Statistical Science
2016, Vol. 31, No. 4, 591–610
DOI: 10.1214/16-STS580

A Review and Comparison of Age–Period–Cohort Models for Cancer Incidence

Theresa R. Smith and Jon Wakefield

Abstract. Age–period–cohort models have been used to examine and forecast cancer incidence and mortality for over three decades. However, the fitting and interpretation of these models requires great care because of the well-known identifiability problem that exists; given any two of age, period, and cohort, the third is determined. In this paper, we review the identifiability problem and models that have been proposed for analysis, from both frequentist and Bayesian standpoints. A number of recent analyses that use age–period–cohort models are described and critiqued before data on cancer incidence in Washington State are analyzed with various models, including a new Bayesian approach based on an identifiable parameterization.

Key words and phrases: Age–period–cohort models, identifiability, random walk priors.

1. INTRODUCTION

Age–period–cohort (APC) models allow the modeling of demographic rates, and in this paper we consider these models in the context of human disease. Time plays an important role in the incidence and progression of most diseases. However, there is no single time scale or group of time scales that accounts for temporal variation in incidence/progression for all health outcomes. In arguing for the careful consideration of the appropriate time scales for health data, Berzuini and Clayton (1994) state that time, "is simply the scale along which other causes operate" and that the role of time in statistical models is to act as a "surrogate or proxy measure" for these unobserved processes that contribute to disease risk. Thus, it is important to have a basic understanding of the dynamics of the disease endpoint when choosing how to incorporate time into a model for disease risk.

APC models account for these processes on three time scales: age, year of diagnosis (period) and year of birth (cohort). The period and cohort effects are both surrogates for exposure to external factors. Period effects include environmental and diagnostic factors. For example, the introduction of a new diagnostic procedure may lead to a jump in disease incidence across all age groups. Cohort effects represent risk factors that change over time and may have a delayed effect on disease outcomes. For example, life style factors such as alcohol and tobacco consumption can manifest themselves as cohort effects.

Cohort analysis and APC models have been a tool used by demographers and sociologists since the late nineteenth and early twentieth centuries when cohort started to be recognized as a key time scale in tracking vital statistics (e.g., fertility or mortality). For a fascinating review of the early developments in APC models, see Keiding (2011). In 1875, Wilhelm Lexis introduced a cohort–age diagram for representing the time scales along which we calculate vital rates (Lexis, 1875, Keiding, 1990). In the mid twentieth century, studying cohort effects though simple techniques such as plots of age-specific mortality by birth year helped medical researchers understand age–time interactions for diseases such as tuberculosis (Frost, 1939, Springett, 1950). Greenberg, Wright and Sheps (1950) introduced a primitive APC model with log incidence

Theresa R. Smith is Senior Research Associate, Lancaster Medical School, Lancaster University, Lancaster LA1 4YG, United Kingdom (e-mail: t.r.smith@bath.ac.uk). Jon Wakefield is Professor, Departments of Statistics and Biostatistics, University of Washington, Seattle, Washington 98195-4322, USA (e-mail: jonno@uw.edu).

rates of syphilis regressed on a nonlinear function of numeric age and categorical period and cohort effects.

The application of APC models to cancer incidence and mortality gained steam in the 1980s with a series of seminal papers including Osmond and Gardner (1982), Holford (1983) and Clayton and Schifflers (1987a, 1987b). APC models continue to be a standard tool in the field of cancer epidemiology: a search of the MEDLINE database for "age–period–cohort + cancer" returns 137 publications from 2010–2014. However, inference with APC models is not straightforward because of the linear dependence in the three time scales; that is, age + year of birth = year of diagnosis (i.e., age + cohort = period). In many contexts, including cancer epidemiology, forecasting is a primary aim; see Brown and Kessler (1988) for an early reference.

Interest in the estimation and identification issues inherent in APC models dates back at least as far as the sociology literature in the 1970s (Mason et al., 1973, Fienberg and Mason, 1979), while specific applications in cancer epidemiology began in the 1980s (Holford, 1983). Clayton and Schifflers discuss in detail age–period and age–cohort models (Clayton and Schifflers, 1987a) and the age–period–cohort model (Clayton and Schifflers, 1987b), including issues of model selection and nonidentifiability. Bayesian APC methods based on intrinsic conditional autoregressive (ICAR) priors date back to Berzuini and Clayton (1994) and Besag et al. (1995). The continued interest in APC modeling is evidenced by two books that have recently been published (Yang and Land, 2013, O'Brien, 2014).

In this paper, we review approaches to APC modeling and inference, with an emphasis on cancer incidence data, and we demonstrate different approaches using cancer data from Washington State. In contrast to other review papers, such as Holford (1991) and Møller et al. (2003), we critique a number of recent applied contributions, in order to demonstrate current misconceptions about what can and cannot be learned about the associations between disease risk and the age, period, and cohort time scales in APC models. We also propose a Bayesian approach based on a parameterization suggested by Kuang, Nielsen and Nielsen (2008b) and Martínez Miranda, Nielsen and Nielsen (2015).

The structure of this paper is as follows. In Section 2, we motivate our discussion by describing breast, cervical and lung cancer incidence data collected in Washington State over the period 1992–2011. Section 3 introduces notation, the basic APC model, the identifiability problem, and previous approaches to addressing the identifiability problem. In Section 4, we discuss extensions to the basic APC model and identifiability concerns within these extensions. Section 5 presents and critiques a collection of APC studies that have appeared in the recent literature. In Section 6, details on parametrization are given, including our Bayesian version of Kuang, Nielsen and Nielsen (2008b) and Martínez Miranda, Nielsen and Nielsen (2015). We return to the Washington State data in Section 7 and compare and contrast a number of proposed models. A discussion concludes the paper in Section 8. Supplement A (Smith and Wakefield, 2016) contains additional analyses, details on various points raised in the paper, and implementation details demonstrated with Danish lung cancer data from the `Epi` package (Carstensen et al., 2014). Supplement B (Smith and Wakefield, 2016) contains `R` code for all the models described in Section 6 applied to the Danish data.

2. MOTIVATING DATA

In this paper, we discuss APC models for modeling time trends in disease count data using cancer incidence data from the Washington State Cancer Registry (WSCR) as our motivating example. Cancer is a collection of many different diseases and illnesses that share the feature of unrestricted cell growth due to genetic changes that allow cancer cells to bypass or circumvent the body's normal growth regulation mechanisms (Escedy and Hunter, 2008). There are two distinct processes to consider when the primary outcome of interest is the number of incident cancer cases (i.e., the number of new cases in a given time frame). The first is the process by which the genetic changes or mutations occur, and the second is the process by which the cancerous cells are identified and the diagnosis of cancer is conferred.

The age and cohort time scales capture heterogeneity in the mutation process. With a few exceptions (e.g., certain leukemias and central nervous system tumors), the risk of cancer increases with age. Age is a surrogate for internal processes such as hormone exposure or the cumulative damage of random genetic mutations when DNA is being copied, but it can also be a proxy for external factors such as cumulative exposure to carcinogens. Cohort can be a surrogate for exposure to carcinogens. For example, inhalation of asbestos is the primary cause of a lung cancer called mesothelioma. Earlier birth cohorts have greater risk of mesothelioma than more recent cohorts because asbestos usage has been phased out (Martínez Miranda,

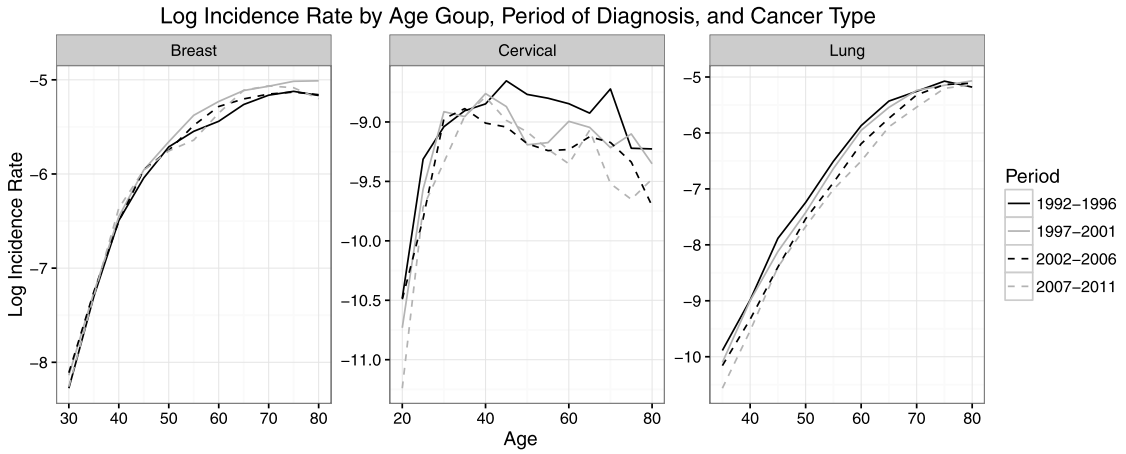

FIG. 1. *Log incidence rates by age group and period.*

Nielsen and Nielsen, 2015). On the other hand, the period time scale captures heterogeneity in the diagnosis process as new diagnostic procedures are introduced.

We compare various APC models using breast, cervical and lung cancer incidence data from the WSCR. The registry contains all reported cases of cancer for the state of Washington, which had a population of approximately 4.9 million inhabitants in 1990 rising to 6.7 million in 2010. The data used in this paper can be downloaded from https://fortress.wa.gov/doh/wscr/WSCR/Query.mvc/Query.

For breast and lung cancer, we treat the data as 11×4 matrices, where 11 is the number of age groups (30–34, ..., 80–84) and 4 is the number of 5-year periods between 1992 and 2011. The risk of cervical cancer starts earlier in life, so the data is a 13×4 matrix with age groups (20–24, ..., 80–84). We restrict our analyses to breast and cervical cancer among females and lung cancer among males. Figures 1 and 2 show the observed incidence rates by age and period and by age and cohort, respectively.

For all three cancers, the incidence rates clearly differ by age. The shape of the curves for breast and lung cancer are typical of noninfectious diseases with incidence rates increasing steadily throughout adulthood. In contrast, the incidence rates of cervical cancer (which is linked with human papilloma virus, a sexually transmitted infection) peaks in the 40–50 age group and then decreases. It is important to note that, without any prior knowledge of the etiology of cancer, the strong observed association of risk increasing with age in any given period can mathematically just as easily be explained as decreasing risk with increasing birth year in any given period, that is, cohort effects.

Age-specific breast cancer rates are similar across periods (Figure 1) and cohorts (Figure 2). It could be

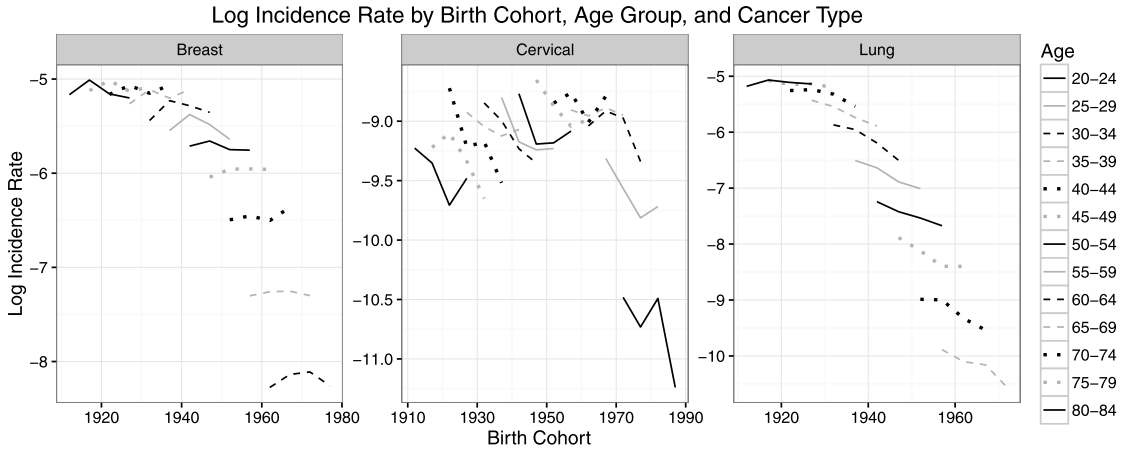

FIG. 2. *Log incidence rates by age group and cohort.*

that the cohort and period effects combine in such a way that marginally the age effects appear the same across periods and cohorts, but we might contextually believe that this is unlikely. For cervical cancer and lung cancer, the age-specific rates decrease with year of diagnosis (period); in Figure 1 the age curves are lower with increasing period but, as noted above, this could be just as well explained by decreasing risk with increasing cohort year in any given diagnosis period; see Figure A.1 in Supplement A. The age-specific rates of cervical and lung cancer also decrease with cohort; while the breast cancer cohort curves in Figure 2 are relatively flat, the cervical and lung curves decrease. It is apparent that cervical cancer is less common than lung and breast cancer; hence, the plots of the raw rates are relatively noisy.

