Published in final edited form as:

Stat Med. 2018 February 10; 37(3): 405–424. doi:10.1002/sim.7519.

Correlated Poisson Models for Age-Period-Cohort Analysis

Pavel Chernyavskiy^{1,2}, Mark P. Little¹, and Philip S. Rosenberg²

¹Radiation Epidemiology Branch, National Cancer Institute, DHHS, NIH, Division of Cancer Epidemiology and Genetics, Bethesda, Maryland 20892-9778, USA

²Biostatistics Branch, National Cancer Institute, DHHS, NIH, Division of Cancer Epidemiology and Genetics, Bethesda, Maryland 20892-9778, USA

Abstract

Age-period-cohort (APC) models are widely used to analyze population-level rates, particularly cancer incidence and mortality. These models are used for descriptive epidemiology, comparative risk analysis, and extrapolating future disease burden. Traditional APC models have two major limitations: 1) they lack parsimony because they require estimation of deviations from linear trends for each level of age, period, and cohort; and 2) rates observed at similar ages, periods, and cohorts are treated as independent, ignoring any correlations between them that may lead to biased parameter estimates and inefficient standard errors. We propose a novel approach to estimation of APC models using a spatially-correlated Poisson model that accounts for over-dispersion and correlations in age, period, and cohort, simultaneously. We treat the outcome of interest as event rates occurring over a grid defined by values of age, period, and cohort. Rates defined in this manner lend themselves to well-established approaches from spatial statistics in which correlation among proximate observations may be modeled using a spatial random effect. Through simulations, we show that in the presence of spatial dependence and over-dispersion: 1) the correlated Poisson model attains lower AIC; 2) the traditional APC model produces biased trend parameter estimates; and 3) the correlated Poisson model corrects most of this bias. We illustrate our approach using brain and breast cancer incidence rates from the Surveillance Epidemiology and End Results Program of the United States. Our approach can be easily extended to accommodate comparative risk analyses and interpolation of cells in the Lexis with sparse data.

Keywords

cancer incidence; h-likelihood; spatial GLMM; age-period-cohort analysis; SEER

Introduction

Age-period-cohort (APC) models are widely used to study population trends in disease incidence and mortality.^{1–5} APC models are especially popular in cancer surveillance where they have been used for comparative risk analyses (e.g., ^{6–10}), and projections of future disease burden (e.g., ^{11,12}). Recent development of a web tool for APC analysis¹³ has substantially improved ease of access to this methodology, which should facilitate its use in a wider range of studies.

Typically, APC models are fitted to a rate matrix indexed by age and calendar period of diagnosis, a so-called Lexis diagram. In studies of incidence, age effects account for natural history, period effects can reflect changes in disease definition or detection due to screening practices, and cohort effects correlate with prevalence of risk factor exposures that vary across birth cohorts.^{14,15} Models are usually specified in terms of either cross-sectional age-period parameters or longitudinal age-cohort parameters, but either way the fitted values are identical. Generically, model parameters include: 1) an intercept, 2) a trend parameter with respect to age, 3) a trend parameter with respect to period or cohort, and 4) deviation parameters (sometimes called “curvature” parameters) that capture non-linear patterns at each unique value of age, period, and cohort.

Despite the popularity of this model, existing methodology does little to address several potential problems that could have a substantial impact on the validity of model outputs. First, there might be correlation between metrically close cells in the Lexis diagram (i.e., similar ages, periods, or cohorts), which might contribute to over-dispersion. APC models usually estimate an overdispersion parameter using quasi-likelihood (e.g.,¹⁶), but this takes no account of possible correlation structure, which, if present, could result in bias and loss of efficiency. Second, rate estimates for extreme birth cohorts are poorly identified because they rely on just one or just a few cells of the Lexis diagram, a particular problem when 1-year age and calendar year intervals are used. Such volatility might be mitigated by borrowing information from nearby cells on the Lexis. Third, for the traditional model, a large Lexis diagram requires estimation of a large number of non-linear deviation parameters, which may impact model efficiency as well as its predictive performance because of over-fitting.

The desire to address these issues motivated development of smoothing methods for the Lexis diagram. One simple method that is widely used in practice consists of tabulating granular 1-year data into larger intervals (typically 3 to 5 years in length) prior to analysis. While this method is easy to implement, important patterns may be lost because of this inherently ad-hoc approach.¹⁴ More sophisticated approaches produce smooth Lexis diagrams using polynomial splines or restricted B-splines^{17,18} for age, period, and cohort deviations. Most recently, Chien et al.¹⁹ developed a smoothing method utilizing two-dimensional Bernstein polynomials estimated in a Bayesian framework. One major drawback of these approaches is that spline coefficients may be difficult to interpret in the context of public health. Furthermore, none of these approaches account for correlations among proximate cells on the Lexis.

In this paper, we propose a novel metrically-correlated Poisson model for age-period-cohort analysis. Our approach treats event rates on a Lexis diagram as data collected on a regular grid with a coordinate system indexed by age, period, and cohort. Defining a grid in this manner allows us to use established distance metrics (e.g., Euclidean) applied to pairs of observations to capture dissimilarity in all three temporal scales simultaneously. To take advantage of similarity among nearby rates, we estimate a spatial random effect with a flexible Matérn covariance function (e.g.,^{20,21}). This approach results in a more parsimonious model because individual age, period, and cohort deviations no longer need be estimated as fixed effects, and can instead be estimated as part of the spatial random effect.

In its most parsimonious form, we specify the mean of the underlying Poisson process using only an intercept and two linear trends: one for age, and another for either calendar period or birth cohort. Even so, our approach allows for the usual post-fit analysis of APC estimable functions (e.g., longitudinal age curve; for others see Table 1 of Rosenberg and Anderson¹) by using linear combinations of fixed effects and an appropriately-summarized spatial random effect. In our analysis, we do not tabulate rates into arbitrary time intervals *a priori*, instead we use the finest data available (1-year intervals). Using a more granular temporal resolution should allow us to capture more complex patterns of disease incidence and mortality.

In addition to adjusting for correlation among similar ages, periods, and cohorts, the correlated Poisson model reduces noise in single-year data and produces a desirable smooth Lexis diagram. The Matérn covariance function ensures that the spatial random effect “borrows information” from nearby observations more so than from distant observations. The degree of smoothing is controlled by the estimated random effect variance and the estimated spatial decay parameter of the Matérn covariance function. Finally, our correlated Poisson model also accounts for any potential overdispersion by estimating the variance of the spatial random effect.

The remainder of this paper is organized as follows. First, we describe two cancer incidence datasets, and provide detailed specifications of both the traditional APC and the correlated Poisson models. Next, we present results of a simulation study, comparing goodness-of-fit of the two models, as well as comparing bias in trend parameter estimates. We conclude by applying both models to analyze incidence of brain cancer and *in situ* estrogen receptor positive (ER+) breast cancer ascertained through the National Cancer Institute’s Surveillance, Epidemiology and End Results (SEER) Program.

Materials and Methods

Cancer incidence data

Cancer incidence was ascertained through the SEER Program’s 13-registry database that covers approximately 26% of the United States population. For our first example, we extracted brain cancer incidence for all races in men and women combined in 1-year age and calendar year intervals for ages 30 to 84, diagnosed from January 1st, 1992 through December 31st, 2012. This dataset consisted of 1155 cells in the corresponding Lexis diagram that included 75 non-overlapping 1-year birth cohorts from 1908 through 1982. For our second example, we extracted *in situ* ER+ breast cancer incidence for post-menopausal non-Hispanic white women ages 50 to 64, diagnosed during the 13-year calendar period from January 1st, 2000 through December 31st, 2012. This dataset consisted of 195 cells in the Lexis with 27 non-overlapping 1-year birth cohorts from 1936 to 1962. The relatively narrow age range for this example was chosen to minimize any potential screening effects and to reflect the well-known differential etiology of breast cancer among different age groups.

