A Novel Approach for Analysis of the Log-Linear Age-Period-Cohort Model: Application to Lung Cancer Incidence

Tengiz Mdzinarishvili, Michael X. Gleason and Simon Sherman
Eppley Cancer Institute, University of Nebraska Medical Center, 986805 Nebraska Medical Center, Omaha, NE.
Email: ssperm@unmc.edu

Abstract: A simple, computationally efficient procedure for analyses of the time period and birth cohort effects on the distribution of the age-specific incidence rates of cancers is proposed. Assuming that cohort effects for neighboring cohorts are almost equal and using the Log-Linear Age-Period-Cohort Model, this procedure allows one to evaluate temporal trends and birth cohort variations of any type of cancer without prior knowledge of the hazard function. This procedure was used to estimate the influence of time period and birth cohort effects on the distribution of the age-specific incidence rates of first primary, microscopically confirmed lung cancer (LC) cases from the SEER9 database. It was shown that since 1975, the time period effect coefficients for men increase up to 1980 and then decrease until 2004. For women, these coefficients increase from 1975 up to 1990 and then remain nearly constant. The LC birth cohort effect coefficients for men and women increase from the cohort of 1890–94 until the cohort of 1925–29, then decrease until the cohort of 1950–54 and then remain almost unchanged. Overall, LC incidence rates, adjusted by period and cohort effects, increase up to the age of about 72–75, turn over, and then fall after the age of 75–78. The peak of the adjusted rates in men is around the age of 77–78, while in women, it is around the age of 72–73. Therefore, these results suggest that the age distribution of the incidence rates in men and women fall at old ages.

Keywords: identifiability problem, temporal trend, cohort effect, cancer incidence, lung cancer
Introduction

It is well recognized that aging plays a fundamental role in the development of cancer in the human adult population. To describe the relationship between cancer incidence rates and the age of cancer presentation several mathematical models have been proposed (see, for example, \textsuperscript{1–5} and references therein). In these models, the distribution of incidence rates are presented as a set of numbers, \( I_j(t_i) \), of new cases of a particular type of cancer that have been diagnosed during a given time period, \( j \), per 100,000 population at each of the considered age intervals, \( t_i \).

Often, five-year-long age intervals are considered. For instance, 100 years of human life span can be divided on 20 five-year intervals: 0–4, 5–9, 10–14, \ldots, 95–99. The center of these intervals can represent the corresponding age. The values \( I_j(t_i) \), representing cancer incidence in each of these age intervals, are called the age-specific cancer incidence rates (or simply cancer rates). These rates can be collected during one or several calendar years. In the mathematical modeling of the age distribution of cancer rates, a five-year age period is commonly used. For instance, \( I_j(t_i) \) may represent numbers of new cancer cases in each of the aforementioned age intervals diagnosed in the \( j \)-th time period that correspond to, for example, 2000–2004. Analyses of this type of data are called cross-sectional studies, because these studies aim to analyze data on people at different ages at the same time period.

The cross-sectional studies of the cancer incidence rate distributions in aging are different from the longitudinal studies of the analogous distributions. In the cross-sectional studies, the number of new cancer diagnoses can be counted simultaneously for different cohorts of people at a given time period, while in the longitudinal studies this data must be obtained for the same cohort of people but in different time periods. Each of these types of studies, cross-sectional vs. longitudinal, has their own advantages and disadvantages. For instance, it is clear that data for cross-sectional studies can be obtained much faster than for the longitudinal studies. In fact, to perform the aforementioned cross-sectional study, one has to collect data over a time period of five years (2000–2004), while for the analogous longitudinal studies, using a cohort of people born, say, in 1905–1909, to get data for all of the considered age intervals one must collect the corresponding incidence rates over 100 years. In addition, studies of cross-sectional data can provide clues to possible time period effects during which data was collected. For instance, implementation of new diagnostic techniques in a particular time period could influence the detection of a given cancer type at earlier ages (age-period effect). On the other hand, longitudinal studies (in contrast to cross-sectional data) can determine the influence of cohort effects on the age distribution of cancer rates (age-cohort effect). For example, dietary and life-style habits characteristic for a given generation of people can affect the cancer incidence rates.

