ESTIMATION AND EXTRAPOLATION OF TIME TRENDS IN REGISTRY DATA—BORROWING STRENGTH FROM RELATED POPULATIONS

BY ANDREA RIEBLER[1,2], LEONHARD HELD[1] AND HÅVARD RUE

University of Zurich, University of Zurich and Norwegian University of Science and Technology

To analyze and project age-specific mortality or morbidity rates age-period-cohort (APC) models are very popular. Bayesian approaches facilitate estimation and improve predictions by assigning smoothing priors to age, period and cohort effects. Adjustments for overdispersion are straightforward using additional random effects. When rates are further stratified, for example, by countries, multivariate APC models can be used, where differences of stratum-specific effects are interpretable as log relative risks. Here, we incorporate correlated stratum-specific smoothing priors and correlated overdispersion parameters into the multivariate APC model, and use Markov chain Monte Carlo and integrated nested Laplace approximations for inference. Compared to a model without correlation, the new approach may lead to more precise relative risk estimates, as shown in an application to chronic obstructive pulmonary disease mortality in three regions of England and Wales. Furthermore, the imputation of missing data for one particular stratum may be improved, since the new approach takes advantage of the remaining strata if the corresponding observations are available there. This is shown in an application to female mortality in Denmark, Sweden and Norway from the 20th century, where we treat for each country in turn either the first or second half of the observations as missing and then impute the omitted data. The projections are compared to those obtained from a univariate APC model and an extended Lee–Carter demographic forecasting approach using the proper Dawid–Sebastiani scoring rule.

1. Introduction. Most developed countries have national health registers to routinely collect demographic rates. Age-period-cohort (APC) models are commonly used to analyze and project mortality or morbidity rates, in which effects related to the age of an individual, calendar time (period) and the generation (cohort) can reasonably be assumed to be present. When several of such register data sets are available, for example, for different countries, each data set could be analyzed separately by a univariate APC model. However, for comparable strata, similar unobservable factors are likely to act on the different time dimensions (age,

period, cohort), so that a multivariate APC analysis may seem more appropriate [Hansell et al. (2003), Hansell (2004), Jacobsen et al. (2004), Riebler and Held (2010)].

A quirk of APC models is the obvious linear dependence of age, period and cohort effects leading to a well-known identifiability problem. Over the last decades several proposals, ranging from the specification of additional identifying restrictions to the definition of estimable functions, have been made to solve the identifiability problem; see, for example, Fienberg and Mason (1979), Osmond and Gardner (1982), Holford (1983), Robertson and Boyle (1986), Clayton and Schifflers (1987), Holford (1992), Fu (2000), Yang, Fu and Land (2004) or Kuang, Nielsen and Nielsen (2008). Provided that at least one set of age, period or cohort effects is forced to be identical across strata (which is often a plausible assumption), differences of stratum-specific effects in the multivariate APC model are identifiable without further identifying restrictions. They can be interpreted as log relative risks, so that heterogeneous time trends, for example, across gender [Riebler et al. (2011)] or geographical regions [Hansell (2004), Riebler and Held (2010)], can be analyzed.

Bayesian APC analyses have become very popular in the last years; see, for example, Nakamura (1986), Berzuini and Clayton (1994), Besag et al. (1995), Ogata et al. (2000), Knorr-Held and Rainer (2001), Bray, Brennan and Boffetta (2001), Bray (2002), Baker and Bray (2005), Schmid and Held (2007), Riebler and Held (2010). As effects adjacent in time are likely to be similar, smoothing priors are typically assumed for age, period and cohort effects. Nakamura (1986) used first-order autoregressive priors, while Berzuini and Clayton (1994) and Besag et al. (1995) proposed to use second-order random walks. The second-order random walk is a discrete-time analogue of a cubic smoothing spline [Fahrmeir and Tutz (2001)]. This prior is defined on the identifiable second differences, a well-known measure of curvature, and penalizes deviations from a linear trend [Fienberg and Mason (1979), Clayton and Schifflers (1987)]. The degree of smoothness is controlled by an unknown smoothing parameter. Using smoothing priors, overfitting cohorts, which by design are sparsely represented, is avoided [Besag et al. (1995)].

When age group and period intervals are of different length, an additional identifiability problem may induce artificial cyclical patterns in the parameter estimates of uni- and multivariate APC models; see Holford (2006) and Riebler and Held (2010), respectively. However, this problem can be solved by applying smoothing functions, such as second-order random walks or penalized splines [Holford (2006)]. The assumed smoothness of age, period and cohort effects can also be exploited for providing projections, as the effects can easily be extrapolated into both the future and past [Knorr-Held and Rainer (2001), Bray (2002)]. In a hierarchical Bayesian model, additional random effects can be included to account for heterogeneity without temporal structure [Berzuini and Clayton (1994), Besag et al. (1995)].

Riebler and Held (2010) assumed independent smoothing priors for stratum-specific effects in multivariate APC models. Here, we propose to link stratum-specific smoothing priors. The new approach leads to a multivariate correlated random walk, where the joint precision matrix is defined as the Kronecker product of the inverse of a uniform correlation matrix and the precision matrix of the univariate second random walk. Inference is done using Markov chain Monte Carlo (MCMC) and integrated nested Laplace approximations [Rue, Martino and Chopin (2009)].

The new specification can be regarded as shrinking the stratum-specific parameters toward some common trend. Indeed, an alternative model formulation would introduce a common period effect, say, modeled via a second order random walk and additionally independent second order random walks for each stratum. While this formulation has two variance parameters, it in fact induces correlation between the stratum-specific increments, which are defined as the sum of the common innovation and the stratum-specific innovations, so that it can be translated into a multivariate random walk with one variance and one correlation parameter.

In time series analysis, the use of multivariate random walks plays a fundamental role in multivariate modeling [Harvey (1990)]. The multivariate random walk is an example of an intrinsic multivariate Gaussian Markov random field (GMRF) model [Rue and Held (2005)]. Multivariate GMRF models with conditional autoregressive (CAR) structure are sometimes called multivariate CAR (MCAR) models; see, for example, Gelfand and Vounatsou (2003) or Carlin and Banerjee (2003). Proper multivariate GMRF models have been introduced by Mardia (1988). Greco and Trivisano (2009) applied MCAR models to handle general forms of spatial dependence occurring in multivariate spatial modeling of area data. Lagazio, Biggeri and Dreassi (2003) and Schmid and Held (2004) used Kronecker product precision matrices to model different types of space–time interactions in spatial APC models [Knorr-Held (2000)]. However, as far as we know, correlated second order random walks have never been used in multivariate APC models.

We further propose the incorporation of correlated overdispersion parameters to model unobserved risk factors without temporal structure but acting simultaneously on the different strata. The use of correlated overdispersion parameters is similar in spirit to seemingly unrelated regressions, where single regression equations are linked by correlated error terms [Harvey (1990)].

Through the introduction of correlation in the prior distribution the effective degrees of freedom are reduced whenever similar behavior in the different strata exists. Hence, the precision of relative risks may be improved. Furthermore, the approach is useful to predict missing records in one particular stratum if the corresponding data are available for the remaining strata. This might be the case for historical data if the collection of demographic rates started not at the same time in different strata. Consider, for example, Switzerland, where each canton (administrative unit) is separately responsible for the implementation of health-policy instruments, so that cancer is registered on a cantonal level [Ess et al. (2010),

Bouchardy, Lutz and Kühni (2011)]. The first Swiss cancer registration system started in 1970 in the canton of Geneva followed by registers in the cantons of Vaud and Neuchâtel in 1974 [Bouchardy, Lutz and Kühni (2011)]. Compared to other cantons without explicit cancer registration, extensive cancer analyses have been performed for these cantons; see, for example, Levi et al. (1993, 1998, 2002), Verkooijhen et al. (2003). Today most cantons have cancer registers and it is planned that within the next years the entire Swiss population will be captured by a cancer registration system [Bouchardy, Lutz and Kühni (2011)]. Our method can be used to impute missing data for cantons with a younger cancer registration system taking advantage of other cantons with a longer collection period. Thus, important insight into cancer progression for all cantons could be gained. A different aspect might be varying collection intervals in different regions, where in some regions data are collected on a yearly basis, say, and in other regions on a five-year basis. Here, the correlated multivariate APC approach may be used to impute the rates for the missing years. In this paper, we demonstrate the ability to impute missing data units in a cross-prediction study of female mortality in Scandinavia.

The paper is organized as follows. Section 2 introduces the two applications presented in this paper. In Section 3 we review multivariate APC models and introduce our extended correlated approach (Section 3.1). Then we present details on the implementation (Section 3.2). In Section 4 we present the results of the two applications. Our findings are summarized in Section 5.

2. Applications.

2.1. *Analysis of heterogeneous time trends in COPD mortality among males in England and Wales.*

We reanalyze male mortality data on chronic obstructive pulmonary disease (COPD) in three regions of England and Wales: Greater London, conurbations excluding Greater London and rural areas (nonconurbations). COPD is one of the most common lung diseases making it hard to breathe as a consequence of limited air flow. One of the main causes of COPD is smoking, but also air pollution, smog, dust and chemical fumes are relevant risk factors. While smoking exerts mainly long-term effects with a lag period of about 20–30 years [Kazerouni et al. (2004)], air pollution can cause both long-term (period or cohort) effects and short-term (period) effects [Sunyer (2001), Dockery and Pope (1994)]. We focus on short-term effects and the relation between marked air pollution events and changes in COPD mortality. For all regions data are available on an annual basis from 1950–1999 for seven age groups: 15–24, 25–34, . . . , 75+ [Hansell et al. (2003), Hansell (2004)]. Riebler and Held (2010) analyzed heterogeneous time trends in these data using an uncorrelated multivariate APC model with common age effects. We will compare their results with those obtained from a model with correlated stratum-specific period, cohort and overdispersion parameters.

