Published in final edited form as:

Stat Med. 2010 May 20; 29(11): 1228–1238. doi:10.1002/sim.3865.

Proportional Hazards Models and Age-Period-Cohort Analysis of Cancer Rates

Philip S. Rosenberg, Ph.D.¹ and William F. Anderson, M.D., M.P.H.¹

¹Biostatistics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Department of Health and Human Services, Rockville MD, USA

Summary

Age-period-cohort (APC) analysis is widely used in cancer epidemiology to model trends in cancer rates. We develop methods for comparative APC analysis of two independent cause-specific hazard rates assuming that an APC model holds for each one. We construct linear hypothesis tests to determine whether the two hazards are absolutely proportional or proportional after stratification by cohort, period, or age. When a given proportional hazards model appears adequate, we derive simple expressions for the relative hazards using identifiable APC parameters. To demonstrate the utility of these new methods, we analyze cancer incidence rates in the United States in blacks versus whites for selected cancers, using data from the National Cancer Institute's Surveillance, Epidemiology, and End Results Program. The examples illustrate that each type of proportionality may be encountered in practice.

Keywords

age-period-cohort models; cause-specific hazards; cancer incidence; descriptive epidemiology; racial disparity

Introduction

Age-period-cohort (APC) analysis is widely used in cancer epidemiology to model trends in cancer incidence and mortality rates^{1–6}. For studies of incidence, age effects in the APC model reflect an underlying age-associated natural history, period effects capture factors that influence persons of all ages concurrently, including innovations in screening and diagnostic practice, and birth cohort effects track the net impact of risk factors that vary in prevalence from one generation to the next. For studies of mortality, APC effects are also modulated by the dissemination of therapeutic advances.

Typically, an APC model is fitted to a matrix of age- and period-specific rates obtained from a population-based cancer registry, such as the National Cancer Institute (NCI) Surveillance, Epidemiology, and End Results (SEER) Program (<http://seer.cancer.gov>). Cohort-specific rates are defined using the relationship that the year of birth equals the calendar period of diagnosis minus the age at diagnosis. Rates can be compiled for any type of cancer that is

Corresponding author: Philip S. Rosenberg, Biostatistics Branch, Division of Cancer Epidemiology and Genetics, National Cancer Institute, 6120 Executive Blvd, Executive Plaza South, Room 8022, Rockville MD 20852-7244. rosenbep@mail.nih.gov; Phone: 301-435-3996; Fax: 301-402-0081.

Contribution: P.S.R. and W.F.A. designed the study; P.S.R. developed the statistical methods; W.F.A. led the assembly of the SEER data; both authors analyzed the data and wrote the paper.

Conflict-of-interest disclosure: The authors have nothing to disclose.

tracked by a registry. Furthermore, the analysis can be stratified according to subgroups defined by sex, ethnicity, geographic region, etc. Hence, rate matrices available for APC analysis describe a wide variety of cause-specific hazards in a target population, and we will refer to the analysis of any one such event rate as a “one-hazard problem.”

Very often the purpose of a study is to make pairwise comparisons of distinct cause-specific hazard rates using APC models. We will refer to this as a “two-hazard problem.” For example, such “comparative” APC studies have considered: ethnicity and breast cancer⁷⁻⁹, regional variation and trends in testicular cancer^{10,11}, possible etiological heterogeneity of testicular seminomas and non-seminomas^{12,13}, patterns of childhood cancer in Germany¹⁴ and in the north-west of England¹⁵, and cohort effects for colorectal cancer across the Nordic countries¹⁶. However, of the ten studies cited above, only three formally assessed the significance of any interactions between age, period, or cohort effects and the study factor of interest^{9,12,15}; the majority relied on qualitative graphical comparisons. It appears that available methodology is limited for model selection and inference *across* two distinct hazard types; ideally, such approaches should complement well-accepted methods for modeling each hazard separately^{4,17,18}.

In this study, we develop one such approach by connecting the APC model to standard paradigms from the statistical analysis of failure time data. We consider the common situation where APC models are fitted to two sets of rates corresponding to distinct event types or populations, and a goal of the analysis is to compare and contrast the respective hazards. When three or more sets of rates are of interest, the approach can be applied pairwise.

Our methods reflect the essential geometry of the rate matrix or Lexis diagram¹⁸. Given two sets of rates defined over the same ages, periods, and cohorts, we develop linear hypothesis tests to determine whether the corresponding hazard rates are absolutely proportional, or proportional after stratification by age, period, or cohort. When a given proportional hazards model appears adequate, we derive corresponding estimators of the relative hazards using identifiable APC parameters.

To demonstrate the utility of these new methods, we compare cancer incidence rates in blacks and whites for selected cancers in the United States using nationally representative SEER data. The examples illustrate that different types of proportionality may be encountered in practice.

Materials and Methods

Cancer Incidence Data

We analyze incidence rates for the following cancers: bladder cancer in black versus white women; colorectal cancer in black versus white women; kidney cancer in black versus white men; pancreas cancer in black versus white women; and oral cancer in black versus white men. These cancers were selected from a larger study that will be presented separately.

We obtained case and population data from NCI's SEER program. SEER integrates cancer incidence data from 17 population-based registries with meticulous and consistent data collection and standards that together cover approximately 26 percent of the US population¹⁹. Our analysis covers the 16-year period from January 1, 1990 through December 31, 2005 using data released by SEER in April 2008. For this analysis, we tabulated the rates into four 4-year time periods (1990-1993, 1994-1997, 1998-2001, 2002-2005) and sixteen 4-year age-at-diagnosis groups (ages 21-24, 25-28, ..., 81-84 years old) spanning nineteen partially overlapping 8-year birth cohorts from 1909 to 1981

(referred to by mid-year of birth). We excluded the youngest or oldest age group whenever the corresponding number of cases was zero.

