Influence of Time Period and Birth Cohort Effects on Age-specific Incidence Rates of Pancreatic and Kidney Cancer

Tengiz Mdzinarishvili
Eppley Cancer Institute, UNMC
tmdzinar@unmc.edu

Michael X. Gleason
Eppley Cancer Institute, UNMC
mgleason@unmc.edu

Simon Sherman
Eppley Cancer Institute, UNMC
Progenomix, Inc.
ssherm@unmc.edu

Abstract
The influence of time period and birth cohort effects on age-specific incidence rates of pancreatic cancer (PC) and kidney cancer (KC) in white males and females was analyzed using SEER 9 data collected during 1975-2004. To estimate these effects, we used our novel approach that is applicable for any arbitrary hazard function. By utilizing the incidence rates of first primary, microscopically confirmed cases observed in six five-year time periods and 17 cohorts, we found turnovers in the PC and KC incidence rate distributions at old ages. For PC, no systematic changes were revealed in both time period and cohort coefficients during the considered time observation. For KC, however, we found systematically increasing trends for time period coefficients, while trends of birth cohort effect coefficients remain unchanged during 1975-2004. The proposed approach can be used in the mathematical modeling of different types of carcinogenesis.

1. Introduction.
Modeling of the age distribution of cancer is often performed using “cross-sectional” data (data collected during a single period of time) without considering time period and birth cohort effects. Also, most studies of cancer in the aging population use “raw” incidence rate data that include cases corresponding to second primary or secondary tumors, and do not omit cases which have not been microscopically confirmed. Ignorance of birth cohort and time period effects and use of raw data can distort the results of modeling of the distribution of cancer incidence rates in aging [1-4].

In fact, studies of lung cancer show that the risk of this cancer increases with age in a cohort-wise fashion, while examination of the cross-sectional rates suggests that this risk declines after a certain age (see, for instance [3]).

More recent studies have also produced controversial results. Harding and colleagues [5] studied raw age-specific cancer incidence rates presented in the National Cancer Institute’s SEER (Surveillance Epidemiology and End Results) database for 20 major organ sites for males and 21 major organ sites for females. They found that according to their cross-sectional data, cancer incidence rates fall at old ages for the vast majority of the examined cancer sites.

On the other hand, Meza and coauthors [6] carried out an adjustment of raw age-specific incidence rates of colorectal and pancreatic cancers for birth cohort and time period effects and did not observe a turnover point at old ages. To adjust time period and birth cohort effects they used the maximum likelihood method where an assumption of the mathematical form of hazard function is necessary. The results obtained by the use of the adjusting procedure [6], however, depend on the mathematical form of the assumed hazard function.

To estimate the influence of time period and birth cohort effects on the age-specific incidence rates of cancer, we proposed a new approach, which is independent of the hazard function. We used this approach to analyze the influence of time period and birth cohort effects on the distribution of the “filtered” (primary, microscopically confirmed cases) age-specific incidence rates of pancreatic cancer (PC) and kidney cancer (KC) in white males and females.

2. Data Preparation and Filtration
To analyze the patterns of the age-specific incidence rate distributions in PC and KC, we used data from the SEER 9 registries that contain cancer data for the following nine locations: Atlanta, Connecticut, Detroit, Hawaii, Iowa, New Mexico, San Francisco-Oakland, Seattle-Puget Sound, and Utah. In the SEER database, each case record contains information on whether this is the first
primary malignant case and whether the case is histopathologically confirmed. All cases where the patient was of known race and gender with a first primary, microscopically confirmed tumor were considered as “filtered” data, and the cases where this filtering was not performed were considered as “raw” data.

For age incidence modeling, we used filtered data, which are expected to be more reliable than raw data. We utilized the incidence rate data expressed per 100,000 persons to the nearest 0.0001 decimal place and age-adjusted by the direct method to the 2000 United States standard population [7, 8].

