A Web Tool for Age–Period–Cohort Analysis of Cancer Incidence and Mortality Rates

Philip S. Rosenberg, David P. Check, and William F. Anderson

Abstract

Background: Age–period–cohort (APC) analysis can inform registry-based studies of cancer incidence and mortality, but concerns about statistical identifiability and interpretability, as well as the learning curves of statistical software packages, have limited its uptake.

Methods: We implemented a panel of easy-to-interpret estimable APC functions and corresponding Wald tests in R code that can be accessed through a user-friendly Web tool.

Results: Input data for the Web tool consist of age-specific numbers of events and person-years over time, in the form of a rate matrix of paired columns. Output functions include model-based estimators of cross-sectional and longitudinal age-specific rates, period and cohort rate ratios that incorporate the overall annual percentage change (net drift), and estimators of the age-specific annual percentage change (local drifts). The Web tool includes built-in examples for teaching and demonstration. User data can be input from a Microsoft Excel worksheet or by uploading a comma-separated–value file. Model outputs can be saved in a variety of formats, including R and Excel.

Conclusions: APC methodology can now be carried out through a freely available user-friendly Web tool. The tool can be accessed at http://analysistools.nci.nih.gov/apc/.

Impact: The Web tool can help cancer surveillance researchers make important discoveries about emerging cancer trends and patterns. Cancer Epidemiol Biomarkers Prev; 23(11); 2296–302. ©2014 AACR.

Introduction

Cancer rates are monitored worldwide to assess the burden of cancer and track cancer trends in populations (1–3). Standard statistical methods include examination of plots of disease rates over time (4, 5) and analysis of directly age-standardized rates (ASR) and estimated annual percentage changes (EAPC) of the ASRs (6, 7). These approaches are descriptive, agnostic, and nonparametric.

Cancer rates are also examined to reveal clues about cancer etiology (8–17), natural history (18–21), and mortality (9, 22–27). Parametric statistical models often play a more prominent role in these studies, especially the age–period–cohort (APC) model (5, 28–33). Nonetheless, many studies have not taken advantage of the APC framework (34). One basic problem is concern about statistical identifiability and corresponding uncertainty about how to interpret APC parameters (especially the so-called deviations). However, we have suggested that these issues reflect a fundamental uncertainty principle intrinsic to all cohort studies, rather than a problem specific to the APC model (34). We and others (5) have also pointed out close connections between estimable functions of APC parameters and standard plots of age-standardized and age-specific disease rates over time. In this regard, estimable APC functions provide a useful parametric framework that complements standard nonparametric descriptive methods.

Several APC functions have proved useful in cancer applications. Anderson and colleagues (35) introduced the longitudinal age curve in their investigation of breast cancer black–white racial disparity. The longitudinal age curve "stitches together" observed cohort-specific age-specific rates, thereby providing a smooth summary curve. Speaks and colleagues (36) considered testicular germ cell tumors and Yang and colleagues (37) studied ovary cancer using a period rate ratio curve. This function describes the relative rate of cancer in any given calendar period versus a referent period, adjusted for age and nonlinear cohort effects. Jemal and colleagues (24) studied lung cancer mortality, Ma and colleagues (38) pancreas cancer mortality, and Rosenberg and colleagues (39) leukemia incidence using a cohort rate ratio curve. This function describes the relative rate of cancer in any given birth cohort versus a referent cohort, adjusted for age and nonlinear period effects. Mbulaiteye and colleagues (40) investigated Burkitt lymphoma, and Chaturvedi and...
colleagues (41) examined oral cavity and oropharyngeal squamous cell carcinoma rates using a quantity that we call local drifts. Local drifts provide a model-based EAPC value for each age group. Anderson and colleagues (42) used all of these functions in their study of breast cancer heterogeneity in Denmark.

Since each of these new functions has proved useful in recent studies, we were motivated to make them readily available. However, many investigators with interesting hypotheses about cancer rates are not experts in statistics or familiar with statistical software packages. Therefore, we developed a freely-available user-friendly Web tool that can be accessed at http://analysistools.nci.nih.gov/apc/. The Web tool provides all of the APC functions described above, along with associated statistical hypothesis tests. In this report, we summarize the Web tool and highlight how it can help identify interesting signals in cancer rates using three illustrative examples from the literature (29, 32, 43). The example data are available through the Web tool, and we encourage potential users to work through the examples online.