2.2. *Extrapolation of overall mortality of Scandinavian females.* All data were obtained from the Human Mortality Database (2011). The number of deaths are stratified by 5-year groups, for all Norwegian, Danish and Swedish women aged 0–84 in the period 1900–1999, leading to 17 age groups (0–4, 5–9, . . . , 80–84) and 20 periods (1900–1904, . . . , 1995–1999). Figure 1 shows the death rates per 1,000 person-years for all three countries stratified by 5-year age groups. To obtain person-years, we used the yearly population sizes available for the same age groups and based on the 1st of January. We used linear interpolation to get mid-year estimates and then added up the resulting quantities to obtain person-years for 1900–1904, . . . , 1995–1999. The rates of all three countries show a very similar progression. The peak in mortality in the 1915–1919 period, present, in particular, among young adults, is supposed to be related to the 1918–1919 Spanish flu pandemic which killed about 50 million people worldwide, with most deaths occurring among young adults [Andreasen, Viboud and Simonsen (2008)]. During the summer of 1918 there were strong influenza waves in Denmark, Sweden and Norway [Andreasen, Viboud and Simonsen (2008), Kolte et al. (2008)].

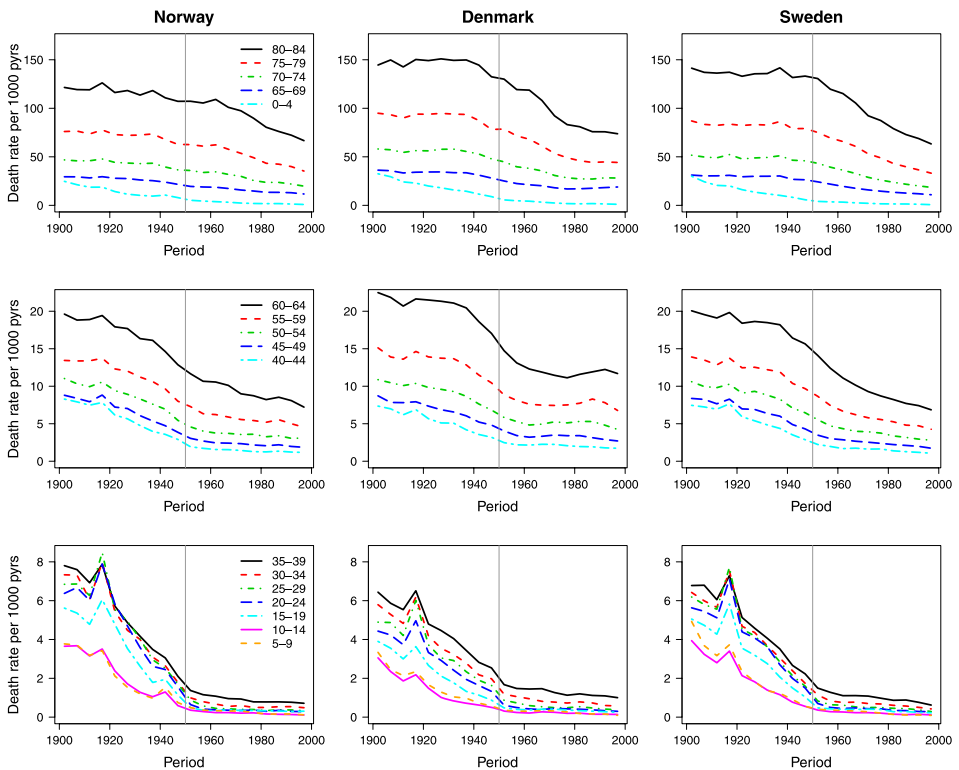FIG. 1. *Female death rates per* 1,000 *person-years* (*pyrs*) *in Norway, Denmark and Sweden by age from* 1900 *to* 1999. *The vertical line divides the time into equally sized parts.*

We divide the calendar period into two equally sized parts (see Figure 1). For either the first or second half of the 20th century, all observations from one particular country are treated as missing. The omitted data are then predicted exploiting the information provided by the complete data sets of the other two countries. This procedure is repeated for all countries and hence termed cross-prediction. Thus, we cannot only assess the ability to project particular events present for all countries, such as the Spanish-flu pattern, but also analyze the prediction quality for years without such specific events. The probabilistic projections are compared to those obtained from a univariate APC model and to those from an extended Lee–Carter demographic forecasting model [Lee and Carter (1992)].

The data and code used in this application are provided in the Supplementary Material [Riebler, Held and Rue (2011)].

3. The correlated multivariate APC model. Let y_{ijr} denote the number of deaths observed for age group i $(i = 1, \ldots, I)$, period j $(j = 1, \ldots, J)$ and stratum r $(r = 1, \ldots, R)$. In both of our applications r represents a geographical region (either a region in England and Wales or a Scandinavian country). Deaths can be regarded as events arising from a Poisson process. Hence, y_{ijr} can be interpreted as the number of events that have occurred during an exposure period of n_{ijr} person-years, in which the occurrence rate is assumed to be λ_{ijr} per person-year. Thus, y_{ijr} is Poisson distributed with rate $n_{ijr}\lambda_{ijr}$, where n_{ijr} is known [Armitage (1966), Brillinger (1986)]. In the most general formulation of the multivariate APC model, the linear predictor is

$$\eta_{ijr} = \log(\lambda_{ijr}) = \mu_r + \theta_{ir} + \varphi_{jr} + \psi_{kr}.$$

Here, μ_r is the stratum-specific intercept, and θ_{ir}, φ_{jr} and ψ_{kr} are stratum-specific age, period and cohort effects, respectively. The cohort index k is a linear function of the age index i and the period index j. If the time interval widths of age group and period are equal, then $k = (I - i) + j$. If age group intervals are M times wider than period intervals, as is the case in the first application (Section 2.1) where $M = 10$, then $k = M \times (I - i) + j$ [Heuer (1997)]. We apply the usual constraints, $\sum_{i=1}^{I} \theta_{ir} = \sum_{j=1}^{J} \varphi_{jr} = \sum_{k=1}^{K} \psi_{kr} = 0$ for $r = 1, \ldots, R$, to ensure identifiability of the stratum-specific intercepts. However, parameter estimates are still not identifiable without imposing additional constraints [Fienberg and Mason (1979), Holford (1983)]. In contrast, second differences of parameter estimates, for example, $\theta_{ir} - 2\theta_{i-1r} + \theta_{i-2r}$, are not affected by the identifiability problem and can be uniquely determined [Fienberg and Mason (1979), Clayton and Schifflers (1987)]. Furthermore, stratum-specific differences, for example, $\theta_{ir_1} - \theta_{ir_2}$ with $r_1 \neq r_2$, are identifiable (absent an additional constraint), provided that at least one of the three time effects (age, period, cohort) is common across strata [Riebler and Held (2010)].

3.1. *Bayesian inference.* In a Bayesian context, we work with a hierarchical model in which prior distributions need to be assigned to all parameters. We use independent flat priors for each stratum-specific intercept μ_r. Riebler and Held (2010) assigned independent smoothing priors to the age effects $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_I)^\top$, each stratum-specific set of period effects $\boldsymbol{\varphi}_r = (\varphi_{1r}, \ldots, \varphi_{Jr})^\top$ and cohort effects $\boldsymbol{\psi}_r = (\psi_{1r}, \ldots, \psi_{Kr})^\top$, $r = 1, \ldots, R$, in a model with common age effects. Consider, for example, the period effects for a specific stratum r. The random walk of second order (RW2) is a smoothing prior based on second differences and penalizes deviations from a linear trend. This improper prior can be written as

$$f(\boldsymbol{\varphi}_r | \kappa_\varphi) \propto \kappa_\varphi^{(J-2)/2} \exp\left(-\frac{\kappa_\varphi}{2} \sum_{j=3}^{J} \left((\varphi_{jr} - \varphi_{(j-1)r}) - (\varphi_{(j-1)r} - \varphi_{(j-2)r})\right)^2\right)$$

$$= \kappa_\varphi^{(J-2)/2} \exp\left(-\frac{1}{2}\boldsymbol{\varphi}_r^\top \mathbf{P}_\varphi \boldsymbol{\varphi}_r\right)$$

with precision matrix $\mathbf{P}_\varphi$, which depends on an unknown precision parameter κ_φ:

$$\mathbf{P}_\varphi = \kappa_\varphi \begin{pmatrix} 1 & -2 & 1 & & & & & \\ -2 & 5 & -4 & 1 & & & & \\ 1 & -4 & 6 & -4 & 1 & & & \\ & \ddots & \ddots & \ddots & \ddots & \ddots & & \\ & & 1 & -4 & 6 & -4 & 1 & \\ & & & 1 & -4 & 5 & -2 & \\ & & & & 1 & -2 & 1 & \end{pmatrix}.$$

Here, we propose the use of correlated smoothing priors for stratum-specific time effects. Let $\mathbf{C} = (1 - \rho)\mathbf{I} + \rho\mathbf{J}$ denote a uniform correlation matrix, where ρ is the unknown correlation parameter, $\mathbf{I}$ the identity matrix and $\mathbf{J}$ a matrix of ones. The random walks of the stratum-specific period effects $\boldsymbol{\varphi}_1, \ldots, \boldsymbol{\varphi}_R$ can be correlated using the stacked vector $\tilde{\boldsymbol{\varphi}} = (\boldsymbol{\varphi}_1^\top, \ldots, \boldsymbol{\varphi}_R^\top)^\top$:

$$f(\tilde{\boldsymbol{\varphi}} | \mathbf{C}_\varphi, \kappa_\varphi) \propto (|\mathbf{C}_\varphi^{-1} \otimes \mathbf{P}_\varphi|^\star)^{1/2} \exp\left(-\tfrac{1}{2}\tilde{\boldsymbol{\varphi}}^\top \{\mathbf{C}_\varphi^{-1} \otimes \mathbf{P}_\varphi\}\tilde{\boldsymbol{\varphi}}\right)$$

$$= |\mathbf{C}_\varphi^{-1}|^{(J-2)/2} \cdot (|\mathbf{P}_\varphi|^\star)^{R/2} \exp\left(-\tfrac{1}{2}\tilde{\boldsymbol{\varphi}}^\top \{\mathbf{C}_\varphi^{-1} \otimes \mathbf{P}_\varphi\}\tilde{\boldsymbol{\varphi}}\right),$$

where $\otimes$ denotes the Kronecker product and $|\cdot|^\star$ the generalized determinant defined as the product of all nonzero eigenvalues. The determinant of $\mathbf{C}_\varphi^{-1}$ is $[(1 + (R-1)\rho_\varphi)(1 - \rho_\varphi)^{R-1}]^{-1}$; see the proof in Appendix A. This formulation corresponds to a multivariate RW2 with correlated increments and is an example for an improper (intrinsic) correlated GMRF [Gelfand and Vounatsou (2003), Rue and Held (2005)].