We will use the following notation. For any given cancer and population group, matrix $\mathbf{Y} = [Y_{pa}, p = 1, \dots, P, a = 1, \dots, A]$ contains the numbers of cancer diagnoses (counts) in calendar period p and age group a and matrix $\mathbf{O} = [O_{pa}, p = 1, \dots, P, a = 1, \dots, A]$ contains the corresponding person-years. The diagonals of matrices $\mathbf{Y}$ and $\mathbf{O}$, from upper right to lower left, represent successive birth cohorts indexed by $c = p - a + A$, from the oldest cohort ($c = 1$) observed only for the oldest age interval $a = A$ during the earliest calendar period $p = 1$, to the youngest cohort ($c = C \equiv p + A - 1$) observed only for the youngest age interval $a = 1$ during the most recent calendar period $p = P$. The observed incidence rates per 100,000 person-years are $\lambda_{pa} = 10^5 Y_{pa} / O_{pa}$. The expected log rates are $\rho_{pa} \equiv \log(E(Y_{pa}) / O_{pa})$. We assume the counts are realizations of Poisson processes with means $O_{pa} E(\lambda_{pa}) 10^{-5}$.

The APC Model

APC analysis assumes an underlying true model for the expected rates in the population with log-linear effects for age, period, and cohort:

$$\rho_{pa} = \alpha_a + \pi_p + \gamma_c. \quad (1)$$

It is computationally convenient to make an orthogonal decomposition of the underlying APC effects and apply standard identifiability conditions¹:

$$\begin{aligned} \rho_{pa} &= \mu + \alpha_L(a - \bar{a}) + \pi_L(p - \bar{p}) + \gamma_L(c - \bar{c}) + \tilde{\alpha}_a + \tilde{\pi}_p + \tilde{\gamma}_c \\ \sum_a \tilde{\alpha}_a &= \sum_a \tilde{\alpha}_a (a - \bar{a}) = 0 \\ \sum_p \tilde{\pi}_p &= \sum_p \tilde{\pi}_p (p - \bar{p}) = 0 \\ \sum_c \tilde{\gamma}_c &= \sum_c \tilde{\gamma}_c (c - \bar{c}) = 0. \end{aligned} \quad (2)$$

In equation (2) $\bar{a} = [(A + 1) / 2]$ and $\bar{p} = [(P + 1) / 2]$, where [...] is the greatest integer function. Also, $\bar{c} = \bar{p} - \bar{a} + A$, so the values of $\bar{a}$, $\bar{p}$, and $\bar{c}$ define convenient (but arbitrary) central or referent indices of age, period, and cohort, respectively.

The intercept μ parameterizes the log rate at the “center” of the table, while α_L , π_L , and γ_L parameterize log-linear trends according to age, period, and cohort, respectively. The parameters $\tilde{\alpha}_a$, $\tilde{\pi}_p$, and $\tilde{\gamma}_c$ represent deviations from the regular trends due to non-linear influences associated with age, period, and cohort, respectively. By construction, the deviations have mean 0 and are orthogonal to the linear trends. It is well recognized that this model is not identifiable: because $c = p - a + A$, the regular trends α_L , π_L , and γ_L are not separable, and model parameters cannot be estimated without imposing one additional linear constraint.

As shown by Holford², if one fits the model using Poisson regression with the additional identifiability constraint that the coefficient of $(p - \bar{p})$ in equation (2) equals 0, it follows that

$$\rho_{pa} = \mu + (\alpha_L + \pi_L)(a - \bar{a}) + (\pi_L + \gamma_L)(c - \bar{c}) + \tilde{\alpha}_a + \tilde{\pi}_p + \tilde{\gamma}_c \quad (3)$$

i.e., the coefficient of the age trend $(a - \bar{a})$ provides an estimate of $(\alpha_L + \pi_L)$, the coefficient of the cohort trend $(c - \bar{c})$ provides an estimate of $(\pi_L + \gamma_L)$, and the difference of these two coefficients provides an estimate of $(\alpha_L - \gamma_L)$.

This solution maximizes the Poisson log likelihood, and gives the same fitted rates as all other constraints that maximize the likelihood. Furthermore, each of the parameters in equation (3) is identifiable. Following conventions we refer to $(\alpha_L + \pi_L)$ as the “longitudinal age trend”²⁰, to $(\pi_L + \gamma_L)$ as the “net drift”⁴ and to $(\alpha_L - \gamma_L) \equiv (\alpha_L + \pi_L) - (\pi_L + \gamma_L)$ as the “cross-sectional age trend”²⁰.

It is important to recognize that application of this particular identifiability constraint in no way imposes the assumption that the true $\pi_L = 0$. Rather, it reflects that the observable effects of π_L cannot be separated from those for α_L and γ_L , but can only be seen indirectly through the estimated values of identifiable quantities such as the net drift, longitudinal age trend, and cross-sectional age trend.

Contrasting Two Groups or Tumor Types: Proportional Hazards and the APC Model

Now we consider the comparison of two sets of independent cause-specific hazard rates with expectations $\{\rho_{pa}^1\}$ versus $\{\rho_{pa}^0\}$, type 1 and type 0 hazards, say, over the same ages $a = 1, \dots, A$, periods $p = 1, \dots, P$, and cohorts $c = 1, \dots, P + A - 1$, assuming that a separate APC model holds for each hazard type. This general setup can be applied in many situations, for example, to compare hazard rates for the same tumor type in two population subgroups, or hazard rates for different tumor types in the same subgroup. In any given application, one hazard rate is the ‘type 1’ hazard, and the other is the ‘type 0’ hazard. For example, in our analysis of bladder cancer in black versus white women (Figure 1A), the incidence of bladder cancer in white women is the ‘type 0’ hazard, and the incidence of bladder cancer in black women is the ‘type 1’ hazard.

We will address the following general questions: Under what circumstances do the hazard rates follow a proportional hazards (PH) model? If PH holds, how does one estimate the relative hazards? Clearly, it might be the case that proportionality holds in an absolute sense for all p and a . In practice, this model might be too restrictive. Therefore, we also consider more flexible models of proportionality. For this purpose, it seems natural to develop PH models that correspond to standard descriptive plots for rates, specifically: rates by age stratified on cohort (Figure 1B), rates by period stratified on age (Figures 1A and 1C), and rates by age stratified on period (Figure 1D)^{18,21,22}. In the next section, we develop each of these PH model from an analytical perspective, and provide illustrative examples.

