previous study (12) in that 1989 mortality data are included and population sizes have been updated on the basis of the 1990 US Census. The current study examines trends from the period 1970–1989, as opposed to the 1969–1988 interval examined in the previous study (mortality data by single year of age are unavailable for all years prior to 1969).

Construction of 2-year mortality rates

In the past, 5-year age and calendar periods have been used routinely in age-period-cohort analyses, presumably because rates are usually reported by 5-year age interval. The birth cohort intervals corresponding to 5-year age and calendar periods, however, are 10 years in length, with considerable overlap between adjacent 10-year intervals. Because shorter birth cohort intervals allow better discrimination of birth cohort risk patterns, age-period-cohort analyses using age and calendar period lengths of less than 5 years should be undertaken whenever feasible. The annual numbers of US breast cancer deaths are available by single year of age. Annual population data are only obtainable by 5-year age intervals, but a catalogue of standard interpolation methods provides techniques with which to estimate 1-year population figures from the 5-year data (19). To estimate 1-year population sizes, we used osculatory interpolation, which is designed to produce interpolated results that have a high degree of smoothness. We used Beers' method based on six-term ordinary formulas, which minimizes the squares of the fifth-order differences and preserves the 5-year population figures (19).

The interpolated population sizes were used to create age-specific rates for 2-year age intervals. With 2-year age intervals and 2-year calendar periods, the corresponding birth cohort intervals are 4 years in length, with 75 percent of births occurring in the middle 2 years of the 4-year interval. In the following exposition, each birth cohort will be identified by the first of the middle 2 years; for example, "the 1905 birth cohort" will refer to the years 1904–1907, with 75 percent of births occurring in the years 1905 and 1906.

Model and identifiability considerations

Table 1 shows the breast cancer mortality rates of US white women by 2-year age interval from age 24 years to age 83 years and by 2-year calendar period from 1970 to 1989. In general, let $R_{ij}$ denote the disease rate for the $i$th of $A$ age intervals and the $j$th of $P$ calendar periods. In conventional age-period-cohort notation, the birth cohort would be indicated by an additional subscript, $c = A + j - i$, but to simplify notation we will suppress the birth cohort subscript. With $A$ age intervals and $P$ calendar periods, the number of birth cohorts is $C = A + P - 1$. In this notation, $c = 1$ corresponds to the earliest birth cohort and $c = C$ corresponds to the most recent birth cohort. In the usual age-period-cohort model, the mean, $E_{ij}$, of the logarithm of the disease rate, $R_{ij}$, is modeled as

$$E_{ij} = \alpha_i + \pi_j + \gamma_c,$$

where the $\alpha_i$'s are the age effects, the $\pi_j$'s are the calendar period effects, and the $\gamma_c$'s are the birth cohort effects.

As is shown in Appendix 1, the equality $c = A + j - i$ guarantees that there are infinitely many sets of