In general, let Y_{apc} be the number of incident cancers in age group a and calendar period p , for birth cohort c , and let O_{apc} be the corresponding number of person-years. The number of

unique age, period, and cohort values will be denoted by A , P , and C , respectively. To relate the estimable functions from our model to those of the traditional APC model, we primarily refer to the case of equally-spaced age, period, and cohort intervals. However, we note that in contrast to the traditional model, our model can easily accommodate age, period, and cohort intervals with unequal lengths. The observed incidence rate per 100,000 person-years

is computed as $10^5 Y_{apc} / O_{apc}$. Figure 7 (panels A and C) shows two Lexis diagrams of observed incidence rates for the brain cancer and breast cancer. Birth cohort index c does not in itself define an orthogonal dimension because it is collinear with ages and periods: $c = p - a$. Furthermore, birth cohorts reflect an essential asymmetry of rectangular Lexis diagrams. Specifically, whereas all ages are observed at all calendar periods and vice versa, cohorts are observed in a variable number of cells with as few times as only once for the youngest and oldest birth cohort.

Traditional APC model

The traditional APC model represents expected incidence rates as a log-linear model with additive age, period, and cohort effects. Because cohort is collinear with age and period, APC models can be written in either the age-period cross-sectional form or the age-cohort longitudinal form (e.g., ⁷). In this paper, we will make use of the longitudinal specification, but the fitted values are identical. Omitting the offset, the linear predictor for the traditional APC model can be written as:

$$\log(\mu_{apc}) = \beta_0 + \beta_a(a - a_0) + \beta_c(c - c_0) + \tilde{\alpha}_a + \tilde{\pi}_p + \tilde{\gamma}_c \quad (1)$$

In equation (1) the intercept β_0 represents the log-rate at an arbitrary set of reference values a_0 and c_0 , $(a - a_0)$ and $(c - c_0)$ are the centered linear effects of age and birth cohort, β_a and β_c are the corresponding trend parameters, and $\tilde{\alpha}_a$, $\tilde{\pi}_p$, and $\tilde{\gamma}_c$ parameterize age, period, and cohort deviations from linear trends. Deviation parameters are constrained for identifiability ^{14,15}, such that: $\sum_a \tilde{\alpha}_a = \sum_p \tilde{\pi}_p = \sum_c \tilde{\gamma}_c = 0$ and $\sum_a \tilde{\alpha}_a(a - \bar{a}) = \sum_p \tilde{\pi}_p(p - \bar{p}) = \sum_c \tilde{\gamma}_c(c - \bar{c}) = 0$. In practice, reference values are often selected to be the mean age and birth cohort values. Parameters are typically estimated by maximizing the Poisson log-likelihood, or quasi-log-likelihood if an over-dispersion parameter appears necessary.

In practical applications, β_a is often called the ‘Longitudinal Age Trend’ (LAT) and β_c the ‘Net Drift’ (e.g., ²²). The Net Drift is often the main parameter of interest in applications to population health and cancer surveillance, reflecting the estimated annual percentage change (EAPC) in age-standardized rates.²³ The difference between the LAT and the Net Drift (i.e., $\beta_a - \beta_c$) is called the ‘Cross-sectional Age Trend’ (CAT). Age, period, and cohort deviations are interesting both on their own, and in combinations with linear trends. In APC terminology, any linear combination of APC parameters is called an estimable function. Rosenberg and Anderson¹ provide an overview of such functions and demonstrate their utility in cancer surveillance research.

Correlated Poisson model

We develop the metrically-correlated or “spatial” APC model with the traditional APC parameterization in mind. To illustrate the most parsimonious case, we include just the intercept and the linear age and cohort trends as fixed effects in equation (3). Non-linear APC deviations, quadratic and/or cubic terms, external covariates, and spline knots can be included as fixed effects or are implicitly absorbed by the random effects. To specify the “spatial” random effects, we first define the age-period-cohort distance for any two cells of the Lexis. We use the Euclidean distance metric, as is most commonly done in geostatistical literature. The L_1 distance metric (i.e., Manhattan distance) and the L_∞ distance metric were both considered, but they produced results nearly identical to those using the Euclidean distance. Thus, for any two ages (a_i and a_j), periods (p_i and p_j), and cohorts (c_i and c_j), the Euclidean age-period-cohort distance is computed as:

$$d_{apc} = \sqrt{(a_i - a_j)^2 + (p_i - p_j)^2 + (c_i - c_j)^2} \quad (2)$$

Omitting the offset, for the correlated Poisson model, the linear predictor is expressed as:

$$\begin{aligned} \log(\mu_{apc}) &= \beta_0 + \beta_a(a - a_0) + \beta_c(c - c_0) + w \\ w &\sim MVN\left(0, \Sigma(d_{apc})\right) \\ \Sigma(d_{apc}; \sigma^2, \nu, \rho) &= \frac{\sigma^2}{2^{\nu-1} \Gamma(\nu)} (\rho d_{apc})^\nu K_\nu(\rho d_{apc}) \end{aligned} \quad (3)$$

In equation (3), w is the spatially-structured random effect, which is assumed to follow a multivariate Normal distribution with mean 0, and variance covariance matrix $\Sigma(d_{apc})$. The covariance matrix is modeled using the Matérn covariance function (e.g.,^{20,21}) evaluated at distances d_{apc} , with parameters ρ , ν , and σ^2 , where K_ν is the modified Bessel function of the second kind of order ν . The variance of w is controlled by σ^2 , and the strength of spatial dependence is controlled by ρ and ν . The rate at which the correlation function decays to 0 is characterized by ρ , with smaller values indicating slower decay. Typically, instead of ρ , the distance at which correlation becomes sufficiently small is reported as the “effective” range. Parameter ν controls the smoothness of the underlying spatial process, specifically corresponding to the number of times it is differentiable when $d_{apc} = 0$. Thus, ρ and ν together control the degree of smoothing induced by the spatial random effect, with large values of ν and small values of ρ resulting in maximal smoothing. The remaining terms in the linear predictor have the same definition as in equation (1).

The Matérn was chosen among other spatial correlation functions because different values of ν allow for great parametric flexibility, which can help model covariances of data separated by small distances.^{20,24} Because the smoothness parameter ν cannot be estimated reliably at the same time as σ^2 and ρ , ν was held constant in all models. Several known values of ν correspond to convenient special cases for the Matérn covariance function; for example, $\nu = \frac{1}{2}$ corresponds to the standard Exponential covariance function, and as $\nu \rightarrow \infty$ the

Matérn converges to the standard Gaussian covariance function. In practice, the shape of the covariance function changes very little for $\nu > 2$, thus ν is typically specified on the interval $(0, 2]$. Given σ^2 , ρ , and ν , the covariance function depends only on the age-period-cohort distance d_{apc} , thus we have assumed the covariance function to be second-order stationary and isotropic.

Parameter Estimation

Parameters for the correlated Poisson model were estimated within the Hierarchical Generalized Linear Models (HGLM) framework, introduced and developed by Lee and Nelder.^{25–28} Our model falls under the Poisson-Normal family in their extended conjugate HGLM class (see Table 2 of Lee and Nelder²⁶). Models with spatially-correlated outcomes have been estimated successfully using the HGLM framework,^{29–33} and have been shown to produce broadly similar results to those obtained in a fully Bayesian framework.^{30,34} We briefly outline the general estimation procedure below, following more detailed description in HGLM literature.

Let $f_{\beta}(y|w)$ be the conditional distribution of Y given random effects w , and let $f_{\theta}(w)$ be the distribution of the random effects. For notation clarity, parameters associated with the random effects (i.e., the Matérn covariance parameters) have been collected into the vector θ . The joint h -likelihood is then defined as:

$$h(\beta, \theta; y, w) = \log f_{\beta}(y|w) + \log f_{\theta}(w) \quad (4)$$

Since we are concerned with modeling rates, we specify $f_{\beta}(y|w)$ to be Poisson with conditional mean μ_{β} , w , and $f_{\theta}(w)$ to be multivariate Normal with variance-covariance matrix $\Sigma(\theta)$.