The PH-A Model: Absolute Proportionality

We will say that proportionality holds *absolutely* (PH-A, “A” for absolute) if

$$\rho_{p,a}^1 = \psi + \rho_{p,a}^0 \text{ for all } p \text{ and } a. \quad (4)$$

When PH-A holds, the quantity $\exp(\psi)$ equals the rate ratio for the type 1 hazard relative to the type 0 hazard, and this value is constant over age, period, and cohort. Clearly, the data are consistent with absolute proportionality if and only if there are no significant differences between any of the APC parameters across hazard types, except for the intercepts.

To illustrate, rates of female bladder cancer in blacks versus whites are consistent with absolute proportionality (Figure 1A). For each age group, plots of age-specific rates

(logarithmically scaled) versus calendar period reveal essentially parallel curves in blacks versus whites, and the gaps between the curves do not vary significantly by age (for clarity, two of 16 age groups are shown in Figure 1A). Using methods described subsequently, the estimated overall black-to-white incidence rate ratio $RR^{1:0}$ is 0.64 (95% Confidence Interval [CI]: 0.56 – 0.72) (Figure 2A).

The PH-L Model: Proportionality within Cohorts

We will say that PH holds on the “natural” or longitudinal time scale of age, if the logarithms of the age-specific hazard rates are shifted by a constant θ_c whenever the experience of the same birth cohort c is evaluated for the type 1 and type 0 hazards. This condition defines a PH model stratified on birth cohort, which we will denote as PH-L (“L” for longitudinal). To illustrate, rates of female colorectal cancer in blacks and whites are consistent with PH-L (Figure 1B; three of 19 birth cohorts are shown). For each cohort, plots of age-specific rates are essentially parallel in blacks versus whites, although the gap between each pair of cohort-specific curves varies by cohort.

In general, because longitudinal follow-up of cohorts corresponds to rates along the diagonals of the rate matrices, for each cohort indexed by $c = 1, \dots, P + A - 1$, the youngest observed age interval is $a_0^c = \max(1, A - c + 1)$, and the oldest observed age interval is $a_1^c = \min(A - c + P, A)$. Therefore, in terms of expected log rates, PH-L holds when

$$\rho_{c+a-A,a}^1 = \theta_c + \rho_{c+a-A,a}^0, a = a_0^c, \dots, a_1^c \quad (5)$$

for all c . The term θ_c equals the logarithm of the rate ratio for the type 1 hazard versus the type 0 hazard, and is a function only of cohort when PH-L holds. Note that PH holds absolutely (PH-A) if PH-L holds and $\theta_c = \theta$ for all c . Hence, if the gaps were all equal in Figure 1B the rates would be PH-A.

Assuming that PH-L holds, the estimated black-to-white incidence rate ratio $RR_c^{1:0}$ for female colorectal cancer (Figure 2B) is increasingly elevated in successive cohorts of black versus white women, up to a peak value of 1.86 (95% CI: 1.6 – 2.1) for the 1945 cohort. In subsequent cohorts the incidence rate ratio declines; no significant excess in black women is apparent in the limited available follow-up of cohorts born since 1969. The model-based estimates in Figure 2B for 19 cohorts concisely summarize the descriptive data, including data for the three cohorts shown in Figure 1B.

The PH-T Model: Proportionality within Age Groups

We will say that PH holds over time (PH-T, “T” for calendar time) if

$$\rho_{p,a}^1 = \delta_a + \rho_{p,a}^0, p = 1, \dots, P \quad (6)$$

for each age group indexed by $a = 1, \dots, A$. If PH-T holds, then a plot of the logarithms of the age-specific rates for the type 1 and type 0 hazards versus calendar time will be parallel whenever the same age groups are compared. The term δ_a equals the logarithm of the rate ratio for the type 1 hazard versus the type 0 hazard, and is a function only of age when PH-T holds. Note that PH-A holds if PH-T holds and $\delta_a = \delta$ for all a .

Rates of male kidney cancer in blacks and whites are consistent with PH-T (Figure 1C; two of 16 age groups are shown). Plots of age-specific rates over calendar time are essentially parallel in blacks versus whites, and the gap between the curves varies by age group. The estimated black-to-white incidence rate ratio $RR_a^{1:0}$ is generally higher in black versus white men, but more so among men in their 40s and 50s than men in their 60s and 70s (Figure 2C).

The PH-X Model: Proportionality within Calendar Periods

We will say that PH holds in cross-section (PH-X, “X” for cross-section), if

$$\rho_{p,a}^1 = \phi_p + \rho_{p,a}^0, a=1, \dots, A \quad (7)$$

For each period indexed by $p = 1, \dots, P$. If PH-X holds, then a plot of the logarithm of the rates versus age for the type 1 and type 0 hazards will be parallel within any given calendar period. The term ϕ_p equals the logarithm of the rate ratio for the type 1 hazard versus the type 0 hazard, and is a function only of period when PH-X holds. Note that PH-A holds if PH-X holds and $\phi_p = \phi$ for all p .

Rates of female pancreas cancer are consistent with PH-X (Figure 1D; two of four periods are shown). The cross-sectional age-specific rates are essentially parallel in blacks and whites within any given calendar period, but the gap varies by period. As shown in Figure 2D, the estimated black-to-white incidence rate ratio $RR_p^{1:0}$ is significantly elevated during each study period, yet declines over time, from a maximum value of 1.63 (95% CI: 1.4 – 2.0) during the 1990 – 1993 period, to a minimum of 1.20 (95% CI: 1.0 – 1.4) during the 2002 – 2005 period.