3. BASICS OF AGE–PERIOD–COHORT MODELS

3.1 Notation

The usual APC model is an additive model for the log rate of incidence or mortality. Let $\mathbf{Y} = \{y_{ij} : i = 1, \ldots, A; j = 1, \ldots, T\}$ be a matrix of the number of cases in each of A age groups and T time points. The cohort index k is a function of age and period. If the age scale and time scale are the same (e.g., 5-year age groups and 5-year time intervals), then $k = A - i + j$ so that $k \in \{1, \ldots, A + T - 1\}$. If the age intervals are M times longer than the time intervals, then the cohort index is $k = M \times (A - i) + j$ so that $k \in \{1, \ldots, M(A - 1) + T\}$ (Heuer, 1997, Knorr-Held and Rainer, 2001). For most of this paper, we focus on equal-width age and period intervals. We discuss additional issues related to unequal intervals in Section 4.2. The cohorts index the diagonals of the age–period matrix, as shown for a small example in Table 1.

Let $\mathbf{N} = \{N_{ij}, i = 1, \ldots, A; j = 1, \ldots, T\}$ be the size of the risk set for each age and time group. Getting precise denominators is not as straightforward as finding good population estimates. For example, a simple way to calculate the denominator for the 30–34 age-band in the 2007–2011 period would be to add together the five mid-year (say July 1) population estimates for those aged 30–34. However, this excludes people who contribute partial-years to the total risk time including those who die in the first half of the year as well as people who turn 30 in the second half of the year. Furthermore, people who turn 35 or die in the second half of the year only contribute partial years to the total risk time but are counted as giving a full year. This complication does not impact inference (and we can use the simple mid-year population estimates) if we can assume constant birth and death rates (or immigration and emigration) within an age–period cell, because the under and over counting balance out. See Section 3.2 of Carstensen (2007) for a more detailed discussion. For this paper, we use the denominators given by WSCR, which are based on the mid-year population estimates produced by the Washington State Office of Financial Management.

TABLE 1

Indices of the age, period, and cohort parameters for equal-width age groups and time intervals. The body of the table shows the cohort index for each age–year combination. The indices in bold indicate the period and cohort effects that need to be forecasted to generate predictions for time periods 6–8. *We will cover forecasting in Section* 4.1

	Period							
Age	1	2	3	4	5	**6**	**7**	**8**
1	5	6	7	8	9	**10**	**11**	**12**
2	4	5	6	7	8	9	**10**	**11**
3	3	4	5	6	7	8	9	**10**
4	2	3	4	5	6	7	8	9
5	1	2	3	4	5	6	7	8

3.2 The Identification Problem in APC Models

The basic APC model is

$$\log \mathsf{E}\left[\frac{y_{ij}}{N_{ij}}\right] = \mu_{ij} = \delta + \alpha_i + \beta_j + \gamma_k. \tag{1}$$

In this model, it is tempting to interpret δ as the overall log rate of incidence and to interpret differences in the age effects (α_i), period effects (β_j), or cohort effects (γ_k) as log relative risks. However, direct interpretation of these effects is difficult because the model is over parametrized. There are two sources of identifiability to consider. The simpler one to account for is that which always occurs in models with factors: with an intercept in the model, we have one more level than is estimable and so a constraint is required. A typical solution is to impose a sum-to-zero constraint. The more insidious form of identifiability arises because of the linear dependence between the three factors, and there is no solution to this problem. Instead, one must make assumptions if one wishes to directly interpret the parameters in equation (1). Further, these assumptions are uncheckable from the raw data alone.

Suppose we group the intercept, age, period and cohort effects into a single vector, $\boldsymbol{\theta}$, where

$$\begin{aligned}\boldsymbol{\theta}' &= \{\delta, \boldsymbol{\alpha}', \boldsymbol{\beta}', \boldsymbol{\gamma}'\} \\ &= \{\delta, \alpha_1, \ldots, \alpha_A, \beta_1, \ldots, \beta_T, \gamma_1, \ldots, \gamma_K\}\end{aligned}$$

with $K = A + T - 1$. We see that, for a suitably defined design matrix $\mathbf{X}$, the vector of log rates is $\boldsymbol{\mu} = \boldsymbol{\mu}(\boldsymbol{\theta}) = \mathbf{X}'\boldsymbol{\theta}$. The matrix $\mathbf{X}$ is rank deficient in this case because the entries corresponding to the cohort effects are linearly dependent on the entries for the age and period effects (see Supplement A, Section B for more details and examples) and because of the general factor problem as described above. Thus, the full set of age, period and cohort effects are not identifiable.

Several authors, beginning with Fienberg and Mason (1979), have discussed the nonidentifiability of the individual terms of the APC model. Kuang, Nielsen and Nielsen (2008b) and Nielsen and Nielsen (2014), following Carstensen (2007), define the identifiability issue from a group theoretic perspective. The overall mean, as given in (1), is invariant to a translation on each set of effects and addition of a linear trend in the age, period and cohort parameters. We call this group of transformations $G = \{g : g\boldsymbol{\theta} = (g\delta, g\boldsymbol{\alpha}, g\boldsymbol{\beta}, g\boldsymbol{\gamma})\}$ where

$$g\delta = \delta - a - b - c + (A-1)d, \tag{2}$$

$$g\boldsymbol{\alpha} = \{\alpha_i + a + (i-1)d\}_{i=1}^{A}, \tag{3}$$

$$g\boldsymbol{\beta} = \{\beta_j + b - (j-1)d\}_{j=1}^{T}, \tag{4}$$

$$g\boldsymbol{\gamma} = \{\gamma_{A-i+j} + c + (A-i+j-1)d\}_{i=1,j=1}^{i=A,j=T} \tag{5}$$

for any real numbers a, b, c, d. An interpretation of these numbers is that a, b, c, are the overall levels of the age, period, cohort effects, respectively, and d is the linear trend. The log rates are invariant with respect to these transformations. That is, for any g, $\mu_{ij}(g\boldsymbol{\theta}) = \mu_{ij}(g\delta, g\alpha_i, g\beta_j, g\gamma_k) = \mu_{ij}(\delta, \alpha_i, \beta_j, \gamma_k) = \mu_{ij}(\boldsymbol{\theta})$. For example,

$$\begin{aligned}&\mu_{ij}(g\delta, g\alpha_i, g\beta_j, g\gamma_k) \\ &\quad= (\delta - a - b - c - (A-1)d) \\ &\qquad+ (\alpha_i + a + (i-1)d) \\ &\qquad+ (\beta_j + b - (j-1)d) \\ &\qquad+ (\gamma_{A-i+j} + c + (A-i+j-1)d) \\ &\quad= \delta + \alpha_i + \beta_j + \gamma_{A-i+j} \\ &\quad= \mu_{ij}(\delta, \alpha_i, \beta_j, \gamma_k).\end{aligned}$$

Since the data likelihood only depends on the age, period and cohort parameters through the log rates, it is also invariant to these groups of transformations. Hence, the full set of age, period, and cohort effects is not identifiable. Intuitively, since cohort (say) is a linear combination of age and period, levels in all three and linear trends are unidentifiable.

3.3 Proposed Remedies to the Identification Problem

The first attempts at fitting the full APC model relied on constraining the age, period and cohort effects to create an identifiable parametrization. A sum-to-zero constraint (i.e., $\sum_i \alpha_i = \sum_j \beta_j = \sum_k \gamma_k = 0$) was a common choice, which gives identifiability of the intercept δ but does not solve the identifiability problem because of the linear relationship between cohort and period and age (Mason et al., 1973, Fienberg and Mason, 1979, Holford, 1983, Rodgers, 1982, Clayton and Schifflers, 1987b). Equations (2)–(5) equate to the rank deficiency of the design matrix being 4 (i.e., we have 4 more parameters than are estimable from the data). The sum-to-zero constraints reduce this number by 3 (effectively fixing a, b, c), and we require one more constraint to produce an identifiable set. One approach is to assume two period or two cohort effects are equal (Mason et al., 1973). For example, Clayton and Schifflers (1987b) consider restricting the first differences of the period effects $(\beta_2 - \beta_1, \beta_3 - \beta_2, \ldots, \beta_T - \beta_{T-1})$ to be zero on average, which is equivalent to the restriction $\beta_1 = \beta_T$. Alternatively, one can restrict a sequential pair of effects to be equal (e.g., $\gamma_1 = \gamma_2$). Holford (1991) rejects these approaches because the estimated effects will depend on which pair of effects are restricted, and generally there is no scientific rationale for choosing, say $\gamma_1 = \gamma_2$ over $\gamma_4 = \gamma_5$. Figure 3 shows estimates of the age, period, and cohort effects for breast cancer in Bayesian ICAR models assuming sum-to-zero constraints and equality of 13 different sequential pairs of cohort effects. Each of the 13 models give identical fitted rates, but interpreting the effect curves leads to different conclusions depending on which cohorts effects are restricted. This is particularly true for the cohort curves, where the linear trend in cohort can appear either increasing or decreasing. In some instances, substantive knowledge is used to place the constraint. For example, Vaccarella et al. (2014), in modeling cervical cancer, constrained incidence rates to be equal at ages 45–49 and 65–69, because cervical cancer incidence is expected to be approximately constant after 45 years of age.

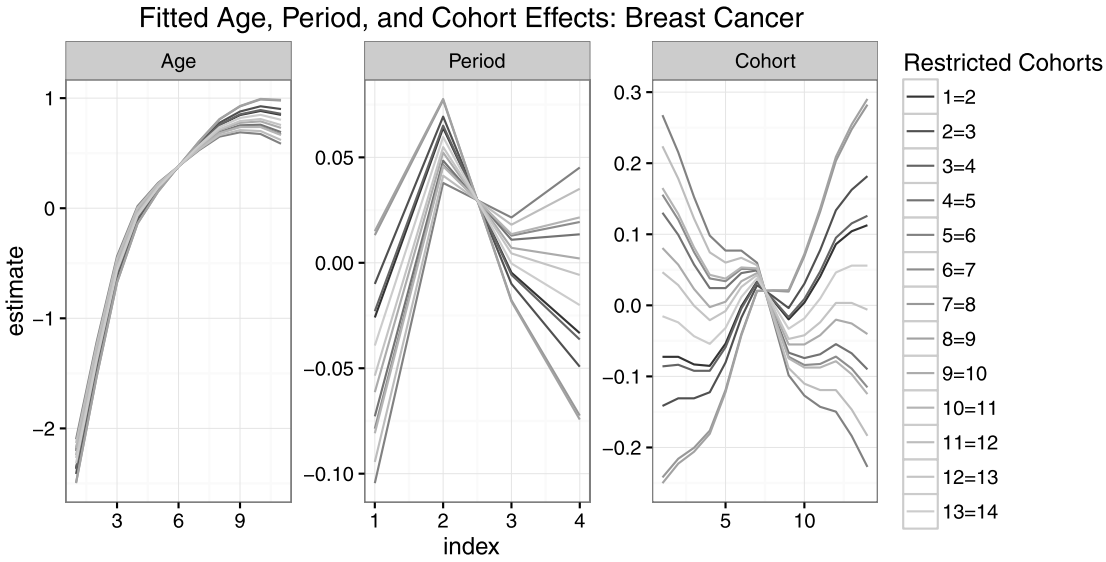

FIG. 3. *Estimates of the age, period and cohort effects under sum-to-zero constraints and equality of different sequential pairs of cohort effects.*

As we will discuss further in Section 6.2, Bayesian methods based on ICAR priors usually include sum-to-zero constraints as well as implicit penalties on the linear trends in age, period and cohort effects (Berzuini and Clayton, 1994, Besag et al., 1995, Knorr-Held and Rainer, 2001). In an extension to modeling mortality across multiple strata (e.g., regions), it can be shown that some contrasts of time effects are identifiable as long as one set of effects are shared across strata (Riebler and Held, 2010, Riebler, Held and Rue, 2012), further discussion is postponed until Section 4.3.

