Detecting a Local Cohort Effect for Cancer Mortality Data Using a Varying Coefficient Model

Tetsuji Tonda¹, Kenichi Satoh², and Ken-ichi Kamo³

¹Faculty of Management and Information Systems, Prefectural University of Hiroshima, Hiroshima, Japan
²Research Institute for Radiation Biology and Medicine, Hiroshima University, Hiroshima, Japan
³Center for Medical Education, Sapporo Medical University, Sapporo, Japan

ABSTRACT

Background: Cancer mortality is increasing with the aging of the population in Japan. Cancer information obtained through feasible methods is therefore becoming the basis for planning effective cancer control programs. There are three time-related factors affecting cancer mortality, of which the cohort effect is one. Past descriptive epidemiologic studies suggest that the cohort effect is not negligible in cancer mortality.

Methods: In this paper, we develop a statistical method for automatically detecting a cohort effect and assessing its statistical significance for cancer mortality data using a varying coefficient model.

Results: The proposed method was applied to liver and lung cancer mortality data on Japanese men for illustration. Our method detected significant positive or negative cohort effects. The relative risk was 1.54 for liver cancer mortality in the cohort born around 1934 and 0.83 for lung cancer in the cohort born around 1939.

Conclusions: Cohort effects detected using the proposed method agree well with previous descriptive epidemiologic findings. In addition, the proposed method is expected to be sensitive enough to detect smaller, previously undetected birth cohort effects.

Key words: birth cohort effect; cancer mortality in Japan; varying coefficient model

INTRODUCTION

In Japan, cancer has been the leading cause of death since 1981, becoming a serious concern in the aging society. Precise trends in cancer risk must be identified to develop efficient cancer control programs.

Three time-dependent factors affect cancer mortality: age, period, and birth cohort. We illustrate these effects with liver cancer mortality data in Japanese men as a typical example. Cancer mortality data can be obtained from the website of the National Cancer Center in Japan.¹ The data are tabulated by 5-year age categories. Figure 1 shows time trends of mortality by age, which allows us to understand the age and period effects. Liver cancer mortality increases with age but has decreased in the recent period.

There are also characteristic local changes for certain periods by age. Figure 2 shows the trend of liver cancer, but with birth year rather than period; local changes occur among subjects whose birth year is around 1935. This is regarded as a birth cohort effect. In many past studies, it has been pointed out that the cohort born in the 1930s has a high risk of liver cancer.²,³ The reason for this is thought to be the high prevalence of hepatitis C virus infection in Japan.⁴,⁵ The cohort effect for liver cancer mortality in Japanese males is easy to be identified, but cohort effects are typically not discernible to the eye.

There are statistical methods for assessing age, period, and cohort effects simultaneously, such as age-period-cohort (APC) analysis.⁶ APC analysis suffers from a model identification problem due to the exact linear dependency among the three variables: cohort = period − age. As a result, it is not generally possible to estimate the three effects separately without additional constraints to identify the model.⁷,⁸ To overcome this problem, several approaches to estimating the three factors have been proposed under various assumptions or constraints. However, differences in assumptions or constraints often produce inconsistent results.⁹ Keyes et al¹⁰ compared three approaches: the traditional constraint-based regression technique,⁶ the Holford model,¹¹,¹² and the median polish technique,¹³ using data on
the obesity prevalence in the United States from 1971–2006.\textsuperscript{14–16} The results of the different approaches regarding cohort effect were not consistent. They considered that this inconsistency was due to differences in conceptual definitions of cohort effect as estimated by the three approaches.

As another approach, Kamo, Satoh, and Tonda\textsuperscript{17} proposed visualizing mortality risk using mortality data tabulated by age and period. They visualized the cancer mortality risk by the surface of the age-period plane and suggested searching for a birth cohort effect based on the surface. The main objective of their method is to elucidate the characteristics of cancer trends using visualization, not to statistically judge whether cohort effects exist or not. While their method is useful for discerning a global trend in age and/or period, it might be difficult to identify a birth cohort effect empirically, except in extreme cases, such as the birth cohort effect with liver cancer mortality in Japanese males.

We developed a statistical method to detect a cohort effect automatically and assess its statistical significance for cancer mortality data using a varying coefficient model. In general, there are two types of cohort effects: global and local. In this paper, we focus on detecting a local change as a cohort effect. It is especially important to describe the birth cohort effect in a statistical model for prediction of future cancer mortality. Automatic detection is important, because evaluation by eyesight carries the risk of errors due to the researcher’s subjective expectations or bias.