Let X be the fixed effects design matrix, Z be the random effects design matrix, $W = (\mu/\eta)^2 V(\mu)^{-1}$ be the GLM weight function, and define the GLM-adjusted variable $Y^* = X\beta + ZW + (Y - \mu)(\eta/\mu)$. Conditional on θ , joint estimates for β and w are obtained by solving the HGLM estimating equations:

$$\begin{bmatrix} X'WX & X'WZ \\ Z'WX & Z'WZ + \Sigma(\theta)^{-1} \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{w} \end{bmatrix} = \begin{bmatrix} X'WY^* \\ Z'WY^* \end{bmatrix} \quad (5)$$

To estimate covariance parameters θ , we maximized the adjusted profile h -likelihood $p_{\beta, w}(h)$ that eliminates both w and β from the h -likelihood:

$$p_{\beta, w}(h) = h - \frac{1}{2} \ln \left| -\frac{1}{2\pi} \frac{\partial^2 h}{\partial(\beta, w) \partial(\beta, w)} \right| \quad (6)$$

where h is the likelihood in equation (4), and $\ln|\cdot|$ denotes the natural logarithm of a matrix determinant. This adjusted profile likelihood $p_{\beta, w}(h)$ is the first-order Laplace approximation to the restricted log-likelihood, after eliminating both fixed and random effects. Noh and

Lee³⁵ found that second-order approximations are slightly more precise, but generally not worth the additional computational cost. The complete model estimation procedure is implemented in R³⁶ within the `spaMM` package³⁷ using the “`HL(1, 1)`” method.^{32,35,38} We provide R code used for the analysis of the breast cancer example as a supplement.

Estimable Functions for the Correlated Poisson Model

Here, we illustrate how to extract familiar estimable functions (e.g., the longitudinal age curve) from the most parsimonious correlated Poisson model, which in addition to the spatial random effect includes only the intercept, the centered age trend, and the centered cohort trend. Let $\hat{\beta}$ denote the estimated vector of these fixed effects, and let $\hat{w}$ denote the vector of estimated spatial random effects, arranged in P blocks of A values: $\hat{w}' = [w_{a1p1}, w_{a2p1}, \dots, w_{Ap1}, w_{a1p2}, \dots, w_{Ap2}, \dots, w_{AP}]$. To compute parameters that mimic the role of traditional APC deviations, we propose to average $\hat{w}$ over two of the three temporal dimensions. For example, to obtain random age deviations ($\hat{w}_a$), we average $\hat{w}$ over periods and cohorts; similarly, to obtain random cohort deviations ($\hat{w}_c$), we average $\hat{w}$ over ages and periods. Once $\hat{w}_a$ and $\hat{w}_c$ are computed, we substitute these estimates into corresponding equations for the estimable functions of interest (e.g., see Rosenberg and Anderson¹ Table 1).

Using mixed models nomenclature, we are in effect redefining APC estimable functions as “predictable” functions, since these are now linear combinations of fixed and random effects. In general, we define the linear predictor for the predictable function of interest as $\eta = K'\beta + M'w$. In the subsequent sections, we explicitly define matrices K and M for the longitudinal age curve and the cohort relative risk. Although we do not address all possible functions of interest, it would be straightforward to extend our approach to other commonly used functions, e.g.: fitted temporal trends, cross-sectional age curve.

Age, period, cohort averaging matrices

Our goal is to specify matrices that average $\hat{w}$ across periods and cohorts to produce random age deviations $\hat{w}_a$, and across ages and periods to produce random cohort deviations $\hat{w}_c$. We define the averaging matrices M_A and M_C as follows:

$$\begin{aligned} M_A &= X_A (X_A' X_A)^{-1} \\ M_C &= X_C (X_C' X_C)^{-1} \end{aligned} \quad (7)$$

where X_A and X_C are the age and cohort design matrices, respectively. Here, we have chosen the usual inner product to project the random effects onto the columns spanned by X_A and X_C . Note that M_A is of dimension $n \times A$, and M_C is of dimension $n \times C$. The averaging matrices can now be applied to compute the random age and cohort deviations as follows: $\hat{w}_a = M_A' \hat{w}$, $\hat{w}_c = M_C' \hat{w}$.

Longitudinal age curve

We define the linear predictor for the longitudinal age curve ($\hat{R}_A$), centered on the reference cohort, as $\eta_{LAC} = K'_A \beta + M'_A w$, with K_A defined by:

$$K_A = \begin{bmatrix} 1 & a_1 & 0 \\ \vdots & \vdots & \vdots \\ 1 & a_A & 0 \end{bmatrix} \quad (8)$$

where $a_1, \dots, a_A$ are the unique ages, and M_A is the age averaging matrix, as defined in (7).

Collecting K_A and M_A into the matrix $L'_A = \begin{bmatrix} K'_A & M'_A \end{bmatrix}$, we compute the variance of $\hat{\eta}_{LAC}$ as:

$$Var(\hat{\eta}_{LAC}) = Var\left(L'_A \begin{bmatrix} \hat{\beta} \\ \hat{w} \end{bmatrix}\right) = L'_A \hat{C} L_A \quad (9)$$

where $\hat{C}$ is the estimated inverse of the left-hand side of the HGLM estimating equations:

$\hat{C} = \begin{bmatrix} X' \hat{W} X & X' \hat{W} Z \\ Z' \hat{W} X & Z' \hat{W} Z + \sum(\hat{\theta})^{-1} \end{bmatrix}^{-1}$. Finally, to produce $\hat{R}_A$ with rates per 10^5 person-years, we apply the inverse-link function to $\hat{\eta}_{LAC}$ and scale by the person-years offset: $\hat{R}_A = 10^5 \exp(\hat{\eta}_{LAC})$. The upper and lower bounds of the asymptotic 95% confidence interval (CI)

are defined as $\hat{R}_A^U = 10^5 \exp\left(\hat{\eta}_{LAC} + 1.96 \sqrt{\text{diag}\left(Var(\hat{\eta}_{LAC})\right)}\right)$, and $\hat{R}_A^L = 10^5 \exp\left(\hat{\eta}_{LAC} - 1.96 \sqrt{\text{diag}\left(Var(\hat{\eta}_{LAC})\right)}\right)$.

Cohort relative risk

We define the linear predictor for the cohort relative risk (CRR), centered on the reference age, as $\eta_{CRR} = K'_C \beta + M'_C w$, with K_C specified as:

$$K_C = \begin{bmatrix} 0 & 0 & c_1 \\ \vdots & \vdots & \vdots \\ 0 & 0 & c_C \end{bmatrix} \quad (10)$$

where $c_1, \dots, c_C$ are the unique birth cohorts, and M_C is the cohort averaging matrix, as

defined in (7). Collecting K_C and M_C into the matrix $L'_C = \begin{bmatrix} K'_C & M'_C \end{bmatrix}$, we compute the variance of $\hat{\eta}_{CRR}$ as:

$$Var(\hat{\eta}_{CRR}) = Var\left(L'_C \begin{bmatrix} \hat{\beta} \\ \hat{w} \end{bmatrix}\right) = L'_C \hat{C} L_C \quad (11)$$

To produce the cohort relative risk, we apply the inverse-link, but do not scale by the offset, thus $CRR = \exp(\eta_{CRR})$. The upper and lower bounds of the asymptotic 95% CI are defined

$$\text{as } CRR^U = \exp\left(\hat{\eta}_{CRR} + 1.96 \sqrt{\text{diag}\left(Var(\hat{\eta}_{CRR})\right)}\right), \text{ and}$$

$$CRR^L = \exp\left(\hat{\eta}_{CRR} - 1.96 \sqrt{\text{diag}\left(Var(\hat{\eta}_{CRR})\right)}\right).$$