To adjust for heterogeneity, which has no temporal structure but is likely to exist in the underlying rates, we introduce stratum-specific latent random effects z_{ijr} into the linear predictor. These overdispersion parameters are typically assumed

to be independent Gaussian variables with mean zero and unknown variance κ_z^{-1}, that is, $z_{ijr} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \kappa_z^{-1})$ [Berzuini and Clayton (1994), Besag et al. (1995)]. However, when interpreting these latent effects as unobserved covariates, it may be plausible that they act partly simultaneously on the different strata. Hence, we propose correlated overdispersion parameters and set $\mathbf{z}_{ij} = (z_{ij1}, \ldots, z_{ijR})^\top \sim \mathcal{N}(0, \kappa_z^{-1}\mathbf{C}_z)$ for all i and j.

All of the up to eight hyperparameters (four precisions and up to four correlations) are treated as unknown. Suitable gamma-hyperpriors are assigned to the precisions. As in Knorr-Held and Rainer (2001), we use Ga(1, 0.00005) for the precisions of age, period and cohort effects and Ga(1, 0.005) for the precision of the overdispersion.

Correlation parameters ρ are reparameterized using the general Fisher's z-transformation [Fisher (1958), page 219]:

$$(1) \qquad \rho = \frac{\exp(\rho^\star) - 1}{\exp(\rho^\star) + R - 1}, \qquad \rho^\star = \log\left(\frac{1 + \rho \cdot (R - 1)}{1 - \rho}\right),$$

where $\rho^\star$ can take any real value. It is worth noting that this transformation ensures that ρ only takes values within the interval $(-1/(R - 1), 1)$, so that $\mathbf{C}$ is positive definite without imposing an additional constraint. Using $R = 2$ in (1), we obtain

$$\rho = \frac{\exp(\rho^\star) - 1}{\exp(\rho^\star) + 1}, \qquad \rho^\star = \log\left(\frac{1 + \rho}{1 - \rho}\right),$$

which is frequently used for constructing confidence intervals for ρ [Konishi (1985)]. Fisher's z-transformation is a variance stabilizing transformation. In a Bayesian context this transformation is of particular interest since the derivative of a variance stabilizing transformation corresponds to Jeffreys' prior for the original parameter [Lehmann (1999), pages 491 and 492]. For example, for $R = 2$, Jeffreys' prior is $\pi(\rho) \propto 1/(1 - \rho^2)$, the derivative of $\log(\frac{1+\rho}{1-\rho})$ [Lindley (1965), pages 215–220].

We assign a normal prior with mean zero and fixed precision $\kappa_{\rho^\star}$ to $\rho^\star$. Thus, the prior probability that ρ is larger than zero is equal to 0.5, independent of R. Figure 2 shows the resulting prior for ρ for three different values of $\kappa_{\rho^\star}$ and three different values of R. For $R = 2$ strata, setting $\kappa_{\rho^\star}$ to 0.2 corresponds to a U-shaped prior, $\kappa_{\rho^\star} = 0.4$ to a roughly uniform prior and $\kappa_{\rho^\star} = 0.8$ to a bump-shaped prior for ρ; compare the first column of Figure 2. Note that $\kappa_{\rho^\star} = 0$ corresponds to the improper Jeffreys' prior. For a larger number of strata, the left boundary for the correlation is shifted toward zero, resulting in an asymmetric prior distribution for ρ, since half of the total density is distributed to a smaller interval, $(-1/(R - 1), 0)$. We use $\kappa_{\rho^\star} = 0.2$, so that sufficient probability mass is assigned to the boundary values as well, making extreme posterior correlation estimates possible.
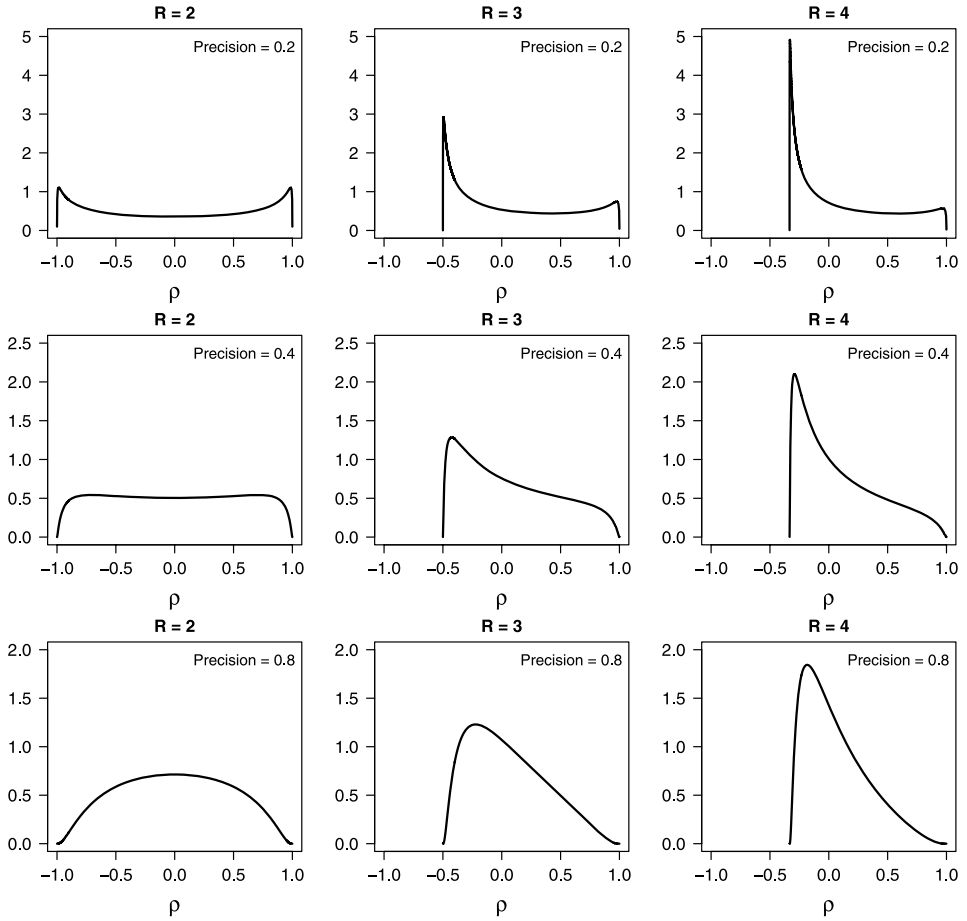
FIG. 2. *Prior distribution for correlation parameter ρ derived from a zero-mean Gaussian distribution for $\rho^\star$ with three different values for the precision $\kappa_{\rho^\star}$* (*top to bottom*: 0.2, 0.4, 0.8) *and three different numbers of strata R* (*left to right*: $R = 2$, $R = 3$, $R = 4$).

3.2. *Implementation.* Bayesian inference for the models presented is not straightforward, since the posterior distribution is not analytically available. The common tool of choice is MCMC sampling. An alternative is integrated nested Laplace approximations (INLAs). To compare these two inference techniques, we implemented correlated multivariate APC models using both MCMC and INLA. In the first application, we apply INLA and MCMC to show the almost perfect coincidence of both approaches. Due to the complexity of the second application resulting in large thinning intervals and burn-in periods, we only present the results of INLA.

3.2.1. *Analysis with MCMC.* Algorithmic routines based on MCMC are implemented in the low-level programming language C using the GMRFLib library

[Rue and Held (2005)]. Following Besag et al. (1995), we reparameterize the model from z_{ijr} to η_{ijr} to obtain multivariate normal full conditional distributions for the stratum-specific intercepts $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_r)^\top$ and all sets of time effects. Block updating allows the proper incorporation of the sum-to-zero constraints for the time effects. It is also possible to omit the sum-to-zero constraint for one set of stratum-specific effects and simultaneously remove the stratum-specific intercepts $\boldsymbol{\mu}$ from the algorithm. For the precisions Gibbs sampling is used as well. The vector $\boldsymbol{\eta}_{ij} = (\eta_{ij1}, \ldots, \eta_{ijR})^\top$ has a nonstandard distribution. It is updated using multivariate Metropolis–Hastings steps with a GMRF proposal distribution based on a second-order Taylor approximation of the log likelihood [Rue and Held (2005), Section 4.4.1]. For the correlation parameters Metropolis–Hastings updates based on a random walk proposal are used, such that acceptance rates around 40% are achieved. In the application to COPD mortality we use a MCMC run of 350,000 iterations, discarding the first 50,000 iterations and storing every 20th sample thereafter, resulting in 15,000 samples. We have routinely examined convergence and mixing diagnostics.

3.2.2. *Analysis with INLA.* Rue, Martino and Chopin (2009) proposed with INLA an alternative deterministic Bayesian inference approach for latent Gaussian random field models. INLA replaces time-consuming MCMC sampling with fast and accurate approximations to the posterior marginal distributions. Some empirical comparison with MCMC results can be found in Rue, Martino and Chopin (2009), Paul et al. (2010) or Schrödle et al. (2011). We incorporated correlated GMRF models into INLA, enabling the analysis of correlated multivariate APC models based on a uniform correlation structure and using the general Fisher's z-transformation. The methodology is integrated in the package INLA (see www.r-inla.org) for R [R Development Core Team (2010)]. For both applications, we use the INLA package built on 14.03.2011.