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>24–25</td>
<td>0.93</td>
<td>(0.5)</td>
<td>0.63</td>
<td>0.42</td>
<td>0.28</td>
<td>0.14</td>
<td>0.00</td>
<td>0.28</td>
<td>0.05</td>
<td>0.28</td>
<td>0.21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>26–27</td>
<td>1.31</td>
<td>(0.1)</td>
<td>1.15</td>
<td>1.30</td>
<td>0.89</td>
<td>0.68</td>
<td>0.50</td>
<td>0.56</td>
<td>0.56</td>
<td>0.59</td>
<td>0.67</td>
<td></td>
<td></td>
</tr>
<tr>
<td>28–29</td>
<td>2.25</td>
<td>(0.1)</td>
<td>2.86</td>
<td>2.56</td>
<td>2.40</td>
<td>1.62</td>
<td>1.60</td>
<td>1.56</td>
<td>1.14</td>
<td>1.10</td>
<td>1.05</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30–31</td>
<td>4.02</td>
<td>(0.1)</td>
<td>4.26</td>
<td>3.98</td>
<td>3.47</td>
<td>3.76</td>
<td>3.50</td>
<td>3.11</td>
<td>3.42</td>
<td>3.19</td>
<td>3.06</td>
<td></td>
<td></td>
</tr>
<tr>
<td>32–33</td>
<td>5.12</td>
<td>(0.1)</td>
<td>5.22</td>
<td>5.12</td>
<td>4.60</td>
<td>5.26</td>
<td>5.28</td>
<td>5.06</td>
<td>5.24</td>
<td>5.06</td>
<td>4.96</td>
<td></td>
<td></td>
</tr>
<tr>
<td>34–35</td>
<td>6.46</td>
<td>(0.1)</td>
<td>6.29</td>
<td>6.17</td>
<td>6.03</td>
<td>6.14</td>
<td>6.14</td>
<td>5.98</td>
<td>5.98</td>
<td>5.89</td>
<td>5.66</td>
<td></td>
<td></td>
</tr>
<tr>
<td>36–37</td>
<td>7.74</td>
<td>(0.1)</td>
<td>7.69</td>
<td>7.69</td>
<td>7.56</td>
<td>7.74</td>
<td>7.77</td>
<td>7.64</td>
<td>7.87</td>
<td>7.79</td>
<td>7.67</td>
<td></td>
<td></td>
</tr>
<tr>
<td>38–39</td>
<td>9.06</td>
<td>(0.1)</td>
<td>9.06</td>
<td>9.08</td>
<td>8.92</td>
<td>9.05</td>
<td>9.12</td>
<td>8.99</td>
<td>9.15</td>
<td>9.07</td>
<td>8.99</td>
<td></td>
<td></td>
</tr>
<tr>
<td>40–41</td>
<td>10.64</td>
<td>(0.1)</td>
<td>10.64</td>
<td>10.69</td>
<td>10.66</td>
<td>10.71</td>
<td>10.76</td>
<td>10.63</td>
<td>10.79</td>
<td>10.71</td>
<td>10.65</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Data were obtained from the National Center for Health Statistics (Hyattsville, Maryland).
† Breast cancer deaths per 100,000 women.
‡ Numbers in parentheses, no. of breast cancer deaths.

tively. It is shown in Appendix 1 that the difference between the slopes, $\beta_2 - \beta_1$, has the same value for all of the infinitely many sets of parameters that specify the mean values associated with a particular model (i.e., the difference is identifiable). As a result, unequivocal inferences can be made regarding changes in the magnitude or direction of birth cohort trends.

Contrasts in birth cohort parameters are defined in Appendix 2 for comparison of the slopes of the linear trends in two eras and quantification of the magnitude of the difference between the two slopes. Each of these contrasts is the difference between linear contrasts defined over the two eras being compared, and is, like the difference in slopes, identifiable. Similar contrasts can be used to compare the slopes in different calendar periods.

RESULTS

Figure 1 shows the maximum likelihood estimates of the birth cohort effects for the breast cancer mortality data presented in table 1. Maximum likelihood estimation for an age-period-cohort model requires the specification of a parameter constraint (15–18), and the estimates in figure 1 were obtained under the constraint that the first and last cohort effects are zero. It is evident that a marked change in the slope of the birth cohort trend occurred around 1925. A different constraint might lead to a curve with very different first-order properties (e.g., declining cohort effects from 1886 to 1920), but the decrease in slope around 1925 would be apparent for each constraint. The previous nonparametric analysis found a significantly increasing birth cohort trend from 1900 to 1914 and a significantly decreasing trend from 1924 to 1938 (12). To compare the slopes of the linear trends in birth cohorts in these two eras using the linear contrasts defined in Appendix 2, we let $h = 8$, corresponding to the 1901 birth cohort, and $k = 20$, corresponding to the 1925 birth cohort. The contrast $C_1$ takes the value $-0.353$ with a standard error of 0.018, and the contrast $C_2$ takes the value $-4.157$ with a standard error of 0.175. Both contrasts indicate a highly significant change in the magnitude of the slope of the linear birth cohort trend, with a moderation of risk for women born from about 1925 to 1940, confirming the results of the nonparametric analysis (12).