Adaptation of the Median Odds Ratio (MOR)

In Generalized Linear Mixed Models (GLMM) applications, random effect variances are best reported on the same scale as fixed effect to facilitate interpretation. To this end, the Median Odds Ratio (*MOR*) has been proposed^{39,40} as a summary measure of variability in the context of GLMM binomial models. Given identical fixed effects, the odds ratio for the difference of two linear predictors is given as $\exp(w_1 - w_2)$, where w_1 and w_2 are any two random effects. The *MOR* is defined as the median of the distribution of $\exp(|w_1 - w_2|)$, assuming $w \sim N(0, \sigma^2 I)$. In this case, the *MOR* can be computed by solving

$$2\Phi\left(\frac{\log(MOR)}{\sqrt{2\sigma^2}}\right) - 1 = 0.5, \text{ where } \Phi(\cdot) \text{ is the standard Normal CDF, resulting in}$$

$$MOR = \exp\left(\sqrt{2\sigma^2} \Phi^{-1}(0.75)\right).$$

In the context of our paper, the *MOR* needs to be adapted in several ways. For a Poisson GLMM, $\exp(w_1 - w_2)$ is the relative risk, thus we refer to the median of the distribution of $\exp(|w_1 - w_2|)$ as median relative risk (*MRR*). In addition, because the random effects in our model are not independently distributed, we need to generalize the standard Normal CDF to a scaled multivariate Normal CDF. Thus, following the derivation of the *MOR*, we define the

MRR as the solution to $2F\left(\frac{\log(MRR)}{\sqrt{2\sigma^2}}\right) - 1 = 0.5$, which produces the following result:

$$MRR = \exp\left(\sqrt{2\hat{\sigma}^2} F^{-1}(0.75, R(\hat{\theta}))\right) \quad (12)$$

where $F(\cdot)$ is the scaled multivariate Normal CDF, and $R(\hat{\theta})$ is the estimated random effects correlation matrix. The `qmvnorm()` function in the `mvtnorm` R package^{41,42} was used to compute $F^{-1}(\cdot)$. Like the *MOR*, the *MRR* can be interpreted as the median risk of a high-risk cell of the Lexis relative to a low-risk cell of the Lexis, for any two randomly selected cells. To make the *MRR* easily comparable to the magnitude of fixed effects parameters, we report the *MRR* as a percentage, i.e. $100\% \times (MRR - 1)$ in Table 1 and Table 2.

Simulation Study

Design

Data for the simulation study were generated under two models: the traditional APC model, and the correlated Poisson. Under the correlated Poisson, data were generated in four scenarios where the magnitude of the random effects variance and the strength of spatial dependence varied. In the fifth scenario, data were generated under the traditional APC model. For each scenario, 1000 realizations were randomly generated from a Poisson distribution, and the two models were fit to each realization. Parameters for the five scenarios were specified as follows:

1. large variance, strong spatial dependence: $\sigma^2 = 0.05, \rho = 1/200$
2. large variance, weak spatial dependence: $\sigma^2 = 0.05, \rho = 1/30$
3. small variance, strong spatial dependence: $\sigma^2 = 0.01, \rho = 1/200$
4. small variance, weak spatial dependence: $\sigma^2 = 0.01, \rho = 1/30$
5. no over-dispersion, no spatial dependence: $\sigma^2 = \rho = 0$

Intuitively, the correlated Poisson model should perform best in scenario 1), with diminishing performance in scenarios 2) through 5). The traditional APC model should perform best in scenario 5), with diminishing performance in scenarios 4) through 1).

The magnitude of σ^2 corresponds to the amount of over-dispersion in the data, whereas the magnitude of ρ corresponds to the strength of spatial dependence induced by the random effect. “Large” and “small” variances were selected to correspond to over-dispersion values greater than 2 and less than 1.5, respectively for our simulated data. Values for “strong” and “weak” spatial dependence were selected to reflect the correlation among rates at maximal distances in the Lexis diagram. For our simulated Lexis, the maximum observed distance was 41.86 years. “Strong” spatial dependence induced a non-negligible correlation of 0.3 at maximal distances, whereas “weak” spatial dependence was selected such that rates were effectively independent (i.e., correlation of ≈ 0.05) at half the maximal distance. The smoothness parameter was kept at a constant value of 0.1.

To design a realistic simulation, fixed effects for both models were specified using parameter estimates derived from the analysis of ER+ in situ breast cancer in white women, ages 50 to 64, over 21 calendar periods from 1992 to 2012. For data generated under the traditional model, fixed effects consisted of 68 total parameters, with: $\beta_0 = -7.470$, $\beta_a = 0.039$, and $\beta_c = 0.023$. For data generated under the correlated Poisson model, fixed effects consisted of 3 parameters, with: $\beta_0 = -7.470$, $\beta_a = 0.035$, and $\beta_c = 0.025$. In a traditional model, the Net Drift can be estimated using the parameter β_c and by computing the slope of cohort-specific RR. We extend this approach to the correlated Poisson model, where the Net Drift was estimated in two ways: 1) using the cohort fixed effect (β_c), and 2) using a fitted slope of cohort-specific RR (denoted by CRR), which was computed using linear combinations of

fixed and random effects. In traditional APC models these two methods produce an identical estimate, thus only one value is reported.

We assessed goodness-of-fit using the following criteria: 1) the classic Akaike's Information Criterion (AIC)^{43,44}; 2) the marginal AIC (mAIC); and 3) the conditional AIC (cAIC). The usual AIC is computed using the maximized Poisson log-likelihood penalized for the number of parameters estimated. cAIC is defined using the log-likelihood conditioned on random effects estimates, and mAIC is defined using the marginal log-likelihood with random effects integrated out.^{45–47} Smaller values for all three information criteria indicate a more preferable model.

Simulation Results

Figures 1 through 3 present the simulation results for the two models - the traditional APC and the correlated Poisson – fit to rates generated under the five scenarios described in the previous section. Except for Scenario 5 (no over-dispersion, no spatial dependence), the correlated Poisson model is selected over the traditional APC in every other scenario according to both the mAIC and the cAIC (Figure 1). Among scenarios where the correlated Poisson is selected, the largest difference in AIC occurs in Scenario 2 (large variance, weak spatial dependence) and the smallest differences occurs in Scenario 4 (small variance, weak spatial dependence). The strength of spatial dependence appreciably impacts model fit only if the variance is large (Scenarios 1 vs. 2); spatial dependence has little impact on model fit when variance is small (Scenarios 3 vs. 4).

Turning to trend parameter (LAT, Net Drift, CAT) estimates (Figure 2), the traditional APC model is unbiased only in Scenario 5 (no over-dispersion, no spatial dependence). The traditional model is most biased when variance is large (Scenarios 1 and 2), and less biased when variance is small (Scenarios 3 and 4). The correlated Poisson model corrects most of the bias: it is unbiased for the LAT, and it is biased slightly upward for the Net Drift in all scenarios, except for Scenario 5. In all scenarios except for Scenario 5, estimating the Net Drift using a slope of the cohort RR produces a less biased estimate than the fixed effect parameter for the correlated Poisson model. Incorrectly fitting the correlated Poisson to rates that are independent and not over-dispersed produces a downward biased LAT and CAT. Spatial dependence does not seem to influence the magnitude of bias in parameter estimates for both models.

Finally, the distribution of dispersion (Figure 3) follows an expected pattern for both models. Across the four scenarios where $\sigma^2 > 0$, the correlated Poisson model produces dispersion of ≈ 1 , whereas the traditional model is over-dispersed. In Scenario 5 (no over-dispersion, no spatial dependence), the traditional model produces dispersion of ≈ 1 and the correlated Poisson model is under-dispersed. Over-dispersion for the traditional APC is greatest in Scenario 2 (large variance, weak spatial dependence) and smallest in Scenario 3 (small variance, large spatial dependence). Increasing the strength of spatial dependence appears to increase the over-dispersion of the traditional model, holding variance constant.