Characterizing the PH Models

In this section we develop necessary and sufficient conditions for each type of proportionality, and derive corresponding expressions for the relative hazards. We assume that a separate APC model holds for both of the type j hazards, $j = 0, 1$, with type-specific values for each APC parameter, so that the expected log rates equal

$$\rho_{pa}^j = \mu^j + (\alpha_L + \pi_L)^j (a - \bar{a}) + (\pi_L + \gamma_L)^j (c - \bar{c}) + \tilde{\alpha}_a^j + \tilde{\pi}_p^j + \tilde{\gamma}_c^j. \quad (8)$$

Clearly, PH-A holds when all identifiable parameters except for the intercepts are equal. Now consider necessary and sufficient conditions for PH-L. From equation (8), it is generally the case that the difference between the logarithms of the type 1 and type 0 hazards equals

$$\begin{aligned} \rho_{c+a-A,a}^1 - \rho_{c+a-A,a}^0 &= \theta_c + \{(\alpha_L + \pi_L)^1 - (\alpha_L + \pi_L)^0\} (a - \bar{a}) + (\tilde{\alpha}_a^1 - \tilde{\alpha}_a^0) + (\tilde{\pi}_{c+a-A}^1 - \tilde{\pi}_{c+a-A}^0) \\ &\quad \text{where} \\ \theta_c &= (\mu^1 - \mu^0) + \{(\pi_L + \gamma_L)^1 - (\pi_L + \gamma_L)^0\} (c - \bar{c}) + (\tilde{\gamma}_c^1 - \tilde{\gamma}_c^0) \end{aligned} \quad (9)$$

In equation (9) the expression for θ_c depends only on c , while the other terms vary with a and p . Hence PH-L holds if and only if the following conditions are true:

$$\begin{aligned}
 (\alpha_L + \pi_L)^1 &= (\alpha_L + \pi_L)^0 \\
 \tilde{\alpha}_a^1 &= \tilde{\alpha}_a^0, a=1, \dots, A \\
 \tilde{\pi}_p^1 &= \tilde{\pi}_p^0, p=1, \dots, P
 \end{aligned}
 \tag{10}$$

In other words, PH-L holds when the longitudinal age trends, the age deviations, and the period deviations are all equal across the type 1 and type 0 hazards. Furthermore, PH holds absolutely (PH-A) when it is also the case that

$$\begin{aligned}
 (\pi_L + \gamma_L)^1 &= (\pi_L + \gamma_L)^0 \\
 \tilde{\gamma}_c^1 &= \tilde{\gamma}_c^0, c=1, \dots, P+A-1
 \end{aligned}
 \tag{11}$$

That is, PH-A holds if PH-L holds, and the net drifts and cohort deviations are also equal. In that case, $\theta_c = \theta = (\mu_1 - \mu_0) = \psi$.

If equations (10) hold but equations (11) do not hold, then the expression for θ_c in equation (9) concisely describes the dependence of the relative hazard on birth cohort. If equations (10) and (11) both hold, then the expression $e^{\mu_1 - \mu_0}$ yields the common value of the relative hazard.

Next, consider conditions for PH-T. By expressing equation (8) in terms of age and period, it can be shown that

$$\begin{aligned}
 \rho_{p,a}^1 - \rho_{p,a}^0 &= \delta_a + \{(\pi_L + \gamma_L)^1 - (\pi_L + \gamma_L)^0\} (p - \bar{p}) + \left(\tilde{\pi}_p^1 - \tilde{\pi}_p^0 \right) + \left(\tilde{\gamma}_{p-a+A}^1 - \tilde{\gamma}_{p-a+A}^0 \right) \\
 \text{where} \\
 \delta_a &= (\mu^1 - \mu^0) + \{(\alpha_L - \gamma_L)^1 - (\alpha_L - \gamma_L)^0\} (a - \bar{a}) + (\tilde{\alpha}_a^1 - \tilde{\alpha}_a^0).
 \end{aligned}
 \tag{12}$$

In equation (12) the expression for δ_a varies only with a , while the other terms depend on p and c . This expression demonstrates that PH-T holds if and only if the following conditions are true:

$$\begin{aligned}
 (\pi_L + \gamma_L)^1 &= (\pi_L + \gamma_L)^0 \\
 \tilde{\pi}_p^1 &= \tilde{\pi}_p^0, p=1, \dots, P \\
 \tilde{\gamma}_c^1 &= \tilde{\gamma}_c^0, c=1, \dots, P+A-1
 \end{aligned}
 \tag{13}$$

In other words, PH-T holds when the net drifts, the period deviations, and the cohort deviations are all equal across the type 1 and type 0 hazards. Furthermore, PH holds absolutely (PH-A) when it is also the case that

$$\begin{aligned}
 (\alpha_L - \gamma_L)^1 &= (\alpha_L - \gamma_L)^0 \\
 \tilde{\alpha}_a^1 &= \tilde{\alpha}_a^0, a=1, \dots, A
 \end{aligned}
 \tag{14}$$

That is, PH-A holds if PH-T holds, and the cross-sectional age trends and age deviations are also equal. In that case, $\delta_a = \delta = (\mu_1 - \mu_0) = \psi$. If equations (13) hold but equations (14) do not hold, then the expression for δ_a in equation (12) concisely describes the dependence of the relative hazard on age.

Finally, consider conditions for PH-X. By rearranging the terms in equation (8), it follows that

$$\rho_{a,p}^1 - \rho_{a,p}^0 = \varphi_p + \{(\alpha_L - \gamma_L)^1 - (\alpha_L - \gamma_L)^0\}(a - \bar{a}) + (\tilde{\alpha}_a^1 - \tilde{\alpha}_a^0) + (\tilde{\gamma}_{p-a+A}^1 - \tilde{\gamma}_{p-a+A}^0)$$

where

$$\varphi_p = (\mu^1 - \mu^0) + \{(\pi_L + \gamma_L)^1 - (\pi_L + \gamma_L)^0\}(p - \bar{p}) + (\tilde{\pi}_p^1 - \tilde{\pi}_p^0). \quad (15)$$