In this connection, it must be emphasized that cross-sectional and longitudinal studies of cancer incidence rates performed independently can provide inconsistent or even confusing results. For instance, recently, using the SEER (Surveillance Epidemiology and End Results) database,\textsuperscript{6} Harding and colleagues\textsuperscript{4} analyzed the distribution of age-specific cancer incidence rates. For the vast majority of the examined cancers they found that the rates collected during three time periods, 1979–83, 1989–93 and 1999–2003 (cross-sectional data) increase up to the age of about 80 years and then fall at the oldest ages. However, the longitudinal data for lung cancer (LC) presented by Holford\textsuperscript{7} showed that the LC cohort risk increases with age, while the LC time period risk falls at old age. Moreover, it was suggested that there was no turnover, if both time period and cohort effects on cancer rates were considered.\textsuperscript{8}

Accounting for time period and cohort effects (age-period-cohort model) represents a main challenge for mathematical modeling of relationship between cancer incidence rates and age of cancer presentation. This is because mathematically this problem falls into a category of so called identifiability problems with multiple estimators.\textsuperscript{9} In general, parameters to be determined (i.e. estimates for time period and cohort effects) as a solution of the considered problem, cannot be unambiguously identified. In other words, multiple estimators can provide equally good solutions for the problem and “true” age-period-cohort effects are difficult (if not impossible) to estimate simultaneously.\textsuperscript{9–15} The only hope for solving this problem (obtain consistent estimation for period and cohort...
effects) is to utilize an additional assumption on the data that is used.

Until recently, mathematical modeling of the relationship between cancer incidence rates and the age of cancer presentation has been performed exclusively using cross-sectional data. To address this shortcoming, we are proposing a simple, computationally efficient procedure for analyses of time period and birth cohort effects on the distribution of the age-specific incidence rates of cancers. Assuming that cohort effects for neighboring cohorts are almost equal and using the Log-Linear Age-Period-Cohort (LLAPC) Model, this procedure allows one to evaluate temporal trends and birth cohort variations of any type of cancer without prior knowledge of the hazard function. The proposed approach was used to analyze the influence of the time period and birth cohort effects on the LC incidence rate distributions. Only first primary, microscopically confirmed cases from the SEER9 database over the period of 1975–2004 were considered. Using a novel approach, which is valid for any hazard function, we demonstrated that the time period trends in men and women are different in LC, while the cohort trends are similar. We also demonstrated that the distribution of these incidence rates falls at old ages, even after accounting for time period and birth cohort effects.

Materials and Methods
We describe a novel, computationally efficient procedure for the analysis of the time period and birth cohort effects in the frame of the LLAPC model. This procedure is tested on the example of LC.

Data preparation and filtration
In our study, we used data from only the SEER registries that correspond to the following nine (SEER9) areas: Atlanta, Connecticut, Detroit, Hawaii, Iowa, New Mexico, San Francisco-Oakland, Seattle-Puget Sound, and Utah. We used these nine registries, rather than the current set of seventeen, because the longitudinal nature of our study requires us to use data dating back two decades when there were only nine registries. First primary, microscopically confirmed LC cases from the SEER9 database for patients with known gender and race were considered to be “filtered” data, whereas the cases where such filtering was not performed were considered to be “raw” data. We used only filtered data that are more reliable and homogeneous than raw data. The incidence rates, \( I(t) \), expressed per 100,000 persons and age-adjusted by the direct method to the 2000 United States standard population, and their standard errors, \( SE \), were utilized. The data were combined in six five-year cross-sectional time periods: 1975–79, …, 2000–2004. The gender-specific incidence rates were grouped into 18 five-year age groups: 17 groups, ranging from 0 to 84 years, and the 18th group that included all cases for ages 85+.

A novel procedure for analysis of the LLAPC model
Table 1 presents approximations of the observed incidence rates, \( I_i(t) \), as the product of a hazard function, \( h(t) \) (which is a function of age \( t \)), the time period effect coefficient, \( v_i \), and the birth cohort effect coefficient, \( u_j \):

\[
I_i(t) = v_i u_j h(t) \quad i = 1, \ldots, n; j = 1, \ldots, m; l = 1, \ldots, k; (1)
\]

where \( i, j, \) and \( l \) denote the given age interval, time period, and cohort, correspondingly; \( n, m, \) and \( k \) are numbers of the age intervals, time periods, and cohorts, correspondingly.

In this table, the approximations of the cross-sectional data for six time periods 1975–79, …, 2000–2004 (index \( j = 1, \ldots, 6 \)) are shown in columns, while the approximations of the incidence rates for the same cohort groups (longitudinal data) are located along diagonals. We used only the data for the groups over age 30 (\( i = 7, \ldots, 18 \)), because the incidence rates for these groups were significant (according to SEER practice, the number of cases should exceed 15 to be statistically significant). We consider 17 birth cohorts (\( l = 1, \ldots, 17 \)), corresponding to birth year ranges of 1890–94, ... 1970–74. From this table one can see that \( l \) can be presented as \( l = j - i + 18 \).

