2021 SISCER: Age-Period-Cohort Modeling and Analysis
Lecture 4: Bayesian Methods

Jon Wakefield

Departments of Statistics and Biostatistics
University of Washington

Outline

Bayesian Age-Period-Cohort Models

Bayesian Modeling

A number of authors, beginning with Berzuini *et al.* (1993), have carried out Bayesian analyses of APC data, using random walk (RW) models.

These are popular and are often referred to as nonparametric smoothing models.

A number of authors have considered various forms of these models including Berzuini and Clayton (1994), Besag *et al.* (1995), Knorr-Held and Rainer (2001), Schmid and Held (2004), Schmid and Held (2007), Riebler and Held (2010), Riebler *et al.* (2012a), Riebler *et al.* (2012b), Smith and Wakefield (2016), Riebler and Held (2017).

We now discuss RW models in an APC context.

Random Walk Models for APC Data

One could just take the prior on the factor levels as independent normals (say), but this is not leveraging the smoothness over the time scales we expect to see.

It makes sense to use RW2 priors on the three time factors for smoothness and since (as discussed above) the second differences (stated for age here)

$$\Delta^2 \alpha_a = \alpha_a - 2\alpha_{a-1} + \alpha_{a-2},$$

are identifiable (i.e., estimable).

Bayesian modeling

The Riebler-Held model is:

$$
\begin{aligned}
y_{ap}|\lambda_{ap} &\sim \text{Poisson}(N_{ap}\lambda_{ap}) \\
\log \lambda_{ap} &= \delta + \alpha_a + \beta_p + \gamma_{A-a+p} + z_{ap}, \\
\delta &\sim \text{N}(m_\delta, s_\delta^2), \\
\boldsymbol{\alpha}|\tau_\alpha^2 \sim \text{RW2}(\tau_\alpha^{-2}), &\quad \tau_\alpha^2 \sim \text{Ga}(a_1, b_1), \\
\boldsymbol{\beta}|\tau_\beta^2 \sim \text{RW2}(\tau_\beta^{-2}), &\quad \tau_\beta^2 \sim \text{Ga}(a_1, b_1), \\
\boldsymbol{\gamma}|\tau_\gamma^2 \sim \text{RW2}(\tau_\gamma^{-2}), &\quad \tau_\gamma^2 \sim \text{Ga}(a_1, b_1), \\
z_{ap}|\tau_z^2 \sim_{iid} \text{N}(0, \tau_z^{-2} I), &\quad \tau_z^2 \sim \text{Ga}(a_2, b_2).
\end{aligned}
$$

The additional unstructured random effect z_{ap} allows for excess-Poisson variation (overdispersion) around the constrained temporal effects.

Bayesian modeling

Previous authors using this formulation have imposed sum-to-zero constraints on the age, period, and cohort effects, but not zero linear trends (Riebler *et al.*, 2012a).

This makes sense because there are no linear terms in the model (whereas there is an intercept).

The posterior distributions of Bayesian APC models can be approximated with MCMC using the stand-alone package `BAMP` (Schmid and Held, 2007) and with integrated nested Laplace approximations (Rue *et al.*, 2009) using the `BAPC` package (Riebler and Held, 2015) – the latter is preferred (much faster).

We fit the RW2 prior Bayesian model to the Danish lung cancer data using the `BAPC` package.