Rather than treat the age, period and cohort functions as factors, one might instead consider using continuous functions. The use of the latter alleviates the identifiability due to factors, but does not remove the linear dependence between the age, period and cohort effects. For example, suppose we assume a model in which the log rates are linear in each of age, period and cohort. If the slopes are denoted s_α, s_β, s_γ, then

$$\begin{aligned}\mu_{ij} &= \delta^\star + is_\alpha + js_\beta + (A - i + j)s_\gamma \\ &= \left(\delta^\star + As_\gamma\right) + i(s_\alpha - s_\gamma) + j(s_\beta + s_\gamma),\end{aligned}$$

so that the slopes are not identifiable. We can fit linear models in age and period or age and cohort, under the assumption that, respectively, there are no cohort or period effects (age is always included in models for cancer since it has such a strong effect). Linear forms are often referred to as "drift" models. Note that an age and period model produces an identical fit to an age and cohort model (Clayton and Schifflers, 1987a).

Penalized splines are one choice of continuous functions. As an example, Carstensen (2007) suggests fitting smoothing splines to produce age, period and cohort functions while imposing a set of constraints for identifiability. The default identification scheme is to choose a reference cohort, so that $f(k_{\text{REF}}) = 0$, and restrict the period function to be zero on average with zero slope. This model can be fit using the `Epi` package in `R` (Carstensen et al., 2014) or in `Stata` (Rutherford, Lambert and Thompson, 2010, Sasieni, 2012). It is appealing to model with continuous functions on each of the time scales, but the transparency of what is estimable is not so clear when such functions are used. The above assumptions on the period effects also seem ad hoc.

The medical statistics and demography literatures contain many other approaches to "solving" the identifiability problem. For example, Robertson and Boyle (1986) propose an approach based on the ability to access individual records, but this approach is based on assumptions also; see Clayton and Schifflers (1987b), page 477. The Lee–Carter model (Lee and Carter, 1992) and its cohort extension (Renshaw and Haberman, 2006) allow for age-specific period trends and age-specific cohort trends but still rely on linear constraints to guarantee a unique solution. The intrinsic estimator approach from demography uses the null space of $\mathbf{X}$ to define the linear constraints (Yang, Fu and Land, 2004), but Luo (2013) argues that this is no more scientifically justified than earlier approaches. Other alternatives include using simpler two-factor models or

a two-factor model with a predictor ("characteristic") in place of the third factor (O'Brien, 2000). For example, a plausible model for lung cancer rates may include age and period factors and the smoking rate as a linear predictor. Replacing the effects of one time scale with an explanatory variable (here, smoking) is a good alternative to the full age–period–cohort model in simple problems where the disease generating process is well understood. However, access to the relevant data may be an issue, and forecasting disease rates would require forecasts of explanatory variables.

In light of the nonidentifiability issues, care must be taken in giving tabular or graphical displays of estimates of the age, period and cohort estimates, as any estimates are conditional on a set of constraints that are (usually) not based on the underlying science and are uncheckable from the data alone. Thus, one can regard displays of the age, period, and cohort effects as merely exploratory tools (Carstensen, 2007) or one can focus on those functions of the time effects that are estimable. Clayton and Schifflers (1987b) show that the second differences of the time effects are invariant to the group of transformations defined in (2)–(5) above. For different choices of a, b, c and d, the second differences are identical. Let Δ represent the difference operator (e.g., $\Delta\alpha_i = \alpha_i - \alpha_{i-1}$) and Δ^2 represent the second-difference operator. Then, for example,

$$\begin{aligned} g\Delta^2\alpha_i &= g\alpha_i - 2g\alpha_{i-1} + g\alpha_{i-2} \\ &= \alpha_i + a + (i-1)d - 2\alpha_{i-1} \\ &\quad - 2a - 2(i-2)d + \alpha_{i-2} + a + (i-3)d \\ &= \Delta^2\alpha_i, \end{aligned}$$

illustrating that second differences are identifiable. The second differences describe the local curvature (or accelerations) in the age, period and cohort effects, and on the exponential scale, they are the ratios of the relative risks for two different time points (Clayton and Schifflers, 1987b).

One approach to "solving" the identifiability problem of APC models is to express the model only in terms of those functions of the age, period and cohort parameters that are estimable. For example, Holford (1983) partitions each set of effects into the linear effect and a curvature effect. Linear combinations of the curvature effects (e.g., the average) are estimable, and some functions of the slopes in the age, period and cohort effects are estimable. Suppose again that s_α, s_β and s_γ, are the linear slopes of the age, period and cohort effects, then functions of the form $us_\alpha + vs_\beta + (v-u)s_\gamma$, for all u, v, are estimable, as we now illustrate. As discussed above, the log rates are invariant to the addition of a linear trend d to the age and cohort effects and subtraction of d from the period effects, i.e., $s^\star_\alpha = s_\alpha + d$, $s^\star_\beta = s_\beta - d$, $s^\star_\gamma = s_\gamma + d$. Then

$$\begin{aligned} us^\star_\alpha + vs^\star_\beta + (v-u)s^\star_\gamma &= us_\alpha + vs_\beta + (v-u)s_\gamma \\ &\quad + ud - vd + (v-u)d \\ &= us_\alpha + vs_\beta + (v-u)s_\gamma. \end{aligned}$$

Hence, some functions of the linear slopes are invariant to any transformation g.

More recently, Rosenberg and Anderson (2011) showed that many epidemiological summaries, such as longitudinal or cross sectional age trends, can be expressed as estimable functions of the parameters in Holford's APC model. These summaries are available in a user-friendly web tool from the National Cancer Institute (http://analysistools.nci.nih.gov/apc/) (Rosenberg, Check and Anderson, 2014). Kuang, Nielsen and Nielsen (2008b) suggested a related parameterization of APC models, which we discuss in more detail in Section 6.1.

4. APC EXTENSIONS

4.1 Forecasting

Forecasts of mortality or incidence are important for allocating public resources and evaluating health policies. Given the identifiability issues, it is desirable to choose a forecasting method that does not depend on the choice of constraints in an ad-hoc identification scheme. Suppose we forecast rates h periods ahead in time for the same set of age groups. Then we want

$$\mu_{i,T+h} = \delta + \alpha_i + \beta_{T+h} + \gamma_{k+h},$$

where $k = A - i + T$. The forecasts therefore depend on projecting the period and cohort effects ahead h steps based on period and cohort effects fitted with the observed data. That is, for some functions f_β and f_γ, $\beta_{T+h} = f_\beta(\beta_{1:T})$ and $\gamma_{k+h} = f_\gamma(\gamma_{1:k})$. If $h < A$, then we project at most h new cohort effects because the rest are estimated from the observed data. For example, Table 1 shows that to predict 3 periods ahead for data with 5 age groups and 5 periods, we require forecasts for the period effect at time $j = 8$ and the cohort effects for $k = 10, 11, 12$.

As it is written, $\mu_{i,T+h}$ is a function of nonidentifiable effects, and so the forecasting functions f_β and f_γ must be carefully chosen so that $\mu(g(\alpha_i, \beta_{T+h},$

$\gamma_{k+h}, \delta)) = \mu(\alpha_i, \beta_{T+h}, \gamma_{k+h}, \delta)$. Two common functions are constant forecasts and linear extrapolation, that is, $f_\beta(\beta_{1:T}) = \beta_T$ and $f_\beta(\beta_{1:T}) = \beta_T + h\Delta\beta_T$. In a Bayesian context, constant forecasts arise from a first-order random walk prior, and linear extrapolation arises from a second-order random walk prior (Rue and Held, 2005).

Kuang, Nielsen and Nielsen (2008a) show that invariant forecasting functions (i.e., functions that give rise to the same forecasts of the log rates regardless of the chosen constraints) are of the form

$$f_\beta(\beta_{1:T}) = \beta_T + h\Delta\beta_T + f\big(\Delta^2\beta_3, \ldots, \Delta^2\beta_T\big)$$

for some function f. If $f(\cdot) = 0$, we recover linear extrapolation, but constant forecasts [i.e., $f_\beta(\beta_{1:T}) = \beta_T$] cannot fit into this form (and hence are not invariant).

4.2 Unequal Age and Period Intervals

In this paper, we focus on APC models for data tabulated by age and period intervals of equal width, but data are often tabulated with unequal intervals, usually with age intervals longer than period intervals. Cancer registries often share data using five-year rather than single-year age groups to avoid reporting or censoring small counts, which pose greater re-identification risk, especially if the data are further stratified by race, ethnicity or small areas. In many cases, single-year or three-year periods are still used. Even in cases where single-age incidence is available, collapsing into five-year age groups may still be used with rare diseases to avoid zeros counts. Thus, it is desirable to have APC models that generalize to unequal age and period intervals.

Unfortunately, there are additional identifiability issues that relate to the fact that cohorts cycle in blocks (Holford, 2006). For example, Table 2 shows an analogue of Table 1 where the age intervals are five times longer than the period intervals. The cohort indices in period column 1 also appear together in period column 6 shifted down one row. This gives rise to extra sources of nonidentifiability because the overall log rate is unchanged if we add a constant to every Mth period effect and subtract the same constant from every Mth cohort effect:

$$\delta + \alpha_i + \beta_j + \gamma_k = \delta + \alpha_i + (\beta_j + v_m I_{[j \equiv m \pmod M]}) + (\gamma_k - v_m I_{[k \equiv m \pmod M]}),$$

where $v_1, \ldots, v_M$ are any real numbers. This is not an issue in the equal age and period interval setting that we consider in this paper because each cohort effect occurs in every period (apart from on the boundaries). Holford (2006) extends his earlier to work to handle unequal intervals by introducing micro and macro time scales where macro time scales correspond to equal-time-duration age–period blocks of the data.

TABLE 2
Cohort index if age intervals are five times longer than period intervals

	Period									
Age	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	21	22	23	24	25	26	27	28	29	30
2	16	17	18	19	20	21	22	23	24	25
3	11	12	13	14	15	16	17	18	19	20
4	6	7	8	9	10	11	12	13	14	15
5	1	2	3	4	5	6	7	8	9	10

4.3 Stratified APC

A second extension of the basic APC model is to incorporate time-independent factors (strata) such as gender, region or disease subtype. In a stratified or multivariate APC model, certain cross-strata relative risks are identifiable up to a multiplicative constant [see (6) below] if one set of effects is shared across strata. Specifically, for the other two sets of nonshared effects, trends in relative risks within the same time index but between strata are identifiable.

In Riebler and Held (2010) and Riebler, Held and Rue (2012), the age effects are assumed to be the same across regions; in Riebler et al. (2012), period effects were shared, again across regions. Let r, $r = 1, \ldots, R$ represent the index for strata, then the most general stratified APC model is

$$\mu_{ijr} = \delta_r + \alpha_{ir} + \beta_{jr} + \gamma_{kr},$$

so that the full parameter space $\boldsymbol{\Theta}$ is an $R \times (1 + A + T + K)$ matrix with rows $(\delta_r, \boldsymbol{\alpha}_r, \boldsymbol{\beta}_r, \boldsymbol{\gamma}_r)$, $r = 1, \ldots, R$. We can apply the group of transformations in equations (2)–(5), with stratum-specific (a_r, b_r, c_r, d_r), to each row of $\boldsymbol{\Theta}$ without effecting the log rates. Hence, $4R$ constraints are required to produce identifiable age, period and cohort effects.

If, for example, the age effects are shared across strata (that is $\alpha_{ir} = \alpha_i$ for all i), then the unidentifiable level of the age effects and the unidentifiable linear trend are common to all strata, that is, $a_r = a$ and

$d_r = d$. Thus, the number of arbitrary constraints defining the group of transformations is reduced from $4R$ to $2R + 2$. As a result, the cross-strata log-relative risks for a given time period or cohort are identifiable up to an additive offset, implying identifiabiblity up to a multiplicative constant on the relative risk scale. For example, the relative risk between strata r_1 and r_2 at time 1 is $\exp(\beta_{1r_1} - \beta_{1r_2})$ if $b_{r_1} = b_{r_2} = 0$, but for a different g, this becomes

$$\begin{aligned} &g(\beta_{jr_1} - \beta_{jr_2}) = \beta_{jr_1} - \beta_{jr_2} + (b_{r_1} - b_{r_2}), \\ &g\big[\exp(\beta_{jr_1} - \beta_{jr_2})\big] \\ &\quad = \exp(\beta_{jr_1} - \beta_{jr_2}) \exp(b_{r_1} - b_{r_2}). \end{aligned} \tag{6}$$

Riebler and Held (2010) and Riebler, Held and Rue (2012) avoid the ambiguity in the additional offset (i.e., $b_{r1} - b_{r2}$) by imposing sum-to-zero constraints, which is natural given their use of ICAR priors. However, even without these constraints, some relationships between the cross-strata relative risks are preserved. That is, if we plot $\exp(\beta_{jr_1} - \beta_{jr_2})$ against j, the y-axis scaling is ambiguous (i.e., can be multiplied by any positive number), but the overall shape (including the existence and direction of linear trends) can be ascertained. Note that all of the identifiability problems discussed in Section 3.2 still persist within each stratum.