Materials and Methods

Overview of the Web tool

Online Help is available (click on Help in the Web tool, or open http://analysistools.nci.nih.gov/apc/help.html). Input data for the Web tool consist of age-specific numbers of events and person-years over time, in the form of a rate matrix of paired columns. Three sample datasets that describe prostate (43), lung (29), and breast cancer (32) mortality are linked to the Web tool (click on Help, then Sample Data, or open http://analysistools.nci.nih.gov/apc/help.html#example). The input page is shown in Fig. 1 for the prostate cancer mortality data (example 1). In general, user data can be input by copy and paste from an Excel worksheet or upload a comma-separated–values (csv) file. As shown in Fig. 1, age groups correspond to rows and calendar periods to columns. The rates are defined by adjacent pairs of columns: the first column of each pair lists the numbers of events by age for a given calendar period, and the second column lists the corresponding person-years. The age and period intervals must all be equal (44), i.e., if 5-year age groups are used, then 5-year calendar periods must also be used. The intervals can range from 1 to 10 years inclusive. Data in this format can easily be obtained from publicly available data resources with cancer case and population data, such as the Surveillance, Epidemiology and End Results (SEER) Program of the NCI (http://www.seer.cancer.gov) and Cancer Incidence in Five Continents (CI5) of the International Agency for Research on Cancer (http://ci5.iarc.fr).

The Web tool fits the APC model and calculates parameters and estimable functions summarized in Table 1. On the website, each function is presented in its own tab in graphical and tabular format, as illustrated in Fig. 2. A number of key hypothesis tests are also provided in the “Wald Tests” tab located in the sidebar on the left-hand side of the Web page. These hypothesis tests are summarized in Table 2.

Results

Interpretation of estimable functions

Using outputs from the Web tool (Table 1), the user can interpret the observed rates as a product of age, period, and cohort effects. Different combinations of functions summarize the patterns longitudinally or prospectively and cross-sectionally. Longitudinally, the expected rate per 100,000 person years among persons born in year \( c \) and followed-up at age \( a \) equals \( R(a|c) = \text{LongAge}(a|c_0) \times \text{CRR}(c|c_0) \times e^{\text{PRR}(a|c)} \). Cross-sectionally, the expected rates by age conditional on period equal \( R(a|p) = \text{CrossAge}(a|p_0) \times \text{PRR}(p|p_0) \times e^{\text{CD}(p-a)} \), and the expected rates by period conditional on age equal \( R(p|a) = \text{FTT}(p|a_0) \times \text{CrossAge}(a|p_0) \times e^{\text{CD}(p-a)} \).

Formally, if we calculate the log-linear regressions of \( R(p|a) \) versus \( a \) (one regression for each age group), the local drifts are equal to the slopes of the regression lines \( \beta_p \) expressed as an estimated annual percentage change \( 100\% \times (e^{\beta_p} - 1) \). From the expression for \( R(p|a) \), it follows that any differences between the local drifts and the net drift are a function of the cohort deviations \( \text{CD}(p-a) \).

Illustrative examples

The built-in examples represent qualitatively different rate patterns. For the prostate cancer example (Sample Data 1 online), the cohort rate ratio curve has a striking inverted-V shape. Consequently, the local drifts (Fig. 2) are highly significant (\( \chi^2 = 137.5 \) on 6 degrees of freedom, \( P \approx 0 \)); the local drift values increase from around 0% per year among men ages 50 to 54 years to around 4% per year among men ages 80 to 84 years. In contrast, for the lung cancer example (Sample Data 2 online), the cohort deviations are not statistically significant; hence, the cross-sectional age-specific rates over time are more or less proportional. Furthermore, because the period deviations are also not statistically significant, the common secular pattern is more or less log-linear.

For the breast cancer example (Sample Data 3 online), the overall net drift of \(-0.289\%\) per year is quite modest. Nonetheless, the cohort rate ratio curve identifies a remarkable moderation in breast cancer mortality among women born after 1926. This cohort trend is responsible for the striking pattern of local drifts: mortality is stable or significantly decreasing over time among women ages 58 to 59 years and younger, but is significantly increasing over time, by as much as 1% per year, among women ages 60 to 61 years and older.

In the prostate and breast cancer examples, cohort deviations are substantially larger than period deviations.
Hence, both sets of rates are well approximated by a multiplicative model in which the longitudinal age curve is modulated up or down according to the values of the cohort rate ratio curve.

**Software**

The computational methods for our Web tool are implemented in R. The R package is freely available through the Web tool (click on Help, then FAQ). The R code runs on a back-end server. Python and JavaScript are used on a front-end server to obtain the user’s input, communicate with the R server, and format the results for the user.