In the present paper, we introduce a varying coefficient model and construct a method for estimating the varying coefficient with its statistical evaluation. We then apply the proposed method to data on liver and lung cancer mortality in Japanese males. Finally, we review possible reasons for the birth cohort effect detected by the proposed method and evaluate the method’s performance with its instructions for use in practice.

**METHODS**

**Varying coefficient model**

Let \((z_a,p, d_{ap})\) denote the set of population person-time and observed number of deaths for age \(a\) during period \(p\). The observed number of deaths is assumed to follow a Poisson distribution,

\[d_{ap} \sim \text{Poisson}(z_{ap} \lambda_{ap}), \quad \log \lambda_{ap} = \beta_0(a,p),\]

where \(\beta_0(a,p)\) is the regression coefficient varying with age \(a\) and period \(p\). Regression coefficients that vary with time, geographical location, or other important covariates are generally called varying coefficients, and the varying coefficient model was proposed by Hastie and Tibshirani.\textsuperscript{18} The varying coefficient transformed to the risk scale, \(\exp(\beta_0(a,p))\), represents a surface of mortality risk on the age-period plane. Note that the mortality rate by age and period is regarded as a crude estimate of \(\exp(\beta_0(a,p))\). Figure 3 shows the mortality rate per 100,000 person-years by 5-year age group and 5-year period with gradations represented by a heat map. Note that a cohort with birth year \(c\) lies on the diagonal line \(p - a = c\) in Figure 3. If a birth cohort effect exists, it appears as a higher or lower diagonal line. However, it is not easy to identify such a trend using the mesh-type mapping in Figure 3. We estimate the varying coefficient to present it as a smooth mapping. There are
several ways to estimate varying coefficients. One is nonparametric method such as kernel smoothing\(^{19,20}\) and geographical weighted regression (GWR).\(^{21}\) Another is parametric method based on an interaction model.\(^{22}\)

The GWR model\(^{21}\) is widely used for spatial data with a continuous outcome. Nakaya et al.\(^{23}\) extended the GWR model to a geographically weighted Poisson regression (GWPR) model for spatial count data. The GWPR is implemented in the GWR4.0 software\(^{24}\) and the spgwr package in R.\(^{25}\)

Regarding the set of age and period as a virtual geographical space, several ways to estimate varying coefficients for longitudinal data with a discrete outcome\(^{26}\) as well as to continuous outcomes. Their method was extended to estimating varying coefficients for continuous outcome. Nakaya et al.\(^{23}\) extended the GWR model to grasp the age-period trend of mortality, but it is often difficult to detect a birth cohort effect visually based solely on the contours. We therefore constructed a parametric model to identify the birth cohort effect automatically using statistical evaluation.

**Estimation of varying coefficient**

Satoh and Yanagihara\(^{22}\) proposed a parametric method for estimating varying coefficients for longitudinal data with continuous outcomes. Their method was extended to longitudinal data with a discrete outcome\(^{26}\) as well as to spatial data.\(^{27-29}\) For cancer mortality data, Kamo, Satoh, and Tonda\(^{17}\) modeled \(\beta_0(a,p)\) by interactions of polynomials in age and period; that is, \(\beta_0(a,p) = \theta^T x(a,p)\), where \(\theta\) is an \(m\)-dimensional vector of unknown parameters and \(x(a,p)\) is an \(m\)-dimensional basis vector whose elements are interactions of polynomials in age and period. For example, interactions of cubic polynomials produce the basis

\[
x(a,p) = (1, a, a^2, a^3, p, ap, a^2p, a^3p, p^2, ap^2, a^2p^2, a^3p^2, a^3p^3).
\]

The interaction terms in \(x(a,p)\) describes not only age and period effects but also a global trend on cohort effect. To model a local change of cohort effect, we here add a basis of normal density with mean \(\mu_c\) and variance \(\sigma_c^2\) into \(\beta_0(a,p)\); that is,

\[
\beta_0(a,p) = \theta^T x(a,p) + \beta_0(\mu_c, \sigma_c^2).
\]

\[
\phi(\mu_c, \sigma_c^2) = \frac{1}{\sqrt{2\pi\sigma_c^2}} \exp\left(-\frac{(\mu - \mu_c)^2}{2\sigma_c^2}\right).
\]

Note that \(\mu_c\) and \(\sigma_c\) denote the center and range of birth cohort.

In this paper, we regard the cohort effect as the local change if the value of \(\sigma_c\) is within a period of about 5% for a whole range of birth year. We then consider the relative risk \(\exp(\beta_0(\mu_c, \sigma_c^2))\) as the cohort effect. A statistical test for a cohort effect is also available based on a test for the usual null hypothesis on \(\beta_c\).