5. CRITIQUE OF RECENT APPLICATIONS OF APC MODELS

We now review several recent uses of APC models from the applied literature. We searched Web of Science for "age–period–cohort" and "cancer" and summarized ten of the most recent papers excluding those.

This exercise is intended to give a sense of the currin statistics-focused journals (e.g., *Statistics in Medicine*). This exercise is intended to give a sense of the current use (and misuse) of APC models, rather than to be a comprehensive review. The methods used in these papers were penalized splines fit using `Stata` (Louie, Mehanna and Sasieni, 2015, Ahmad, Ormiston-Smith and Sasieni, 2015, Valery, Laversanne and Bray, 2015) or the `Epi` package (Ho et al., 2015), Bayesian ICAR models fit using `BAMP` (Zheng et al., 2015) or `R-INLA` (Papoila et al., 2014), or models with a net drift and curvature terms (Seoane-Mato et al., 2014, Ananth et al., 2014, Cervantes-Amat et al., 2015, Tzeng and Lee, 2015). Most authors acknowledge the difficulties in simultaneously estimating effects along the three time scales; however, we found examples of interpretations or conclusions that depend on the unidentifiable age, period and cohort effects.

A common pitfall of several papers is that they give plots of the age, period, cohort effects and then interpret the plots in terms of linear trends. Seoane-Mato et al. (2014) compare the effects of age, period and cohort on mortality in three related cancers (oral, pharyngeal and oesophageal) for men and women using the methods in Osmond and Gardner (1982) and Holford (1991) to estimate net drift and curvature terms in period and cohort. They claim that, "there are many components of the information depicted graphically that do not vary across the different solutions, for example, the shape of the cohort effect (local changes or curvature) is independent of the solution chosen. Hence, where a trend is observed in the cohort effect, generally speaking this remains relatively unchanged in the different solutions." However, as we illustrated in Figure 3, linear trends in age, period or cohort effects can be substantially different even when local curvature is the same. Cervantes-Amat et al. (2015) analyze cervical cancer mortality in Spain using a model similar to that in Seoane-Mato et al. (2014) (age effects, overall linear drift and curvature in period and cohort). Again, they give interpretations relying on linear trends in these effects such as, "At about this time—the early 1940s—the probability of dying due to cervical cancer in Spanish women began to increase by birth cohort until 1962, when the risk again moved sharply downward."

Ananth et al. (2014) compare rates of placental abruption in seven countries. Using the `Epi` package, they fit a model with age effects, overall linear drift, and curvature in period and cohort, and plot the resulting curves (their Figure 3). With reference to this figure, the authors directly interpret the period effects saying, "In all seven countries, there was a slight temporal increase in the rate of abruption in the 1980s, and either a flattening of the rate (in the US) or a temporal decline, especially among the European countries." Tzeng and Lee (2015) use the APC model in (Holford, 1983) to forecast mortality from hepatocellular carcinoma (a form of liver cancer). They plot and give interpretations of the linear trends in the effects of all three time scales.

Ho et al. (2015) investigate temporal change in breast cancer mortality in Taiwan and use the sequential fitting approach discussed in Section 7 of Carstensen (2007). The age, period and cohort effects are fit using smoothing splines with the `Epi` package

and with the number of knots chosen by AIC. Age, period and cohort effects are plotted and interpreted in terms of linear trends, for example, "Period effects suggested slightly raised RRs in the periods 1976–80 and 1991–2000," a statement that cannot be made from the data alone.

In contrast, we were encouraged to find several papers that did not exhibit these misunderstandings of what is and is not identifiable from APC models. Louie, Mehanna and Sasieni (2015) and Ahmad, Ormiston-Smith and Sasieni (2015) use the penalized splines approach to fit APC models in `Stata` (Sasieni, 2012). Louie, Mehanna and Sasieni (2015) use APC models to project age and sex specific incidence rates of head and neck cancer occurring at eight different anatomical sites. They use a model with smoothing splines in each set of effects and net drift. Ahmad, Ormiston-Smith and Sasieni (2015) use the same approach to project lifetime risk of dying from cancer in Great Britain. Neither paper contains plots of age, period or cohort effects. Instead the plots and discussion are based on age–period and age–cohort plots similar to Figures 1 and 2.

Papoila et al. (2014) and Zheng et al. (2015) use Bayesian APC models with ICAR priors. Papoila et al. (2014) use stratified APC model for stomach cancer in 109 counties in Southern Portugal from 1990–2006, stratified by region and gender. The model includes gender-specific intercept, gender-specific age, gender-specific period effects, region-specific cohort effects, spatially-structured and unstructured regional effects and a region–period interaction term. Sum-to-zero constraints are used for all but the region–period interaction terms, where a much larger set of constraints is needed. The model is fit using the `R-INLA` software (Rue, Martino and Chopin, 2009), and male versus female relative risks are calculated by age and period. Zheng et al. (2015) use Bayesian APC models for incidence and mortality of childhood cancers in China using the `BAMP` software (Schmid and Held, 2007). Plots of individual age, period and cohort effects are avoided in both of these papers.

Although many of the papers we considered compare, rates across stratum such as regions, sexes or cancer types, Papoila et al. (2014) is the only paper to use joint, stratified APC model rather than post-hoc comparisons of individual analyses within each stratum. Because the cohort effects are assumed to be constant for males and females in Papoila et al. (2014), trends in the relative risk between two stratum across different time points (such as their Figure 2) do not depend upon an arbitrary choice of constraints.

6. PARAMETRIZATIONS OF THE APC MODEL

6.1 The Nielsen Parametrization of the APC Model

Kuang, Nielsen and Nielsen (2008a, 2008b), Nielsen and Nielsen (2014) and Martínez Miranda, Nielsen and Nielsen (2015) parametrize the APC model in terms of three initial log rates and the full set of second differences for data with equal-width age and time intervals. Kuang, Nielsen and Nielsen (2008b) construct a mapping from the rates at three initial time points using age–cohort indices (i.e., μ_{ik} instead of μ_{ij}). Here, we focus on a parametrization in Martínez Miranda, Nielsen and Nielsen (2015) based on age–period indexing. We refer to this as the MNN model. From Martínez Miranda, Nielsen and Nielsen (2015) equation (4.3), the overall log rate can be written as

$$\begin{aligned}
\mu_{ij} &= \mu_{A1} + (i-A)(\mu_{A1} - \mu_{A-1,1}) \\
&\quad + (j-1)(\mu_{A2} - \mu_{A1}) + a_{ij}, \\
a_{ij} &= \sum_{t=i}^{A-2} \sum_{s=t}^{A-2} \Delta^2 \alpha_{s+2} + \sum_{t=3}^{j} \sum_{s=3}^{t} \Delta^2 \beta_s \\
&\quad + \sum_{t=3}^{A-i+j} \sum_{s=3}^{t} \Delta^2 \gamma_s \\
&= \sum_{s=3}^{A} 1_{s \geq i+2} \big(s - (i+1)\big) \Delta^2 \alpha_s \\
&\quad + \sum_{s=3}^{j} (j-s+1) \Delta^2 \beta_s \\
&\quad + \sum_{s=3}^{A-i+j} (A-i+j-s+1) \Delta^2 \gamma_s.
\end{aligned} \tag{7}$$

This means that, for a design matrix $\mathbf{M}$, we can write $\boldsymbol{\mu} = \mathbf{M}'\boldsymbol{\xi}$, where

$$\begin{aligned}
\boldsymbol{\xi} = \{&\mu_{A1}, \mu_{A1} - \mu_{A-1,1}, \mu_{A2} - \mu_{A1}, \Delta^2 \alpha_3, \ldots, \\
&\Delta^2 \alpha_A, \Delta^2 \beta_3, \ldots, \Delta^2 \beta_T, \Delta^2 \gamma_3, \ldots, \Delta^2 \gamma_{A+T-1}\}.
\end{aligned}$$

The entries in $\mathbf{M}$ are equal to the multiplicative factors in (7). For example, for $A = T = 3$, the mapping

is

$$\begin{pmatrix}\mu_{11}\\ \mu_{12}\\ \mu_{13}\\ \mu_{21}\\ \mu_{22}\\ \mu_{23}\\ \mu_{31}\\ \mu_{32}\\ \mu_{33}\end{pmatrix} = \begin{pmatrix}1 & -2 & 0 & 1 & 0 & 1 & 0 & 0\\ 1 & -2 & 1 & 1 & 0 & 2 & 1 & 0\\ 1 & -2 & 2 & 1 & 1 & 3 & 2 & 1\\ 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0\\ 1 & -1 & 1 & 0 & 0 & 1 & 0 & 0\\ 1 & -1 & 2 & 0 & 1 & 2 & 1 & 0\\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0\\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0\\ 1 & 0 & 2 & 0 & 1 & 1 & 0 & 0\end{pmatrix} \cdot \begin{pmatrix}\mu_{31}\\ \mu_{31}-\mu_{21}\\ \mu_{32}-\mu_{31}\\ \Delta^2\alpha_3\\ \Delta^2\beta_3\\ \Delta^2\gamma_3\\ \Delta^2\gamma_4\\ \Delta^2\gamma_5\end{pmatrix}.$$

Note this parametrization has four fewer parameters than the model with an overall rate and the full set of age, period and cohort effects. This makes sense because the group transformation g, given by (2)–(5), is defined by four real numbers. Kuang, Nielsen and Nielsen (2008b) show that the parameter $\boldsymbol{\xi}$ is identifiable in that $\mu(\boldsymbol{\xi}) = \mu(\boldsymbol{\xi}^\star)$ only if $\boldsymbol{\xi} = \boldsymbol{\xi}^\star$. The parameters $\boldsymbol{\xi}$ are easily estimated via standard Poisson regression where the columns of $\mathbf{M}$ are treated as the predictors. This can be done directly or using the `apc` package (Nielsen, 2014).

If we re-write $\mu_{A1} - \mu_{A-1,1}$ as $\Delta\alpha_A - \Delta\gamma_2$, and $\mu_{A2} - \mu_{A1}$ as $\Delta\beta_2 + \Delta\gamma_2$, we see that this parametrization is similar to the model Holford (1983) proposed with linear and curvature effects. In Holford's model all the curvature effects and linear combinations of the slopes of the form $us_\alpha + vs_\beta + (v-u)s_\gamma$ are estimable. Letting $(u, v) = (1, 0)$ we see that $s_\alpha - s_\gamma$ is identifiable and letting $(u, v) = (0, 1)$ shows $s_\beta + s_\gamma$ is identifiable. Just as the set of estimable functions for the linear trends in Holford's model is infinite, the choice of the three initial points in the Nielsen parametrizations (which are, equivalently, functions of first differences) is not unique. Instead of $\{\mu_{A1}, \mu_{A-1,1}, \mu_{A2}\}$ we can choose any three $\{\mu_{i_1 j_1}, \mu_{i_2 j_2}, \mu_{i_3 j_3}\}$ as long as the indices of the three points are not linearly dependent (see Corollary 1 in Kuang, Nielsen and Nielsen, 2008b). Nielsen and Nielsen (2014) choose initial points based on the median age and cohort levels and not on the extremes, as in earlier papers. Using this guideline, as an example, the baseline rates reflect the rates in the middle of Table 1 rather than the corners.

We can also fit a MNN model in a Bayesian paradigm by specifying priors directly on $\boldsymbol{\xi}$. One sensible way to do this is do this is as follows:

$$\begin{aligned}\pi(\mu_{i_1 j_1}, \mu_{i_2 j_2}, \mu_{i_3 j_3}) &\propto 1,\\ \Delta^2\alpha_3, \ldots, \Delta^2\alpha_A | \tau_\alpha^2 &\sim \mathsf{N}(0, \tau_\alpha^{-2}\mathbf{I}),\\ \Delta^2\beta_3, \ldots, \Delta^2\beta_T | \tau_\beta^2 &\sim \mathsf{N}(0, \tau_\beta^{-2}\mathbf{I}),\\ \Delta^2\gamma_3, \ldots, \Delta^2\gamma_{A+T-1} | \tau_\gamma^2 &\sim \mathsf{N}(0, \tau_\gamma^{-2}\mathbf{I}),\\ \tau_\theta &\sim \pi_\phi, \quad \phi = \alpha, \beta, \gamma.\end{aligned}$$

We use an improper flat prior on the three initial points because we want the analysis to be invariant to which set of points we select. Note that there are only four time periods ($T = 4$) in our motivating example, meaning we only have two second differences for the period effects. As a result, estimating τ_β^2 is challenging and sensitive to the prior specification. We discuss this further in Section 7.3, and we include an example with $T = 11$ time periods in Supplement A, Section C.