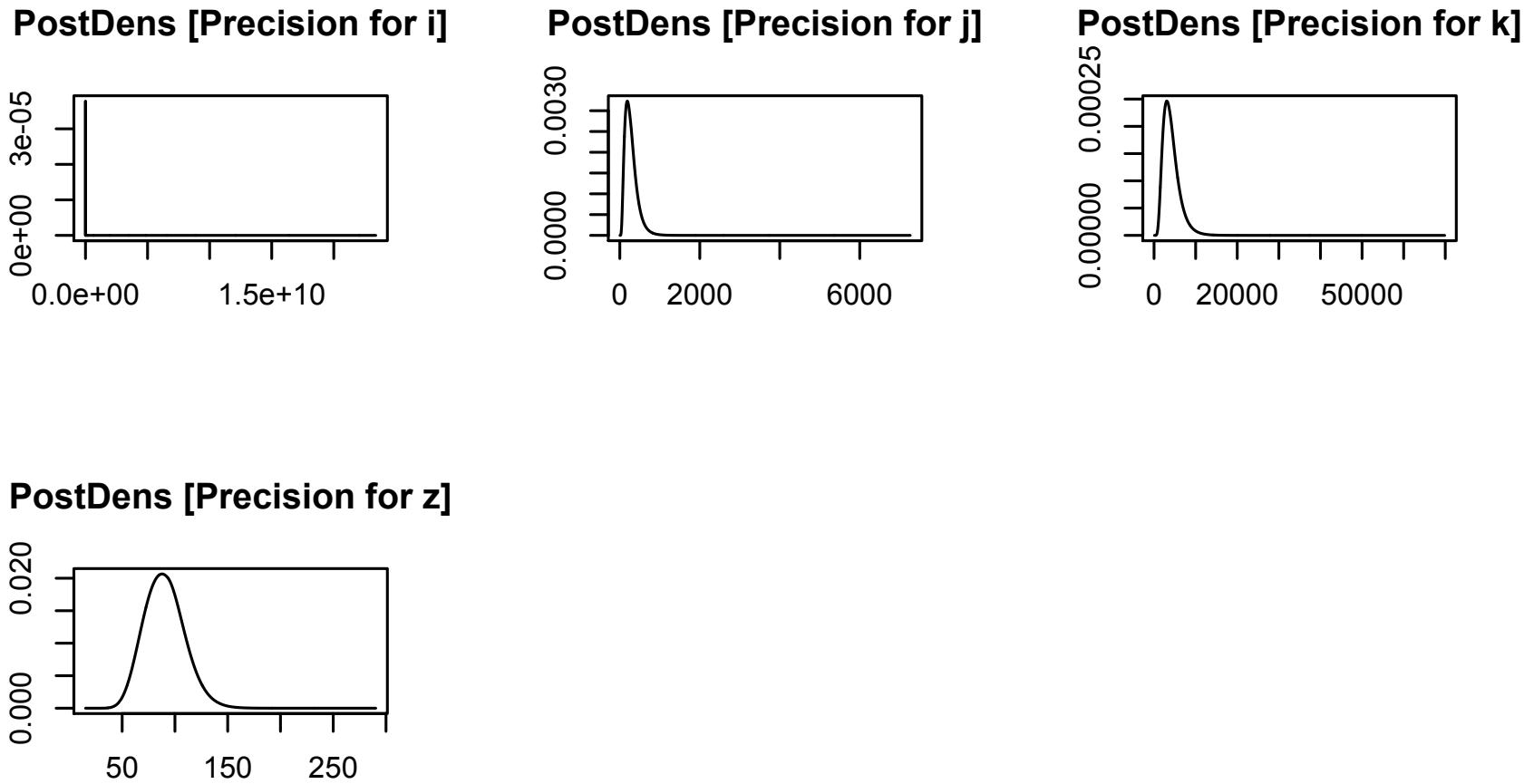

Figure 1: Posterior distributions for the precision parameters, τ_α^2, τ_β^2, τ_γ^2, τ_z^2, from the `BAPC` package.

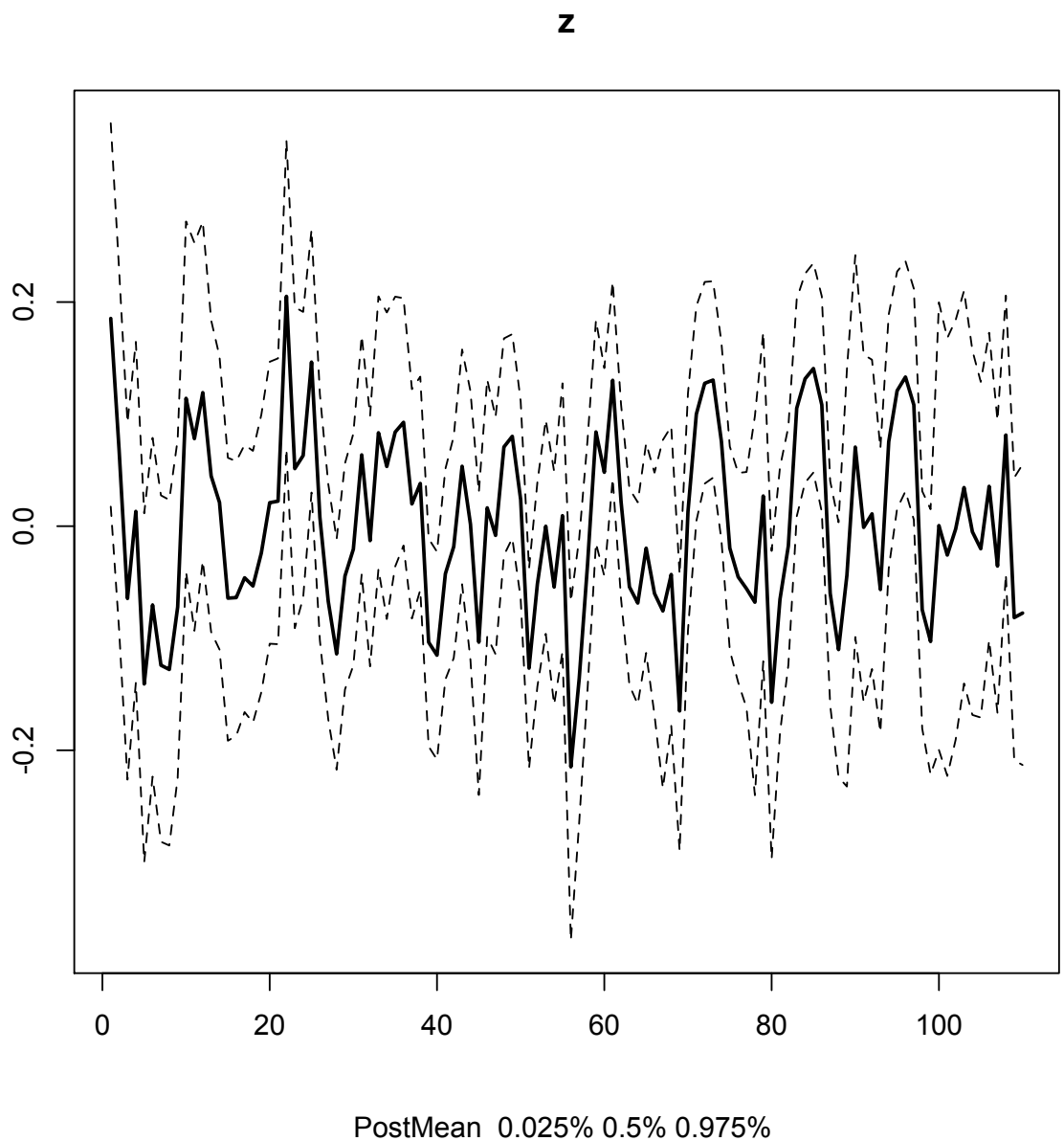

Figure 2: Posterior distributions for the overdispersion parameters z_{ap}, from the `BAPC` package.