Application

The Application section is organized as follows. First, we compare the estimates and the estimable/predictable functions for the traditional model and the most parsimonious correlated Poisson (with only 3 fixed effects) for two real data examples: brain cancer incidence, and ER+ *in-situ* breast cancer incidence. For each example, three correlated Poisson models with different values of the Matérn smoothness parameter are estimated. Second, we examine the effect of varying the fixed effects structure of the correlated Poisson on the resultant parameter estimates.

Brain cancer incidence, ages 30 – 84, 1992–2012

All three correlated Poisson models produce smaller mAIC and cAIC values compared to the traditional model (Table 1). Among the correlated Poisson models, the model with a smooth spatial surface ($\nu=2$) produces the smallest mAIC and cAIC values. However, an examination of diagnostic plots revealed non-trivial deviations in the random effects quantile-quantile (Q-Q) plot and nonsensical leverage values for both Models 2 and 3. For these two models, the algorithm to compute the quantile of the multivariate Normal distribution did not converge, thus MRR values are reported as Did Not Converge (DNC). Because of this, Model 1 with $\nu=0.1$ was selected among the models shown in Table 1.

Compared to the traditional model, Model 1 has smaller LAT and CAT values and a nearly identical Net Drift along with longer confidence intervals (CI) for all parameters. The difference in the LAT and greater estimated uncertainty are clearly visible in the flatter longitudinal age curve with wider CI produced by the correlated Poisson model (Figure 4A). The cohort RR curve of the correlated Poisson model generally follows that of the traditional model (Figure 4B). Smoothing induced by the spatial random becomes clearly visible for younger birth cohorts (>1970) with relatively few cases, where the correlated Poisson model does not suffer from the volatility of the traditional model. Adjusted for age, brain cancer incidence was decreasing by 0.81% (1.41%, 0.22%) per year in this population. The decrease in brain cancer incidence rates occurred for every age group, except for the oldest age group (79 to 84-year-olds), where incidence has been increasing (Figure 5).

The effective range for Model 1 was estimated to be very large (Table 1), indicating that rates are highly correlated even at maximal distances within the Lexis. Compared to the Lexis diagram of observed incidence rates (Figure 7A), the fitted Lexis diagram (Figure 7B) is generally smooth, revealing that much of the variation in brain cancer incidence can be attributed to age. In fact, incidence rates change little between ages 30 to 50 and ages 70 to 84, with the greatest rate of change occurring between ages 50 and 70.

ER+ *in-situ* breast cancer incidence, ages 50 – 64, 2000–2012

In contrast to brain cancer incidence, ER+ breast cancer incidence is relatively insensitive to the Matérn smoothness parameter (Table 2). Additionally, all three correlated Poisson models produced acceptable diagnostic plots. Model 3 produces the smallest mAIC and cAIC among the correlated Poisson models, with both values smaller than the AIC of the traditional model, and thus it was selected among the models in Table 2.

The LAT and Net Drift estimates for Model 3 differ substantially from those of the traditional model. According the Model 3, ER+ breast cancer incidence rates do not change significantly with age and are stable over time in this population. These findings are consistent with a flat longitudinal age curve (Figure 4C) and cohort RR curve with confidence bounds that generally overlap with $RR=1$ (Figure 4D). Age-group-specific trends over time and birth cohort (Figure 6) reveal complex patterns, with the youngest age group (50 to 54-year-olds) having higher incidence than the second youngest age group (55 to 59-year-olds) since 2007. Incidence rates have been declining for all three age groups after reaching a peak in 2010.

The effective range for Model 3 was estimated to be 5.9 years (Table 2), indicating that rates more than 6 years apart are essentially independent. Even with a small effective range, the correlated Poisson model still produced smoothed rates, whereas it would be difficult to discern meaningful patterns from the observed Lexis (Figure 7 C–D). The smoothed Lexis reveals uniformly high incidence for ages > 60 , with two birth cohorts (women born in 1950–51) having substantially lower rates.

Different fixed effects structures for the correlated Poisson

We examined three fixed effects structures for brain and breast cancer incidence: 1) most parsimonious model (intercept, linear age trend, linear cohort trend); 2) most parsimonious model with cubic age, period, and cohort splines; and 3) most parsimonious model with all traditional APC deviation parameters. In fitting these three models we evaluated the full spectrum of fixed effects: from the most parsimonious with only 3 fixed effects, to the least parsimonious with 148 and 52 fixed effects for brain and breast cancer, respectively. The model with cubic splines served as a “middle ground” alternative to modeling deviations from linear trends as entirely fixed or entirely random.

For brain cancer, including more fixed effects substantially reduced the estimated random effects standard deviation and effective range (Table 3). There was a 2.3-times decrease in the estimated random effects standard deviation between the most parsimonious and the least parsimonious model; the effective range decreased 7.2-times. As random effects variability decreased, so did the length of the fixed effects CI, resulting in more precise estimates. According to mAIC and cAIC, the model with age, period, and cohort cubic splines would be selected among the models in Table 3.

For breast cancer, models were generally less sensitive to changing fixed effects structure. The estimated random effects standard deviation decreased by 20% between the most parsimonious and least parsimonious models (Table 4). There was a greater impact on the effective range, however: the estimate was 2.3-times smaller, resulting in a model with nearly-independent random effects in the least parsimonious case. The most parsimonious model and the model with cubic splines had nearly identical mAIC and cAIC values, without clear evidence as to which of these should be selected.

Discussion

In this manuscript, we proposed a novel approach to APC analysis using spatially-structured random effects specified on the Lexis diagram. By treating rates in this manner, we can fit a metrically-correlated Poisson model, which uses a flexible Matérn correlation function to model spatial dependence among nearby rates in the Lexis. Despite having only an intercept and two linear trends as fixed effects in its most parsimonious form, we have shown how to construct familiar APC estimable functions (e.g., the longitudinal age curve) using our proposed model, which are now comprised of linear combinations of fixed and random effects. Our model can be used in a broad range of applications because it can be estimated using any R package designed for spatial GLMMs (e.g., *spaMM*³⁷). Thus, it would be advisable to compare our model against the traditional APC model in any application where non-negligible over-dispersion is present.

Our proposed approach provides a flexible multi-level modeling framework that can be easily extended to enable formal comparative analyses. For example, one might wish to compare incidence rates between several population-level groups (e.g., different ethnicities, tumor characteristics). To do so, one could add a grouping variable and interaction terms to the fixed effects of the correlated Poisson model. Spatial structure could be specified as identical across groups, or it could be free to vary between groups. The significance of interaction terms and/or group-specific spatial structure could then be tested using the appropriate information criteria.

Traditional APC models may struggle with sparse event counts, for example: in rates for rare cancers, or rates obtained by stratification by a grouping variable. In fact, the volatility in cohort RR for the traditional model (e.g., Figure 4 panel B) is likely due to sparse brain cancer incidence in the young birth cohorts. The correlated Poisson model appears particularly well-suited for this type of analysis due to its ability to “borrow strength” from nearby cells in the Lexis diagram and produce more stable estimates. In addition, if so desired, the task of imputing missing cell counts is relatively straightforward, because it falls under spatial interpolation or “kriging” – a common type of analysis in geo-statistics.

We also explored the impact of changing the number of fixed effects in correlated Poisson models by fitting a model with only an intercept and two trends, a model with all traditional APC deviations, and the “middle ground” model with an intercept, two trends, and three cubic splines. As expected, increasing the number of fixed effects reduced the amount of information carried by the random effects, decreasing both the estimated standard deviation and the effective range. The model with cubic splines provided an attractive alternative to modeling deviations from linear trends as wholly fixed or wholly random in the brain cancer example, but not in the breast cancer example. It is possible that using splines in correlated Poisson models improve fit for large Lexis diagrams (as in the brain cancer case), but not small Lexis diagrams. Using a traditional APC model with an additional spatial random effect appears unnecessary: information carried by the deviation parameters and the spatial random effect is essentially redundant and this model was never selected over its alternatives.