We used SEER 9 data collected between 1975 and 2004. To smooth out random fluctuations, the data were combined in six five-year cross-sectional time periods: 1975-1979, 1980-1984, 1985-1989, 1990-1994, 1995-1999, and 2000-2004. For PC and KC, the gender-specific incidence rates were grouped into 18 five-year age intervals (or groups): 17 groups, ranging from 0 to 84 years old, and the 18th group that included all cases for ages 85 and over. For each of these intervals, \( i \), the corresponding \( I(t_i) \) and standard errors \( (SE_i) \) were obtained by processing the SEER data according to SEER’s rate algorithms. For each age interval, the values of the coefficient of variance were also determined as: 

\[
CV_i = \frac{SE_i}{I(t_i)}
\]

### 3. Novel Approach for Estimation of Time Period and Birth Cohort Effects

Cancer incidence rates by age also depend on the epoch of birth and year of diagnosis (birth cohort and time period). For cancer registry data, age-cohort-period (log-linear) models are usually used [6, 9]. These models assume that each observed incidence rate, \( I(t_i) \), can be approximated as a product of the corresponding coefficient of period time effect, \( v_i \), coefficient of birth cohort effect, \( u_{ij} \), and the hazard function, \( h(t) \):

\[
I_{y}(t_i) = v_j u_i h(t_i)
\]

where \( i, j, \) and \( l \) are indexes of the age, time period and birth cohort groups, correspondingly; and \( t = t_i \) is the midpoint of the corresponding age group. Indices \( i \) and \( j \) determine index \( l \) (see below). The mathematical form of the hazard function is assumed to be known a priori.

The birth cohort and time period adjustments can be performed by applying a maximum likelihood method to Equation 1 (assuming that the number of incident cases follows a Poisson distribution) for simultaneous derivation of time period and birth cohort effect coefficients and hazard function parameters. One coefficient will be arbitrarily anchored for time period coefficient (\( v = 1 \)) and one coefficient will be arbitrarily anchored for birth cohort effect coefficient (\( u = 1 \)). It is clear that the adjusting procedure depends on an unknown hazard function and anchoring coefficient indexes. We propose an approach of the assessment of coefficients which does not need any information about the hazard function. The adjusting procedure is described below using Tables 1 and 2.

Table 1 schematically presents the incidence rates of cross-sectional data collected in 1975-80, 1980-84, 1985-1989, 1990-1994, 1995-1999, and 2000-2004 \((j = 1, \ldots, 6, \) six time periods). In this table, the incidence rates of the same cohort groups are located along diagonals. We used data for the age groups over age 30 (index \( i = 7, \ldots, 18 \)), because the incidence rates for these groups are statistically significant (according to the SEER practice, the number of cases should exceed 15 to be statistically significant). We consider 17 cohorts (index \( l = 1, \ldots, 17 \)). The first cohort includes patients that were born in the years of 1970-1974, while the 17th cohort is formed from patients born in 1890-1894. From Table 2, it can be seen that:

\[
i - j = l
\]

In Table 1, each of these 17 birth cohorts is marked by an arrow linking the diagonal cells, in which the cancer incidence rates observed for this group in each time period are presented.

Table 2 schematically presents the observed incidence rates as a product of the hazard function, \( h(t) \), and the corresponding time period and birth cohort coefficients, \( v \) and \( u \). As can be seen on the corresponding diagonals in Table 2, the same time period coefficients \( v_l \) are located in one column \( (u_l) \), while the cohort coefficients are along the corresponding diagonals.