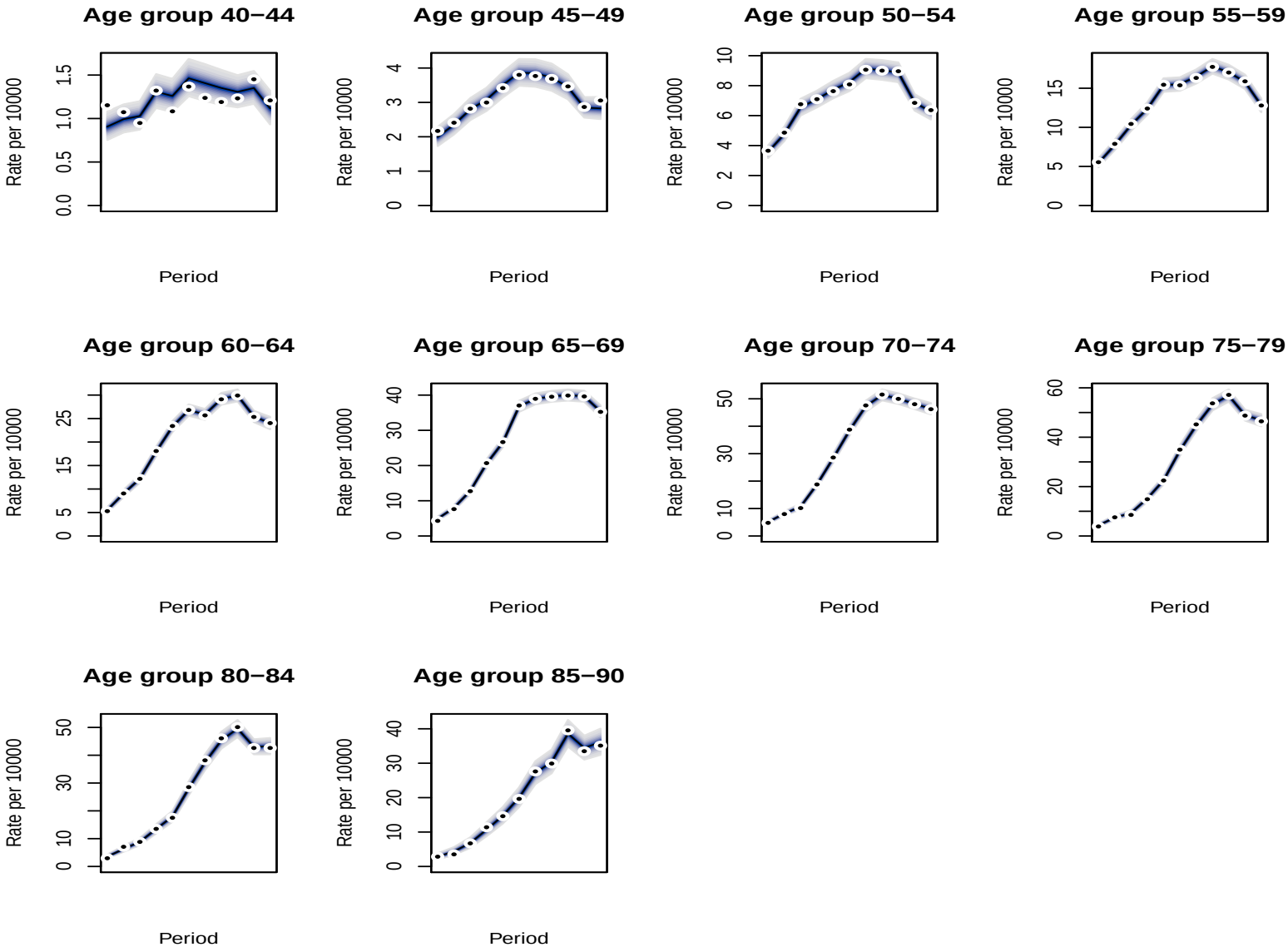

Figure 3: Observed and fitted, with interval estimates, from the `BAPC` package.

A Bayesian Version of the MNN Parameterization

Smith and Wakefield (2016) develop a Bayesian version of the MMNN parameterization.

They specify priors directly on $\boldsymbol{\theta}$:

$$\begin{aligned}
\pi(\eta_{i_1 j_1}, \eta_{i_2 j_2}, \eta_{i_3 j_3}) &\propto 1 \\
\Delta^2\alpha_3, \ldots, \Delta^2\alpha_A | \tau_\alpha^2 &\sim \mathrm{N}(0, \tau_\alpha^{-2}\mathbf{I}) \\
\Delta^2\beta_3, \ldots, \Delta^2\beta_P | \tau_\beta^2 &\sim \mathrm{N}(0, \tau_\beta^{-2}\mathbf{I}) \\
\Delta^2\gamma_3, \ldots, \Delta^2\gamma_{A+P-1} | \tau_\gamma^2 &\sim \mathrm{N}(0, \tau_\gamma^{-2}\mathbf{I})
\end{aligned}$$

with gamma priors on the precisions.