The trend in calendar period effects was examined for evidence of a change in slope in the 1980s compared with the 1970s. The contrast,

$$2\pi_{10} + \pi_9 - \pi_7 - 2\pi_6 - (2\pi_5 + \pi_4 - \pi_2 - 2\pi_1),$$

takes the value 0.168 with a standard error of 0.017, indicating an increase in the calendar period slope in the 1980s.

Clayton and Schifflers (18) presented age-period-cohort analyses of Japanese female breast cancer mortality using 5-year age intervals and calendar periods. To determine whether the Japanese rates provide evi-

![FIGURE 1](https://academic.oup.com/aje/article-abstract/143/1/85/63315)
must be interpreted cautiously; however, this apparent

\[ C_3 = 3\gamma_{k+3} + \gamma_{k+2} - \gamma_{k+1} - 3\gamma_k \]

was evaluated for \( h = 5 \) and \( k = 10 \). The birth cohort indexed by \( h = 5 \) is the birth cohort centered around 1900, while the birth cohort indexed by \( k = 10 \) is the birth cohort centered around 1925; thus, the eras being compared using \( C_3 \) in the Japanese data are the same as those compared in the US data using contrasts \( C_1 \) and \( C_2 \). Based on maximum likelihood estimation under the parameter constraint that the first and last cohort effects are zero, we calculate for the Japanese data that \( C_3 = -0.568 \) with an estimated standard error of 0.123. It can be easily verified that \( C_3 \), being an identifiable contrast, takes the same value for the three sets of maximum likelihood estimates for the Japanese data reported previously (18), each previous set of estimators being based on a different parameter constraint than that used in the current analysis. The contrast indicates that the Japanese breast cancer mortality data, like the US breast cancer mortality data, provide evidence of a moderation of mortality risk beginning with women born around 1925. This moderation of the birth cohort trend in the Japanese data is not evident from an examination of previously presented identifiable contrasts (18).

Previously proposed local curvature contrasts provided evidence in Japanese breast cancer mortality rates of two sudden changes in the birth cohort trend (18). One change occurred at the turn of the century. Applying the contrast \( C_3 \) to the Japanese data, taking \( h = 2 \) (i.e., the 1885 birth cohort) and \( k = 6 \) (i.e., the 1905 birth cohort), gives \( C_3 = 1.402 \) with a standard error of 0.126, confirming a marked increase in birth cohort slope. Applying \( C_2 \) to the US data with \( h = 2 \) (i.e., the 1889 cohort) and \( k = 10 \) (i.e., the 1905 cohort) gives \( C_2 = 0.467 \) with a standard error of 0.185. Thus, the birth cohort slope also increased somewhat in US women around the turn of the century, but to a much smaller extent than in Japanese women. A second sudden change in the Japanese birth cohort trend which occurred around 1935 was indicated previously (18). Evaluating the contrast

\[ C_4 = \gamma_{k+2} - \gamma_k - (\gamma_{h+2} - \gamma_h) \]

at \( h = 10 \) (i.e., the 1925 cohort) and \( k = 12 \) (i.e., the 1935 cohort) gives \( C_4 = 0.180 \) with a standard error of 0.059. Trends in the most recent birth cohort effects must be interpreted cautiously; however, this apparent increase in the Japanese birth cohort slope following 1935, if confirmed with more recent mortality data, would mark a clear departure from the US birth cohort trend, which gives no evidence of an increase in slope after 1920.

**DISCUSSION**

We have shown that, in spite of the identifiability problems inherent in age-period-cohort models, useful inferences can be made regarding trends in age, calendar period, or birth cohort effects. A change in the magnitude of the slope of a long term trend can be identified unambiguously, and such a change can have important interpretations regarding disease etiology. Although first-order properties of trends (e.g., the magnitude of the slope of the trend) are usually of primary interest in investigations of disease rates, unequivocal inferences cannot be based on the magnitude or direction of trends in parameters in age-period-cohort models (15–18).