This expression demonstrates that PH-X holds if and only if the following conditions are true:

$$\begin{aligned} (\alpha_L - \gamma_L)^1 &= (\alpha_L - \gamma_L)^0 \\ \tilde{\alpha}_a^1 &= \tilde{\alpha}_a^0, a = 1, \dots, A \\ \tilde{\gamma}_c^1 &= \tilde{\gamma}_c^0, c = 1, \dots, P+A-1 \end{aligned} \quad (16)$$

In other words, PH-X holds when the cross-sectional age trends, the age deviations, and the cohort deviations are all equal across the type 1 and type 0 hazards. Furthermore, PH holds absolutely (PH-A) when it is also the case that

$$\begin{aligned} (\pi_L + \gamma_L)^1 &= (\pi_L + \gamma_L)^0 \\ \tilde{\pi}_p^1 &= \tilde{\pi}_p^0, p = 1, \dots, P \end{aligned} \quad (17)$$

That is, PH-A holds if PH-X holds, and the net drifts and period deviations are also equal. In that case, $\phi_p = \phi = (\mu_1 - \mu_0) = \psi$. If equations (16) hold but equations (17) do not hold, then the expression for ϕ_p in equation (15) concisely describes the dependence of the relative hazard on period.

As summarized in Table 1, each PH model requires specific linear constraints on the APC parameters. For each model, a simple expression is available for the relative hazard that depends only on the remaining parameters, which are free to vary.

Goodness-of-Fit Tests

It is convenient to develop Wald tests for the PH hypotheses because the required ingredients are standard outputs of an APC analysis. For each test, the number of degrees of freedom corresponds to the number of parameters that must be equal under the model (Table 1). For example, PH-L requires that the longitudinal age trends, the age deviations, and the period deviations are all equal across the type 1 and type 0 hazards, resulting in $1 + (A - 2) + (P - 2) = (A + P - 3)$ *df*. Corresponding *df* are $2(A + P) - 5$ for PH-A, $A + 2P - 4$ for PH-T, and $2A + P - 4$ for PH-X.

The vector of model parameters for the type g hazards, $g = 0, 1$, equals

$\Psi^g = [\mu^g, (\alpha_L + \pi_L)^g, (\pi_L + \gamma_L)^g, \tilde{\alpha}_{(2:A-1)}^g, \tilde{\pi}_{(2:P-1)}^g, \tilde{\gamma}_{(2:P+A-2)}^g]'$. The parameters are asymptotically normally distributed. The first and last age, period, and cohort deviations are excluded from Ψ^g because these are determined from the others through the identifiability constraints shown in equation (2). The variance-covariance matrix $\text{Var}(\Psi^g)$ can be estimated from the Poisson log-likelihoods for the numbers of type 0 and type 1 events. If over-dispersion is suspected for one or both sets of hazard rates, the variance-covariance matrices can be scaled by the dispersion parameters $(\hat{\sigma}^g)^2 = D(\mathbf{Y}^g, \hat{\mathbf{Y}}^g) / (PA - 2(P + A - 1))$, where $D(\dots)$ is the usual deviance statistic²³. Using the independence assumption, PH-A, PH-L, PH-T, and PH-X are linear hypotheses that can be evaluated using a Wald Chi-Square test with the general form

$$\chi^2_{df} = (\hat{\Psi}^1 - \hat{\Psi}^0)' \mathbf{C}'_{PH} \left[(\hat{\sigma}^1)^2 \text{Var}(\hat{\Psi}^1) + (\hat{\sigma}^0)^2 \text{Var}(\hat{\Psi}^0) \right]^{-1} \mathbf{C}_{PH} (\hat{\Psi}^1 - \hat{\Psi}^0)$$

where $\mathbf{C}_{PH}$ is a $2(P + A + 1) \times 2(P + A + 1)$ contrast matrix with a 1 on the diagonal for each parameter that is being compared across hazard types, and 0 elsewhere, and $df = \text{trace}(\mathbf{C}_{PH})$ is the degrees of freedom for the test. Alternatively, likelihood ratio tests (LRT) can be constructed. Details are presented in Appendix A.1.

To illustrate the Wald tests, we summarize test results for the examples presented in Figure 1. Rates of female bladder cancer in blacks and whites (Figure 1A) are consistent with PH-A, because none of the four Wald tests for proportionality flag any significant lack-of-fit ($\chi^2_{15}(\text{PH} - \text{L}) = 14.07$, $\chi^2_{18}(\text{PH} - \text{T}) = 9.76$, $\chi^2_{28}(\text{PH} - \text{X}) = 30.17$, and $\chi^2_{31}(\text{PH} - \text{A}) = 34.87$, all $P \geq 0.29$). Rates of female colorectal cancer in blacks and whites are consistent with PH-L (Figure 1B) because departures from PH-L are not significant ($\chi^2_{17}(\text{PH} - \text{L}) = 16.06$, $P = 0.52$), yet departures from each of the three other PH models are significant, each at $P < 10^{-3}$. Rates of male kidney cancer in blacks and whites are consistent with PH-T (Figure 1C); $P = 0.75$ for PH-T versus $P < 0.01$ for each of the other three tests. Rates of female pancreas cancer are consistent with PH-X (Figure 1D); $P = 0.14$ for PH-X versus $P \leq 0.02$ for each of the other three tests.

In practice, non-proportionality is also a common finding. Rates of oral cancer in black versus white men provide an example. The non-proportionality is particularly striking when the rates are stratified by period (Figure 3A; the curves cross for the earlier of the two calendar periods shown). Lack-of-fit is convincingly detected by each of the 4 Wald tests, for example, $\chi^2_{32}(\text{PH} - \text{X}) = 240.87$, $P \approx 0$.

Model Selection

As illustrated above, the following logic can be applied to select a model. The data are consistent with PH-A when none of the four PH tests detects significant lack-of-fit. When two of the three hypotheses PH-L, PH-T, or PH-X are rejected but one is not, the data are consistent with the latter model. None of the models provides a completely adequate fit if all four PH hypotheses are rejected. When this is the case we say the rate ratios are significantly heterogeneous. Importantly, one cannot use the tests to prove that a given model holds. Rather, absence of a significant difference, especially when the significance level is well above the usual 5%, constitutes evidence that the hypothesis provides a reasonable working model.