The improper flat prior on the three initial points is because we want the analysis to be invariant to which set of points we select.

Forecasting

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.

We require, for the a-th age group,

$$\eta_{a,P+h} = \delta + \alpha_a + \beta_{P+h} + \gamma_{A-a+P+h}.$$

In general, the forecasts depend on projecting the period and cohort effects ahead by h steps, based on period and cohort effects estimated from the observed data.

Forecasting

That is, for some functions f_β and f_γ, we require,

$$\beta_{P+h} = f_\beta(\boldsymbol{\beta}_{1:P})$$

and

$$\gamma_{A-a+P+h} = f_\gamma(\boldsymbol{\gamma}_{1:A-a+P}).$$

Table 1 shows what is required to predict 1, 2, 3 periods ahead for data with $A = 5$ age groups and $P = 5$ periods.

For example, for a $h = 2$ prediction for age group $a = 2$ at $p = 7$ we need to estimate $\beta_{P+2} = \beta_7$ and $\gamma_{A-a+P+2} = \gamma_{10}$.

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

Table 1: Illustration of projections required for forecasts at $p = 6, 7, 8$.

Forecasting

As written, $\eta_{a,P+h}$ is a function of non-identifiable effects, and so the forecasting functions f_β and f_γ must be carefully chosen so that

$$\eta(g(\alpha_a, \beta_{P+h}, \gamma_{A-a+P+h}, \delta)) = \eta(\alpha_a, \beta_{P+h}, \gamma_{A-a+P+h}, \delta).$$

Two common functions are constant forecasts,

$$f_\beta(\beta_{1:P}) = \beta_P$$

and linear extrapolation,

$$f_\beta(\beta_{1:P}) = \beta_P + h\Delta\beta_P.$$

Forecasting

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 *et al.* (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

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

for some function f.

If $f(\cdot) = 0$, we recover linear extrapolation, but constant forecasts (i.e., $f_\beta(\beta_{1:P}) = \beta_P$) cannot fit into this form (and hence are not invariant) — don't use RW1 models.

We show forecasts from the `BAPC` package.