The Web tool can save the outputs in a text file, Excel Workbook, or R Workspace. After fitting a model, select the desired output format from the drop-down menu (Fig. 2), then click on the Download button. The text output files were generated by the R package and are displayed in tabs in the browser window. Users can copy and paste this output to their own files. Many users may prefer to download the outputs in the form of an Excel Workbook. Each function tab on the website appears as a function sheet in the workbook, which provides model outputs in tabular and graphical form, precisely as they appear on the Web. The table of Wald tests is also included, along with the net drift and log-linear APC parameters. The inputs and outputs can also be saved as R Workspaces, which can be opened in R for downstream analyses in the R environment.

**Discussion**

We developed a user-friendly Web tool that makes APC analysis broadly available without the need for any
programming. Our R package is also freely available. The functions in the Web tool have proven utility in registry-based studies of cancer incidence and mortality. The tool includes three built-in examples for teaching and demonstration, and it can output all of the results in several formats, including Excel.

Net drift is perhaps the single most important parameter in the APC model, so the drift value is prominently displayed by our Web tool. One important yet perhaps under-appreciated etiological implication of drift is made clear by our Web tool: whenever there is drift, the cross-sectional and longitudinal age curves will diverge. The mathematical reason is that the cross-sectional age trend (CAT) in the former is always equal to the longitudinal age trend (LAT) in the latter minus the log-linear net drift coefficient (45, Table 1). The intuitive reason is that the cross-sectional age curve reflects the experience of older cohorts at older ages and younger cohorts at younger ages. When there is a progressive increase in rates from older to younger generations, the cross-sectional age curve gives the false impression of falling rates with advancing age at diagnosis (46), and conversely. Therefore, the cross-sectional age curve should be interpreted cautiously as a surrogate for the longitudinal age-associated natural history unless there is little net drift. The bias can be severe whenever there is substantial drift of say 1% per year or more, as in the lung cancer example. We prefer to make inferences about age-associated natural history using the longitudinal age curve (35, 40, 47–51), which is now broadly available through our Web tool.

One of the most novel components of our Web tool is the local drifts and associated Wald test for heterogeneity. In our view, after net drift, the test for significance of the local drifts is the second most important number coming out of an APC analysis. It can be shown that the local drifts are determined from the slope (derivative) of the cohort rate ratio curve. In other words, local drifts are a consequence of trends in birth cohort effects. If the local drifts are significant (because of birth cohort effects), an important implication is a single summary age-standardized rate curve and EAPC value cannot adequately describe the time trends in every age group. The Web tool for the first time allows the user to test for this situation. In our experience, local drifts are often quite heterogeneous, and striking local drift patterns such as those illustrated by the prostate and breast cancer examples are not uncommon.

<table>
<thead>
<tr>
<th>Nomenclature</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Net drift</td>
<td>APC analog of the EAPC of the ASR, log-linear component of FTT(p</td>
</tr>
<tr>
<td>CAT = LAT – net drift</td>
<td>Cross-sectional age trend; log-linear trend in CrossAge(a</td>
</tr>
<tr>
<td>LAT = CAT + net drift</td>
<td>Longitudinal age trend; log-linear trend in LongAge(a</td>
</tr>
<tr>
<td>Age deviations, AD(a)</td>
<td>Nonlinear age effects incorporated into LongAge(a</td>
</tr>
<tr>
<td>Period deviations, PD(p)</td>
<td>Nonlinear period effects incorporated into FTT(p</td>
</tr>
<tr>
<td>Cohort deviations, CD(c)</td>
<td>Nonlinear cohort effects incorporated into CRR(c</td>
</tr>
<tr>
<td>Fitted temporal trends, FTT(p</td>
<td>a</td>
</tr>
<tr>
<td>Cross-sectional age curve, CrossAge(a</td>
<td>p</td>
</tr>
<tr>
<td>Longitudinal age curve, LongAge(a</td>
<td>c</td>
</tr>
<tr>
<td>Ratio of longitudinal versus cross-sectional age curves Long2CrossRR(a</td>
<td>c</td>
</tr>
</tbody>
</table>

*For the APC model defined over |j| age groups and |p| calendar periods with equal intervals. In the Web tool, the central age group, calendar period, and birth cohort define the reference values |a|0|, |p|0|, and |c|0| respectively. When there is an even number of age, period, or cohort categories, the reference value is the lower of the two central values. In the R package, reference values can be set to any observed age, period, and cohort – period-age in the rate matrix.
Age, period, and cohort deviations are also provided by the Web tool. These quantities measure curvature, which describes local changes in trends, independently of the magnitude or direction of the overall trend. There are situations where it is desirable to base inferences on these quantities (36, 38), especially in comparative analysis.