Assuming that the numbers of cases have a Poisson distribution and the mathematical form of the hazard function is known a priori and the LLAPC model is used, one can make adjustments by using the maximum likelihood method for assessing the birth cohort and time period effect coefficients as well as parameters of the hazard function. These coefficients can be estimated by anchoring one time...
| Age group | Period of observation | Birth cohort |
|-----------|----------------------|--------------|
|           | 1975–79   | 1980–84   | 1985–89   | 1990–94   | 1995–99   | 2000–04   |                  |
| i | mp, t, | j = 1 | j = 2 | j = 3 | j = 4 | j = 5 | j = 6 | I, l Years, 19xx |
| 1 | 2.5 |                  |         |       |       |       |       |    |
| 7 | 32.5 | v₃u₁₂ h(t₁) | v₂u₁₃ h(t₁) | v₃u₁₄ h(t₁) | v₄u₁₅ h(t₁) | v₅u₁₆ h(t₁) | v₆u₁₇ h(t₁) | 17 | 70–74 |
| 8 | 37.5 | v₄u₁₀ h(t₁) | v₃u₁₁ h(t₁) | v₄u₁₂ h(t₁) | v₅u₁₃ h(t₁) | v₆u₁₄ h(t₁) | v₇u₁₅ h(t₁) | 16 | 65–69 |
| 9 | 42.5 | v₅u₁₀ h(t₁) | v₄u₁₁ h(t₁) | v₅u₁₂ h(t₁) | v₆u₁₃ h(t₁) | v₇u₁₄ h(t₁) | v₈u₁₅ h(t₁) | 15 | 60–64 |
| 10 | 47.5 | v₆u₉ h(t₁) | v₅u₁₀ h(t₁) | v₆u₁₁ h(t₁) | v₇u₁₂ h(t₁) | v₈u₁₃ h(t₁) | v₉u₁₄ h(t₁) | 14 | 55–59 |
| 11 | 52.5 | v₇u₈ h(t₁) | v₆u₉ h(t₁) | v₇u₁₀ h(t₁) | v₈u₁₁ h(t₁) | v₉u₁₂ h(t₁) | v₁₀u₁₃ h(t₁) | 13 | 50–54 |
| 12 | 57.5 | v₈u₇ h(t₁) | v₇u₈ h(t₁) | v₈u₉ h(t₁) | v₉u₁₀ h(t₁) | v₁₀u₁₁ h(t₁) | v₁₁u₁₂ h(t₁) | 12 | 45–49 |
| 13 | 62.5 | v₉u₆ h(t₁) | v₈u₇ h(t₁) | v₉u₈ h(t₁) | v₁₀u₉ h(t₁) | v₁₁u₁₀ h(t₁) | v₁₂u₁₁ h(t₁) | 11 | 40–44 |
| 14 | 67.5 | v₁₀u₅ h(t₁) | v₉u₆ h(t₁) | v₁₀u₇ h(t₁) | v₁₁u₈ h(t₁) | v₁₂u₉ h(t₁) | v₁₃u₁₀ h(t₁) | 10 | 35–39 |
| 15 | 72.5 | v₁₁u₄ h(t₁) | v₁₀u₅ h(t₁) | v₁₁u₆ h(t₁) | v₁₂u₇ h(t₁) | v₁₃u₈ h(t₁) | v₁₄u₉ h(t₁) | 9 | 30–34 |
| 16 | 77.5 | v₁₂u₃ h(t₁) | v₁₁u₄ h(t₁) | v₁₂u₅ h(t₁) | v₁₃u₆ h(t₁) | v₁₄u₇ h(t₁) | v₁₅u₈ h(t₁) | 8 | 25–29 |
| 17 | 82.5 | v₁₃u₂ h(t₁) | v₁₂u₃ h(t₁) | v₁₃u₄ h(t₁) | v₁₄u₅ h(t₁) | v₁₅u₆ h(t₁) | v₁₆u₇ h(t₁) | 7 | 20–24 |
| 18 | 87.5⁺ | v₁₄u₁ h(t₁) | v₁₃u₂ h(t₁) | v₁₄u₃ h(t₁) | v₁₅u₄ h(t₁) | v₁₆u₅ h(t₁) | v₁₇u₆ h(t₁) | 6 | 15–19 |

| 1890–94 | 1895–99 | 1900–04 | 1905–09 | 1910–14 |

The results of this procedure depend on the hazard function, and also the time period and cohort, to which the coefficients are anchored.