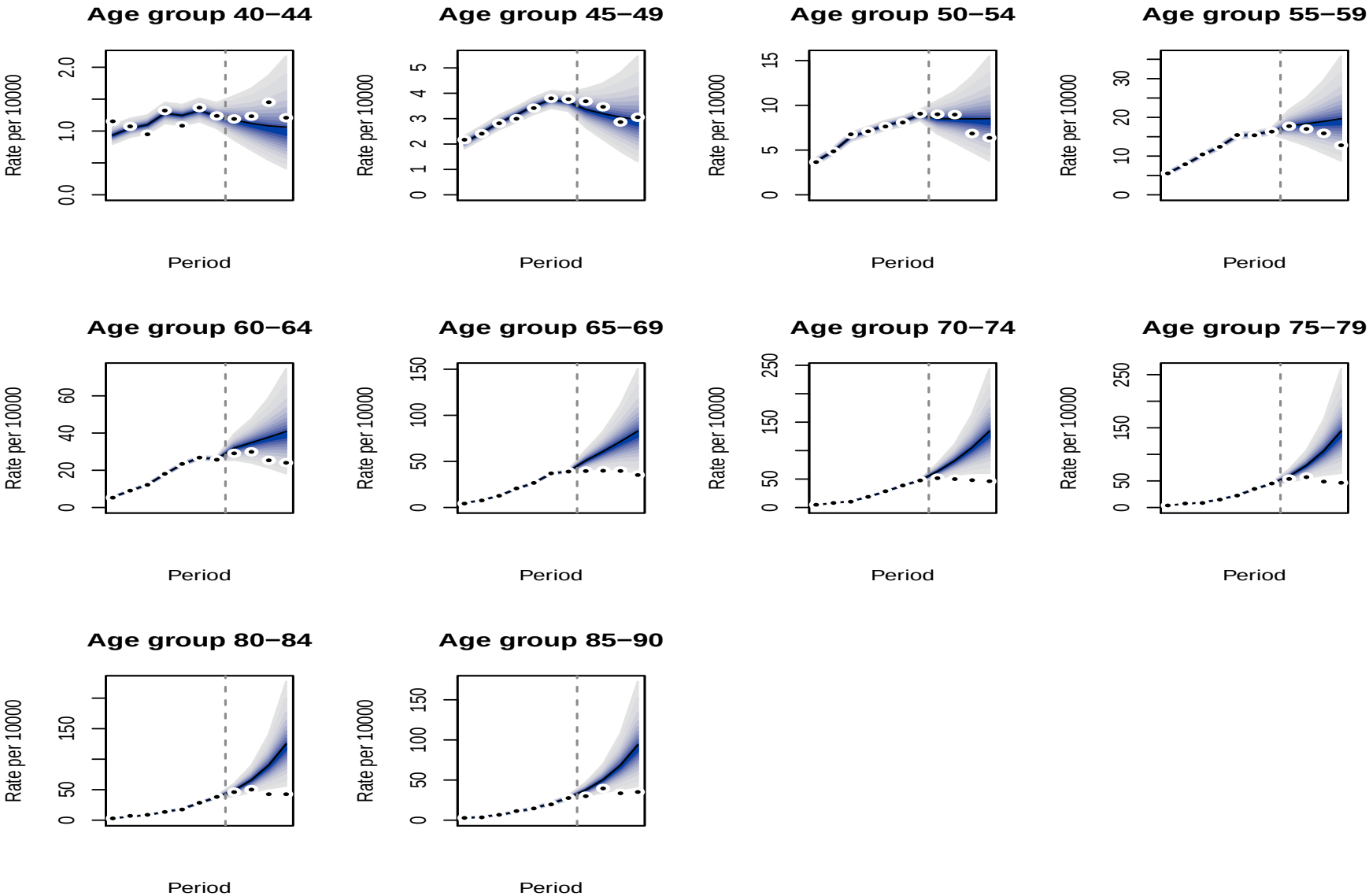

Figure 4: Observed, fitted and predictions, and then forecasts for the last 4 periods, with the model fitting based on all data but these last 4 periods.

The Lee-Carter Model

Lee-Carter Model

The Lee-Carter (LC) model (Lee and Carter, 1992) is a popular forecasting model.

The basic Poisson LC model is:

$$
\begin{aligned}
Y_{ap}|\lambda_{ap} &\sim \text{Poisson}(N_{ap}\lambda_{ap}) \\
\log \lambda_{ap} &= \alpha_a + \underbrace{\beta_a \kappa_p}_{\text{age-period interaction}} + \epsilon_{ap} \\
\epsilon_{ap} &\sim_{iid} \text{N}(0, \sigma_\epsilon^2)
\end{aligned}
$$

In this model:

- α_a is an overall age-profile, average over the study period,
- a period-specific effect κ_p,
- β_a are adjustments to the period pattern for different age groups,
- constraints are required for identifiability – typically $\sum_a \beta_a = 1$ and $\sum_p \kappa_p = 0$,
- $\epsilon_{a,p}$ allow for overdispersion.

Lee-Carter Model

The basic idea is to take a time-invariant pattern (κ_p) and perturb in am age-specific fashion.

To forecast mortality rates, a random walk has been proposed:

$$\kappa_p = \phi_p + \kappa_{p-1} + \upsilon_p$$

The Poisson extension of the original LC model is due to Brouhns *et al.* (2002), with a Bayesian implementation (using MCMC) by Czado *et al.* (2005).

Lee-Carter Model Extensions

Renshaw and Haberman (2006) proposed a cohort extension to allow for both age-specific period trends and age-specific cohort trends – see the `ilc` package.

Lee-Carter models can be fitted in the `Epi` package via the `LCa.fit` function.

So far as Bayesian versions are concerned, Wiśniowski *et al.* (2015) propose a very general version of the LC model by forecasting the age patterns of fertility, mortality, immigration, and emigration within a cohort projection model. For implementation, MCMC is used within OpenBUGS.

It is not currently possible to implement LC models (or extensions) in INLA, because the product term of unknown parameters (known as a bilinear model[1]) does not satisfy the requirement a latent Gaussian model.

[1] A bilinear interaction is where the slope of a regression line for Y and X changes as a linear function of a third variable, Z

Concluding Remarks

APC Models: Concluding Remarks

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 (1987, p. 477).

The intrinsic estimator approach from demography uses the null space of $\boldsymbol{x}$ to define the linear constraints (Yang *et al.*, 2004); this approach is described in detail in a book-length treatment (Yang and Land, 2013).

But Luo (2013) argues that this is no more scientifically justified than earlier approaches.

APC Models: Concluding Remarks

One may fix aspects of the curves, based on the context, in order to examine period and cohort effects, but one must be aware that this choice is an assumption and cannot be confirmed/refuted by the data.

For example, when studying lung cancer one may assume that cohort effects are strong, since they reflect smoking behavior amongst different groups.

In contrast, when studying breast cancer, period effects may reflect the introduction of screening across all age groups.

APC Models: Concluding Remarks

The Bayesian framework for APC models has been extended to a spatial context, see Riebler and Held (2010); Riebler *et al.* (2012a).

Work needed on unequally-spaced groupings, e.g., age in 5 year intervals, period in 1 years.

In my own research, I'm looking to analyze child mortality using APC models. Issues include: modeling on unequal time scales; how to combine with spatial smoothing models; how to account for survey design.

Final Summary: While we can obtain estimates by age and period (say), after fitting a sequence of models, setting the constituent levels or apportioning the trend to period or cohort effects is not possible without uncheckable assumptions. Hence, it is best to just examine the fitted values, using a "good" parameterization such as that suggested by Kuang *et al.* (2008b). A lot is made of the second order terms that are identifiable, but these are very difficult to interpret.

Appendix: Forecasting the MMNN way

Meseothelioma Example

Martínez Miranda *et al.* (2015) discuss forecasting of mesothelioma mortality.

Meseothelioma is a lung cancer that is associated with exposure to asbestos.

They focus on mesothelioma deaths of males in the age range to 25–89 years owing to sparsity in the more extreme age groups.