In our simulation study, the correlated Poisson with only three fixed effects was selected over the traditional model in all scenarios where any over-dispersion and even weak spatial dependence were present. In these cases, the traditional model produced parameter estimates that were substantially biased, whereas the correlated Poisson model corrected most of this bias. However, the correlated Poisson is not appropriate for rates that are known to be independent and/or are lacking evidence of over-dispersion. For independent rates without over-dispersion, the traditional model is unbiased with efficient standard errors, whereas the correlated Poisson results in standard errors that are too large (i.e., under-dispersion) and produces biased parameter estimates. In the case of under-dispersion, our model may be inadequate because the Poisson distribution may not fully account for the excess zeros, which are likely to occur in under-dispersed data with high spatial correlation.³³ Thus, an extension of our methods to zero-inflated Poisson and/or negative binomial models may be a fruitful avenue for future research.

While the simulation study provides strong evidence in favor of the correlated model, it also shows that some bias remains. We suspect that this remaining bias may be a consequence of the first-order Laplace approximation that is used to compute the objective function $p_{\beta,\omega}(h)$. To our knowledge, a comprehensive evaluation of the Laplace approximation in spatially structured models has not yet been performed, and merits further study. For count data in general, first-order Laplace approximation has been shown to out-perform Penalized Quasi-likelihood (PQL), whereas Gauss-Hermite quadrature has been deemed to be infeasible when the number of random effects is large^{35,48} – as is the case with spatially-structured data. Joe⁴⁸ found that the first-order Laplace approximation has increasing bias with decreasing observations per level of random effect, with 2 observations being the smallest value he considered. In spatially structured data, we typically obtain only a single observation at each location, making our case more difficult than the “worst case” considered by Joe. Bias in parameter estimates may suggest a role for the use of fully Bayesian inference procedures in this setting.

Our simulations show that cAIC was always smaller than mAIC, which is consistent with the findings of Ha et al.,⁴⁷ who determined that cAIC tends to select more complex models. The difference between the two information criteria in our settings appears to be large when random effects carry a lot of information – for example, large variance, strong spatial dependence. To this end, cAIC appears to under-penalize the likelihood by treating the estimated random effects as known. Furthermore, using mAIC for model selection is akin to what Vaida and Blanchard⁴⁶ call “population-focus” model selection, where the population in our context is the complete Lexis diagram.

The effective range was estimated to be quite large and unrealistic in several models, for example: Models 1 and 2 (Table 1). It is known that the range parameter is poorly identified by the data, with a likelihood surface that is particularly flat,²⁴ therefore large estimates should not be entirely unexpected. Furthermore, it has been shown that even a misspecified Matérn can be asymptotically optimal for interpolation,^{20,49} thus this unrealistic estimate should not immediately cause concern. Despite these challenges, Kaufman and Shaby recently showed⁵⁰ it is still better to estimate ρ , with better estimates expected when the effective range is small relative to the spatial extent of the data (i.e., falls within the maximal

observed distance). It is possible that the true effective range for the brain cancer example is beyond the maximum observed distance, producing a nonsensical estimate of the effective range. In contrast, the effective range for the breast cancer example was both realistic and relatively insensitive to model specification.

There are several potentially useful methodological extensions of our model in terms of accounting for spatial structure, and the corresponding computational methods. These are discussed briefly here to highlight possible future research directions. First, there are several ways to compute the distance between cells in a Lexis besides the usual Euclidean distance. Instead, one could use a more flexible distance metric (e.g., Minkowski) to compute the age-period-cohort distance d_{apc} . In a brief sensitivity analysis (not shown) we found that using Manhattan distances (i.e., L_1 -norm) had no appreciable impact on model estimates and inferences.

Second, here we assumed the spatial structure to be isotropic and stationary on the Lexis. Although these are simplifying assumptions, estimation of a non-stationary and/or anisotropic spatial random effect is well beyond the scope of this paper. It is plausible, however, that the spatial correlation in the APC model context is anisotropic. Diagonals on the Lexis diagram (e.g., Figure 7) follow identical birth cohorts, which may translate into stronger spatial dependence along a 45-degree angle from the origin, and weaker spatial dependence in other directions. Additionally, the assumption of isotropy and stationarity may become unrealistic as the Lexis diagram grows, as may have been the case with our brain cancer example. A large Lexis has greater potential for localized patterns to occur and thus a non-stationary covariance function may be more appropriate. Thus, we posit that a follow-up study may be useful to formally assess the nature of anisotropy for spatially-correlated models in the context of APC analysis.

Third, instead of treating age, period, and cohort values as coordinates, one could think of rates as data collected on a regular lattice and proceed to define a spatial adjacency matrix to use along with Conditional/Simultaneous Auto-regressive models (CAR/SAR). This approach may be particularly computationally attractive for large datasets (e.g., > 5000 cells in the Lexis diagram), because CAR/SAR models can be parameterized using sparse matrices. We note that data of this size may frequently arise in comparative risk analyses, where we observe the entire Lexis diagram for each comparison group. However, it is possible that the spatial dependence induced by these types of models could be overly simplistic and inappropriate for most APC analyses, thus further study is needed.

Finally, one could argue that fully Bayesian estimation may be attractive because it would properly account for uncertainty in the spatial covariance parameters. Indeed, Bayesian computation is popular across spatial statistics, but the lack of clear consensus regarding prior specification and convergence diagnostics can still be challenging. More studies are needed to determine which combination of distance matrix specification (point-referenced vs. adjacency-based), nature of spatial dependence (isotropic vs. anisotropic), and estimation methodology (likelihood-based vs. Bayesian) produces the most favorable results in the context of APC analysis.

Acknowledgments

This work was supported by the Intramural Research Program of the National Institutes of Health, the National Cancer Institute, Division of Cancer Epidemiology and Genetics. The authors are grateful for the insightful suggestions of two referees and an associate editor who helped improve this manuscript.