The number of deaths is observed at \(n\) combinations of age and period. Then, let \([(a_j, p_j), j = 1, \ldots, n]\) be a set of observed age and period, where \(a_j\) and \(p_j\) denote the \(j\)th combination of age and period. For a fixed \(\mu_c\) and \(\sigma_c\), the unknown parameters \(\theta\) and \(\beta_c\) are estimated by maximizing the log-likelihood:

\[
\begin{align*}
\theta_{\hat{\beta}_c} &= \arg \max \ell(\theta, \beta_c|\mu_c, \sigma_c^2), \\
\ell(\theta, \beta_c|\mu_c, \sigma_c^2) &= \sum_{j=1}^{n} \log f(d_{a_j, p_j}|z_{a_j, p_j}),
\end{align*}
\]

where \(\log f(d|z) = d(\log z + \beta_0(a,p)) - z\exp(\beta_0(a,p)) - \log d!\). In addition, \(\beta_0(a,p) = \hat{\theta}^T x(a,p) + \beta_c(\mu_c, \sigma_c^2)\) is obtained. Using \(-2\ell(\hat{\theta}, \hat{\beta}_c|\mu_c, \sigma_c^2)\) as an index of goodness-of-fit, \(\mu_c\) and \(\sigma_c\) are determined by minimizing \(-2\ell(\hat{\theta}, \hat{\beta}_c|\mu_c, \sigma_c^2)\).

**RESULTS**

We evaluated the performance of our proposed method for detecting a cohort effect using liver and lung cancer mortality in Japanese males. We choose these two cancers as examples because past epidemiological studies already reported the possibility of birth cohort effects.

**Liver cancer mortality**

As mentioned previously, many past studies have noted that the cohort born around 1935 has a high risk of liver cancer. We used our method to search for a cohort effect automatically. We estimated \(\beta_0(a,p)\) using \(x(a,p)\) based on the interactions of cubic polynomials basis. Minimizing the values of \(-2\ell(\hat{\theta}, \hat{\beta}_c|\mu_c, \sigma_c^2)\) for several sets of \(\mu_c\) and \(\sigma_c\), the best-fitted model was \((\mu_c, \sigma_c) = (1934, 4)\).

Table 1 gives the estimated parameters corresponding to \(\theta\) for the interactions of cubic polynomials basis and \(\beta_c\) for \((\mu_c, \sigma_c) = (1934, 4)\). Because the estimate of \(\beta_c\) was about 4.35 and the statistical test of the null hypothesis that \(\beta_c = 0\) was rejected with a high level of significance, we were able to declare a significant positive effect (increased risk) for the birth cohort around
1934. Figure 4 shows the relative risks corresponding to this birth cohort effect. The maximum relative risk was about 1.54. Figure 5 shows the estimated cancer mortality surface describing a smooth mapping against Figure 3. The diagonal line denotes the 1934 birth cohort, which was detected as the center of cohort. The positive birth cohort effect can be seen around the diagonal line for the 1934 birth cohort.

### Lung cancer mortality

A previous descriptive epidemiologic study noted the possibility of a small birth cohort effect with a local peak around the late 1920s and a declining trend until the late 1930s. Figure 6 and Figure 7 show time trends of lung cancer mortality in Japanese males by age. The horizontal axes in Figure 6 and Figure 7 are calendar year and birth year, respectively. From these figures, it is difficult to identify a cohort effect intuitively unless one is a highly experienced epidemiologist. Therefore, we applied our method to search for a cohort effect automatically. Figure 8 shows the data tabulated by age and period colored with the corresponding heat map for lung cancer mortality in Japanese males. We estimated $\beta_c$ using $x(a,p)$ based on the interactions of cubic polynomials basis. Minimizing the values of $-2\ell(\hat{\theta}, \hat{\beta}_c | \mu_c, \sigma^2)$ for several sets of $\mu_c$ and $\sigma_c$, the best-fitted model was $(\mu_c, \sigma_c) = (1939, 3)$. Table 2 gives the estimated parameters corresponding to $\theta$ for the interactions of cubic polynomials basis and $\beta_c$ for $(\mu_c, \sigma_c) = (1939, 3)$. Because the estimate of $\beta_c$ was about $-1.38$ and a statistical test of the hypothesis $\beta_c = 0$ was highly significant, we were able to declare a significant negative effect (decreased risk) around the 1939 birth cohort. Figure 9 shows the relative risks in the detected birth cohort effect. The minimum relative risk for the birth cohort was about 0.83. Figure 10 shows the estimated cancer mortality surface describing a smooth mapping against Figure 8. The diagonal line denotes the 1939
birth cohort, which was detected as the center of the cohort. The negative birth cohort effect can be seen around the diagonal line for the 1939 birth cohort.