Various investigators have calculated cohort and period rate ratio curves by imposing different constraints on the parameters in the APC model (8, 9, 17, 52). In our Web tool, both the cohort rate ratio curve and the period rate ratio curve incorporate the entire value of the net drift. Importantly, in our software, the cohort deviations are constructed to be orthogonal to the linear trend in cohort over the entire set of observed rates. This definition ensures that various products of age, period, and cohort functions provided by our software are mathematically equivalent to the fitted rates. Because the fitted rates track the observed rates as closely as possible, the estimable functions provided by our software are most consistent with the observed data.

An important caveat is that the functions in our Web tool were designed to highlight key signals in cancer rates, but it remains up to the investigator to propose plausible explanations. Also, the tool does not produce standard descriptive plots as these can be generated by numerous other software applications, including Excel. In the future, we hope to develop a companion Web tool for comparative analysis of two sets of rates (45).

In summary, our Web tool for APC analysis provides a suite of APC functions and parameters that complement traditional descriptive approaches. It is very simple to cut and paste case and population data into our Web tool or...
upload the data from a csv file. All of the outputs can be downloaded in a number of formats that enable further downstream analysis. Our hope is that our Web tool will help cancer surveillance researchers make important discoveries about emerging cancer trends and patterns.

Disclosure of Potential Conflicts of Interest
No potential conflicts of interest were disclosed by the authors.

Acknowledgments
The Web tool was made possible thanks to the expert support of computer scientists at the NCI Center for Biomedical Informatics and Information Technology (CBIIT). The authors thank the NCI CBIIT team for implementing the APC analysis Web tool: Robert Shirley (NCI CBIIT), Sue Pan (NCI CBIIT), Larry Bren, Brent Coffey, Shaun Einolf, Sula Rajapakse (Leidos Biomedical Research, Inc.; NCI CBIIT Dev Team Contractor), and Cuong Nguyen (SRA International, Inc.; NCI CBIIT Systems Team Contractor).

Grant Support
This research was supported entirely by the Intramural Research Program of the NIH, NCI, Division of Cancer Epidemiology and Genetics. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received March 18, 2014; revised July 17, 2014; accepted August 11, 2014; published OnlineFirst August 21, 2014.

References


Table 2. Wald χ² tests for estimable functions in the APC model

<table>
<thead>
<tr>
<th>Null hypothesis</th>
<th>Implications</th>
<th>Degrees of freedom</th>
</tr>
</thead>
<tbody>
<tr>
<td>Net drift equals 0</td>
<td>Fitted temporal trends are stable over time.</td>
<td>1</td>
</tr>
<tr>
<td>All age deviations equal 0</td>
<td>Fitted longitudinal and cross-sectional age curves are proportional.</td>
<td>A – 2</td>
</tr>
<tr>
<td>All period deviations equal 0</td>
<td>Fitted temporal trends and period rate ratios are log-linear.</td>
<td>P – 2</td>
</tr>
<tr>
<td>All cohort deviations equal 0</td>
<td>Cohort rate ratios are log-linear; all local drifts equal the net drift.</td>
<td>C – 2</td>
</tr>
<tr>
<td>All period rate ratios equal 1</td>
<td>Net drift is 0 and fitted temporal trends are constant; Cross-sectional age curve describes age incidence pattern in every period.</td>
<td>P – 1</td>
</tr>
<tr>
<td>All cohort rate ratios equal 1</td>
<td>Net drift is 0, and all local drifts are 0; Longitudinal age curve describes age incidence in every cohort.</td>
<td>C – 1</td>
</tr>
<tr>
<td>All local drifts equal the net drift</td>
<td>Temporal trends are the same in every age group.</td>
<td>A – 1 if A = P, A otherwise</td>
</tr>
</tbody>
</table>

*For APC model defined over A age groups, P calendar periods, and C = P + A – 1 birth cohorts.


A Web Tool for Age–Period–Cohort Analysis of Cancer Incidence and Mortality Rates

Philip S. Rosenberg, David P. Check and William F. Anderson


Updated version
Access the most recent version of this article at: doi:10.1158/1055-9965.EPI-14-0300

Cited articles
This article cites 50 articles, 11 of which you can access for free at: http://cebp.aacrjournals.org/content/23/11/2296.full#ref-list-1

Citing articles
This article has been cited by 3 HighWire-hosted articles. Access the articles at: http://cebp.aacrjournals.org/content/23/11/2296.full#related-urls

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.