6.2 Bayesian Intrinsic Autoregressive Models

Several authors have suggested a Bayesian formulation of the APC model with first- or second-order random walk priors on each of the three sets of effects (Berzuini and Clayton, 1994, Besag et al., 1995, Knorr-Held and Rainer, 2001, Riebler, Held and Rue, 2012). The second-order random walk (RW-2) prior follows from assuming that the second differences are independent, identically distributed Gaussian random variates. This in turn implies a Gaussian Markov random field (GMRF) wherein the conditional mean for an effect at a given time depends on the effect at the previous two and following two time points. That is, if $\boldsymbol{\alpha} | \tau_\alpha^2 \sim \text{RW-2}(\tau_\alpha^2)$, where τ_α^2 is the precision:

$$\begin{aligned}&\Delta^2\alpha_i | \tau_\alpha^2 \sim \mathsf{N}(0, \tau_\alpha^{-2}\mathbf{I})\\ &\quad\Longrightarrow\quad \pi(\boldsymbol{\alpha} | \tau_\alpha^2) \propto \tau_\alpha^{n-2} \exp\left(\frac{\tau_\alpha^2}{2}\sum_{i=3}^{A}(\alpha_i - 2\alpha_{i-1} + \alpha_{i-2})^2\right),\\ &\quad\Longrightarrow\quad \alpha_i \mid \boldsymbol{\alpha}_{-i}, \tau_\alpha^2 \sim \mathsf{N}\left(\frac{2}{3}(\alpha_{i-1} + \alpha_{i+1}) - \frac{1}{6}(\alpha_{i-2} + \alpha_{i+2}), \frac{1}{6\tau_\alpha^2}\right),\\ &\qquad\qquad i = 3, \ldots, A-2.\end{aligned}$$

The RW-2 prior does not have a representation as a proper multivariate normal distribution: the precision matrix implied by the full conditionals is of rank $n-2$. This class of prior is often called an *intrinsic* GMRF and, when there is an intercept and slopes in the RW-2 model, the impropriety is usually addressed by conditioning on a set of linear constraints spanning the null space of the implied precision matrix so that the improper density is equivalent to a proper density on a lower dimensional space (Rue and Held, 2005, Chapter 3). Specifically, for RW-2 models there is a sum-to-zero constraint and a zero slope constraint. For example, for the age effects, this is equivalent to $\boldsymbol{L}\boldsymbol{\alpha} = \mathbf{0}$ where the ith column of $\boldsymbol{L}$ is $\{1, i\}$, $i = 1, \ldots, A$. In the APC context, these constraints give a model which is not over-parameterized but do not yield interpretable intercepts or slopes, since the data cannot inform on these.

Knorr-Held and Rainer (2001) considered both RW-2 and RW-1 priors for the age, period and cohort effects (the RW-1 model arises from assuming the first differences are independent normal) but found that the RW-2 model gives better forecasts. The predictive distribution for future values under the RW-1 prior has a mean equal to the most recent observation. As discussed in Section 4.1, predictions of $\mu_{i,T+h}$ are not invariant to the group of transformations defined in (2)–(5) when constant forecasts for β_{T+h} or $\gamma_{A+T-1+h}$ are used. In contrast, the standard forecast from a RW-2 model (linear extrapolation from the two most recent observations) does give invariant forecasts of the log rates.

The usual specification for a Bayesian APC model using RW-2 priors is (Besag et al., 1995, Knorr-Held and Rainer, 2001):

$$\begin{aligned}
y_{ij}|\mu_{ij} &\sim \mathsf{Poi}\big(N_{ij}\exp(\mu_{ij})\big),\\
\mu_{ij} &= \delta + \alpha_i + \beta_j + \gamma_k + z_{ij},\\
\delta &\sim \mathsf{N}\big(m_\delta, s_\delta^2\big),\\
\boldsymbol{\alpha}|\tau_\alpha^2 &\sim \text{RW-2}\big(\tau_\alpha^2\big),\\
\boldsymbol{\beta}|\tau_\beta^2 &\sim \text{RW-2}\big(\tau_\beta^2\big),\\
\boldsymbol{\gamma}|\tau_\gamma^2 &\sim \text{RW-2}\big(\tau_\gamma^2\big),\\
z|\tau_z^2 &\sim \mathsf{N}\big(0, \tau_z^{-2} I\big),\\
\tau_\theta &\sim \pi_\phi, \quad \phi = \alpha, \beta, \gamma, z.
\end{aligned}$$

The additional unstructured random effect z_{ij} allows for heterogeneity around the constrained temporal effects and simplifies computation if using Markov chain Monte Carlo (MCMC) methods to fit the model since Gibbs sampling steps are available. Previous authors using this formulation have imposed sum-to-zero constraints on the age, period and cohort effects, but not zero linear trends (Riebler, Held and Rue, 2012). This makes sense because there are no linear terms in the model.

An alternative model uses the RW-2 prior with the full set of constraints and adds a linear slope in age, period and/or cohort. Borrowing from the spirit of the Nielsen parametrization, we incorporate a linear trend in age and time of diagnosis (period):

$$\begin{aligned}
y_{ij}|\mu_{ij} &\sim \mathsf{Poi}\big(N_{ij}\exp(\mu_{ij})\big),\\
\mu_{ij} &= \delta + \nu_1 i + \nu_2 j + \alpha_i + \beta_j + \gamma_k + z_{ij}.
\end{aligned} \tag{8}$$

So rather than take the log rate at a single point and two differences of log rates, we have specified the overall mean and two slopes.

For the remainder of this paper, we refer to the model with only sum-to-zero constraints as the *classical* RW-2 model and the fully constrained RW-2 model with linear trends in age and period as the *new* RW-2 model.

The RW-2 APC model is attractive because of its ease of computation and because it does not require equal age–period intervals; whereas, extending the MNN parameterization to incorporate unequal intervals is an open problem.

6.3 Variance Component Prior Specification

In this section, we will consider prior specification for the variance components in the APC model in the RW2 formulation. We place priors on the three components, α, β, γ, z independently. We first consider the term $\exp(\alpha_i)$, which is the relative risk associated with a unit (say, 5 year) increase in age. Marginally, this quantity is difficult to specify, since we need to think about the absolute scale, but, conditionally it is more straightforward which is an advantage of the RW2 (conditional) prior. We can interpret a RW2 model either directionally or undirectionally. Directionally, we have

$$\mathsf{E}[\alpha_i|\alpha_{i-1}, \alpha_{i-2}] = 2\alpha_{i-1} + \alpha_{i-2}, \tag{9}$$

with $\text{Prec}(\alpha_i|\alpha_{i-1}, \alpha_{i-2}) = \tau_\alpha^2/2$, for $i = 3, \ldots, A$. This mean form can be interpreted as fitting a line to the two previous time points and extrapolating one time unit forward. Undirectionally,

$$\begin{aligned}
&\mathsf{E}[\alpha_i|\alpha_{i-1}, \alpha_{i-2}, \alpha_{i+1}, \alpha_{i+2}]\\
&\quad = \frac{4}{6}(\alpha_{i+1} + \alpha_{i-1}) - \frac{1}{6}(\alpha_{i+2} + \alpha_{i-2}),
\end{aligned} \tag{10}$$

with $\text{Prec}(\alpha_i|\alpha_{i-1},\alpha_{i-2},\alpha_{i+1},\alpha_{i+2}) = 6\tau_\alpha^2$, for $i = 3,\ldots,A-2$. An interpretation is of fitting a quadratic through the 4 points either side of the point of interest and then interpolating. When it comes to prior specification for τ_α^2, either form can be used, but we find it more natural to use the directional case and consider the deviation we expect in the relative risk from the linear extrapolation from the last two points. Naturally, if the priors are specified consistently, a prior interval around the mean in (9) will be wider than one around the mean in (10) (which is reflected in the different multipliers on the precisions).

We place a $\text{Ga}(a,b)$ prior on the precision τ_α^2, so that the marginal distribution of the deviation from $\mathsf{E}[\alpha_i|\alpha_{i-1},\alpha_{i-2}]$ is distributed as a Student's t random variable with location 0, scale $2b_\alpha/a_\alpha$ and $d_\alpha = 2a_\alpha$ degrees of freedom (recall that the standard deviation of the normal deviation is $\sqrt{2}/\tau_\alpha$). Our overall strategy is to fix the degrees of freedom and then specify a range for the log relative risk deviations from the mean. A similar rationale is used in Fong, Rue and Wakefield (2010), Section 4.2, for a marginal random effect. We fix the degrees of freedom as $d_\alpha = 2$ (i.e., $a_\alpha = 1$) to give an exponential distribution of rate $2b_\alpha$ for τ_α. We solve for b_α by assuming the relative risk lies within $\exp(\pm c_\alpha)$ of $\exp(2\alpha_{i-1}+\alpha_{i-2})$ with probability q. Let $t^{d_\alpha}_{1-(1-q)/2}$ be the $100\times q$th quantile of a standard (i.e., zero mean, scale 1) Student's t random variable with d_α degrees of freedom; then consideration of the endpoint of the interval gives (for general d_α) $b_\alpha = c_\alpha^2 d_\alpha/[4(t^{d_\alpha}_{1-(1-q)/2})^2]$.

For the "over-dispersion" precision τ_z^2 we use the same procedure, though the approach is a little more straightforward since z_{ij} is a marginal random effect. If $\exp(\pm c_z)$ is the range that we believe, a priori, the over-dispersion will fall within, then we are led to $b_z = c_z^2 d_z/[2(t^{d_z}_{1-(1-q)/2})^2]$. We typically have less information on the size of the over-dispersion and choose slightly wider intervals than for the other components.

In general, we might believe that changes in the age parameters will be larger than the period or cohort parameters. In the analyses in Section 7, we take $a_\alpha = a_\beta = a_\gamma = 1$ and $q = 0.95$ and solve for $b_\alpha = b_\beta = b_\gamma$ based on $c_\alpha = \log(1.2)$, $c_\beta = \log(1.1)$, $c_\gamma = \log(1.1)$. These correspond to deviations from one-step-ahead predictions that are between $(0.83, 1.2)$, $(0.91, 1.1)$ and $(0.91, 1.1)$, respectively. For the measurement error, we take $a_z = 1$ and $c_z = \log(1.5)$, which corresponds to independent deviations between $(0.67, 1.5)$.

6.4 Implementation

In general, the posterior distributions of Bayesian APC models with RW-1 or RW-2 priors can be approximated with Markov chain Monte Carlo (MCMC) using the stand-alone packages `BAMP` (Schmid and Held, 2007) and `WinBUGS` (Lunn et al., 2000) or with integrated nested Laplace approximations (Rue, Martino and Chopin, 2009) using the recent `BAPC R` package (Riebler and Held, 2016). By default, the RW-2 model in the `R-INLA` implementation only uses sum-to-zero constraints; however, constraints on the overall linear trends can be added. The Bayesian MNN model can be fit using INLA or MCMC, and we demonstrate both here. Fitting this model in `R-INLA` is possible though not as simple as fitting the RW-2 models because the mean is given in terms of linear combinations of random effects. See Supplement A, Section C and Supplement B for more details.

For the MCMC, we sample from the posterior distribution $\pi(\boldsymbol{\xi},\tau_\alpha^2,\tau_\beta^2,\tau_\gamma^2 \mid y)$ using a combination of Gibbs sampling for the precisions and Metropolis-within-Gibbs steps for $\boldsymbol{\xi}$. For the Metropolis updates, we use a version of Metropolis–Hastings known as the Metropolis-adjusted Langevin algorithm (MALA). MALA was first proposed by Roberts and Tweedie (1996) and can have better mixing properties than the random walk Metropolis algorithm. Suppose the target distribution is $\pi(\mathbf{u})$ then the MALA proposal, given current state $\mathbf{u}_s$, is

$$\mathbf{u}' \sim \mathsf{N}\left(\mathbf{u}_s + \frac{h}{2}\boldsymbol{\Sigma}\nabla\log\pi(\mathbf{u}_s), h\boldsymbol{\Sigma}\right),$$

where $\nabla\log\pi(\mathbf{u}_s)$ is the gradient of the target distribution evaluated at the current state of the chain and $\boldsymbol{\Sigma}$ is an estimate of the posterior variance of u (e.g., one could use the asymptotic variance of the MLE of $\mathbf{u}$) scaled by $1.65^2/\dim(\mathbf{u})$. We do not pre-select the proposal scale h but instead adaptively modify h to achieve a sensible acceptance rate for the sampler.