One interesting question is whether two sets of rates that are not PH-A can satisfy two of the three restricted PH models, PH-L and PH-T, say. This cannot be so, as shown in Appendix A.2. In principle, therefore, the existence (or not) of a parsimonious summary of the relative hazards can be determined from the data (Table 1). Unfortunately, experience shows that the goodness-of-fit tests may not clearly indicate which model is best.

Uncertainty about the parameter estimates lies at the root of this problem. Figure 4 illustrates the situation assuming no differences in age, period, or cohort deviations, so the model depends exclusively on the trend parameters. The difference between the net drifts is plotted on the y -axis, the difference between the longitudinal age trends is plotted on the x -axis, and the line $z = \{(x, y) \in \mathbb{R}^2 : y - x = 0\}$ corresponds to equality of the cross-sectional age trends $(\alpha_L - \gamma_L)^j, j = 0, 1$. PH-L holds if the differences between the expected parameters fall on the y -axis, PH-T if the values fall on the x -axis, PH-X if the values fall on line z , and PH-A if the values lie at the origin. Otherwise PH does not hold.

Uncertainty about the trend parameters can be quantified by a joint confidence ellipse. It is very important that the uncertainty not be under-stated, therefore, in routine practice we almost always allow for a separate over-dispersion parameter for each hazard type. When the confidence region intersects the origin, or one and only one of the y -axis, x -axis, or z -line, then the corresponding choice of model is unambiguously PH-A, PH-L, PH-T, or PH-X, respectively (Figure 4). Similarly, if the region *excludes* the x axis, y axis, or line z , there is strong evidence for *heterogeneity* because the data support none of the PH models. The best-fitting model will be *ambiguous* when the confidence ellipse intersects any two of the three lines, for example, $x = 0$ and $y = 0$. In these three scenarios, two of four PH models can be rejected with confidence, but the remaining two are both consistent with the data.

Even when PH does not hold, the fitted values $\{\hat{\rho}_{pa}^1\}$ and $\{\hat{\rho}_{pa}^0\}$ will always be less noisy than the raw data. In this situation it is still valid to consider the rate ratios as a joint function of age and period (or of age and cohort). For this purpose, patterns may be seen more clearly in the model-based estimates $\widehat{RR}_{pa}^{1:0} = \exp(\hat{\rho}_{pa}^1 - \hat{\rho}_{pa}^0)$ than the corresponding empirical estimates $(RR_{pa}^{Obs})^{1:0} = (Y_{pa}^1 / O_{pa}^1) / (Y_{pa}^0 / O_{pa}^0)$. This is also a valid approach when the best-fit model is ambiguous.

To illustrate, fitted rate ratios reveal a clear secular pattern in oral cancer incidence in black versus white men (Figure 3B). During the initial study period (1990 – 1993), incidence was significantly higher in black men ages 37 through 64 years compared to white men, whereas rates in younger and older black and white men were similar. By the final study period (2002 – 2005), the rate ratios had moderated for the high-risk age groups, from initial values of about 2, to significantly lower final values of about 1.2.

Discussion

The APC model is widely used in comparative studies of cancer rates. Until now, methods have been limited for hypothesis testing and rate ratio estimation across subgroups or event types. In this study, we identified close connections between comparative APC analysis and proportional hazards models for failure time data. We defined four types of proportionality (PH-A, PH-L, PH-T, and PH-X), and showed that when specific parameters in the APC models are equal, the rates must be proportional in one sense or another.

Descriptive analysis is always complicated by the fact that three inter-related time scales must be considered (age, period, and cohort). *A priori*, it is not clear whether one or another

descriptive presentation will most clearly reveal the key trends. This problem is especially acute in a comparative study of two or more hazard types. The methods we present here can help to organize the results of a comparative study. A longitudinal presentation may be clearest when the rate ratios vary by cohort (PH-L), a secular presentation when the rate ratios vary by age (PH-T), and a cross-sectional presentation when the rate ratios vary by period (PH-X). Furthermore, fitted rates and fitted rate ratios are always useful, regardless of whether the hazard types are proportional or not. We illustrated our new method using five examples selected from a larger study of racial differences in cancer incidence. In the entire study, about one half of the cancers were proportional in one sense or another. For a fraction of the remaining cancers, the rate ratios were clearly heterogeneous, which also provides etiological clues. Therefore, we believe the models will be useful in practice.

Comparative APC analysis has important limitations. Firstly, the approach shares all the intrinsic limitations of standard descriptive analysis. Secondly, it can be difficult to identify a single best-fitting model. A third, more subtle issue is that the identifiability problem limits the specificity of conclusions that can be derived from contrasts between the linear trends²⁴.

A number of technical issues remain. The optimal strategy for estimation and testing is unclear when the data contain comparatively little signal, for example, when one or both hazards describe rare events. In these situations, extension of our approach using spline functions might be useful²⁵. Power and sample size formulae would also be helpful. Conversely, when the signal-to-noise ratio is high, one is more likely to reject a model based on a small P -value, even if the model fits the data quite well. For these situations, it would be useful to develop objective procedures to identify models that hold reasonably well in an approximate sense. The statistical tests are sensitive to assumptions about the mean-to-variance relationship. Approaches that estimate the variance function from the data²⁶ might be more robust to departures from the standard Poisson assumption than the approach used here, which allowed for a single over-dispersion parameter for each hazard type. Additional methodological development may also be fruitful for modeling the K -group problem with $K \geq 3$, and more generally, for modeling systematic variation of APC parameters as a function of population-level covariates, and for simultaneous modeling²⁷ of large numbers of subgroups and tumor types.

In summary, the proportional hazards framework can help to organize and clarify comparative age-period-cohort analysis. We anticipate that methodological improvements will advance the scope and power of the models. Nonetheless, the relatively simple approach described here is likely to remain appealing, because it relies only on standard outputs from Poisson regression, coupled with simple Wald or likelihood ratio tests.