Let us choose anchor coefficients, for example \( v_6 = 1 \) \((SE(v_6) = 0)\) and \( u_{12} = 1 \) \((SE(u_{12}) = 0)\). By element-by-element division of each column by the 6th column of Table 2, we obtain:

\[
\frac{I_{i,j}(t_i)}{I_{i,6}(t_i)} = \frac{v_j u_{i-j}}{v_6 u_{i-6}}, \quad i = 7, \ldots, 18; \quad j = 1, \ldots, 5
\]

If the birth cohort effect is absent (i.e. all coefficients \( u = 1 \), age-time period model), we have:

\[
\frac{I_{i,j}(t_i)}{I_{i,6}(t_i)} = v_j, \quad i = 7, \ldots, 18; \quad j = 1, \ldots, 5.
\]

After averaging through index \( i \), estimates of time period coefficients are:

\[
v_j = \frac{1}{12} \sum_{i=7}^{18} \frac{I_{i,j}(t_i)}{I_{i,6}(t_i)}, \quad j = 1, \ldots, 5.
\]
### Table 1. Presentation of the observed age-specific incidence rates for 17 birth cohorts during six time periods.

<table>
<thead>
<tr>
<th>Period of observation</th>
<th>Birth cohort</th>
</tr>
</thead>
<tbody>
<tr>
<td>1974-79</td>
<td>1980-84</td>
</tr>
<tr>
<td>1985-89</td>
<td>1990-94</td>
</tr>
<tr>
<td>1995-99</td>
<td>2000-04</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>index, i</td>
<td>j = 1</td>
<td>j = 2</td>
<td>j = 3</td>
<td>j = 4</td>
<td>j = 5</td>
<td>j = 6</td>
</tr>
<tr>
<td>1</td>
<td>2.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>32.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>37.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>42.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>47.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>52.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>57.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>62.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>67.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>72.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>77.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>82.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>87.5+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 2. Presentation of the observed incidence rates as the product of the hazard function, \( h(t) \), and the corresponding time period \( (-) \) and birth cohort \( (\cdot) \) coefficients.

<table>
<thead>
<tr>
<th>Period of observation</th>
<th>Birth cohort</th>
</tr>
</thead>
<tbody>
<tr>
<td>1974-79</td>
<td>1980-84</td>
</tr>
<tr>
<td>1985-89</td>
<td>1990-94</td>
</tr>
<tr>
<td>1995-99</td>
<td>2000-04</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>index, i</td>
<td>j = 1</td>
<td>j = 2</td>
<td>j = 3</td>
<td>j = 4</td>
<td>j = 5</td>
<td>j = 6</td>
</tr>
<tr>
<td>1</td>
<td>2.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>32.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>37.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>42.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>47.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>52.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>57.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>62.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>67.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>72.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>77.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>82.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>87.5+</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Note that the values for \( v_j, j = 1, \ldots, 5 \) are estimated with different precision and the most effective estimates were weighted means of the observed values. The weighting problem needs special consideration that is out of the scope of this work and will be addressed elsewhere. Here we restrict ourselves to arithmetic means. The same note concerns the estimates of the cohort effect coefficients \( u \). Calculations showed that such a simplification does not impact the results significantly [data not shown].

In Table 1 age-specific incidence rates can be considered as approximately normally distributed variables (approximation is valid for Poisson data not shown)

For estimation of the birth cohort coefficient ratios let us consider the incidence rate ratios with the following indexes (note \( i - j = l \) from Table 2:

\[
\frac{I_{i,j}(t_i)}{I_{i,j+1}(t_i)} = \frac{u_j v_{j+1}}{u_{j+1} v_j}
\]

(7)

and:

\[
I_{i,j}(t_i) = \frac{u_j v_j}{u_{j+1} v_{j+1}}
\]

(8)

and:

\[
I_{i,j}(t_i) = \frac{u_j}{u_{j+1}}
\]

(9)

If the time period effect is absent (i.e. all coefficients \( v = 1 \), age-cohort model), we have:

\[
I_{i,j}(t_i) = \frac{u_j}{u_{j+1}}
\]

(10)

By averaging through index \( i \) (Equations 9 and 10), we can find estimates (denoted by an asterisk) of ratios of the cohort effect coefficients.