4. Results.

4.1. *COPD mortality among males in England and Wales.* We compared the uncorrelated model with joint age-effects, and region-specific period and cohort effect presented by Riebler and Held (2010) with three different correlated formulations: (1) Region-specific period and region-specific cohort effects are correlated; (2) the overdispersion parameters are correlated; (3) region-specific period effects, region-specific cohort effects and overdispersion parameters are correlated. To make the models comparable, we used, in contrast to Riebler and Held (2010), the same precision for the independent priors of region-specific period effects and also the same precision for the independent priors of region-specific cohort effects. For all models MCMC and INLA produce virtually identical results; see Figure 3 for a comparison of precision and correlation estimates in model 3. The running time of INLA was always less than the computation time with MCMC. Inspecting
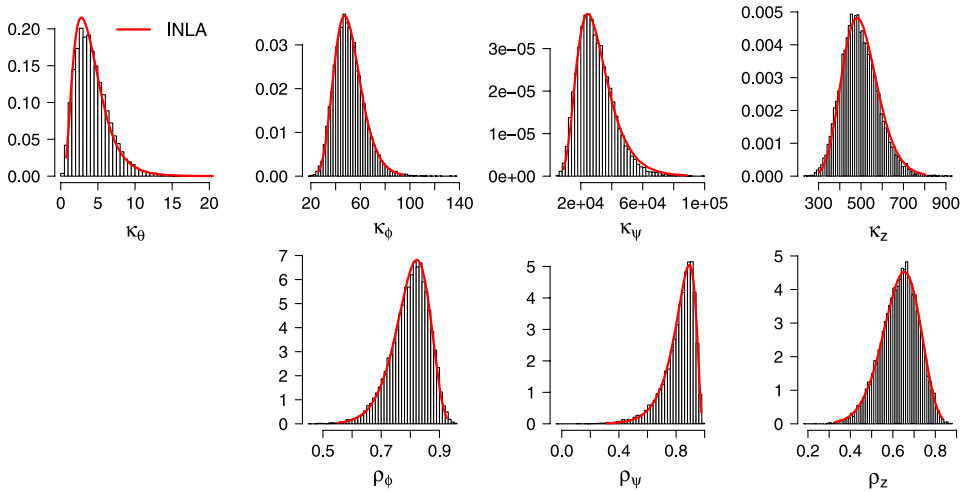
FIG. 3. *Approximated posterior marginal densities (solid red line) of precision and correlation parameters for the multivariate model with joint age effects and correlated period, cohort and overdispersion parameters obtained from INLA. Moreover, the corresponding histograms of 15,000 MCMC samples obtained from a run with 50,000 burn-in iterations and a thinning of 20 are shown.*

the log marginal likelihood returned by INLA, the model with correlated period and cohort effects, and correlated overdispersion parameters, was classified as the best model. Despite the improper random walk prior, the log-marginal likelihood can be used here, as the models are based on the same underlying latent structure and only differ by the inclusion of correlation. Furthermore, the correlation estimates ρ_φ, ρ_ψ and ρ_z of model 3 (Figure 3) are clearly different from zero, confirming the between-region dependence.

Figure 4 compares the estimates of average relative risks obtained from MCMC for the models with uncorrelated and correlated region-specific effects and overdispersion parameters, respectively. The estimates are relative to nonconurbations, where the mortality rates tend to be the lowest. The results of both models are very similar. The average relative risk of period effects shows the typical year-to-year variation, with higher values in years of known air pollution events, such as the "Great Smog" in London in 1952. In the average relative risks of cohort effects, different smoking behavior may be visible. For a detailed interpretation of the relative risks we refer to Riebler and Held (2010). Due to fewer observations, the credible intervals are getting wider for younger birth cohorts. However, adjusting for correlation improves the precision of the relative risks estimates, in particular, for younger birth cohorts. The average posterior standard deviation in the correlated approach is about 20% and 25% smaller for the average relative risks of the period effects and cohort effects, respectively.
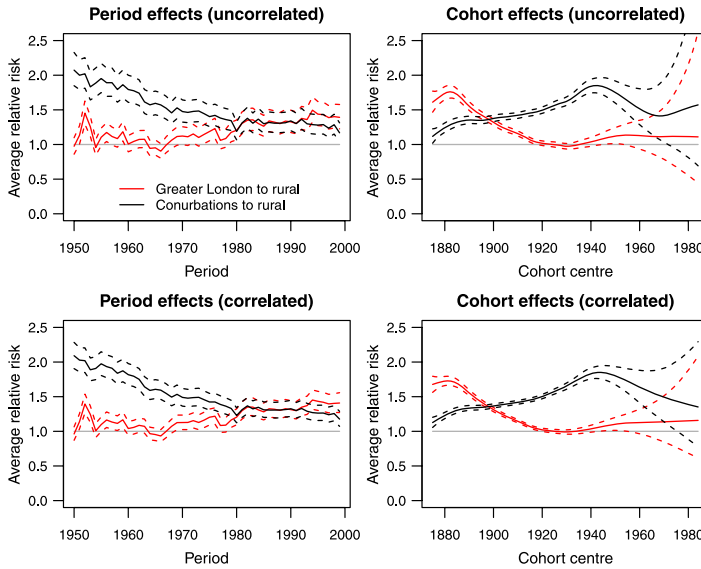
FIG. 4. *Average relative risk of death for Greater London and conurbations excluding Greater London compared with nonconurbations, analyzed by a multivariate model with joint age effects and no correlations across other parameters* (*top*), *and a multivariate model with joint age effects and correlated period, cohort and overdispersion parameters* (*bottom*). *Shown are the median estimates within* 95% *pointwise credible bands.*

4.2. *Extrapolation of overall mortality of Scandinavian females.* We will first briefly introduce the basic and the extended Lee–Carter model considered. Then we will present the results of the predictive model assessment and compare the projections obtained by the different approaches.

4.2.1. *The quasi-Poisson version of the Lee–Carter model.* The Lee–Carter model, introduced by Lee and Carter (1992) to forecast mortality in the U.S., is one of the best-known methods for mortality forecasting and often used as a reference [Booth (2006), Booth et al. (2006)]. It assumes a log-bilinear form

$$\log \frac{y_{ij}}{n_{ij}} = \alpha_i + \beta_i \kappa_j + \varepsilon_{ij},$$

where α_i describes the average shape of the age profile, β_i the age-specific mortality change from this pattern with time-varying trend κ_j, and ε_{ij} are homoskedastic centered error terms. The parameters are constrained to $\sum_j \kappa_j = 0$ and $\sum_i \beta_i = 1$. Forecasting using this model proceeds in two steps: (1) the model coefficients are estimated; (2) the time trend κ_j is extrapolated based on an ARIMA(0,1,0) time-series model, that is, a random walk with drift. This forecasted trend is used to derive the projected age-specific mortality rates based on the estimates for α_i and β_i from step 1. Note that only the uncertainty in the time trend κ_j is taken into

account in the projected rates, so that not all variability is captured [Lee and Carter (1992), Butt and Haberman (2009)].

The Lee–Carter model was further developed and embedded in a quasi-Poisson regression model by Brouhns, Denuit and Vermunt (2002). We used the ilc-package [Butt and Haberman (2009)] in R to generate univariate predictions for the country under consideration based on this extended model. Since the implementation does not allow to project into the past, we reversed the time-scale when predicting data of the first half of the 20th century.

4.2.2. *Predictive model assessment.* Each of the three models (Lee–Carter, univariate APC and multivariate APC; with the latter two abbreviated as APC and cMAPC, resp.) generates, for each of the six scenarios of the cross-prediction procedure, $10 \times 17 = 170$ probabilistic forecasts for country $r^\star$ under consideration. We used the mean squared error score to assess the concentration of the predictive distribution (sharpness). To assess the statistical consistency between the distributional forecasts and the observations (calibration), we calculated prediction intervals at various levels and computed the empirical coverage probabilities, that is, the proportion of the prediction intervals that cover the observed number of cases. To combine sharpness and calibration in one measure, we further report the Dawid–Sebastiani scoring rule (DSS) defined as

$$\mathrm{DSS} = \left(\frac{y_{ijr^\star} - \mu_{ijr^\star}}{\sigma_{ijr^\star}} \right)^2 + 2\log(\sigma_{ijr^\star}),$$

where $y_{ijr^\star}$ is the observation that realizes, $\mu_{ijr^\star}$ the mean and $\sigma_{ijr^\star}$ the standard deviation of the predictive distribution [Gneiting and Raftery (2007)]. This score has been proposed as a proper alternative to the predictive model choice criterion of Gelfand and Ghosh (1998) and was also used by Czado, Gneiting and Held (2009) to assess the predictive quality of a univariate APC analysis applied to cancer incidence in Germany. To calculate these quantities, we need to post-process the results returned by INLA and the ilc-package.

For the univariate and multivariate APC analysis, INLA returns posterior summary estimates and posterior marginal densities for the linear predictor $\eta_{ijr^\star}$, with $i = 1, \ldots, 17$ and $j = 1, \ldots, 10$ or $j = 11, \ldots, 20$, depending on whether we project the first or last half of the 20th century. The corresponding estimates for $\lambda_{ijr^\star}$ are straightforward to derive. For the univariate Lee–Carter analysis of country $r^\star$, the ilc-package returns the predicted mortality rate $\lambda_{ijr^\star}$ and the prediction intervals (symmetric on the log-scale) at a predefined level.

We need the mean $\mu_{ijr^\star} = \mathbb{E}(y_{ijr^\star})$ of the predictive distribution for computing the DSS and the mean squared error score. Using the law of iterated expectations [Billingsley (1986), Theorem 34.4], $\mu_{ijr^\star}$ can be derived. With $y_{ijr^\star}|\lambda_{ijr^\star} \sim \mathrm{Po}(n_{ijr^\star} \cdot \lambda_{ijr^\star})$, it follows

$$\mu_{ijr^\star} = \mathbb{E}(\mathbb{E}(y_{ijr^\star}|\lambda_{ijr^\star})) = \mathbb{E}(n_{ijr^\star} \cdot \lambda_{ijr^\star}) = n_{ijr^\star} \cdot \mathbb{E}(\lambda_{ijr^\star}).$$

Analogously, the variance $\sigma^2_{ijr\star} = \text{Var}(y_{ijr\star})$ follows from the law of total variance as

$$\begin{aligned}
\sigma^2_{ijr\star} &= \mathbb{E}(\text{Var}(y_{ijr\star}|\lambda_{ijr\star})) + \text{Var}(\mathbb{E}(y_{ijr\star}|\lambda_{ijr\star})) \\
&= \mathbb{E}(n_{ijr\star} \cdot \lambda_{ijr\star}) + \text{Var}(n_{ijr\star} \cdot \lambda_{ijr\star}) \\
&= n_{ijr\star} \cdot \mathbb{E}(\lambda_{ijr\star}) + n^2_{ijr\star} \text{Var}(\lambda_{ijr\star})
\end{aligned}$$

for INLA. Under a quasi-Poisson approach with $\text{Var}(y_{ijr}|\lambda_{ijr}) = \phi \cdot n_{ijr} \cdot \lambda_{ijr}$ we need to explicitly incorporate the overdispersion parameter ϕ, so $\sigma^2_{ijr} = \phi \cdot n_{ijr}\mathbb{E}(\lambda_{ijr}) + n^2_{ijr} \text{Var}(\lambda_{ijr})$. Here, we used the total lack of fit as ϕ; compare Booth, Maindonald and Smith (2002).