Thus, the data are an age-period array with $A = 65$ age levels and $P = 41$ periods. The total number of deaths is 31,902, with the annual observed number of deaths peaking at 1774 in 2007.

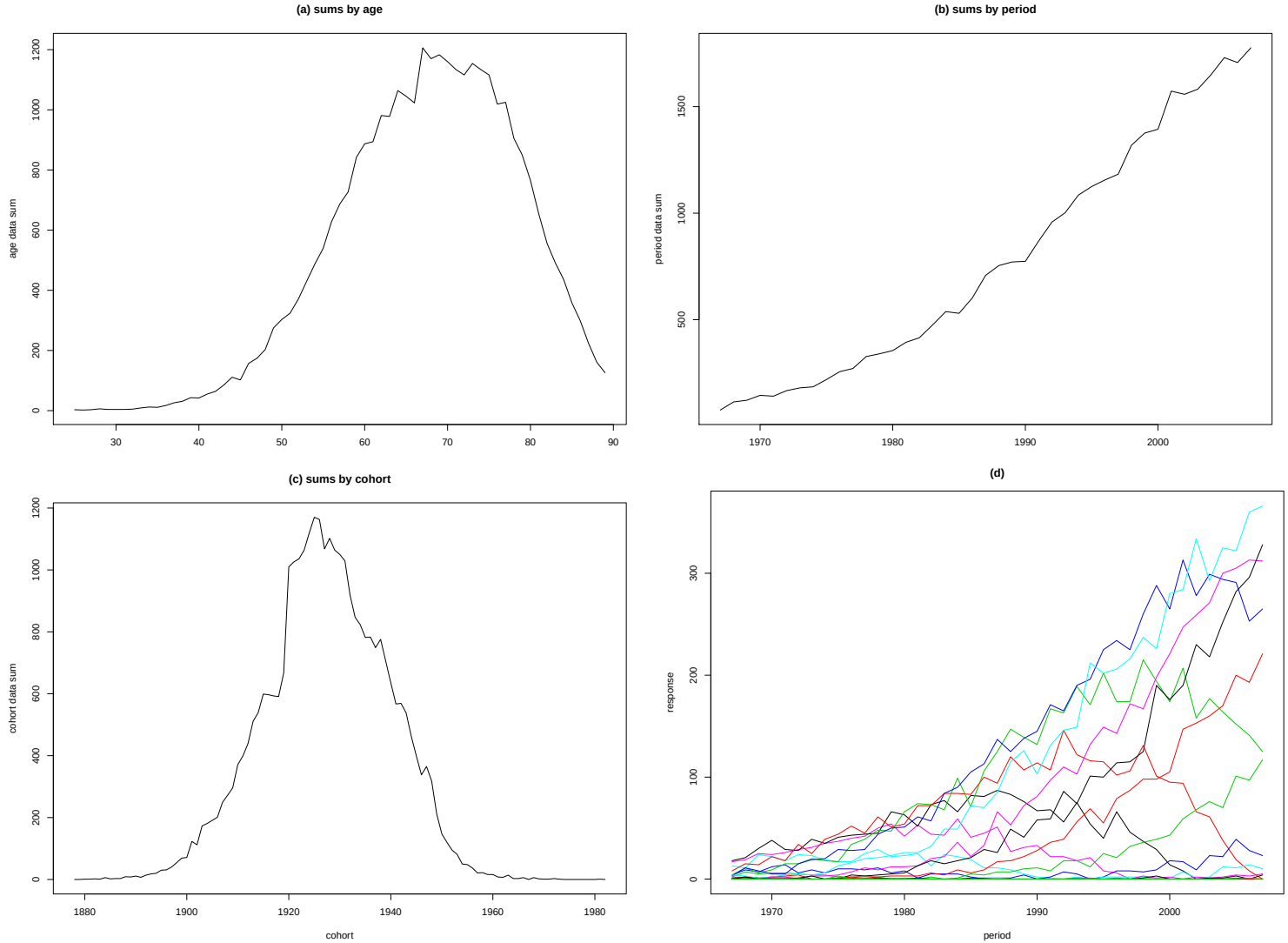

Figure 5: Observed deaths by (a) period, (b) age, (c) cohort, and (d) log-cumulative deaths by 5-year age and cohort group: viewing the curves from top left to bottom right they represent the cohorts 1923-1927, 1928-1932, etc.

Meseothelioma Example

The APC model is adequate (when compared to the saturated model).

Also some evidence to that the AC model may be reasonable when compared to APC model.

From Martínez Miranda *et al.* (2015), “The decision is therefore marginal so the data are not sufficiently informative to tell whether a period effect is needed or not. Thus, from an inferential viewpoint we cannot draw strong conclusions about the period effect. However, from a forecasting viewpoint parsimony is often useful so the period effect will be dropped”.

Meseothelioma Example

Figure 6 shows forecasts based on different cohorts; the dots indicate the observed counts of mesothelioma deaths by period.

The top curve represents forecasts of the total number of deaths among those cohorts in which the men were born in 1966 and before.

The next curve includes cohorts until 1952 and the bottom curve cohorts until 1937.

Figure 7 shows forecasts based based on data up to the given (period) date – this was done to compare with forecasts carried out in the literature, at particular times (calendar years).