Below we describe a procedure that provides results independent of the hazard function. The hazard function values, presented in Table 1, can be canceled out by dividing the corresponding elements of the neighboring columns with indices \( j \) and \( j + 1 \) or \( j + 1 \) and \( j \). Then from (1), one can obtain a pair of systems:

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

\( i = 1, \ldots, 18; \ j = 1, \ldots, 5; \ l = j - i + 18 \)
\[
I_{i,j+i}(t_i) = \frac{v_{j+i} u_{j+i}}{I_{i,j}(t_i)} \\
i = 7, \ldots, 18; j = 1, \ldots, 5; l = j - i + 18 \tag{3}
\]

In (2) and (3), \( I_{i,j}(t) \) can be considered as the observed values of the normally distributed variables with known standard errors, \( SE_{i,j} \) (Because we are using only the incidence rates with numbers of cases larger than 15, the normal distribution can be used instead of the Poisson distribution). Coefficients of variation for these rates can be estimated as:

\[
CV_{i,j} = \frac{SE_{i,j}}{I_{i,j}(t_i)} \\
i = 7, \ldots, 18; j = 1, \ldots, 6; l = j - i + 18 \tag{4}
\]

Note: (2) provides 12 \( \times \) 5 conditional equations for assessing five ratios of the time period coefficients \((v_j/v_{j+1}, j = 1, \ldots, 5)\), and 16 ratios of the cohort effect coefficients \((u_i/u_{i+1}, I = 1, \ldots, 16)\). Analogously, (3) provides 12 \( \times \) 5 conditional equations for assessing five ratios of the time period coefficients \((v_j/v_{j+1}, j = 1, \ldots, 5)\), and 16 ratios of the cohort effect coefficients \((u_i/u_{i+1}, I = 1, \ldots, 16)\). Here, a problem of parameter identifiability arises.\(^7\) In particular, these systems do not have a single set of best estimates of \(v_j/v_{j+1}\) and \(u_i/u_{i+1}\). In fact, suppose we obtained the best estimates of \(v_j/v_{j+1}\) and \(u_i/u_{i+1}\). Now, if we multiply the estimates of \(v_j/v_{j+1}\) by a constant and divide the estimates of \(u_i/u_{i+1}\) on this constant, the derived set of new estimates will also be as good as the initially assumed “best” estimates.

In order to solve this identifiability problem, additional assumptions are required.\(^9\)\(^-\)\(^15\) Assuming that any pair of the neighboring cohorts has the cohort effect coefficient ratio close to 1, these ratios can be set equal to 1 in (2) and (3). The rationale behind this assumption is that the adjacent cohorts usually overlap in time intervals and thus values of their cohort effect coefficient should be close. Now for estimating five ratios, \(v_j/v_{j+1}\) and five ratios \(v_{j+1}/v_j\), one has a pair of systems:

\[
I_{i,j+i}(t_i) = \frac{v_{j+i} u_{j+i}}{I_{i,j}(t_i)} \\
i = 7, \ldots, 18; j = 1, \ldots, 5. \tag{5}
\]

\[
I_{i,j+i}(t_i) = \frac{v_{j+i} u_{j+i}}{I_{i,j}(t_i)} \\
i = 7, \ldots, 18; j = 1, \ldots, 5. \tag{6}
\]

When coefficients of variation (4) are small, the standard errors of the ratios, \(I_{i,j}(t)/I_{j,i+1}(t_i)\), \(I_{i,j+i}(t_i)/I_{i,j}(t_i)\) can be calculated by the standard rules of error propagation.\(^20\)

\[
SE^2 \left( \frac{I_{i,j}(t_i)}{I_{i,j+i}(t_i)} \right) = \left( \frac{I_{i,j}(t_i)}{I_{i,j+i}(t_i)} \right)^2 \left( \frac{SE^2_{i,j}(t_i)}{I^2_{i,j}(t_i)} + \frac{SE^2_{i,j+i}(t_i)}{I^2_{i,j+i}(t_i)} \right) \tag{7}
\]

\[
SE^2 \left( \frac{I_{i,j+i}(t_i)}{I_{i,j}(t_i)} \right) = \left( \frac{I_{i,j+i}(t_i)}{I_{i,j}(t_i)} \right)^2 \left( \frac{SE^2_{i,j+i}(t_i)}{I^2_{i,j+i}(t_i)} + \frac{SE^2_{i,j}(t_i)}{I^2_{i,j}(t_i)} \right) \tag{8}
\]