To obtain posterior predictive quantiles, the missing Poisson variation was added to the predicted mortality rates. When using INLA, this is done by numerical integration over the predicted posterior marginal of $\lambda_{ijr\star}$. Since we do not obtain the posterior marginal of $\lambda_{ijr\star}$ using the ilc-package, we used Monte Carlo sampling instead. To be more precise, we generated $N = 100{,}000$ samples for the linear predictor $\eta_{ijr\star}$ from a normal distribution with mean $\log(\hat{\lambda}_{ijr\star})$ and variance derived from the symmetric prediction intervals on log-scale. Then, we generated for each sample $\eta^{(s)}_{ijr\star}$, $s = 1, \ldots, N$, one sample $y^{(s)}_{ijr\star}$ from a negative binomial distribution with density

$$f(y) = \frac{\Gamma(y+d)}{\Gamma(d)\Gamma(y+1)}\left(\frac{d}{m+d}\right)^d \left(\frac{m}{m+d}\right)^y,$$

where $\mathbb{E}(y) = m$ and $\text{Var}(y) = m(1+m/d)$. To match the mean and variance of the quasi-Poisson distribution, we set $m^{(s)} = n_{ijr\star}\exp(\eta^{(s)}_{ijr\star})$ and $d^{(s)} = m^{(s)}/(\phi - 1)$ for each sample $\eta^{(s)}_{ijr\star}$. Subsequently, quantiles at different prediction levels could be extracted from the samples.

Table 1 shows for all models the mean squared error (MSE) score, the empirical coverage probabilities and the mean DSS averaged over all 170 projections. For five of the six scenarios and especially when predicting the first 10 periods, the correlated multivariate APC model is clearly the best model. Although the prediction intervals are sometimes too large, as indicated by larger empirical coverage probabilities than the nominal level, the empirical coverage is mostly closer to the nominal level than for the other two approaches. Regarding mean DSS and empirical coverage, the univariate APC model also performs mostly better than the extended Lee–Carter approach. In particular, predicting the first half of the 20th century, the extended Lee–Carter approach showed severe deficits. It was classified as the best model regarding all predictive assessment criteria only when predicting the second half of the 20th century for Norway. Inspecting the posterior correlation estimates of the cMAPC model (Table 2), we observe that for this scenario the posterior correlations between country-specific period effects and also between

Mean Dawid–Sebastiani score ($\overline{DSS}$), mean squared error score (MSE), empirical coverage probabilities for all models predicting female mortality of one country either for the first or second half of the 20th century

	Measure	1900–1949			1950–1999		
		Lee–Carter	APC	cMAPC	Lee–Carter	APC	cMAPC
		NORWAY					
	$\overline{DSS}$	232.4	38.9	17.9	13.5	16.8	15.0
	MSE	3.49e+06	7.21e+06	1.73e+06	3.10e+06	1.92e+07	1.50e+07
Level	95%	19	66	70	91	96	78
	80%	13	52	35	79	73	54
	50%	9	39	11	52	35	37
		DENMARK					
	$\overline{DSS}$	50.5	41.2	14.7	22.2	17.1	13.7
	MSE	3.37e+06	2.30e+07	3.56e+06	2.03e+07	6.43e+07	1.80e+07
Level	95%	24	58	96	66	95	99
	80%	17	50	93	51	92	83
	50%	6	15	79	30	64	65
		SWEDEN					
	$\overline{DSS}$	249.6	43.4	15.4	24.7	18.5	13.6
	MSE	4.45e+07	9.48e+07	9.04e+06	9.43e+07	2.04e+08	3.05e+06
Level	95%	11	64	99	70	97	95
	80%	8	28	94	62	95	72
	50%	4	8	55	41	75	54

country-specific cohort effects are lower than in the other scenarios. In contrast, the correlation between country-specific overdispersion parameters is quite high.

TABLE 2
Median and 95% credible interval for all correlation parameters in the correlated multivariate APC model

| Predicted period	Predicted country	Age	Period	Cohort	Overdispersion
1900–1949	Norway	$_{0.978}0.991_{0.996}$	$_{0.81}0.96_{0.99}$	$_{0.57}0.82_{0.94}$	$_{0.91}0.93_{0.94}$
	Denmark	$_{0.994}0.998_{0.999}$	$_{0.35}0.87_{0.99}$	$_{0.44}0.76_{0.92}$	$_{0.83}0.86_{0.89}$
	Sweden	$_{0.984}0.994_{0.997}$	$_{0.70}0.95_{0.99}$	$_{0.58}0.84_{0.95}$	$_{0.75}0.79_{0.84}$
1950–1999	Norway	$_{0.969}0.986_{0.994}$	$_{0.01}0.51_{0.85}$	$_{0.28}0.62_{0.84}$	$_{0.90}0.92_{0.94}$
	Denmark	$_{0.964}0.984_{0.994}$	$_{0.18}0.75_{0.96}$	$_{0.89}0.97_{0.99}$	$_{0.82}0.85_{0.88}$
	Sweden	$_{0.976}0.990_{0.996}$	$_{0.62}0.96_{1.00}$	$_{0.60}0.83_{0.94}$	$_{0.77}0.81_{0.84}$
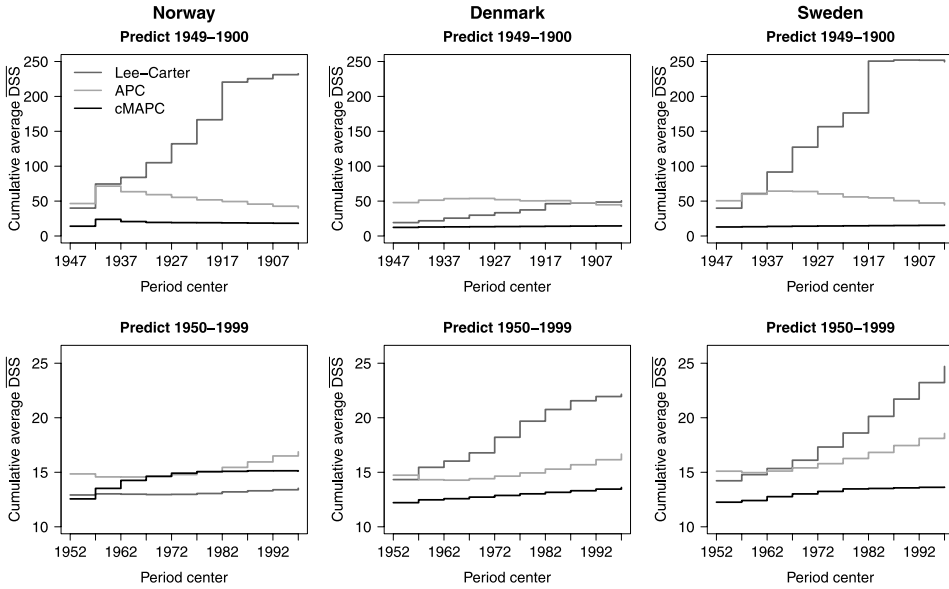
FIG. 5.   *Cumulative average of mean Dawid–Sebastiani scores across age groups.*

To compare the performance change from short-term to long-term forecasts, Figure 5 shows the cumulative average $\overline{\mathrm{DSS}}_j$, where $\overline{\mathrm{DSS}}_j$ denotes the mean DSS across age group at period j. Except for predicting death rates in Norway from 1950–1999, the curve for the cMAPC model is always below those of the two univariate approaches. Predicting the periods in the first half of the 20th century, the cumulative average $\overline{\mathrm{DSS}}_j$ of the extended Lee–Carter model is constantly increasing, indicating a lower projection quality with increasing time. The largest jump occurs for the period 1915–1919 with the Spanish flu. In contrast, the score of the univariate APC model decreases when predicting more periods, while the score for the correlated multivariate APC model stays fairly constant. Predicting the periods in the second half of the 20th century, the cumulative average $\overline{\mathrm{DSS}}_j$ slowly increases for all models. However, except for Norway, the cumulative score of the extended Lee–Carter model shows larger jumps from one period to the next.