\[
\begin{align*}
\left( \frac{u_1}{u_2} \right)^* &= \frac{I_{2,6}(t_i)}{I_{5,6}(t_i)} \\
\left( \frac{u_2}{u_3} \right)^* &= \frac{1}{2} \left( \frac{I_{5,7}(t_i) + I_{7,8}(t_i)}{I_{7,8}(t_i)} \right) \\
\left( \frac{u_3}{u_4} \right)^* &= \frac{1}{3} \left( \frac{I_{7,9}(t_i) + I_{9,10}(t_i) + I_{10,12}(t_i)}{I_{9,10}(t_i)} \right) \\
\left( \frac{u_4}{u_5} \right)^* &= \frac{1}{4} \left( \frac{I_{9,11}(t_i) + I_{11,12}(t_i) + \ldots + I_{10,12}(t_i)}{I_{10,12}(t_i)} \right) \\
\left( \frac{u_5}{u_6} \right)^* &= \frac{1}{5} \left( \frac{I_{11,13}(t_i) + I_{13,14}(t_i) + \ldots + I_{12,14}(t_i)}{I_{12,14}(t_i)} \right)
\end{align*}
\]

The standard errors of \( v_j^* \) can be calculated using Equations 4 and 6 and standard rules of error propagation [10]. When errors of two observed variables are distributed normally and their coefficients of variation are small, the errors of their ratio will also be distributed normally [11].

\[
\begin{align*}
l &= 6, \quad i = 7, \ldots, 11 \\
l &= 7, \quad i = 8, \ldots, 12 \\
l &= 8, \quad i = 9, \ldots, 13 \\
l &= 9, \quad i = 10, \ldots, 14 \\
l &= 10, \quad i = 11, \ldots, 15 \\
l &= 11, \quad i = 12, \ldots, 16 \\
l &= 12, \quad i = 13, \ldots, 17
\end{align*}
\]
By setting $u_{12} = 1$, from Equations 11 and 12 we can estimate step-by-step all cohort coefficients:

$$u_{13}^* = \frac{u_{13}}{u_{12}}$$

$$u_{10}^* = u_{11} \left( \frac{u_{10}}{u_{11}} \right)$$

$$u_9^* = \frac{u_9}{u_{10}}$$

$$u_8^* = \frac{u_8}{u_9}$$

$$u_7^* = \frac{u_7}{u_8}$$

$$u_6^* = u_7 \left( \frac{u_6}{u_7} \right)$$

$$u_5^* = \frac{u_5}{u_6}$$

$$u_4^* = \frac{u_4}{u_5}$$

$$u_3^* = \frac{u_3}{u_4}$$

$$u_2^* = \frac{u_2}{u_3}$$

$$u_1^* = \frac{u_1}{u_2}$$

(13)

and:

$$u_{13}^* = \frac{u_{13}}{u_{12}}$$

$$u_{10}^* = u_{11} \left( \frac{u_{10}}{u_{11}} \right)$$

$$u_9^* = \frac{u_9}{u_{10}}$$

$$u_8^* = \frac{u_8}{u_9}$$

$$u_7^* = \frac{u_7}{u_8}$$

$$u_6^* = u_7 \left( \frac{u_6}{u_7} \right)$$

$$u_5^* = \frac{u_5}{u_6}$$

$$u_4^* = \frac{u_4}{u_5}$$

$$u_3^* = \frac{u_3}{u_4}$$

$$u_2^* = \frac{u_2}{u_3}$$

$$u_1^* = \frac{u_1}{u_2}$$

(14)

The standard errors of $u_l$ ($l = 1, ..., 11$ and $l = 13, ..., 17$) can be calculated using Equations 6 and 11 to 14 by standard rules of error propagation [10].

Evaluation of time period and birth cohort coefficients were performed for the age-period-cohort model using an iterative technique analogous to one proposed by Luebeck and Moolgavkar [9]. Initially, we assumed that the birth cohort effect is absent ($v = 1$) and evaluated coefficients of the time period effect, $v$, using the system of equations (Equation 5). Then, we fixed the obtained time period coefficients and corrected the observed incidence rates, dividing them by the coefficients presented in Equation 1. Then, we estimated the birth cohort effect coefficients, $v$, using the system of equations (Equations 11 to 14) with incidence rates that already had been corrected for the time effect. Assuming $v = 1$ in Equation 1, and using the estimated birth cohort effect coefficients $v$, the incidence rates can be corrected one more time. Numerical experiments showed that one iteration was sufficient to obtain stable estimates for time period and birth cohort coefficients [data not shown].