It can be shown that when the numerators and denominators in (5) and (6) are normally distributed and their coefficients of variation are small, then the ratios presented in (5) and (6) will be also normally distributed. In fact, let us assume that we have random variables \(A_1 = a_1 + \varepsilon_1\) and \(A_2 = a_2 + \varepsilon_2\), where \(a_1 \neq 0\) and \(a_2 \neq 0\) are constants and \(\varepsilon_1\) and \(\varepsilon_2\) are normally distributed random variables with zero means and standard deviations \(\sigma_1\) and \(\sigma_2\), correspondingly. When coefficients of variation are small (i.e. \(\sigma_1/a_1 << 1\) and \(\sigma_2/a_2 << 1\)), then one can express the \(A_1/A_2\) and \(A_2/A_1\) ratios in the bivariate Taylor series around \(a_1/a_2\) and \(a_2/a_1\), and consider their linear approximations:

\[
A_1 = a_1 + \varepsilon_1 = a_1 + \frac{1}{a_2} \varepsilon_1 - a_2 \frac{a_1}{a_2} \varepsilon_2 \\
A_2 = a_2 + \varepsilon_2 = a_2 + \frac{1}{a_1} \varepsilon_2 - a_1 \frac{a_2}{a_1} \varepsilon_1
\]

Because \(\varepsilon_1\) and \(\varepsilon_2\) are normally distributed variables, these linear combinations will be also normally distributed.

In the considered incidence rate data, coefficients of variation are less than 0.1, therefore the errors of the observed incidence rate ratios of the systems (5) and (6) can be considered as normally distributed. For estimation of \(v_j/v_{j+1}\) and \(v_{j+1}/v_j\), a least squares
method can be applied and the most efficient estimates for these ratios are the weighted means of the observed values \( I_{i,j}(t_{i})/I_{i,j+1}(t_{i}) \) and \( I_{i,j+1}(t_{i})/I_{i,j}(t_{i}) \) averaged through index \( i \) correspondingly (weights are given as reciprocals of the square of their standard errors). The \( SE \) of the estimates \((v_{i}/v_{j})^{*}\) and \((v_{j+1}/v_{j})^{*}\) can also be calculated in a standard way (here and below asterisks (*) designate the corresponding estimates).

After anchoring any time period coefficient (for example, assuming that \( v_{6} = 1 \) and \( SE(v_{6}) = 0 \), one can obtain step by step the following estimates of \( v_{j}^{*} \) derived by the estimated ratios \((v_{j+1}/v_{j})^{*}\):

\[
\begin{align*}
\tau_{5}^{*} &= \left( \frac{v_{3}}{v_{6}} \right)^{*}; \quad \tau_{4}^{*} = \left( \frac{v_{4}}{v_{5}} \right)^{*} v_{3}^{*}; \quad \tau_{3}^{*} = \left( \frac{v_{3}}{v_{4}} \right)^{*} v_{4}^{*}; \\
\tau_{2}^{*} &= \left( \frac{v_{2}}{v_{3}} \right)^{*} v_{3}^{*}; \quad \tau_{1}^{*} = \left( \frac{v_{1}}{v_{2}} \right)^{*} v_{2}^{*} \\
&= (9)
\end{align*}
\]

For the other anchored time period coefficients, the estimates of \( v_{j}^{*} \) can be derived analogously by means of the estimates \((v_{j+1}/v_{j})^{*}\) and \((v_{j+1}/v_{j})^{*}\). The \( SE \) of \( v_{j}^{*} \) can also be calculated by the standard rules of error propagation.

After estimating the time period coefficients, the incidence rates can be corrected for the time effects and the following system can be obtained from (2) and (3):

\[
\begin{align*}
I_{i,j}^{T}(t_{i}) &= \frac{u_{i}}{u_{i+1}}, \quad l = 1, \ldots, 17; \quad j = 1, \ldots, 5; \quad i = j - 1 + 18 \quad (10) \\
I_{i,j}^{T}(t_{i}) &= \frac{u_{i+1}}{u_{i}}, \quad l = 1, \ldots, 17; \quad j = 1, \ldots, 5; \quad i = j - 1 + 218 \quad (11)
\end{align*}
\]