Appendix

A.1 Likelihood Ratio Tests of PH-A, PH-L, PH-T, and PH-X

Likelihood ratio tests (LRT) for each hypothesis can be constructed by fitting three Poisson regression models. The first two models are separate APC fits for the type g hazards, $g = 0, 1$, obtained using a standard design matrix for APC analysis as described by Holford¹. The third model is a joint fit, applied to a concatenation of the event rate data for each hazard type, using the standard APC design matrix extended with additional columns corresponding to a main effect for hazard type, (i.e. 1 if $g = 1, 0$ otherwise), and interaction terms between the hazard type indicator and those columns of the APC design matrix that correspond to parameters that are allowed to *differ* under the hypothesis of interest. For example, for PH-L, the design matrix includes the usual APC contrasts, plus a main effect

for hazard type, plus interaction terms for hazard type by net drift and hazard type by cohort deviations. The likelihood ratio test statistic equals the difference of the deviance for the combined fit, minus the sum of the deviances for the separate fits. To account for possible over-dispersion, a common value of the over-dispersion parameter $\sigma_{0,1}$ can be estimated from a fourth joint-fit model that includes all interaction terms between APC parameters and group.

A.2 PH-L, PH-T, and PH-X are mutually exclusive

Here we show that PH-L, PH-T, and PH-X require mutually exclusive constraints on the APC parameters. For example, consider the possibility that two sets of rates which are not PH-A nonetheless satisfy two of the three restricted models, PH-L and PH-T, say. If the rates are PH-L but not PH-A, then either the net drifts, or the cohort deviations, or both, differ significantly across the hazard types (see equations (10) and (11)). However, if both of equations (11) cannot hold, then equations (13) must be violated, therefore, PH-T cannot also hold. Similarly, equations (16) must also be violated, therefore, PH-X cannot hold.

Using the same logic, by enumerating all possible combinations of equality or inequality of the APC trends and deviations across hazard types, it can be shown that except for the case of absolute proportionality, if the expected rates are proportional longitudinally they cannot be proportional over time or in cross-section, if the expected rates are proportional over time they cannot be proportional longitudinally or in cross-section, and if the expected rates are proportional in cross-section they cannot be proportional longitudinally or over time.

Acknowledgments

This research was supported by the Intramural Research Program of the NIH, National Cancer Institute, Division of Cancer Epidemiology and Genetics. The authors gratefully acknowledge the assistance of Christina McIntosh and Julia Tse for data management and analysis. Julia Tse was supported as a research fellow by the NCI Cancer Research and Training Award (CRTA) Program, and Christina McIntosh was supported as a research fellow by the National Heart Lung and Blood Institute's Biomedical Research Training Program for Underrepresented Groups (BRTUG). The authors also acknowledge helpful comments from two reviewers and the associate editor, which led to substantial improvements in the manuscript.

References

1. Holford TR. The estimation of age, period and cohort effects for vital rates. *Biometrics*. 1983; 39:311–324. [PubMed: 6626659]
2. Holford TR. Understanding the effects of age, period, and cohort on incidence and mortality rates. *AnnuRevPublic Health*. 1991; 12:425–457.
3. Holford, TR.; Armitage, P.; Colton, T. Age-Period-Cohort Analysis Encyclopedia of Biostatistics. 2nd. Vol. Vol. John Wiley & Sons, Ltd.; 2005. p. 82-99.
4. Clayton D, Schifflers E. Models for temporal variation in cancer rates. II: Age-period-cohort models. *StatMed*. 1987; 6:469–481.
5. Tarone RE, Chu KC. Evaluation of birth cohort patterns in population disease rates. *AmJEpidemiol*. 1996; 143:85–91.
6. Robertson C, Boyle P. Age-period-cohort analysis of chronic disease rates. I: Modelling approach. *Stat Med*. 1998; 17:1305–1323. [PubMed: 9682322]
7. Chu KC, Tarone RE, Brawley OW. Breast cancer trends of black women compared with white women. *Arch Fam Med*. 1999; 8:521–528. [PubMed: 10575392]
8. Jatoi I, Anderson WF, Rao SR, Devesa SS. Breast cancer trends among black and white women in the United States. *J Clin Oncol*. 2005; 23:7836–7841. [PubMed: 16258086]
9. Sim X, Ali RA, Wedren S, et al. Ethnic differences in the time trend of female breast cancer incidence: Singapore, 1968-2002. *BMC Cancer*. 2006; 6:261. [PubMed: 17078893]