4. Time Period and Birth Cohort Adjustment for KC and PC

The proposed approach was applied for the estimation of time period and birth cohort effects for KC and PC in white males and females using SEER 9 filtered data. The obtained results are presented on Figures 1 and 2.

Panels A and B of Figure 1 show age distributions of the KC incidence rate in males without and with adjustments, correspondingly. Time periods of observations are presented in legends. As can be seen from these panels, the adjusted incidence rate distribution exhibits a tighter pattern compared to the unadjusted one. The distribution increases monotonically starting from the age of ~30, reaches its maximum at ~72-73 and then falls at older ages. As is seen from Figures 1 and 2, adjustment procedure does not change the PC and KC cancer onset ages, which can be obtained directly from the SEER data.
Panels C and D of Figure 1 show age distributions of the KC incidence rates in females without and with adjustments, correspondingly. This distribution also increases monotonically from the age of ~30, reaches its maximum at ~77-78 and then falls at older ages. The gender-specific differences between the corresponding incidence rate distributions are in ages at which the maximum of the distributions are reached and in the incidence rate maximum values. For males, the maximum of about 50 cases per 100,000 population is reached at ~72-73, while for females the maximum incidence rate (~30 cases per 100,000 population) is reached at ~77-78.

Panels A and B of Figure 2 show age distributions of the PC incidence rates in males without and with adjustments, correspondingly. Time periods of observations are presented in legends. As can be seen from these panels, the adjusted incidence rate distribution also exhibits a tighter pattern compared to the unadjusted one. The distribution rises starting from age ~30, reaches its maximum at ~77-78 and then falls at older ages.

**Figure 1. Influence of time period and birth cohort effects on the distribution of age-specific incidence rates of KC in white males and females.**
Panels C and D of Figure 2 show the age distributions of the PC incidence rates in females without and with adjustments, correspondingly. This distribution also rises from the age of ~30, reaches its maximum at ~77-78 and then falls at old ages. The gender-specific differences between the corresponding incidence rate distributions of PC are not as pronounced as ones for KC. For males, the maximum of about 45 cases per 100,000 population is reached at ~77-78, while for females the maximum incidence rate ~40 cases per 100,000 population is reached at the same age as for males.

In contrast to the existing approaches, the proposed method does not require any a priori knowledge on the mathematical form of the hazard function. To solve the identifiability problem, we assumed that the cohort effect is absent. (This assumption, however, appears to be too strong and in our ongoing work, we are focusing on the use of a weaker assumption.)

We utilized the proposed approach to analyze incidence rates of KC and PC in white males and females using the SEER 9 data on the first primary, microscopically confirmed cases. We observed the turnover in the PC and KC incidence rate distributions at old ages. The systematically increasing trends for time period coefficients were shown for KC in both males and females, while the trends of birth cohort effect coefficients remained unchanged. In the case of PC, no systematic changes in time period and cohort effect coefficients were observed in males and females.

5. Conclusion and Future Directions

In this work, we proposed a new approach to estimate effects of time period and birth cohort coefficients on age-specific incidence rates of cancers. This approach is aimed to solve the well-known computational problem of identifiability [1-3, 9], when defining the corresponding coefficients of the time period effect, $v_i$, and birth cohort effect, $u_i$, and the hazard function parameters, $h(t_i)$.
6. Acknowledgments

We would like to thank Dr. Leo Kinarsky for fruitful discussion and helpful remarks. We acknowledge the use of the UNMC Bioinformatics and Molecular Modeling Facility, partially supported by Nebraska Research Initiative and NIH Cancer Center Support Grant (3 P30 CA036727).

7. References