Table 3 summarizes the three models discussed in Sections 6.1 and 6.2 by the form the main effects, the prior distributions on the age, period and cohort effects, and the estimation procedure.

7. COMPARISONS OF PROPOSED MODELS

We compare the Bayesian models based on parameter estimates of the second differences and on forecasts. One of the main purposes of fitting temporal models for disease incidence and mortality is to provide forecasts for the number of cases or deaths. For example,

TABLE 3
Summary of models that are compared in Section 7

Model Name	Main Effects	Age, Period, and Cohort Effects	Estimation
MNN-MLE	$\mu_{i_1,j_1}, \mu_{i_2,j_2}, \mu_{i_3,j_3}$	No penalization via priors	MLE
MNN-INLA	$\mu_{i_1,j_1}, \mu_{i_2,j_2}, \mu_{i_3,j_3}$	Direct iid normal prior on second differences	INLA
MNN-MCMC	$\mu_{i_1,j_1}, \mu_{i_2,j_2}, \mu_{i_3,j_3}$	Direct iid normal prior on second differences	MCMC
RW2-Class	δ	RW-2 with sum-to-zero constraints	INLA
RW2-New	δ, linear term in age and period	RW-2 with sum-to-zero and zero-linear-trends constraints	INLA

every January, the American Cancer Society and National Cancer Institute produce national and state-level forecasts for cancer mortality for the current year and for the next 3 years (Ghosh et al., 2007, Zhu et al., 2012, 2014). These projections are used for planning tasks such as estimating the total cost of cancer in the United States.

7.1 Parameter Estimates

Figures 4–6 show the estimated double differences based on the maximum likelihood estimates under the MNN model, the new RW-2 model with factor and linear effects in age and period and full linear constraints (8) and the Bayesian version of the MNN model fit via MCMC. We expect some differences in the estimated second differences between the MNN and RW2 models because the main effects are defined differently (see Table 3). However, estimates from the Bayesian models are very similar and, compared to the maximum likelihood estimates, attenuated toward zero. These relationships are consistent when using the classical RW-2 model (with only sum-to-zero constraints) or the Bayesian MNN model fit using `R-INLA` (see Supplement A, Section A for full results). The shrinkage of the estimates toward zero is most evident in the cohort effects for cervical cancer, which is expected given that cervical cancer is rare compared to breast and lung cancer. In practice, it may be more realistic to use a simpler model for cervical cancer, such as an age–period or an age–cohort model, or use substantive knowledge to constrain rates, as in Vaccarella et al. (2014).

In Figure 7, we plot the exponentiated double differences (which we refer to as ratios, since they correspond to ratios of two adjacent relative risks) for age against age index and for cohort against cohort index. We plot both the MLEs and the smoothed versions, and interpret the latter. To reiterate, the double differences are accelerations,so positive values could indicate an increasing trend that is becoming more pronounced or a decreasing trend that is becoming less steep. The age effects for cervical cancer are consistent with a flattening out of risk, as discussed by Vaccarella et al.

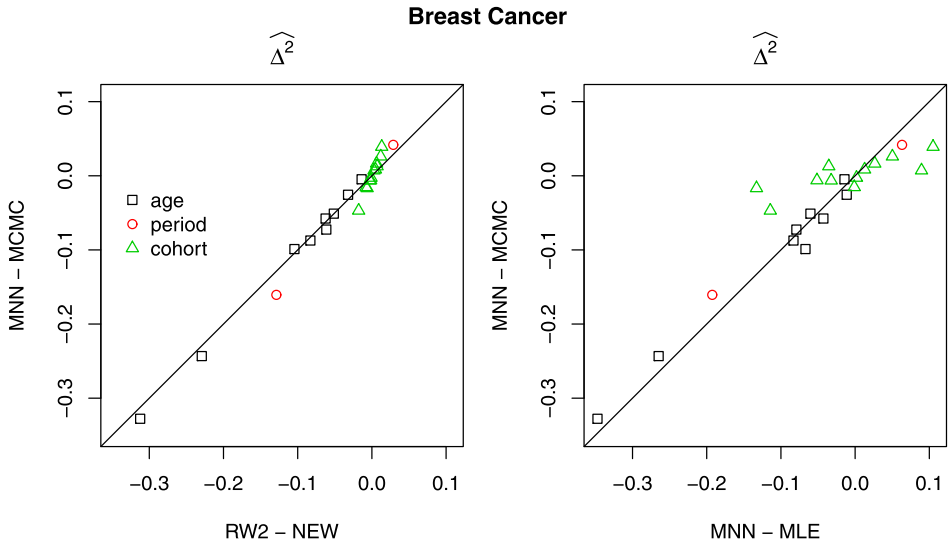

FIG. 4. *Estimates of double differences with breast cancer data.*

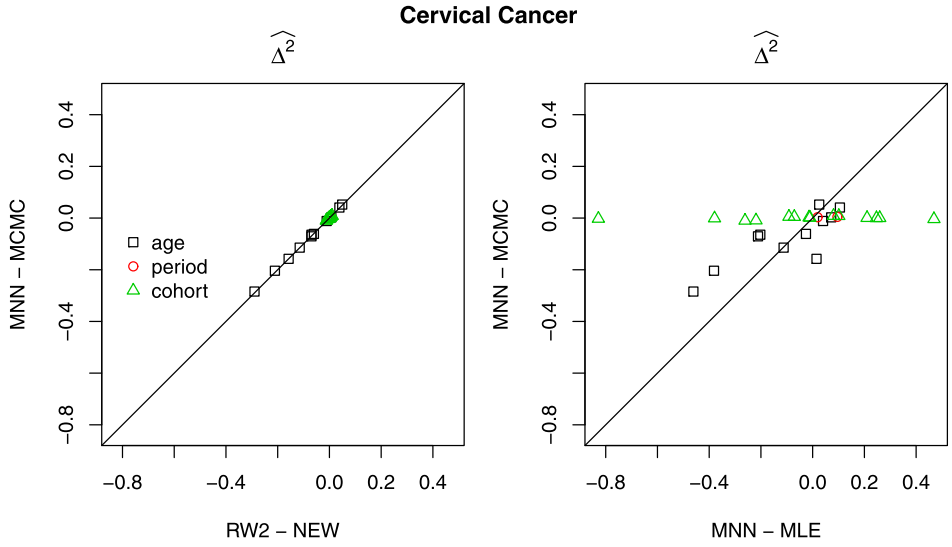

FIG. 5. *Estimates of double differences with cervical cancer data.*

(2014). The cervical cohort double differences are all close to 1, so that the underlying trend is linear in cohort. A positive linear trend could correspond to an increase in the risk of HPV exposure (but this cannot be gleaned from the data, rather from substantive knowledge). The lung cancer cohort ratios are also relatively flat.

We do not plot the period double differences, as there are only two of them. However, for cervical and lung cancer the estimates of $\exp(\Delta^2\beta_3)$ and $\exp(\Delta^2\beta_4)$ are close to one, suggesting that there is no acceleration (deviation from a linear trend) in these cancers. We emphasize that this says nothing about whether the rates of lung and cervical cancer are increasing, decreasing, or flat in time. On the other hand, for breast cancer $\Delta^2\beta_3$ and $\Delta^2\beta_4$ are away from zero, raising the possibility there is nonlinearity in the period effects. In this case, a 95% credible interval for $\Delta^2\beta_3$ under the MNN-MCMC model is $(-0.24, -0.08)$ and so significantly less than zero, indicating a deceleration in risk during the first three periods.

Table 4 shows the posterior medians of the standard deviation τ^{-1} for each set of effects for the three cancers and four models. In all cases, the variability in the age effect is larger than the variability in the period and cohort effect. This is unsurprising given the ex-

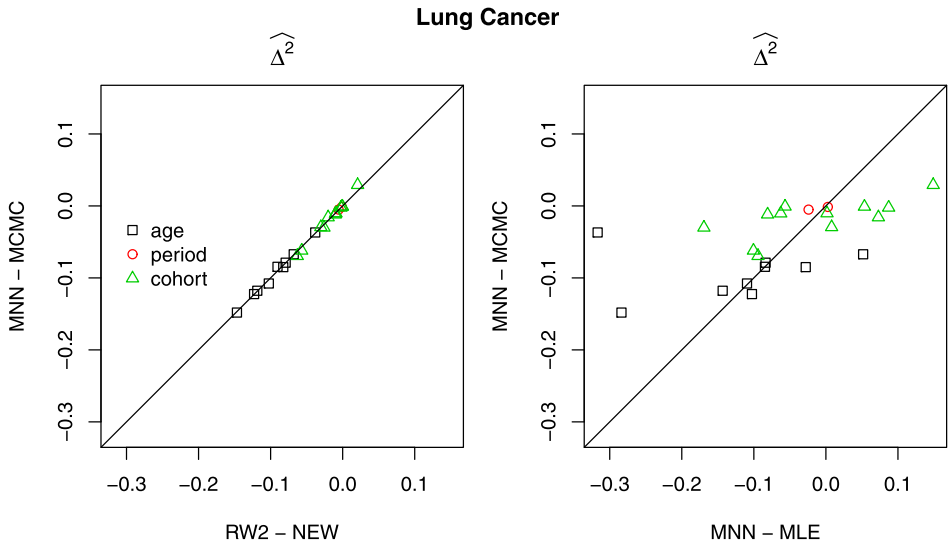

FIG. 6. *Estimates of double differences with lung cancer data.*

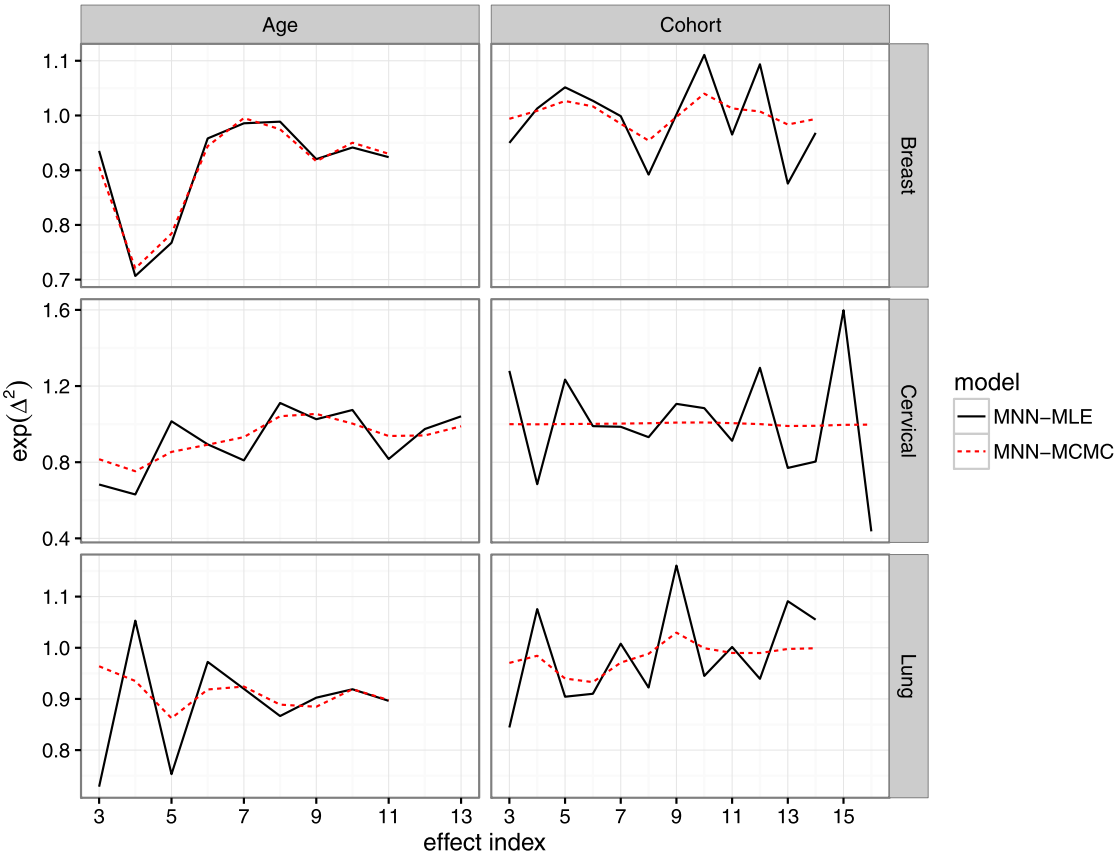

FIG. 7. *Exponentiated double differences for age and cohort effects, for breast, cervical and lung cancer.*

ploratory plots and our understanding that cancer risk differs substantially by age.