**DISCUSSION**

In this paper, we deal with the data for cancer mortality. Analysis using data for cancer mortality is suitable for the first proposal of a new statistical procedure because the quality of the data is quite high in Japan. If established epidemiological knowledge for the cohort effect can be detected using this method, then this method can be considered to work well. Moreover, this method may useful in evaluating incidence of

**Figure 6.** Lung cancer mortality trend in Japanese males by age. The horizontal axis is period, and the vertical axis is the mortality rate per 100,000 person years.

**Figure 7.** Lung cancer mortality trend for Japanese males by age. The horizontal axis is birth year, and the vertical axis is the mortality rate per 100,000 person years.

Table 2. The estimated parameters corresponding to $\theta$ and $\beta_c$ in the case $(\mu_c, \sigma_c) = (1939, 3)$ for lung cancer mortality in Japanese males

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>Std. Error</th>
<th>z-value</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta^a$ (Intercept)</td>
<td>$-6.8839$</td>
<td>$0.0057$</td>
<td>$-1200.643$</td>
<td>$&lt;0.001$</td>
</tr>
<tr>
<td>$a$</td>
<td>$1.2338$</td>
<td>$0.0081$</td>
<td>$151.528$</td>
<td>$&lt;0.001$</td>
</tr>
<tr>
<td>$a^2$</td>
<td>$-0.1676$</td>
<td>$0.0038$</td>
<td>$-44.632$</td>
<td>$&lt;0.001$</td>
</tr>
<tr>
<td>$a^3$</td>
<td>$-0.0491$</td>
<td>$0.0029$</td>
<td>$-16.700$</td>
<td>$&lt;0.001$</td>
</tr>
<tr>
<td>$p$</td>
<td>$0.1238$</td>
<td>$0.0067$</td>
<td>$18.364$</td>
<td>$&lt;0.001$</td>
</tr>
<tr>
<td>$p^2$</td>
<td>$-0.0649$</td>
<td>$0.0020$</td>
<td>$-32.567$</td>
<td>$&lt;0.001$</td>
</tr>
<tr>
<td>$p^3$</td>
<td>$0.0003$</td>
<td>$0.0014$</td>
<td>$0.218$</td>
<td>$0.827$</td>
</tr>
<tr>
<td>$ap$</td>
<td>$0.0187$</td>
<td>$0.0086$</td>
<td>$2.179$</td>
<td>$0.029$</td>
</tr>
<tr>
<td>$ap^2$</td>
<td>$-0.0291$</td>
<td>$0.0031$</td>
<td>$-9.295$</td>
<td>$&lt;0.001$</td>
</tr>
<tr>
<td>$ap^3$</td>
<td>$0.0079$</td>
<td>$0.0020$</td>
<td>$4.002$</td>
<td>$&lt;0.001$</td>
</tr>
<tr>
<td>$a^2p$</td>
<td>$0.0436$</td>
<td>$0.0045$</td>
<td>$9.672$</td>
<td>$&lt;0.001$</td>
</tr>
<tr>
<td>$a^2p^2$</td>
<td>$-0.0123$</td>
<td>$0.0015$</td>
<td>$-8.438$</td>
<td>$&lt;0.001$</td>
</tr>
<tr>
<td>$a^2p^3$</td>
<td>$-0.0007$</td>
<td>$0.0010$</td>
<td>$-0.724$</td>
<td>$0.469$</td>
</tr>
<tr>
<td>$ap^3$</td>
<td>$0.0134$</td>
<td>$0.0032$</td>
<td>$4.127$</td>
<td>$0.001$</td>
</tr>
<tr>
<td>$ap^2$</td>
<td>$0.0039$</td>
<td>$0.0012$</td>
<td>$3.271$</td>
<td>$0.001$</td>
</tr>
<tr>
<td>$ap$</td>
<td>$-0.0007$</td>
<td>$0.0008$</td>
<td>$-0.875$</td>
<td>$0.382$</td>
</tr>
<tr>
<td>$\beta_c$</td>
<td>$-1.3787$</td>
<td>$0.0835$</td>
<td>$-16.517$</td>
<td>$&lt;0.001$</td>
</tr>
</tbody>
</table>

*The elements of $\theta$ are coefficients of interactions in the cubic polynomial basis.*

J Epidemiol 2015;25(10):639-646
cancer or other diseases. From the viewpoint of cancer control programs, information on not only cancer mortality but also cancer incidence is particularly important. However, since cancer incidence data tend to have issues with completeness and quality, we chose a more robust dataset for our evaluation.\(^\text{32}\) Similarly, when we apply this method to other diseases, we have to pay attention to the properties of the data.