10. Bergstrom R, Adami HO, Mohnen M, et al. Increase in testicular cancer incidence in six European countries: a birth cohort phenomenon. *J Natl Cancer Inst.* 1996; 88:727–733. [PubMed: 8637026]
11. Verhoeven R, Houterman S, Kiemeny B, Koldewijn E, Coebergh JW. Testicular cancer: marked birth cohort effects on incidence and a decline in mortality in southern Netherlands since 1970. *Int J Cancer.* 2008; 122:639–642. [PubMed: 17764111]
12. Liu S, Semenciw R, Waters C, Wen SW, Mery LS, Mao Y. Clues to the aetiological heterogeneity of testicular seminomas and non-seminomas: time trends and age-period-cohort effects. *Int J Epidemiol.* 2000; 29:826–831. [PubMed: 11034964]
13. Bray F, Richiardi L, Ekbom A, et al. Do testicular seminoma and nonseminoma share the same etiology? Evidence from an age-period-cohort analysis of incidence trends in eight European countries. *Cancer Epidemiol Biomarkers Prev.* 2006; 15:652–658. [PubMed: 16614105]
14. Spix C, Eletr D, Blettner M, Kaatsch P. Temporal trends in the incidence rate of childhood cancer in Germany 1987–2004. *Int J Cancer.* 2008; 122:1859–1867. [PubMed: 18076067]
15. McNally RJ, Cairns DP, Eden OB, Kelsey AM, Taylor GM, Birch JM. Examination of temporal trends in the incidence of childhood leukaemias and lymphomas provides aetiological clues. *Leukemia.* 2001; 15:1612–1618. [PubMed: 11587220]
16. Svensson E, Moller B, Tretli S, et al. Early life events and later risk of colorectal cancer: age-period-cohort modelling in the Nordic countries and Estonia. *Cancer Causes Control.* 2005; 16:215–223. [PubMed: 15947873]
17. Clayton D, Schifflers E. Models for temporal variation in cancer rates. I: Age-period and age-cohort models. *StatMed.* 1987; 6:449–467.
18. Carstensen B. Age-period-cohort models for the Lexis diagram. *Stat Med.* 2007; 26:3018–3045. [PubMed: 17177166]
19. SEER-17. Surveillance, Epidemiology, and End Results (SEER) Program (www.seer.cancer.gov) SEER*Stat Database: Incidence-SEER 17 Regs Limited-Use, Nov 2007 Sub (1973–2005) <Single Ages to 85+, Katrina/Rita Population Adjustment> -Linked To County Attributes - Total U.S., 1969–2005 Counties, National Cancer Institute, DCCPS, Surveillance Research Program, Cancer Statistics Branch, released April 2008, based on November 2007 submission. 2008
20. Robertson C, Gandini S, Boyle P. Age-period-cohort models: a comparative study of available methodologies. *J Clin Epidemiol.* 1999; 52:569–583. [PubMed: 10408997]
21. Devesa SS, Donaldson J, Fears T. Graphical presentation of trends in rates. *Am J Epidemiol.* 1995; 141:300–304. [PubMed: 7840107]
22. Robertson C, Boyle P. Age-period-cohort models of chronic disease rates. II: Graphical approaches. *Stat Med.* 1998; 17:1325–1339. [PubMed: 9682323]
23. McCullagh, P.; Nelder, JA. *Generalized Linear Models.* London: Chapman and Hall; 1983.
24. Reimers LL, Anderson WF, Rosenberg PS, Henson DE, Castle PE. Etiological heterogeneity for cervical carcinoma by histopathological type, using comparative age-period-cohort (APC) models. *Cancer Epidemiol Biomarkers Prev.* 2009 in the press.
25. Heuer C. Modeling of time trends and interactions in vital rates using restricted regression splines. *Biometrics.* 1997; 53:161–177. [PubMed: 9147590]
26. Breslow NE. Extra-poisson variation in log-linear models. *Applied Statistics-Journal of the Royal Statistical Society Series C.* 1984; 33:38–44.
27. Robertson C, Ecob R. Simultaneous modelling of time trends and regional variation in mortality rates. *Int J Epidemiol.* 1999; 28:955–963. [PubMed: 10597997]

Figure 1.

Descriptive analysis of cancer incidence rates in blacks and whites, for selected cancers, using data from the NCI SEER database. **A)** Female bladder cancer; data for APC analysis included all 2186 cases in blacks and 29,148 cases in whites; panel shows age-specific rates over time for two selected age groups. **B)** Female colorectal cancer; APC analysis included 17,460 cases in blacks and 129,975 cases in whites; panel shows longitudinal age-specific incidence rates for three selected birth cohorts. **C)** Male kidney cancer; APC analysis included 5,161 cases in blacks and 44,050 cases in whites; panel shows age-specific rates over time for two selected age groups. **D)** Female pancreas cancer; APC analysis included 4,316 cases in blacks and 28,463 cases in whites; panel shows cross-sectional age-specific rates for two selected calendar periods.

Figure 2. Black-versus-white incidence rate ratios and 95% point-wise confidence limits for selected cancers, estimated from SEER using comparative APC analysis. Models were fitted to data summarized in Figure 1. **A)** Female bladder cancer; panel shows fitted rate ratio from PH-A model. **B)** Female colorectal cancer, fitted rate ratios from PH-L model versus birth cohort. **C)** Male kidney cancer, fitted rate ratios from PH-T model versus age at diagnosis. **D)** Female pancreas cancer, fitted rate ratios from PH-X model versus period of diagnosis.

Figure 3.

Black-versus white incidence rates and incidence rate ratios for oral cancer in males, by age and calendar year, based on 5,516 cases in blacks and 42,468 cases in whites in SEER, using comparative APC analysis. **A)** Observed rates by age for two selected calendar periods. **B)** Fitted rate ratios, with corresponding 95% point-wise confidence limits, derived using separate APC models. Estimates are shown for the first and last calendar periods included in the analyses.

Figure 4.
The model selection problem in comparative APC analysis.

Table 1
Proportional Hazards Models for Cancer Rates

Hypothesis	Equality of Age-Period-Cohort Parameters	Goodness-of-Fit Test *, <i>df</i>	log Rate Ratio for Type 1 versus Type 0 Hazards given Hypothesis is True
Absolute PH (PH-A)	All except Intercept	$2(A + P) - 5$	$\log RR^{1:0} = \psi = (\mu^1 - \mu^0)$
PH by Age given Cohort (PH-L)	longitudinal age trends, age and period deviations	$A + P - 3$	$\log RR_c^{1:0} = \theta_c = (\mu^1 - \mu^0) + \{(\pi_L + \gamma_L)^1 - (\pi_L + \gamma_L)^0\} (c - \bar{c}) + (\tilde{\gamma}_c^1 - \tilde{\gamma}_c^0)$
PH by Period given Age (PH-T)	net drifts, period and cohort deviations	$A + 2P - 4$	$\log RR_a^{1:0} = \delta_a = (\mu^1 - \mu^0) + \{(\alpha_L - \gamma_L)^1 - (\alpha_L - \gamma_L)^0\} (a - \bar{a}) + (\tilde{\alpha}_a^1 - \tilde{\alpha}_a^0)$
PH by Age given Period (PH-X)	cross-sectional age trends, age and cohort deviations	$2A + P - 4$	$\log RR_p^{1:0} = \varphi_p = (\mu^1 - \mu^0) + \{(\pi_L + \gamma_L)^1 - (\pi_L + \gamma_L)^0\} (p - \bar{p}) + (\tilde{\pi}_p^1 - \tilde{\pi}_p^0)$

* The Hypothesis is rejected if the test is significant.