A recent analysis of Swedish cancer incidence rates interpreted increasing birth cohort parameter estimates with successive birth cohorts as evidence that cancer risk was increasing because of recent increasing exposure to carcinogenic influences (20). As Appendix 1 shows, however, such inferences based on the direction of the birth cohort trend are tenuous at best. Only if the parameter constraint used in the maximum likelihood estimation procedure holds exactly in the underlying true population rate structure will the direction of the trend be informative (16), and the validity of the constraint can never be known with certainty. Although it was not noted by the Swedish investigators (20), there is a decrease in the slope of the birth cohort trend in the male Swedish cancer incidence data beginning with birth cohorts of the 1940s. This diminution of the rate of increase in rates with recent cohorts, which would seem to be at odds with the published interpretation of the cohort trend, could be evaluated using methods developed in the current paper.

Interpretation of the Swedish incidence data for women is complicated by a marked moderation of the risk of reproductive cancers beginning with cohorts born in the 1920s (20). Interestingly, because breast cancer would represent a major component of the reproductive cancers, it appears that breast cancer rates for Swedish women give evidence of the same moderation in risk beginning with cohorts born in the mid-1920s that is seen in US and Japanese breast cancer mortality rates. Thus, it appears that the factor or factors responsible for the moderation of breast cancer mortality rates observed in US women born from about 1925 through 1940 played a similar role in Japan, Sweden, and possibly many other countries. It would be useful to determine the extent to which the
moderation of breast cancer risk with birth cohorts beginning in the mid-1920s is a common feature in international data, and to determine any factors other than worldwide changes in childbearing patterns following World War II which could account for the timing of the moderation. For example, a portion of the prepubescent years of women born between 1925 and 1940 would have coincided with the Great Depression or World War II, and thus trends in energy intake early in life (21) may have contributed to the observed moderation of birth cohort risk.

Our analysis was based on mortality rates. Therefore, the possible impact of improved medical intervention on the survival of breast cancer patients must be considered. Improvements in breast cancer survival are unlikely to have had a major impact on population breast cancer mortality rates over the period considered in this study (22). Such improvements usually affect all or several age groups simultaneously, and thus typically appear in an age-period-cohort model as a decrease in the calendar period trend. The documented increase in the slope of the calendar period trend in the 1980s rules out a major contribution of improved medical treatment to the recent mortality rate decreases in women under the age of 60 (12). The fact that breast cancer mortality rate trends have been driven predominantly by birth cohort trends over the last three decades suggests that changes in breast cancer mortality rates have been influenced largely by changes in etiologic factors.

REFERENCES


APPENDIX 1

Because of the equality \( i - j + c = A \), it follows for any arbitrarily chosen constant, \( \delta \), that \( \delta (i - j + c - A) = 0 \). Thus, for any such \( \delta \), any constants \( a_1 \), \( a_2 \), and \( a_3 \) chosen such that \( a_1 - a_2 + a_3 = A \) and any set of parameters, \( \alpha_i \), \( \pi_j \), and \( \gamma_c \), which specify the underlying expected values, \( E_{i,j} \), defining the model, the set of parameters given by

\[
\alpha_i^* = \alpha_i + \delta (i - a_i) \\
\pi_j^* = \pi_j - \delta (j - a_j) \\
\gamma_c^* = \gamma_c + \delta (c - a_j)
\]

specifies exactly the same expected values for every rate (i.e., for every \( i \) and \( j \), \( \alpha_i^* + \pi_j^* + \gamma_c^* = \alpha_i + \pi_j + \gamma_c \)). Thus, there are infinitely many sets of parameters that specify exactly the same model.
For a parameter to be identifiable, it must take the same value (i.e., be invariant) for all transformations of the above type (i.e., for all choices of \( a_1, a_2, \) and \( a_3 \)). Suppose now that the slope of the linear trend in birth cohort effects differs in different eras (that is, there are cohort indices \( c_1 < c_2 \)) such that