Meseothelioma Example

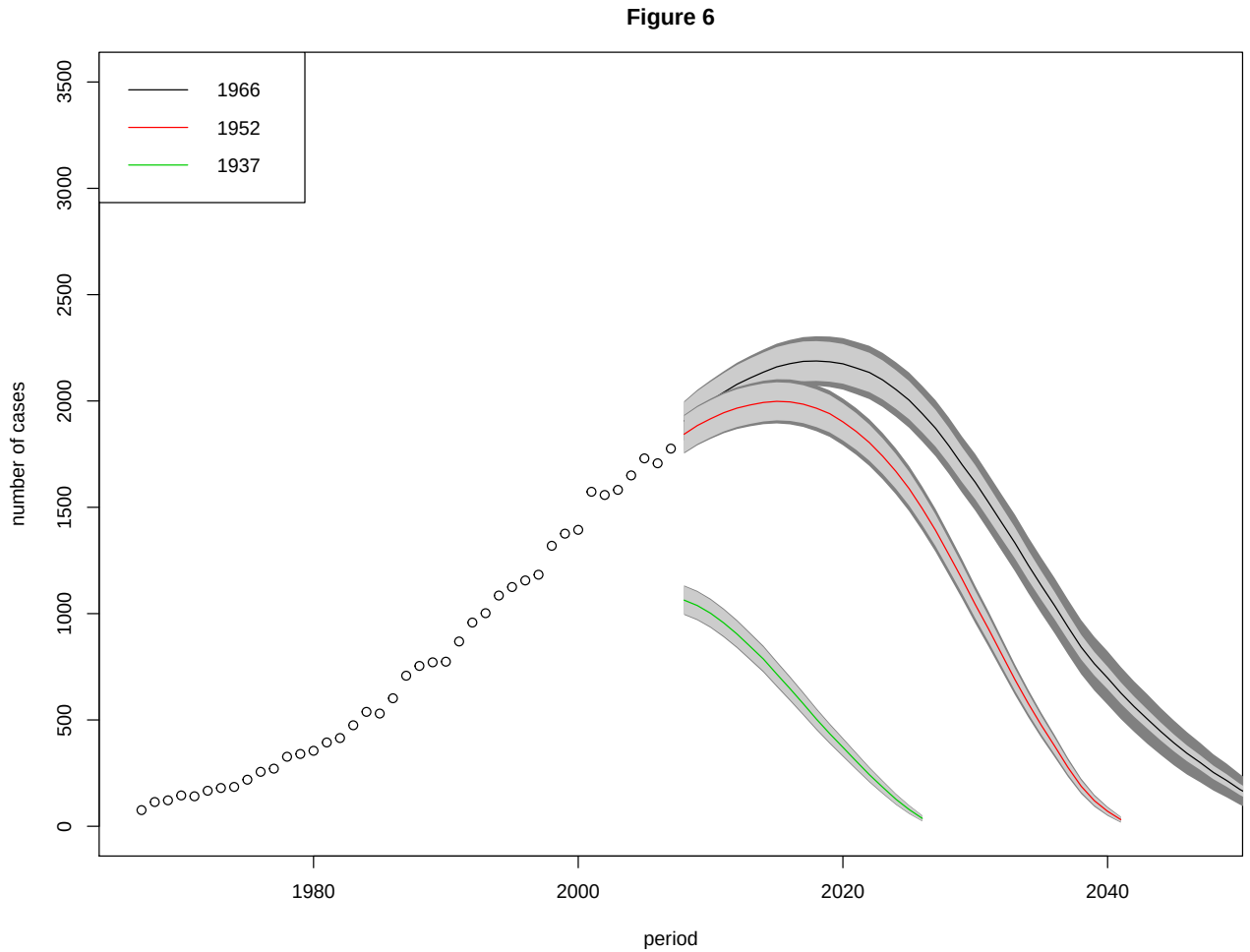

Figure 6: Forecasts of annual numbers of deaths based on the full sample and decomposed by cohort contribution (age-cohort model).