TABLE 4
Standard deviation estimates for the age, period and cohort effects under the two different RW-2 models and the MNN model fit using MCMC or `R-INLA`

Cancer	Model	τ_α^{-1}	τ_β^{-1}	τ_γ^{-1}	τ_z^{-1}
Breast	RW2-Class	0.15	0.047	0.015	0.052
	RW2-New	0.15	0.084	0.020	0.049
	MNN-INLA	0.15	0.091	0.031	
	MNN-MCMC	0.15	0.091	0.031	
Cervical	RW2-Class	0.16	0.018	0.018	0.068
	RW2-New	0.18	0.028	0.026	0.068
	MNN-INLA	0.17	0.018	0.018	
	MNN-MCMC	0.17	0.018	0.018	
Lung	RW2-Class	0.12	0.017	0.045	0.050
	RW2-New	0.13	0.024	0.051	0.050
	MNN-INLA	0.11	0.017	0.048	
	MNN-MCMC	0.12	0.017	0.048	

In these examples, there is agreement between the estimates variance components of the MNN model from INLA and MCMC. In general, we would expect close correspondence between MCMC and INLA but there are cases when INLA can be inaccurate, in particular with binary data (Fong, Rue and Wakefield, 2010, Ferkingstad and Rue, 2015).

7.2 Forecasts

Forecasting cases of disease requires predicting rates and denominators as well as taking into account the Poisson variability in the number of cases. Here, we compare the Bayesian models based on a simpler task of predicting incidence rate one period ahead. This is equivalent to predicting $E(Y)$ assuming the denominators are fixed. We use a linear extrapolation scheme, which is the implicit forecast for RW-2 priors. This forecast is automatically given in `R-INLA` if the responses for the additional time periods are included as missing data. For the MNN-MCMC model, we first convert $\boldsymbol{\xi}$ to δ, $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ by choosing any ad hoc identification scheme (e.g., $\alpha_1 = \beta_1 = \gamma_1 = \gamma_2 = 0$).

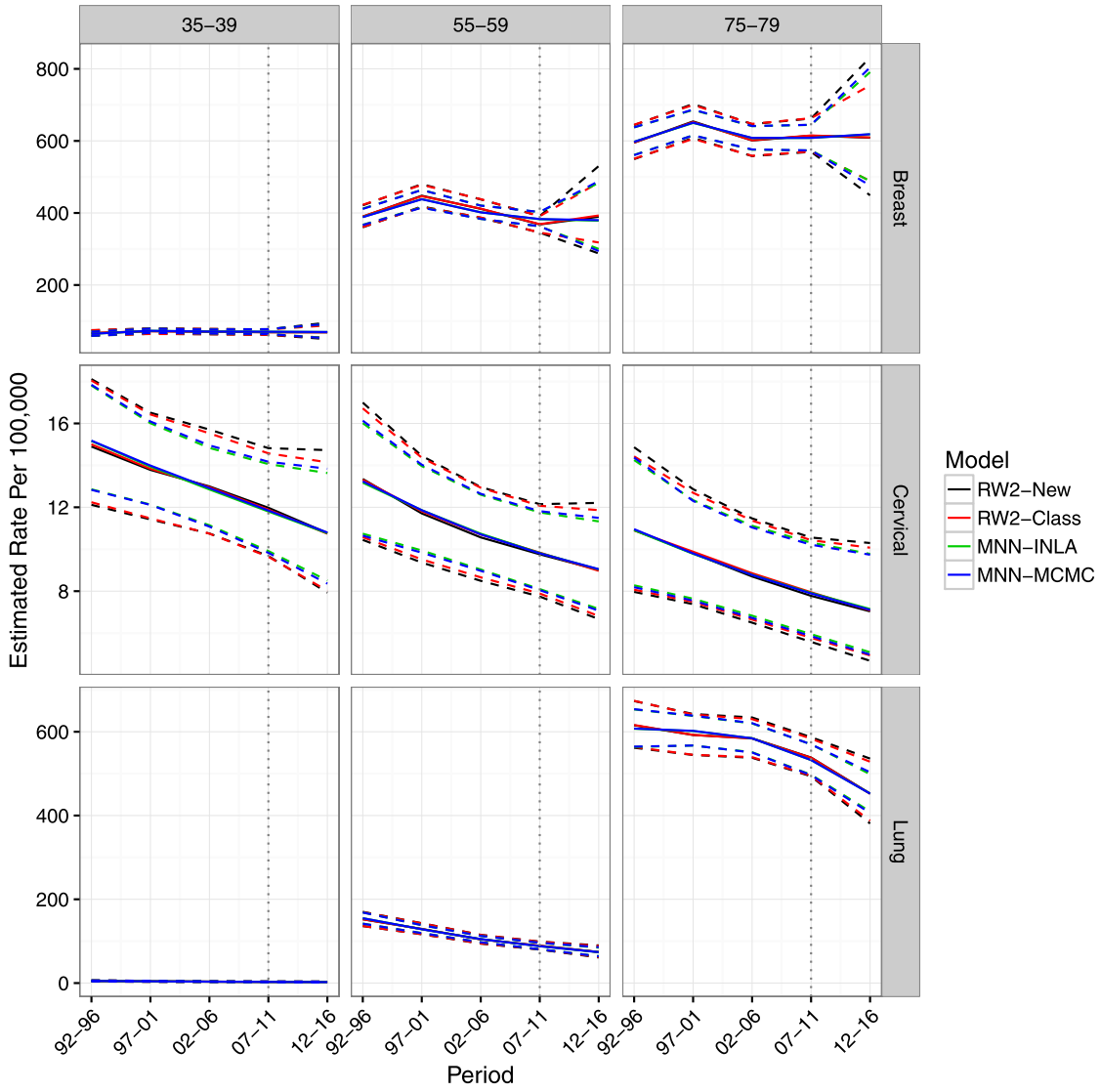

FIG. 8. *Posterior medians and point-wise* 95% *credible intervals for age-specific incidence rates per* 100,000 *for three age groups including forecasts for* 2012–2016.

We then forecast the additional period and cohort effects as

$$\beta_{T+1}|\beta_{1:T}, \tau_\beta^2 \sim \mathsf{N}(\beta_T + \Delta\beta_T, \tau_\beta^{-2}),$$

$$\gamma_{A+T}|\gamma_{1:(A+T-1)}, \tau_\gamma^2 \sim \mathsf{N}(\gamma_{A+T-1} + \Delta\gamma_{A+T-1}, \tau_\gamma^{-2}).$$

Further details are given in Supplement A, Section D.

Figure 8 shows point estimates and 95% quantile-based intervals for the incidence rates over time (including a future period) for three age groups and each cancer under the two different RW-2 models and the MNN model, fit using MCMC or INLA. The point estimates from all four models are virtually identical, but the intervals from the new RW-2 model are slightly larger, which is not surprising given that the posterior medians for τ_β^{-1} in particular were larger for all three cancers under the new RW-2 model. We predict that the age-specific rates for cervical cancer and lung cancer will continue to decrease. While the point-level predictions for the rates of breast cancer show decreases or leveling-off, the intervals for the 2012–2016 rates cover a wide range of possible trajectories. Plots for all age groups and forecasted age-specific rate curves are in Supplement A, Section A.

7.3 Prior Sensitivity

Now we turn to the issue of prior sensitivity. For the variance components, estimates of the precision of the

period effects τ_β^2 are sensitive to the choice of prior in our example (where we only have four time periods and, therefore, only two second differences in the RW-2 models). As a result, the width of the intervals for the predictions are also sensitive to the priors (see Figure E.2 in Supplement A). However, there is very little sensitivity to the priors on τ_α^2 and τ_γ^2 when keeping a fixed prior for τ_β^2 (see Figure E.3).

For data with a larger number of time periods, we expect that the posterior distributions will not be overly sensitive to the prior specification. To investigate this, we carried out the same sensitivity analysis using data on Danish lung caner incidence with ten 5-year age-groups and eleven 5-year periods; these data are available in the `Epi` package (Carstensen et al., 2014). We found that the predictions and estimates of the variance components are not sensitive to the prior specification for this example with more periods. See Supplement A, Section E for full results.

For the models with fixed effects, we chose a flat prior so that we have invariance to the choice of initial three points. If we instead use a Gaussian prior, then the estimates of the second differences do depend on the choice of initial three points (Supplement A, Section E). As the variance of the Gaussian prior increases, the estimates of the second differences become less sensitive to the choice of the three points, as we would expect.

8. DISCUSSION

In this paper, we have reviewed approaches to APC modeling and inference and have critiqued a number of recent applied contributions. We found several examples of authors over-interpreting plots of age, period and cohort effects even when they were apparently familiar with the identifiability problems. Hence, if such plots are included, the limitations of interpretation must be clearly stated. In our example, we speculated on causes of particular patterns, but emphasized that these speculations were based on biological/substantive arguments rather than on the observed data. We strongly suggest using models based on an identifiable parametrization or focusing on summaries of the fitted and forecasted rates, with the latter obtained from an invariant parameterization.

In this paper, we developed a Bayesian APC model based on one such invariant parameterization originally suggested by Kuang, Nielsen and Nielsen (2008b) and Martínez Miranda, Nielsen and Nielsen (2015). We compared this to the usual specification for a Bayesian APC model using RW-2 priors and found the methods agree in the key identifiable quantities (double differences and forecasts).

Research into APC models is not limited to the identifiability problems discussed in this review. If individual-level records are available, then one possible approach is to model each of age, period and cohort effects using the continuous time analog of the RW-2 model, which is the solution of a stochastic differential equation and has close connections with smoothing splines and integrated Wiener processes (Lindgren and Rue, 2008). In this paper, we have concentrated on the log link, but alternative *power* link functions have been used, with a power of 0.2 being found to perform well in an empirical evaluation (Møller et al., 2003). The good performance stemmed from the power model providing a leveling off of the exponential growth associated with the log link, which was found to sometimes lead to overestimation when prediction was carried out. All of the identifiability issues discussed in the paper still hold for the power model; but the constraints described here can be incorporated into these models, and implemented in INLA.

ACKNOWLEDGEMENTS

The authors would like to thank Dr. Mahesh Cheteri, of the Washington State Cancer Registry, for assistance in obtaining the data analyzed in this paper, Andrea Riebler for helpful comments on earlier drafts, and the reviewers for their valuable suggestions. The first author was funded by MRC, BBSRC, ESRC and NERC under the Environmental & Social Ecology of Human Infectious Diseases (ESEI) programme. Grant reference: G1100799/1. The second author was supported by R01CA095994 from the National Institutes of Health.

SUPPLEMENTARY MATERIAL

Supplement A: Supplementary Material (DOI: 10.1214/16-STS580SUPPA; .pdf). *Additional figures and details*: A. Additional figures, B. Design matrices, C. Code, D. Forecasting details, E. Sensitivity analysis.

Supplement B: R Code (DOI: 10.1214/16-STS580SUPPB; .zip). *APC_Models_LungDK.R*: Contains code for running the four models, comparing estimates of the double differences, forecasts, and priors using the Danish lung cancer data from Carstensen (2007). *Functions.R*: Helper functions for the MNN parameterization. *MCMC_Functions.R*: functions for running the MCMC.

REFERENCES

AHMAD, A. S., ORMISTON-SMITH, N. and SASIENI, P. D. (2015). Trends in the lifetime risk of developing cancer in Great Britain: Comparison of risk for those born from 1930 to 1960. *Br. J. Cancer* **112** 943–947.

ANANTH, C. V., KEYES, K. M., HAMILTON, A., GISSLER, M., WU, C., LIU, S., LUQUE-FERNANDEZ, M. A., SKJÆRVEN, R., WILLIAMS, M. A. and TIKKANEN, M. (2014). An international contrast of rates of placental abruption: An age–period–cohort analysis. *PLoS ONE* **10** e0125246.

BERZUINI, C. and CLAYTON, D. (1994). Bayesian analysis of survival on multiple time scales. *Stat. Med.* **13** 823–838.

BESAG, J., GREEN, P., HIGDON, D. and MENGERSEN, K. (1995). Bayesian computation and stochastic systems. *Statist. Sci.* **10** 3–66. MR1349818

BROWN, C. C. and KESSLER, L. G. (1988). Projections of lung cancer mortality in the United States: 1985–2025. *J. Natl. Cancer Inst.* **80** 43–51.