4.2.3. *Projections.*   The median projected death rates per 1,000 person-years together with 80% pointwise prediction intervals for Norwegian women obtained from all three models are shown in Figure 6 for the first half and Figure 7 for the second half of the 20th century. Furthermore, the true death rates of Norwegian women are added to the figures. For all models and especially for the univariate APC model, the prediction intervals are getting wider as prediction time goes on. While the projections of the two univariate approaches are almost straight lines, different temporal patterns across age groups can be seen for the correlated model. The Spanish flu was especially well captured by the correlated approach. Since
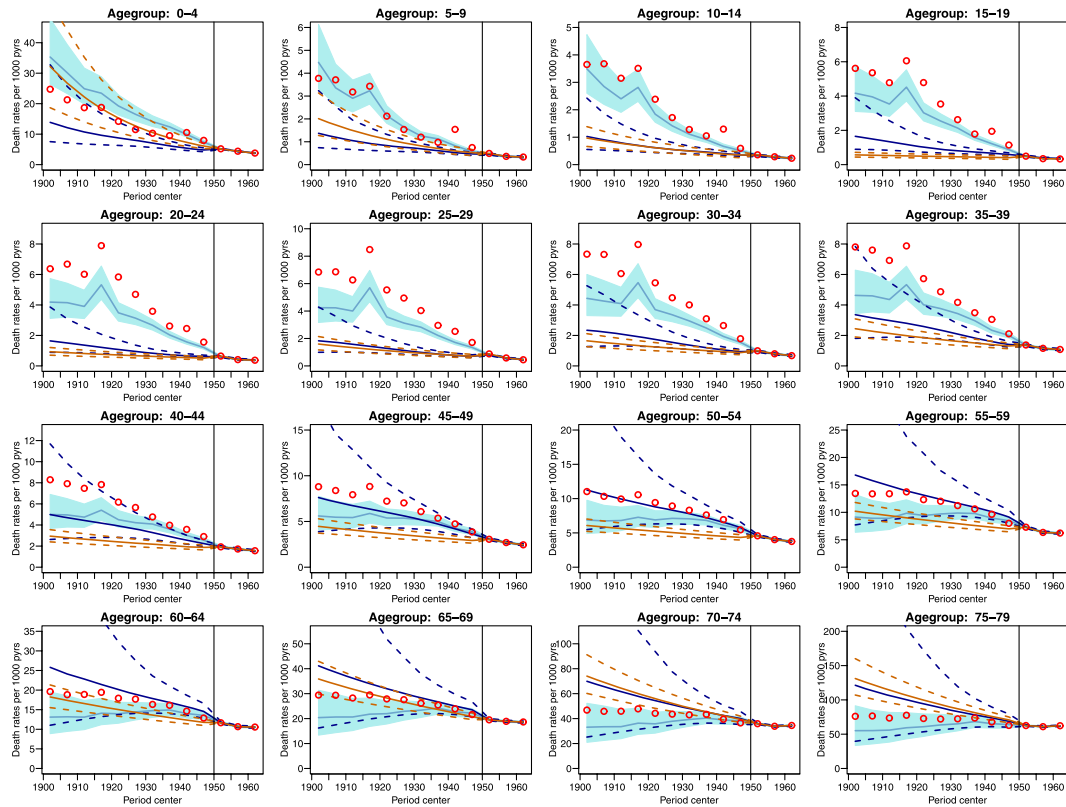
FIG. 6. *Median predicted mortality rates per* 1,000 *person-years* (*pyrs*) *of Norwegian women within* 80% *CI regions for the years* 1900–1949. *Shown are the results from an extended Lee–Carter model* (*orange lines*), *a univariate APC model* (*dark blue lines*) *and a multivariate APC model with age, period, cohort and overdispersion parameters correlated across regions* (*light blue shaded*). *In addition, the true mortality rates for Norwegian women* (*red ○*) *are shown.*
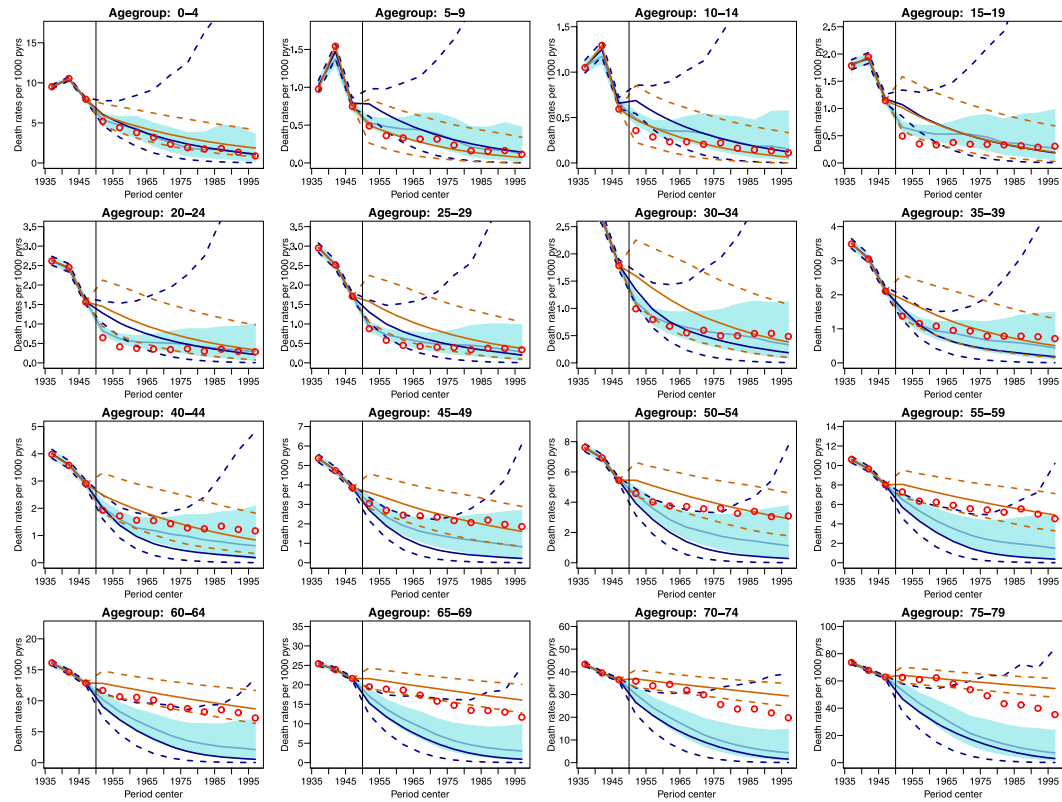
FIG. 7.   *Median predicted mortality rates per* 1,000 *person-years* (*pyrs*) *of Norwegian women within* 80% *CI regions for the years* 1950–1999. *Shown are the results from an extended Lee–Carter model* (*orange lines*), *a univariate APC model* (*dark blue lines*) *and a multivariate APC model with age, period, cohort and overdispersion parameters correlated across regions* (*light blue shaded*). *In addition, the true mortality rates for Norwegian women* (*red ○*) *are shown.*

the Spanish flu did not affect all age groups, this event (also present for Denmark and Sweden) was captured and transferred to the projections due to the correlated overdispersion parameters and not because of the correlated period effects, as one might intuitively guess. Predicting the second half of the 20th century, the projections of the extended Lee–Carter model agree very well with the true observations. In particular, the rates for ages over 60 are well projected, where the cMAPC model tends to underestimate the overall death rate.

Projections for Danish and Swedish women are shown in Figures 8, 9, 10 and 11 in Appendix B. Here, the projections of the correlated approach coincide for all scenarios and all age groups very well with the observed rates. In contrast, the extended Lee–Carter approach tends to underestimate rates for younger age groups when predicting the first half, and to overestimate rates for older age groups when predicting the second half of the 20th century.

5. Discussion. In this paper we proposed the use of correlated smoothing priors and correlated overdispersion parameters for multivariate Bayesian APC approaches analyzing mortality or morbidity rates stratified by age, period, cohort and one further variate. The specification of correlated smoothing priors involves a Kronecker product precision structure for the outcome-specific time effects, that is, age, period and/or cohort effects. We implemented correlated multivariate APC models based on a uniform correlation structure in MCMC and INLA. In the first application we analyzed COPD mortality among males in England and Wales using MCMC and INLA, and compared the results of an ordinary multivariate APC model with those obtained from different correlated model formulations. A comparison of MCMC and INLA showed virtually identical results. As indicated by the log marginal likelihood, the formulation with both correlated overdispersion and correlated stratum-specific period and cohort effects was classified as the best. As shown in the relative risk estimates, the correlated model structure improved the precision of the relative risk estimates especially for younger birth cohorts.

In a second application on overall mortality of Danish, Swedish and Norwegian women in the 20th century, we performed a cross-prediction study. We illustrated the good predictive quality of the correlated approach when imputing missing data units for one country if these units are available for the other countries. As focus was set on projections, which are an estimable function in (multivariate) APC models, we were able to consider the most flexible model with country-specific age, period, cohort and overdispersion parameters that were all correlated across countries. In total, we considered six scenarios treating in turn for a particular country either the first or second half of the data as missing and subsequently predicting the omitted data units. We compared the projections to those obtained from a univariate APC model and a Lee–Carter approach embedded into a quasi-Poisson model using the proper Dawid–Sebastiani scoring rule. Since only the correlated formulation can take advantage from the complete tables of the remaining two countries, it was classified as the best model in five of the six scenarios. Furthermore, we

observed that the predictive quality stayed almost constant when increasing the number of periods to predict, which was not the case for the two univariate approaches. Thus, the correlated approach outperformed both univariate methods in short and long-term projections.

In real life, long-term projections of mortality or disease rates into the future are difficult to make using our proposed approach, since there will be no data from comparable time-series available. For short-term predictions, data for some strata may be already available, while for others they are still missing. Here, the correlation approach will be useful to forecast the missing units.

For the simultaneous projection of several strata, Li and Lee (2005) extended the original log-linear Lee–Carter model. In the simplest extension, they assigned each stratum its own age pattern, while assuming shared age-specific patterns of mortality change and a shared time trend. Future values are then predicted for the shared time trend based on an ARIMA process. Incorporating both shared and stratum-specific parameters, this model seems to be similar in spirit to an uncorrelated multivariate APC model [Riebler and Held (2010)]. In a more complex model, Li and Lee (2005) included an additional stratum-specific bilinear term to allow for differences between the rate of change in mortality in a particular stratum and the rate of change implied by the common bilinear term. However, in contrast to the correlated approach presented here, those extensions cannot take advantage of data units missing in one stratum but present for the remaining. By contrast, our approach could be equally used to impute data for all strata simultaneously, benefitting from the periods where complete data existed. Furthermore, we can use stratum-specific effects for all parameters. Information from the remaining strata is borrowed by incorporating correlation. In a Bayesian setting the inclusion of correlation between parameters is straightforward via the prior distributions, whereas in a frequentist setting this seems to be more complex.

Another interesting field of application is similar in spirit to the inference on collapsed margins, proposed by Byers and Besag (2000). In the context of collapsed margins, complete data are available on several risk factors, but a subsequent analysis indicates that information on an additional variable is relevant. For this variable the numbers of persons at risk are available but not the numbers of cases. Byers and Besag (2000) propose a Bayesian approach to estimate the effect of the variable. In multivariate APC models it might be that multiple data sets are only available for a specific period in time, while, before and/or after this date, data only exist for the conjunction of outcomes. A typical example could be Germany, which was formerly united, then separated and now united again. Using age-specific data on the population sizes from 1990 up to now for East and West Germany separately, it may be possible to project mortality rates for both individual parts, by exploiting the correlation present when they were divided. Thus, the observations for the conjunction of both parts could be separated. However, further investigations are required to explore the applicability.