Meseothelioma Example

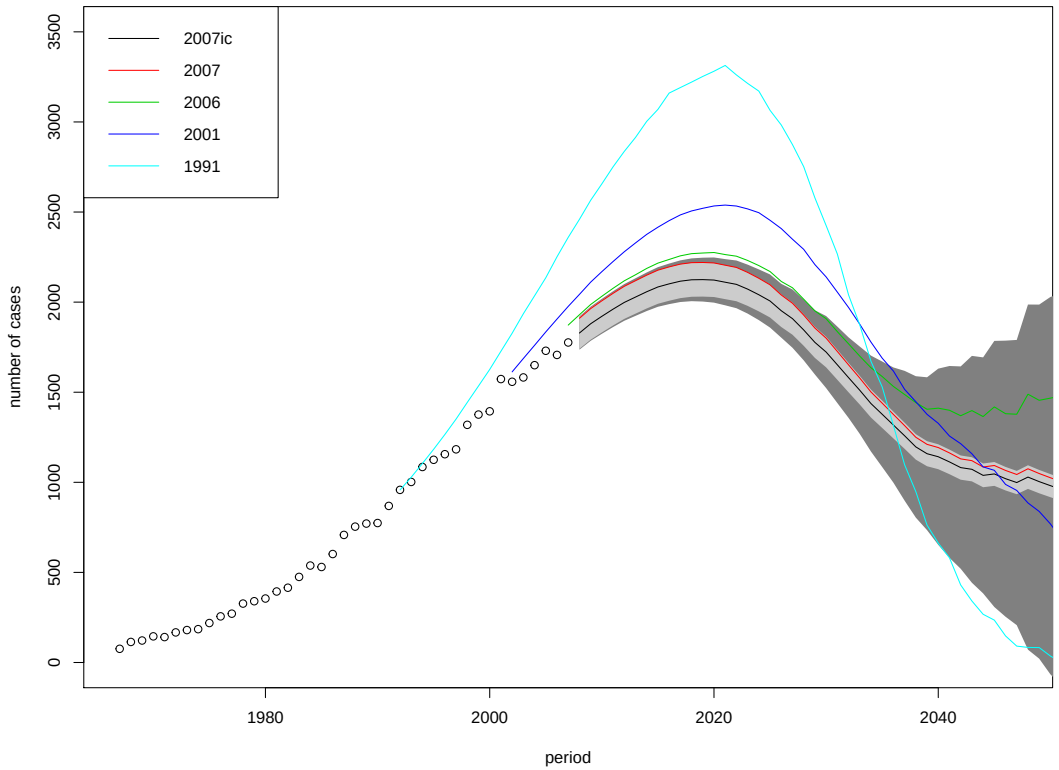

Figure 7: Recursive forecasts and forecasts of annual number of deaths (age-cohort model).

References

Berzuini, C. and Clayton, D. (1994). Bayesian analysis of survival on multiple time scales. *Statistics in Medicine*, **13**, 823–838.

Berzuini, C., Clayton, D., and Bernardinelli, L. (1993). Bayesian inference on the lexis diagram. *Bulletin of the Interntional Statistical Institute*, **50**, 149–164.

Besag, J., Green, P., Higdon, D., and Mengersen, K. (1995). Bayesian computation and stochastic systems (with discussion). *Statistical Science*, **10**, 3–66.

Brouhns, N., Denuit, M., and Vermunt, J. K. (2002). A poisson log-bilinear regression approach to the construction of projected lifetables. *Insurance: Mathematics and Economics*, **31**, 373–393.

Clayton, D. and Schifflers, E. (1987). Models for temporal variation in cancer rates. II: age–period–cohort models. *Statistics in medicine*, **6**, 469–481.

Czado, C., Delwarde, A., and Denuit, M. (2005). Bayesian poisson log-bilinear mortality projections. *Insurance: Mathematics and Economics*, **36**(3), 260–284.

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. (2008a). Forecasting with the age-period-cohort model and the extended chain-ladder model. *Biometrika*, **95**, 987–991.

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

Lee, R. and Carter, L. (1992). Modeling and forecasting US mortality. *Journal of the American Statistical Association*, **87**, 659–671.

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., Nielsen, B., and Nielsen, J. (2015). Inference and forecasting in the age–period–cohort model with unknown exposure with an application to mesothelioma mortality. *Journal of the Royal Statistical Society: Series A*, **278**, 29–55.

Renshaw, A. and Haberman, S. (2006). A cohort-based extension to the Lee-Carter model for mortality reduction factors. *Insurance: Mathematics and Economics*, **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. (2015). Projecting the future burden of cancer - Bayesian age-period-cohort analysis for routine use. Technical report, Norwegian University of Science and Technology.

Riebler, A. and Held, L. (2017). Projecting the future burden of cancer: Bayesian age–period–cohort analysis with integrated nested Laplace approximations. *Biometrical Journal*, **59**, 531–549.

Riebler, A., Held, L., and Rue, H. (2012a). Estimation and extrapolation of time trends in registry data – borrowing strength from related populations. *Annals of Applied Statistics*, **6**, 304–333.

Riebler, A., Held, L., Rue, H., and Bopp, M. (2012b). Gender-specific differences and the impact of family integration on time trends in age-stratified Swiss suicide rates. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, **175**, 473–490.

Robertson, C. and Boyle, P. (1986). Age, period and cohort models: The use of individual records. *Statistics in Medicine*, **5**, 527–538.

Rue, H. and Held, L. (2005). *Gaussian Markov Random Fields*, volume 104 of *Monographs on Statistics and Applied Probability*. Chapman & Hall/CRC, Boca Raton, FL.

Rue, H., Martino, S., and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested

Laplace approximations. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **71**, 319–392.

Schmid, V. and Held, L. (2004). Bayesian extrapolation of space–time trends in cancer registry data. *Biometrics*, **60**, 1034–1042.

Schmid, V. and Held, L. (2007). Bayesian age-period-cohort modeling and prediction-BAMP. *Journal of Statistical Software*, **21**, 1–15.

Smith, T. and Wakefield, J. (2016). A review and comparison of age-period-cohort models for cancer incidence. *Statistical Science*, **32**, 165–175.

Wiśniowski, A., Smith, P. W., Bijak, J., Raymer, J., and Forster, J. J. (2015). Bayesian population forecasting: extending the lee-carter method. *Demography*, **52**, 1035–1059.

Yang, Y. and Land, K. C. (2013). *Age-Period-Cohort Analysis: New Models, Methods, and Empirical Applications*. Chapman and Hall/CRC Press.

Yang, Y., Fu, W., and Land, K. (2004). A methodological comparison of age-period-cohort models: The intrinsic estimator and conventional generalized linear models. *Sociological Methodology*, **34**, 75–110.