References

1. Rosenberg PS, Anderson WF. Age-period-cohort models in cancer surveillance research: ready for prime time? *Cancer Epidemiol Biomarkers Prev.* 2011; 20(7):1263–1268. [PubMed: 21610223]
2. Shiels MS, Chernyavskiy P, Anderson WF, et al. Diverging Trends in Premature Mortality in the U.S. by Sex, Race and Ethnicity in the 21st Century. *The Lancet.* 2017; 389(10073):1043–1054.
3. Wang Z, Yu C, Wang J, Bao J, Gao X, Xiang H. Age-period-cohort analysis of suicide mortality by gender among white and black Americans, 1983–2012. *Int J Equity Health.* 2016; 15(1):107–116. [PubMed: 27412030]
4. Kramer MR, Valderrama AL, Casper ML. Decomposing Black-White Disparities in Heart Disease Mortality in the United States, 1973–2010: An Age-Period-Cohort Analysis. *Am J Epidemiol.* 2015; 182(4):302–312. [PubMed: 26199382]
5. McNally RJ, Blakey K, James PW, Gomez Pozo B, Basta NO, Hale J. Increasing incidence of thyroid cancer in Great Britain, 1976–2005: age-period-cohort analysis. *Eur J Epidemiol.* 2012; 27(8):615–622. [PubMed: 22760704]
6. 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):261–272. [PubMed: 17078893]
7. Rosenberg PS, Anderson WF. Proportional hazards models and age-period-cohort analysis of cancer rates. *Stat Med.* 2010; 29(11):1228–1238. [PubMed: 20209480]
8. Kelly SP, Rosenberg PS, Anderson WF, et al. Trends in the Incidence of Fatal Prostate Cancer in the United States by Race. *Eur Urol.* 2017; 71(2):195–201. [PubMed: 27476048]
9. Chernyavskiy P, Little MP, Rosenberg PS. A unified approach for assessing heterogeneity in age-period-cohort model parameters using random effects. *Stat Methods Med Res.* 2017 962280217713033.
10. Siegel RL, Fedewa SA, Anderson WF, et al. Colorectal Cancer Incidence Patterns in the United States, 1974–2013. *J Natl Cancer Inst.* 2017; 109(8):1–6.
11. Rosenberg PS, Barker KA, Anderson WF. Estrogen Receptor Status and the Future Burden of Invasive and In Situ Breast Cancers in the United States. *J Natl Cancer Inst.* 2015; 107(9):1–7.
12. Petrick JL, Kelly SP, Altekruse SF, McGlynn KA, Rosenberg PS. Future of Hepatocellular Carcinoma Incidence in the United States Forecast Through 2030. *J Clin Oncol.* 2016; 34(15):1787–1794. [PubMed: 27044939]
13. Rosenberg PS, Check DP, Anderson WF. A web tool for age-period-cohort analysis of cancer incidence and mortality rates. *Cancer Epidemiol Biomarkers Prev.* 2014; 23(11):2296–2302. [PubMed: 25146089]
14. Holford TR. Understanding the effects of Age, Period, and Cohort on incidence and mortality rates. *Annual Review of Public Health.* 1991; 12:425–457.
15. Holford TR. The estimation of age, period and cohort effects for vital rates. *Biometrics.* 1983; 39(2):311–324. [PubMed: 6626659]
16. Maccullagh, P., Nelder, JA. *Generalized Linear Models.* New York, NY: Chapman and Hall; 1992.
17. Carstensen B. Age-period-cohort models for the Lexis diagram. *Stat Med.* 2007; 26(15):3018–3045. [PubMed: 17177166]
18. Heuer C. Modeling of time trends and interactions in vital rates using restricted regression splines. *Biometrics.* 1997; 53(1):161–177. [PubMed: 9147590]
19. Chien LC, Wu YJ, Hsiung CA, Wang LH, Chang IS. Smoothed Lexis Diagrams with Applications to Lung and Breast Cancer Trends in Taiwan. *Journal of the American Statistical Association.* 2015; 110(511):1000–1012.
20. Stein, ML. *Interpolation of Spatial Data: Some Theory for Kriging.* New York: Springer; 1999.

21. Banerjee, S., Carlin, BP., Gelfand, AE. Hierarchical Modeling and Analysis for Spatial Data. 2. CRC Press; 2015.
22. Robertson C, Gandini S, Boyle P. Age-period-cohort models: a comparative study of available methodologies. *Journal of Clinical Epidemiology*. 1999; 52(6):569–583. [PubMed: 10408997]
23. Fay MP, Tiwari RC, Feuer EJ, Zou Z. Estimating average annual percent change for disease rates without assuming constant change. *Biometrics*. 2006; 62(3):847–854. [PubMed: 16984328]
24. Zhang H. Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. *Journal of the American Statistical Association*. 2004; 99(465):250–261.
25. Lee Y, Nelder JA. Hierarchical Generalized Linear Models. *Journal of the Royal Statistical Society (Series B)*. 1996; 58(4):619–678.
26. Lee Y, Nelder JA. Hierarchical generalised linear models: A synthesis of generalised linear models, random-effect models and structured dispersions. *Biometrika*. 2001; 88(4):987–1006.
27. Lee Y, Nelder JA. Likelihood Inference for Models with Unobservables: Another View. *Statistical Science*. 2009; 24(3):255–269.
28. Lee, Y., Nelder, JA., Pawitan, Y. Generalized Linear Models with Random Effects Unified Analysis via H-likelihood. Boca Raton: Chapman & Hall/CRC; 2006.
29. Lee Y, Nelder JA. Modelling and analysing correlated non-normal data. *Statistical Modelling*. 2001; 1(1):3–16.
30. Lee YD, Yun S, Lee Y. Analyzing weather effects on airborne particulate matter with HGLM. *Environmetrics*. 2003; 14:687–697.
31. Dutta S, Mondal D. An h-likelihood method for spatial mixed linear models based on intrinsic auto-regressions. *J R Stat Soc B*. 2015; 77(3):699–726.
32. Lee W, Lee Y. Modification of REML algorithm for HGLMs. *Stat Comput*. 2012; 22:959–966.
33. Lee Y, Alam M, Noh M, Rönnegård L, Skarin A. Spatial modeling of data with excessive zeros applied to reindeer pellet-group counts. *Ecology and Evolution*. 2016; 6(19):7047–7056. [PubMed: 28725382]
34. Jang JM, Lee Y, Lawson A, Browne WJ. A comparison of the hierarchical likelihood and Bayesian approaches to spatial epidemiological modelling. *Environmetrics*. 2007; 18:809–821.
35. Noh M, Lee Y. REML estimation for binary data in GLMMs. *Journal of Multivariate Analysis*. 2007; 98(5):896–915.
36. R: A Language and Environment for Statistical Computing [computer program]. Vienna, Austria: R Foundation for Statistical Computing; 2015.
37. Rousset F, Ferdy JB. Testing environmental and genetic effects in the presence of spatial autocorrelation. *Ecography*. 2014; 37(8):781–790.
38. Yun S, Lee Y. Comparison of hierarchical and marginal likelihood estimators for binary outcomes. *Computational Statistics & Data Analysis*. 2004; 45(3):639–650.
39. Larsen K, Merlo J. Appropriate assessment of neighborhood effects on individual health: integrating random and fixed effects in multilevel logistic regression. *American journal of epidemiology*. 2005; 161(1):81–88. [PubMed: 15615918]
40. Larsen K, Petersen JH, Budtz-Jørgensen E, Endahl L. Interpreting parameters in the logistic regression model with random effects. *Biometrics*. 2000; 56(3):909–914. [PubMed: 10985236]
41. mvtnorm: Multivariate Normal and t Distributions. [computer program]. Version 1.0–52016.
42. Genz, A., Bretz, F. Computation of Multivariate Normal and t Probabilities. Springer; Berlin Heidelberg: 2009.
43. Akaike H. A new look at the statistical model identification. *IEEE transactions on automatic control*. 1974; 19(6):716–723.
44. Akaike H. Likelihood of a model and information criteria. *J Econometrics*. 1981; 16(1):3–14.
45. Donohue MC, Overholser R, Xu R, Vaida F. Conditional Akaike information under generalized linear and proportional hazards mixed models. *Biometrika*. 2011; 98(3):685–700. [PubMed: 22822261]
46. Vaida F, Blanchard S. Conditional Akaike information for mixed-effects models. *Biometrika*. 2005; 92(2):351–370.

47. Ha ID, Lee Y, MacKenzie G. Model selection for multi-component frailty models. *Stat Med*. 2007; 26(26):4790–4807. [PubMed: 17476647]
48. Joe H. Accuracy of Laplace approximation for discrete response mixed models. *Computational Statistics & Data Analysis*. 2008; 52(12):5066–5074.
49. Zhang H, Wang Y. Kriging and cross-validation for massive spatial data. *Environmetrics*. 2010; 21(3–4):290–304.
50. Kaufman C, Shaby B. The role of the range parameter for estimation and prediction in geostatistics. *Biometrika*. 2013; 100(2):473–484.

Figure 1.

Figure 2.

Figure 3.

Figure 4.

Figure 5.

Figure 6.

Figure 7.

Table 1

Parameter estimates and goodness-of-fit criteria for correlated Poisson and traditional APC models, fit to brain cancer incidence rates. Longitudinal Age Trend (LAT), Cross-sectional Age Trend (CAT), and Net Drift with 95% confidence intervals in parentheses. Random effect SD is reported on the linear predictor scale. MRR is the median relative risk between a cell of the Lexis at higher risk and lower risk, selected at random. mAIC is the marginal AIC, cAIC is the conditional AIC, and R^2 corresponds to the squared correlation between observed and fitted counts.