The proposed methodology can only be applied to data stratified by one further variate. For analyzing mortality rates stratified by more than one further variate, a conditional approach using a multinomial logistic regression model has been proposed [Held and Riebler (2011)]. However, the incorporation of correlation has not yet been considered.

A disadvantage of the proposed methodology might be that it is essentially additive in age, period and cohort, so that interactions between the time dimensions cannot be explicitly modeled. Currie, Durban and Eilers (2004) proposed two-dimensional smoothing to address this problem in the analysis of an individual registry data set. Biatat and Currie (2010) started to extend this work and proposed a model to compare various mortality tables by assuming a common two-dimensional P-spline surface and additional one-dimensional smoothing functions for age and period.

In general, the use of a Kronecker product structure is a promising area for further research, as different correlation structures can easily be combined with different precision matrices. Based on the uniform correlation structure INLA can, by now, correlate a wide range of other GMRF models as components of more general additive regression models. Examples are as follows: nonparametric seasonal models, continuous-time random walks or models with a user specified precision matrix. However, the uniform correlation structure is rather restrictive and may only be plausible for a few outcomes. Future work encompasses the integration of other correlation structures, for example, depending on the distance between units, so that the approach can be extended to the space–time context, for example. Furthermore, we are investigating the use of correlated two-dimensional smoothing priors in INLA to incorporate interactions between time-dimensions into the multivariate APC model.

APPENDIX A: UNIFORM CORRELATION STRUCTURE

Let $\mathbf{C}$ be an $R \times R$ correlation matrix with uniform correlation structure, so that $\mathbf{C} = (1 - \rho)\mathbf{I} + \rho\mathbf{J}$:

$$\mathbf{C} = \begin{pmatrix} 1 & \rho & \cdots & \rho \\ \rho & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \rho \\ \rho & \cdots & \rho & 1 \end{pmatrix},$$

where ρ is the correlation parameter, $\mathbf{I}$ denotes the $R \times R$ identity matrix and $\mathbf{J}$ an $R \times R$ matrix of ones. Then the inverse $\mathbf{C}^{-1}$ is given by

$$\mathbf{C}^{-1} = \begin{pmatrix} a & b & \cdots & b \\ b & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & b \\ b & \cdots & b & a \end{pmatrix}$$

with

$$a = -\frac{(R-2) \cdot \rho + 1}{(\rho - 1)\{(R-1) \cdot \rho + 1\}},$$

$$b = \frac{\rho}{(\rho - 1)\{(R-1) \cdot \rho + 1\}}.$$

PROOF. If $\mathbf{C}^{-1}\mathbf{C} = \mathbf{I}$, then $\mathbf{C}^{-1}$ is the inverse of $\mathbf{C}$. For the diagonal elements of $\mathbf{C}^{-1}\mathbf{C}$ it follows that

$$(\mathbf{C}^{-1}\mathbf{C})_{(i,i)} = a + (R-1) \cdot b \cdot \rho$$

$$= \frac{-(R-2) \cdot \rho - 1 + (R-1) \cdot \rho^2}{(\rho - 1)\{(R-1) \cdot \rho + 1\}}$$

$$= \frac{-R\rho + 2\rho - 1 + R\rho^2 - \rho^2}{R\rho^2 - \rho^2 - R\rho + \rho + \rho - 1} = 1$$

for all $i = 1, \ldots, R$. For the nondiagonal elements, that is, $i \neq j$, we get

$$(\mathbf{C}^{-1}\mathbf{C})_{(i,j)} = a \cdot \rho + b + (R-2) \cdot b \cdot \rho$$

$$= \frac{\{-(R-2) \cdot \rho - 1\}\rho + \rho + (R-2) \cdot \rho^2}{(\rho - 1)\{(R-1) \cdot \rho + 1\}}$$

$$= \frac{-R\rho^2 + 2\rho^2 - \rho + \rho + R\rho^2 - 2\rho^2}{(\rho - 1)\{(R-1) \cdot \rho + 1\}} = 0. \qquad \square$$

The determinant $|\mathbf{C}^{-1}|$ is given by

$$|\mathbf{C}^{-1}| = |\mathbf{C}|^{-1} = \left[(1 + (R-1)\rho)(1-\rho)^{R-1}\right]^{-1}.$$

PROOF. We show that $|\mathbf{C}| = (1 + (R-1)\rho)(1-\rho)^{R-1}$, as the inverse case follows immediately. Remember that $|\mathbf{C}| = |\mathbf{I} - \rho\mathbf{I} + \rho\mathbf{J}|$. The identity matrix has R times the eigenvalue 1. The matrix $\mathbf{J}$ has once the eigenvalue R and $R - 1$ times the eigenvalue 0. Since both matrices ($\mathbf{I}$ and $\mathbf{J}$) share the same eigenvectors, the eigenvalues for $\mathbf{C}$ are $(1 - \rho + \rho \cdot R)$ and $(1 - \rho)$ with multiplicity $R - 1$, so that the determinant of $\mathbf{C}$, the product of the eigenvalues, is

$$|\mathbf{C}| = (1 - \rho + \rho \cdot R)(1-\rho)^{R-1} = (1 + (R-1)\rho)(1-\rho)^{R-1}. \qquad \square$$

APPENDIX B: PROJECTION FOR DENMARK AND SWEDEN

The median projected death rates per 1,000 person-years together with 80% pointwise prediction intervals for Danish and Swedish women obtained from all three models are shown in Figures 8 and 10 for the first half and Figures 9 and 11 for the second half of the 20th century.
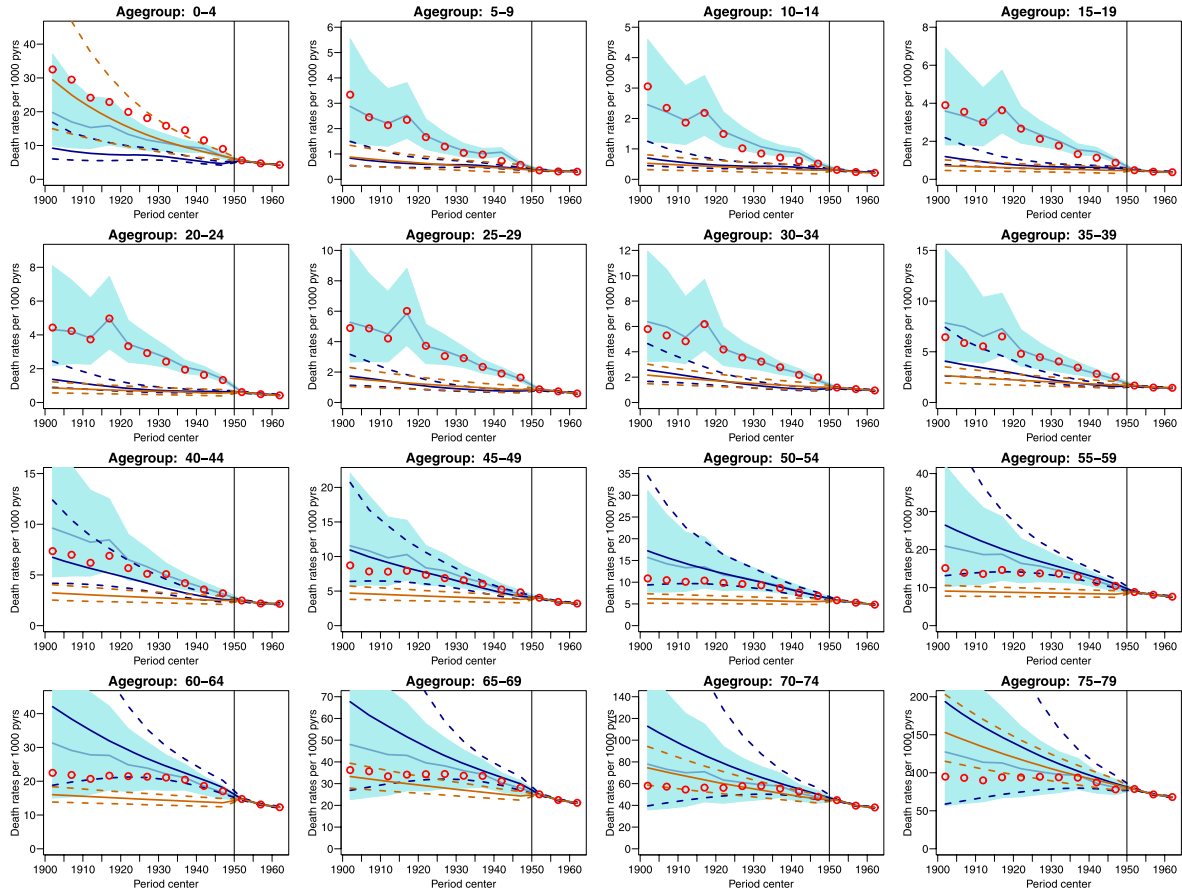
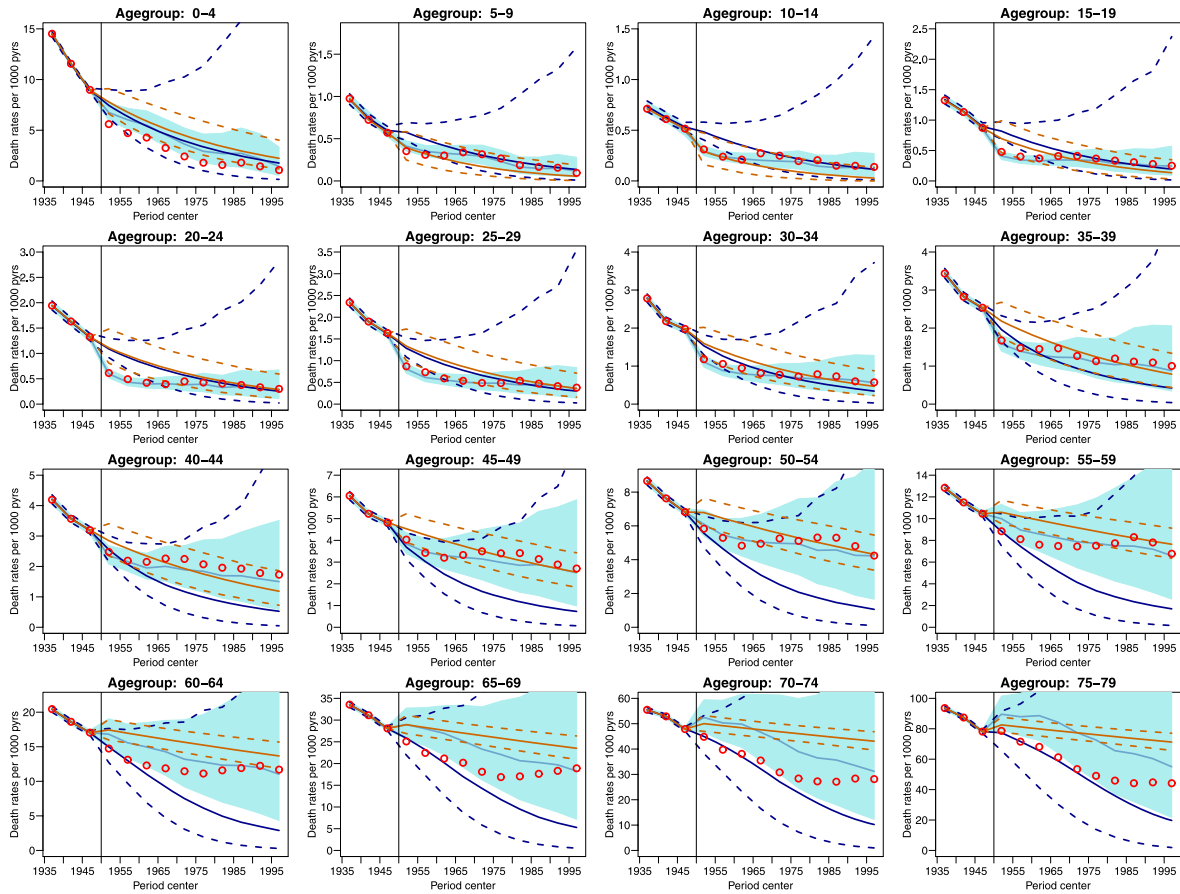FIG. 8. *Same quantities as in Figure* 6 *but for Danish women* (*red* ∘).