Here \( I_{i,j}^{T}(t_{i}) \) denotes incidence rates corrected for the time effects. By the standard rules of error propagation, one can calculate the \( SE \) of the incidence rates ratios presented on the left side of (10) and (11). Now, there are \( 12 \times 5 \) conditional equations for assessing 16 ratios of the cohort effect coefficients \( u_{l}/u_{l+1} \) and the same number of equations for assessing \( u_{l+1}/u_{l} \). Similarly, as for ratios of time period coefficients, the ratios \( u_{i}/u_{i+1} \) and \( u_{i+1}/u_{i} \) can be estimated by the weighted means of the incidence rate ratios of the left side of system (10) and (11). Weights should be given according to the \( SE \) of the incidence rate ratios (reciprocal of squares of the \( SE \)). Then, by choosing an anchored cohort coefficient, say \( u_{l} = 1 \), \( SE(u_{l}) = 0 \), all cohort coefficients and their \( SE \) can be estimated step-by-step by a procedure analogous to one used for the time period coefficients:

\[
\begin{align*}
& \tau_{10}^{*} = \left( \frac{u_{10}}{u_{9}} \right)^{*}, \quad \tau_{11}^{*} = \left( \frac{u_{11}}{u_{10}} \right)^{*} u_{10}^{*}, \ldots, \\
& \tau_{17}^{*} = \left( \frac{u_{17}}{u_{16}} \right)^{*} u_{16}^{*} \quad (12)
\end{align*}
\]

and

\[
\begin{align*}
& \tau_{6}^{*} = \left( \frac{u_{6}}{u_{5}} \right)^{*}, \quad \tau_{7}^{*} = \left( \frac{u_{7}}{u_{6}} \right)^{*} u_{6}^{*}, \ldots, \\
& \tau_{1}^{*} = \left( \frac{u_{1}}{u_{2}} \right)^{*} u_{2}^{*} \quad (13)
\end{align*}
\]

After evaluating time period and cohort effect coefficients, one can divide the initial incidence rates, \( I_{i,j}(t_{i}) \), with the product of \( v_{j}^{*} \) and \( u_{j}^{*} \), to obtain the incidence rates corrected for time period and cohort effects:

\[
\begin{align*}
I_{i,j}^{*}(t_{i}) &= \frac{I_{i,j}(t_{i})}{v_{j} u_{j}}; \quad i = 7, \ldots, 18; \quad j = 1, \ldots, 6; \quad l = j - 1 + 18 \quad (14)
\end{align*}
\]

The aforementioned approach looks similar to one used in.\(^{18}\) However, the approach used in\(^{18}\) is based on the assumption that the birth cohort effects are absent. This allows the authors of\(^{18}\) to evaluate coefficients \( v_{j} \). Then, using the obtained time period effects, they correct the observed incidence rates and after that, estimate the \( u_{l} \) coefficients. In contrast to\(^{18}\) our approach uses the assumption that cohort effects for any two neighboring cohorts are almost equal. As can be seen, this assumption is not as strong as one used in.\(^{18}\) In addition, our procedure can assess time period and cohort coefficients without knowing the mathematical form of the hazard function, while
the procedure described in\(^{18}\) requires prior knowledge of this function.

**Results and Discussion**
Table 2 shows the time period distributions (presented in columns) of the first primary, microscopically confirmed incidence LC rates for women. The observed patterns of the cross-sectional data are shown along columns and longitudinal data along diagonals. The cross-sectional and longitudinal data for the three consecutive cohorts that contain observations for the elderly exhibit turnovers at old age. Analogous observations for men also have turnovers (data not shown).

**Table 2.** Filtered LC incidence rates for women collected in the SEER9 database during 1975–2004. Arrows show cohorts that turn over.