\[
\gamma_c = \theta_1 + \beta_1 c
\]

for \( c \leq c_1 \), and

\[
\gamma_c = \theta_2 + \beta_2 c
\]

for \( c \geq c_2 \). To see that the difference in slopes, \( \beta_2 - \beta_1 \), is identifiable, note that under the transformation

\[
\gamma_c^* = \gamma_c + \delta (c - a_3),
\]

\[
\begin{align*}
\gamma_c^* &= \theta_1 + \beta_1 c + \delta (c - a_3) \\
&= \theta_1 - \delta a_3 + (\beta_1 + \delta) c
\end{align*}
\]

for \( c \leq c_1 \) and

\[
\gamma_c^* = \theta_2 + \beta_2 c + \delta (c - a_3)
\]

\[
= \theta_2 - \delta a_3 + (\beta_2 + \delta) c
\]

for \( c \geq c_2 \). The slopes for the transformed parameters, \( \beta_1^* = \beta_1 + \delta \) and \( \beta_2^* = \beta_2 + \delta \), are not equal to the slopes for the original parameters, \( \beta_1 \) and \( \beta_2 \). That is, the slopes are not invariant under the transformation (i.e., are not identifiable). Since \( \delta \) can take any value, it follows that different sets of maximum likelihood estimates can indicate different magnitudes or even directions for trends in birth cohort effects, as demonstrated in the analysis of Japanese breast cancer mortality data (18). Regardless of the value taken by \( \delta \), however, it follows that \( \beta_2^* - \beta_1^* = \beta_2 - \beta_1 \), so the difference in slopes is identifiable. Analogous arguments demonstrate that similar results hold for changes in the slope of age or calendar period effects.

**APPENDIX 2**

To obtain identifiable parameters for detection of changes in the slope of the linear trend in birth cohort risk in different eras, we need to derive contrasts that are invariant under the transformation defined in Appendix 1, i.e., \( \gamma_c^* = \gamma_c + \delta (c - a_3) \). In the previous analysis of US breast cancer mortality, we considered trends over intervals of eight consecutive birth cohorts (12). One simple invariant contrast for comparing slopes between two disjoint blocks of eight consecutive birth cohorts is

\[
C_1 = \gamma_{k+7} - \gamma_k - (\gamma_{h+7} - \gamma_h),
\]

where \( h + 7 \leq k \). This contrast is analogous to the local curvature contrast, \( \gamma_{c+1} - 2\gamma_c + \gamma_{c-1} \), given previously (18). More analogous to the previously applied nonparametric method for investigating trends over eight consecutive cohorts, in that all relevant birth cohorts contribute to the analysis, is the difference in linear contrasts,

\[
C_2 = 7\gamma_{k+7} + 5\gamma_{k+6} + 3\gamma_{k+5} + \gamma_{k+4} - \gamma_{k+3} - 3\gamma_{k+2} - 5\gamma_{k+1} - 7\gamma_k
\]

\[
- (7\gamma_{h+7} + 5\gamma_{h+6} + 3\gamma_{h+5} + \gamma_{h+4} - \gamma_{h+3} - 3\gamma_{h+2} - 5\gamma_{h+1} - 7\gamma_h).
\]

Any such difference in linear contrasts can be shown to be identifiable (i.e., to be invariant under the transformation from \( \gamma_c \) to \( \gamma_c^* \)) and can be expressed algebraically as a linear combination of the parameters—that is, \( C = s'\gamma \), where \( s \) is the vector of coefficients defining the contrast and \( \gamma \) is the vector of birth cohort effects. If \( \mathbf{V}_\gamma \) denotes the estimated covariance matrix for the maximum likelihood estimates of the birth cohort effects, then the variance of \( C \) can be calculated as \( s'\mathbf{V}_\gamma s \).

---