The proposed method detected a significant positive effect for the cohort born around 1934 for liver cancer mortality in Japanese males. This result agrees well with previous epidemiologic studies.\(^\text{2,3,33}\) Yoshimi and Sobue\(^\text{2}\) and Inamura and Sobue\(^\text{3}\) reported that liver cancer mortality exhibits the most notable birth cohort effect among cancer sites in Japan. As mentioned previously, this positive cohort effect is attributed to the markedly high prevalence of hepatitis C virus infection in the cohort born around 1935.\(^\text{4,5}\)

For lung cancer mortality in Japanese males, the proposed method detected a significant negative effect for the cohort born around 1939. A possible reason for such an effect was discussed by Marugame et al.\(^\text{34}\) It is known that cigarette smoking relates to lung cancer. From the end of World War II to the beginning of the Japanese post-World War II economic expansion, Japan experienced an extreme shortage of cigarettes. Therefore, men born during the 1930s had less opportunity to begin smoking during adolescence, so their cohort showed a corresponding dip in smoking prevalence. It is difficult to detect this birth cohort effect directly from the figures without careful observation, demonstrating the utility of the proposed method. Takahashi et al\(^\text{35}\) analyzed mortality data on lung cancer in Japan using an APC model based on that of Holford.\(^\text{11,36}\) Based on a figure showing changing patterns of non-linear birth cohort effect, they also suggested the existence of a local change at the end of the 1930s. However, their suggestion is based on visual consideration of the figure without rigorous statistical evaluation.

We discuss the instructions for use of the proposed method in practice using cancer mortality data on other sites. Applying the proposed method to stomach cancer mortality in Japanese males, the optimal center and range of cohort were \((\mu_c, \sigma_c^2) = (1920, 21)\). This range of cohort seems to be too wide to be considered a local change. We therefore consider that no local birth cohort effect exists. As mentioned previously, there are two types of cohort effects: global trends, which denote a gradual change over time, and local changes within a short time. In the proposed method, the interaction terms of the function \(x(a, p)\) is regarded as modeling a global trend on cohort effect. The normal density basis, \(\phi(\mu_c, \sigma^2_c)\), models a local change on cohort effect, if such a change exists. However, if no local change exists, \(\phi(\mu_c, \sigma^2_c)\) works for fitting a global trend on cohort effect together with the interaction terms of \(x(a, p)\), and the range of the cohort, \(\sigma^2_c\), tends to be large. In this case, it is natural to consider that no local change exists. For cancer mortality at other sites, except for the rectum, the results were similar to that for stomach cancer in that the optimal range of the cohort tended to be wide. These results show no cohort effect for cancer mortality at most sites. The findings for rectal cancer showed that the optimal center and range of cohort was \((\mu_c, \sigma_c) = (1921, 7)\). The range of the cohort seems to be not so wide, suggesting the potential existence of a cohort effect. Note that the estimate of \(\beta_c\) was about \(-4.24\), and the minimum relative risk with birth cohort was about 0.79.
In many past epidemiologic studies, cohort effects have been inferred by evaluating longitudinal behavior. However, such a classical method is subject to error due to the researcher’s bias or subjectivity influencing the results. One solution to this problem is to judge the effect automatically using a statistical method, which allows for objective detection of a cohort effect. In our analyses of liver and lung cancer in Japanese males, we successfully identified birth cohort effects automatically, and our results agree with those of past epidemiologic findings. It is expected that this method will prove useful in identifying small or previously undetected cohort effects. The method should also prove useful for developing a statistical model to predict future cancer mortality.

In this paper we focused on detecting a single birth cohort effect that changes locally. In previous descriptive studies, some epidemiologists noted the possibility of multiple birth cohort effects; for example, lung cancer in males shows another cohort effect in the late 1920s, which can be seen in Figure 10 as a second candidate cohort effect. Our method should be revised to account for multiple cohort effects in the future.

ONLINE ONLY MATERIAL

Abstract in Japanese.

ACKNOWLEDGEMENTS

The authors are grateful to the referees for his/her valuable comments and suggestions which improved an earlier version of the manuscript. Part of this research was supported by the Ministry of Education, Culture, Sports, Science and Technology, Scientific Research (C) #26460754 (2014–2016), #26330043 (2014–2016), #24590801 (2012–2016), and #26330042 (2016–2017). This work was carried out at the Joint Usage/Research Center (RIRBM), Hiroshima University.

Conflicts of interest: None declared.

REFERENCES