| Age group | Period of observation | Birth cohort |
|-----------|-----------------------|--------------|
|           | 1975–79 | 1980–84 | 1985–89 | 1990–94 | 1995–99 | 2000–04 |         |
| \(i\) | mp, \(t_i\) | \(j = 1\) | \(j = 2\) | \(j = 3\) | \(j = 4\) | \(j = 5\) | \(j = 6\) |
| 1   | 2.5   |       |       |       |       |       |       |
| ... | ...   |       |       |       |       |       |       |
| 7   | 32.5  | 4.4   | 3.0   | 2.6   | 2.5   | 2.4   | 2.4   |
| 8   | 37.5  | 12.7  | 10.5  | 8.1   | 7.6   | 7.0   | 6.8   |
| 9   | 42.5  | 32.4  | 25.9  | 20.1  | 18.6  | 16.5  | 16.4  |
| 10  | 47.5  | 59.7  | 47.5  | 42.5  | 40.5  | 39.0  | 40.4  |
| 11  | 52.5  | 87.6  | 82.2  | 72.8  | 69.0  | 65.5  | 68.3  |
| 12  | 57.5  | \(133.8\) | 115.9 | 110.4 | 107.8 | 103.6 | 110.2 |
| 13  | 62.5  | \(173.7\) | 159.3 | 150.2 | 141.9 | 146.4 | 151.9 |
| 14  | 67.5  | \(197.0\) | 193.2 | 182.0 | 185.2 | 191.6 | 189.9 |
| 15  | 72.5  | 212.0  | 203.7 | 201.6 | 204.7 | 208.3 | 218.3 |
| 16  | 77.5  | 184.2  | 189.1 | 195.5 | 195.7 | \(200.2\) | 212.8 |
| 17  | 82.5  | 143.6  | 153.9 | 159.5 | \(157.4\) | \(162.2\) | \(175.3\) |
| 18  | 87.5+ | 102.3  | 102.3 | 92.4  | 82.3  | \(84.0\) | \(84.2\) |
| \(1\) | \(2\) | \(3\) | \(4\) | \(5\) | \(6\) | \(7\) | \(8\) |
| \(1890–94\) | \(1895–99\) | \(1900–04\) | \(1905–09\) | \(1910–14\) |
The longitudinal patterns shown in Table 2 are different from those presented in. In, using the raw LC incidence rates for women collected in Connecticut during the years of 1940–1984, it was shown that the longitudinal risks always increased with age (see Table 1 in\textsuperscript{7}). This discrepancy can be explained by the fact that in contrast to,\textsuperscript{7} where raw data was used, we analyzed filtered SEER9 data collected during 1975–2004.

Figures 1A and 2A show the distributions of the filtered incidence rates for men and women, correspondingly.

After estimating the time period and birth cohort effects, the incidence rates were adjusted to the 2000–2004 time period and to the 1945–1949 birth cohort. The adjusted rates are shown in Figures 1B and 2B. As can be seen, the adjusted LC incidence rate distributions for both men and women have turnovers at old ages. The adjusted rates for men and women increase (starting from the age of 30), reach the maximum and then fall at old ages. The differences are only in the age at which the distributions reach the maximum, and in the maximum values of the corresponding incidence rates. For men, this maximum is near the age of 77–78, while for women, it is near the age of 72–73. These patterns are different from the linear patterns (up to the age of 85) obtained in\textsuperscript{8} by accounting for time period and cohort effects on cancer. Again, this discrepancy can be explained by the use of raw data and an \textit{a priori} assumed form of the hazard function for the time period and birth cohort adjustments used in,\textsuperscript{8} whereas we utilized only filtered data and our approach is independent of the hazard function.

Panels C and D of Figures 1 and 2 show the changes of the time period and cohort effect coefficients for men and women, correspondingly. The
time period effect coefficients for men increase from the year 1975 to 1980 and then decrease until 2004. For women, these effects increase from 1975 to 1990 and then remain nearly constant. The birth cohort effect coefficients for men and women are similar; they increase from the cohort of 1890–94 until the cohort of 1925–29, then decrease until the cohort of 1950–54 and after that remain almost unchanged. It is possible that the observed temporal differences of the LC rates in men and women can be explained by the gender-specific smoking habits as it was suggested in\textsuperscript{21} (see also references in that paper).

**Conclusion**

For analyses of the time period and birth cohort effects on the distribution of the age-specific incidence rates of cancers, a simple, computationally efficient procedure, which does not require any prior knowledge of the hazard function, was proposed. Our approach uses the LLAPC model and assumes that cohort effects for neighboring cohorts are almost equal. The proposed procedure was used for analyzing the influence of the time period and birth cohort effects on the LC incidence rate distributions. However, this procedure can be applied for different types of cancers as well as for epidemiological studies of chronic diseases.

We found that the incidence rates of first primary, microscopically confirmed LC cases from the SEER\textsuperscript{9} database, adjusted by period and cohort effects, increase for both women and men, then turn over (at ages of about 72–73 and 77–78 for women and men, correspondingly) and fall at older ages. Thus, by utilizing the longitudinal and cross-sectional data and by accounting for time period and cohort effects, we have demonstrated that the LC incidence rates have a turnover at old ages, and the age at which this turnover

**Figure 2.** Age-specific incidence rates, time period and cohort effect coefficients for LC in women. A) Rates prior to adjustment. B) Rates after adjustment (age-specific weighted means are connected by the use of the spline function). C) Variation of time period effect coefficients with time period index, anchored at index 6. D) Variation of cohort effect coefficients with cohort index, anchored at index 9. In C and D, error bars denote 95% confidence intervals.
takes place, is gender-specific. The explanation of this phenomenon should be a subject for future studies.