CARSTENSEN, B. (2007). Age–period–cohort models for the Lexis diagram. *Stat. Med.* **26** 3018–3045. MR2370989

CARSTENSEN, B., PLUMMER, M., LAARA, E. and HILLS, M. (2014). Epi: A package for statistical analysis in epidemiology. Available at http://CRAN.R-project.org/package=Epi, R package version 1.1.67.

CERVANTES-AMAT, M., LÓPEZ-ABENTE, G., ARAGONÉS, N., POLLÁN, M., PASTOR-BARRIUSO, R. and PÉREZ-GÓMEZ, B. (2015). The end of the decline in cervical cancer mortality in Spain: Trends across the period 1981–2012. *BMC Cancer* **15** 287.

CLAYTON, D. and SCHIFFLERS, E. (1987a). Models for temporal variation in cancer rates. I: Age–period and age–cohort models. *Stat. Med.* **6** 449–467.

CLAYTON, D. and SCHIFFLERS, E. (1987b). Models for temporal variation in cancer rates. II: Age–period–cohort models. *Stat. Med.* **6** 469–481.

ESCEDY, J. and HUNTER, D. (2008). The origin of cancer. In *Textbook of Cancer Epidemiology*, 2nd ed. (H. O. Adami, D. Hunter and D. Trichopoulos, eds.). Oxford Univ. Press, Oxford.

FERKINGSTAD, E. and RUE, H. (2015). Improving the INLA approach for approximate Bayesian inference for latent Gaussian models. *Electron. J. Stat.* **9** 2706–2731. MR3433587

FIENBERG, S. E. and MASON, W. M. (1979). Identification and estimation of age–period–cohort models in the analysis of discrete archival data. *Sociol. Method.* **10** 1–67.

FONG, Y., RUE, H. and WAKEFIELD, J. C. (2010). Bayesian inference for generalized linear mixed models. *Biostatistics* **11** 397–412.

FROST, W. H. (1939). The age selection of mortality from tuberculosis in successive decades. *American Journal of Hygiene* **30** 91–96. Reprinted in *American Journal of Epidemiology* **141** (1995) 4–9.

GHOSH, K., TIWARI, R. C., FEUER, E. J., CRONIN, K. and JEMAL, A. (2007). Predicting US cancer mortality counts using state space models. In *Computational Methods in Biomedical Research* (R. Khattree and D. N. Naik, eds.) 131–151.

GREENBERG, B. G., WRIGHT, J. J. and SHEPS, C. G. (1950). A technique for analyzing some factors affecting the incidence of syphilis. *J. Amer. Statist. Assoc.* **45** 373–399.

HEUER, C. (1997). Modeling of time trends and interactions in vital rates using restricted regression splines. *Biometrics* **53** 161–177.

HO, M.-L., HSIAO, Y.-H., SU, S.-Y., CHOU, M.-C. and LIAW, Y.-P. (2015). Mortality of breast cancer in Taiwan, 1971–2010: Temporal changes and an age–period–cohort analysis. *Journal of Obstetrics & Gynaecology* **35** 60–63.

HOLFORD, T. R. (1983). The estimation of age, period and cohort effects for vital rates. *Biometrics* **39** 311–324. MR0714415

HOLFORD, T. R. (1991). Understanding the effects of age, period, and cohort on incidence and mortality rates. *Annu. Rev. Public Health* **12** 425–457.

HOLFORD, T. R. (2006). Approaches to fitting age–period–cohort models with unequal intervals. *Stat. Med.* **25** 977–993. MR2225187

KEIDING, N. (1990). Statistical inference in the Lexis diagram. *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.* **332** 487–509. MR1084720

KEIDING, N. (2011). Age–period–cohort analysis in the 1870s: Diagrams, stereograms, and the basic differential equation. *Canad. J. Statist.* **39** 405–420. MR2842421

KNORR-HELD, L. and RAINER, E. (2001). Projections of lung cancer mortality in West Germany: A case study in Bayesian prediction. *Biostatistics* **2** 109–129.

KUANG, D., NIELSEN, B. and NIELSEN, J. P. (2008a). Forecasting with the age–period–cohort model and the extended chain-ladder model. *Biometrika* **95** 987–991. MR2461225

KUANG, D., NIELSEN, B. and NIELSEN, J. P. (2008b). Identification of the age–period–cohort model and the extended chain-ladder model. *Biometrika* **95** 979–986. MR2461224

LEE, R. D. and CARTER, L. R. (1992). Modeling and forecasting US mortality. *J. Amer. Statist. Assoc.* **87** 659–671.

LEXIS, W. H. R. A. (1875). *Einleitung in die Theorie der Bevölkerungsstatistik*. KJ Trübner.

LINDGREN, F. and RUE, H. (2008). On the second-order random walk model for irregular locations. *Scand. J. Stat.* **35** 691–700. MR2468870

LOUIE, K. S., MEHANNA, H. and SASIENI, P. (2015). Trends in head and neck cancers in England from 1995 to 2011 and projections up to 2025. *Oral Oncology* **51** 341–348.

LUNN, D. J., THOMAS, A., BEST, N. and SPIEGELHALTER, D. (2000). WinBUGS—a Bayesian modelling framework: Concepts, structure, and extensibility. *Stat. Comput.* **10** 325–337.

LUO, L. (2013). Assessing validity and application scope of the intrinsic estimator approach to the age–period–cohort problem. *Demography* **50** 1945–1967.

MARTÍNEZ MIRANDA, M. D., NIELSEN, B. and NIELSEN, J. P. (2015). Inference and forecasting in the age–period–cohort model with unknown exposure with an application to mesothelioma mortality. *J. Roy. Statist. Soc. Ser. A* **178** 29–55. MR3291760

MASON, K. O., MASON, W. M., WINSBOROUGH, H. H. and POOLE, W. K. (1973). Some methodological issues in cohort analysis of archival data. *Am. Sociol. Rev.* **38** 242–258.

MØLLER, B., FEKJÆR, H., HAKULINEN, T., SIGVALDASON, H., STORM, H. H., TALBÄCK, M. and HALDORSEN, T. (2003). Prediction of cancer incidence in the Nordic countries: Empirical comparison of different approaches. *Stat. Med.* **22** 2751–2766.

NIELSEN, B. (2014). apc: A package for age–period–cohort analysis. Available at http://CRAN.R-project.org/package=apc, R package version 1.0.

NIELSEN, B. and NIELSEN, J. P. (2014). Identification and forecasting in mortality models. *Sci. World J.* **2014** 347043.

O'BRIEN, R. M. (2000). Age period cohort characteristic models. *Soc. Sci. Res.* **29** 123–139.

O'BRIEN, R. (2014). *Age–Period–Cohort Models: Approaches and Analyses with Aggregate Data*. Chapman & Hall/CRC Press, London.

OSMOND, C. and GARDNER, M. J. (1982). Age, period and cohort models applied to cancer mortality rates. *Stat. Med.* **1** 245–259.

PAPOILA, A. L., RIEBLER, A., AMARAL-TURKMAN, A., SÃO-JOÃO, R., RIBEIRO, C., GERALDES, C. and MIRANDA, A. (2014). Stomach cancer incidence in Southern Portugal 1998–2006: A spatio-temporal analysis. *Biom. J.* **56** 403–415. MR3256497

RENSHAW, A. E. and HABERMAN, S. (2006). A cohort-based extension to the Lee–Carter model for mortality reduction factors. *Insurance Math. Econom.* **38** 556–570.

RIEBLER, A. and HELD, L. (2010). The analysis of heterogeneous time trends in multivariate age–period–cohort models. *Biostatistics* **11** 57–69.

RIEBLER, A. and HELD, L. (2016). Projecting the future burden of cancer: Bayesian age–period–cohort analysis with integrated nested Laplace approximations. *Biom. J.* To appear.

RIEBLER, A., HELD, L. and RUE, H. (2012). Estimation and extrapolation of time trends in registry data—borrowing strength from related populations. *Ann. Appl. Stat.* **6** 304–333. MR2951539

RIEBLER, A., HELD, L., RUE, H. and BOPP, M. (2012). Gender-specific differences and the impact of family integration on time trends in age-stratified Swiss suicide rates. *J. Roy. Statist. Soc. Ser. A* **175** 473–490. MR2905049

ROBERTS, G. O. and TWEEDIE, R. L. (1996). Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli* **2** 341–363. MR1440273

ROBERTSON, C. and BOYLE, P. (1986). Age, period and cohort models: The use of individual records. *Stat. Med.* **5** 527–538.

RODGERS, W. L. (1982). Estimable functions of age, period, and cohort effects. *Am. Sociol. Rev.* **47** 774–787.

ROSENBERG, P. S. and ANDERSON, W. F. (2011). Age–period–cohort models in cancer surveillance research: Ready for prime time? *Cancer Epidemiol. Biomark. Prev.* **20** 1263–1268.

ROSENBERG, P. S., CHECK, D. P. and ANDERSON, W. F. (2014). A web tool for age–period–cohort analysis of cancer incidence and mortality rates. *Cancer Epidemiol. Biomark. Prev.* **23** 2296–2302.

RUE, H. and HELD, L. (2005). *Gaussian Markov Random Fields: Theory and Applications. Monographs on Statistics and Applied Probability* **104**. Chapman & Hall/CRC, Boca Raton, FL. MR2130347

RUE, H., MARTINO, S. and CHOPIN, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **71** 319–392. MR2649602

RUTHERFORD, M. J., LAMBERT, P. C. and THOMPSON, J. R. (2010). Age–period–cohort modeling. *Stata J.* **10** 606–627.

SASIENI, P. D. (2012). Age–period–cohort models in Stata. *Stata J.* **12** 45–60.

SCHMID, V. J. and HELD, L. (2007). Bayesian age–period–cohort modeling and prediction—BAMP. *J. Stat. Softw.* **21** 1–15.

SEOANE-MATO, D., ARAGONÉS, N., FERRERAS, E., GARCÍA-PÉREZ, J., CERVANTES-AMAT, M., FERNÁNDEZ-NAVARRO, P., PASTOR-BARRIUSO, R. and LÓPEZ-ABENTE, G. (2014). Trends in oral cavity, pharyngeal, oesophageal and gastric cancer mortality rates in Spain, 1952–2006: An age–period–cohort analysis. *BMC Cancer* **14** 254.

SMITH, T. R. and WAKEFIELD, J. (2016). Supplement to "A review and comparison of age–period–cohort models for cancer incidence." DOI:10.1214/16-STS580SUPPA, DOI:10.1214/16-STS580SUPPB.

SPRINGETT, V. H. (1950). A comparative study of tuberculosis mortality rates. *J. Hyg.* **48** 361–395.

TZENG, I. S. and LEE, W. C. (2015). Forecasting hepatocellular carcinoma mortality in Taiwan using an age–period–cohort model. *Asia-Pacific Journal of Public Health* **27** NP65–NP73.

VACCARELLA, S., FRANCESCHI, S., ENGHOLM, G., LÖNNBERG, S., KHAN, S. and BRAY, F. (2014). 50 years of screening in the Nordic countries: Quantifying the effects on cervical cancer incidence. *Br. J. Cancer* **111** 965–969.

VALERY, P. C., LAVERSANNE, M. and BRAY, F. (2015). Bone cancer incidence by morphological subtype: A global assessment. *Cancer Causes Control* **26** 1127–1139.

YANG, Y., FU, W. J. and LAND, K. C. (2004). A methodological comparison of age–period–cohort models: The intrinsic estimator and conventional generalized linear models. *Sociol. Method.* **34** 75–110.

YANG, Y. and LAND, K. C. (2013). *Age–Period–Cohort Analysis: New Models, Methods, and Empirical Applications*. CRC Press, Boca Raton, FL. MR3026791

ZHENG, R., PENG, X., ZENG, H., ZHANG, S., CHEN, T., WANG, H. and CHEN, W. (2015). Incidence, mortality and survival of childhood cancer in China during 2000–2010 period: A population-based study. *Cancer Letters* **363** 176–180.

ZHU, L., PICKLE, L. W., GHOSH, K., NAISHADHAM, D., PORTIER, K., CHEN, H., KIM, H., ZOU, J., CUCINELLI, Z., KOHLER, B., EDWARDS, B. K., KING, J., FEUER, E. J. and JEMAL, A. (2012). Predicting US- and state-level cancer counts for the current calendar year. *Cancer* **118** 1100–1109.

ZHU, L., PICKLE, L. W., ZOU, Z. and CUCINELLI, J. (2014). Trends and patterns of childhood cancer incidence in the United States, 1995–2010. *Stat. Interface* **7** 121–134. MR3197576