Parameter	Correlated Poisson			Traditional APC*
	Model 1	Model 2	Model 3	
LAT $\beta_a(\%)^{\dagger}$	2.97 (2.30, 3.64)	2.53 (0.90, 4.20)	1.48 (−1.62, 4.68)	3.42 (3.24, 3.60)
Net Drift $\beta_c(\%)^{\dagger}$	−0.81 (−1.41, −0.22)	−0.75 (−2.24, 0.76)	−0.83 (−3.64, 2.07)	−0.85 (−1.02, −0.69)
CAT $\beta_a - \beta_c(\%)^{\dagger}$	3.81 (3.40, 4.23)	3.31 (2.04, 4.59)	2.33 (0.11, 4.59)	4.31 (4.24, 4.39)
Random effect SD σ	0.16 (0.14, 0.17)	0.92 (0.80, 1.07)	0.55 (0.17, 0.74)	
MRR (%) ^{‡‡}	81.06	<i>DNC</i>	<i>DNC</i>	
Effective range [†] (years)	795	2126	89	
Smoothness $\nu(fixed)$	0.1	0.5	2	
Deviance	782.09	1060.86	1199.65	1067.20
AIC				7538.01
mAIC	7533.15	7491.70	7483.83	
cAIC	7468.60	7428.66	7428.19	
R^2	0.83	0.75	0.71	0.74

* Estimates for 145 age, period, cohort deviations were excluded from the table in the interest of clarity.

[†] Effective range used here is the distance at which correlation becomes < 0.10

[‡] Reported in terms of annual percent change, i.e. $100\% \times (RR - 1)$

^{‡‡} Correlation matrix approximated using singular value decomposition with 1000 of the 1155 columns to ensure computational feasibility using the `qmvnorm()` function in R

Table 2

Parameter estimates and goodness-of-fit for correlated Poisson and traditional APC models, fit to ER+ *in-situ* breast cancer incidence rates. Longitudinal Age Trend (LAT), Cross-sectional Age Trend (CAT), and Net Drift with 95% confidence intervals in parentheses. Random effect SD is reported on the linear predictor scale. MRR is the median relative risk between a cell of the Lexis at higher risk and lower risk, selected at random. mAIC is the marginal AIC, cAIC is the conditional AIC, and R^2 corresponds to the squared correlation between observed and fitted counts.

Parameter	Correlated Poisson			Traditional APC*
	Model 1	Model 2	Model 3	
LAT $\beta_a(\%)^{\ddagger}$	0.53 (−1.23, 2.32)	0.24 (−1.91, 2.43)	0.69 (−0.77, 2.18)	0.96 (0.44, 1.49)
Net Drift $\beta_c(\%)^{\ddagger}$	−0.99 (−2.40, 0.45)	−1.22 (−2.86, 0.44)	−0.86 (−1.95, 0.25)	−0.66 (−1.06, −0.27)
CAT $\beta_a - \beta_c(\%)^{\ddagger}$	1.53 (0.18, 2.91)	1.48 (−0.04, 3.03)	1.56 (0.57, 2.56)	1.55 (1.20, 1.89)
Random effect SD σ	0.16 (0.14, 0.19)	0.13 (0.11, 0.16)	0.12 (0.10, 0.14)	
MRR $(\%)^{\ddagger}$	74.39	68.83	61.17	
Effective range ⁺ (years)	262	8.2	5.9	
Smoothness $\nu(fixed)$	0.1	0.5	2	
Deviance	98.59	129.95	133.92	241.71
AIC				1582.85
mAIC	1575.50	1568.10	1567.07	
cAIC	1527.04	1521.86	1521.26	
R^2	0.89	0.83	0.82	0.65

* Estimates for 49 age, period, cohort deviations were excluded from the table in the interest of clarity.

⁺ Effective range used here is the distance at which correlation becomes < 0.10

[‡] Reported in terms of annual percent change, i.e. $100\% \times (RR - 1)$

Table 3

Parameter estimates and goodness-of-fit criteria for correlated Poisson models with varying fixed effects structures fit to brain cancer incidence rates. Longitudinal Age Trend (LAT), Cross-sectional Age Trend (CAT), and Net Drift with 95% confidence intervals in parentheses. Random effect SD is reported on the linear predictor scale. MRR is the median relative risk between a cell of the Lexis at higher risk and lower risk, selected at random. mAIC is the marginal AIC, cAIC is the conditional AIC, and R^2 corresponds to the squared correlation between observed and fitted counts.

Parameter	Correlated Poisson		
	Int, age, cohort trend	Int, age, cohort trend + age, period, cohort cubic splines	Int, age, cohort trend + all traditional APC deviations
LAT $\beta_a(\%)^{\dagger}$	2.97 (2.30, 3.64)	2.96 (2.44, 3.48)	3.44 (3.01, 3.88)
Net Drift $\beta_c(\%)^{\dagger}$	-0.81 (-1.41, -0.22)	-1.03 (-1.50, -0.56)	-0.81 (-1.21, -0.42)
CAT $\beta_a - \beta_c(\%)^{\dagger}$	3.81 (3.40, 4.23)	4.03 (3.65, 4.42)	4.29 (4.04, 4.55)
Random effect SD σ	0.16 (0.14, 0.17)	0.10 (0.09, 0.12)	0.07 (0.06, 0.09)
MRR $(\%)^{\ddagger\ddagger}$	81.06	47.38	37.34
Effective range ⁺ (years)	795	795	110
Smoothness $\nu(fixed)$	0.1	0.1	0.1
Deviance	782.09	982.45	890.36
mAIC	7533.15	7468.21	7568.00
cAIC	7468.60	7434.21	7526.61
R^2	0.83	0.77	0.79

⁺Effective range used here is the distance at which correlation becomes < 0.10

[†]Reported in terms of annual percent change, i.e. $100\% \times (RR - 1)$

^{‡‡}Correlation matrix approximated using 1000 of the 1155 singular values to ensure computational feasibility using the `qmvnorm()` function in R

Table 4

Parameter estimates and goodness-of-fit criteria for correlated Poisson models with varying fixed effects structures fit to ER+ *in-situ* breast cancer incidence rates. Longitudinal Age Trend (LAT), Cross-sectional Age Trend (CAT), and Net Drift with 95% confidence intervals in parentheses. Random effect SD is reported on the linear predictor scale. MRR is the median relative risk between a cell of the Lexis at higher risk and lower risk, selected at random. mAIC is the marginal AIC, cAIC is the conditional AIC, and R^2 corresponds to the squared correlation between observed and fitted counts.

Parameter	Correlated Poisson		
	Int, age, cohort trend	Int, age, cohort trend + age, period, cohort cubic splines	Int, age, cohort trend + all traditional APC deviations
LAT β_a (%) [†]	0.69 (−0.77, 2.18)	−0.15 (−1.52, 1.22)	1.04 (−0.05, 2.15)
Net Drift β_c (%) [†]	−0.86 (−1.95, 0.25)	−1.47 (−2.62, −0.31)	−0.62 (−1.44, 0.20)
CAT $\beta_a - \beta_c$ (%) [†]	1.56 (0.57, 2.56)	1.33 (0.16, 2.52)	1.68 (0.95, 2.41)
Random effect SD σ	0.12 (0.10, 0.14)	0.09 (0.08, 0.11)	0.10 (0.08, 0.12)
MRR (%) [†]	61.17	48.23	49.79
Effective range ⁺ (years)	5.9	2.5	2.6
Smoothness ν (fixed)	2	2	2
Deviance	133.92	123.95	101.40
mAIC	1567.07	1561.00	1599.42
cAIC	1521.26	1521.31	1542.47
R^2	0.82	0.84	0.86

⁺ Effective range used here is the distance at which correlation becomes < 0.10

[†] Reported in terms of annual percent change, i.e. $100\% \times (RR - 1)$