Disclosures
The authors report no conflicts of interest.

References
1. Armitage P, Doll R. The age distribution of cancer and a multistage theory of carcinogenesis. Br J Cancer. 1954;8:1–12.
2. Cook PJ, Doll R, Fellingham SA. A mathematical model for the age distribution of cancer in man. Int J Cancer. 1969;4:93–112.
3. Pompei F, Wilson R. Age distribution of cancer: the incidence turnover at old age. Hum Ecol Risk Assess. 2001;7:1619–50.
4. Harding C, Pompei F, Lee E, Wilson R. Cancer Suppression at Old Age. Cancer Res. 2008;68:4465–78.
5. Mdzinarishvili T, Gleason MX, Kinarsky L, Sherman S. A generalized beta model for age distribution of cancers: application to pancreatic and kidney cancer. Cancer Inform. 2009;7:183–97.
6. Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov) Limited-Use Data (1973–2004), National Cancer Institute, DCCPS, Surveillance Research Program, Cancer Statistics Branch, released April 2007, based on the November 2006 submission.
7. Holford TR. Understanding the effects of age, period, and cohort on incidence and mortality rates. Am Rev Public Health. 1991;12:425–57.
8. Meza R, Jean J, Moolgavkar SH, Luebeck EG. Age-specific incidence of cancer: phases, transitions, and biological implications. Proc Natl Acad Sci U S A. 2008;105:16284–9.
9. Fu WJA. Smoothing Cohort Model in Age-Period-Cohort Analysis with Applications to Homicide Arrest Rates Lung Cancer Mortality Rates. Sociol Method Res. 2008;36:327–61.
10. Mason KO, Mason WM, Winsborough HH, Poole WK. Some Methodological Issues in the Cohort Analysis of Archival Data. Am Sociol Rev. 1973;38:242–58.
11. Rogers WL. Estimable Functions in Age, Period and Cohort Effects. Am Sociol Rev. 1982;47:774–87.
12. Clayton D, Schifflers E. Models for temporal variation in cancer rates. I: age-period and age-cohort models. Statistics in Medicine. 1987;6:449–67.
13. Clayton D, Schifflers E. Models for temporal variation in cancer rates. II: age-period-cohort models. Statistics in Medicine. 1987;6:469–81.
14. Moolgavkar SH, Lee JAH, Stevens RG. Analysis of vital statistical data. In: Rothman K, Greenland S, editors. Modern Epidemiology, 2nd Ed. Lippincott-Raven, PA; 1998:482–97.
15. Yang Y, Fu WJ, Land K. A Methodological Comparison of Age-Period-Cohort Models: The Intrinsic Estimator and Conventional Generalized Linear Models. Socio Methods. 2004;34:75–110.
16. Ries LA, Young JL, Keel GE, Eisner MP, Lin YD, Horner MJ (editors). SEER Surveillance Monograph: Cancer Survival Among Adults: U.S. SEER Program, 1988–2001, Patient and Tumor Characteristics. National Cancer Institute, SEER Program, NIH Pub. No. 07-6215, Bethesda, MD, 2007.
17. Surveillance, Epidemiology, and End Results (SEER) Program. Standard Populations (Millions) for Age-Adjustment [cited 2009 Feb 2]. Available from: http://seer.cancer.gov/stdpopulations/stdpop.singleagesfthr99.txt.
18. Luebeck EG, Moolgavkar SH. Multistage carcinogenesis and the incidence of colorectal cancer. Proc Natl Acad Sci U S A. 2002;99:15095–100.
19. Moolgavkar SH, Meza R, Turin J. Pleural and peritoneal mesotheliomas in SEER: age effects and temporal trends, 1973–2005. Cancer Causes Control. 2009;20(6):935–44.
20. Lindberg V. Guide to uncertainties and error propagation. Rochester, NY; c1999–2003 [updated 2003 Aug; cited 2009 Feb 2]. Available from: http://www.rit.edu/cos/physics/uncertainties/Uncertainties.html.
21. Zheng T, Holford TR, Boyle P, et al. Time trend and the age-period-cohort effect on the incidence of histologic types of lung cancer in Connecticut, 1960–1989. Cancer. 2006;74(5):1556–67.