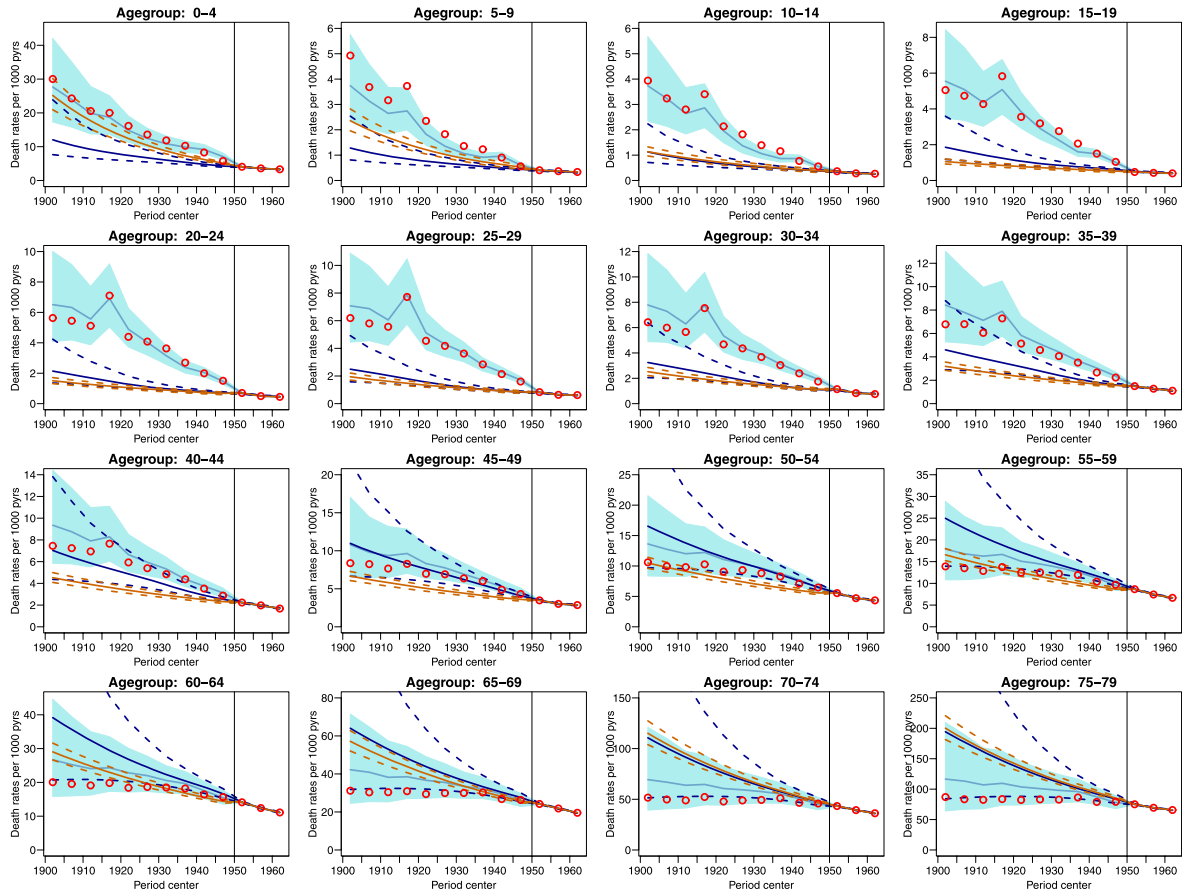FIG. 9. *Same quantities as in Figure 7 but for Danish women (red ○).*

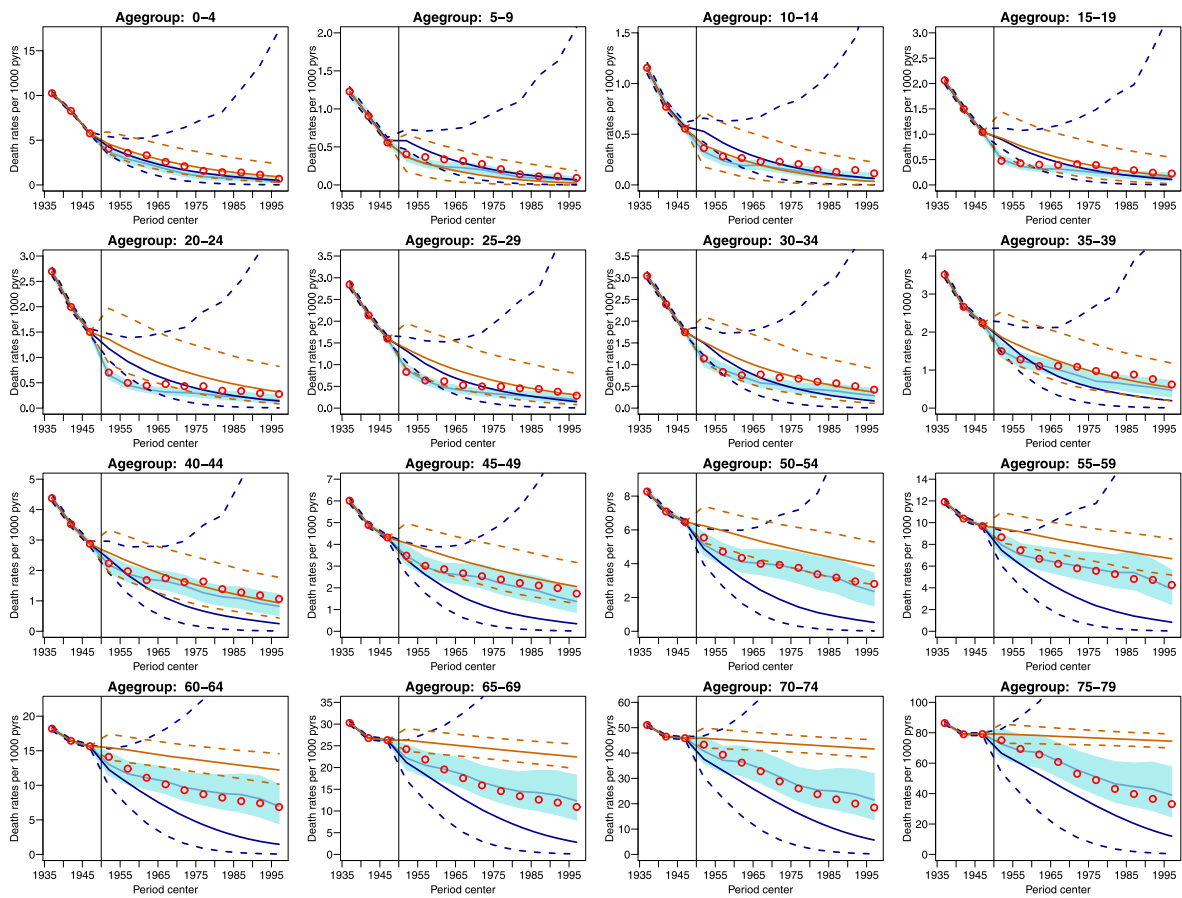FIG. 10.   *Same quantities as in Figure 6 but for Swedish women (red ○).*

FIG. 11. *Same quantities as in Figure 7 but for Swedish women* (*red* ∘).

Acknowledgments.   We would like to thank the Editor and referees for their helpful comments and suggestions that improved the paper.

SUPPLEMENTARY MATERIAL

Code repository for the cross-prediction study of overall mortality of Scandinavian women (DOI: 10.1214/11-AOAS498SUPP; .zip). This repository archives the data, R-code and results for the cross-prediction study of overall mortality of Scandinavian women presented in Section 4.2. In particular, it contains code to make Table 1 and Figures 5–11.

REFERENCES

ANDREASEN, V., VIBOUD, C. and SIMONSEN, L. (2008). Epidemiologic characterization of the 1918 influenza pandemic summer wave in Copenhagen: Implications for pandemic control strategies. *J. Infect. Dis.* **197** 270–278.

ARMITAGE, P. (1966). The chi-square test for heterogeneity of proportions, after adjustment for stratification. *J. Roy. Statist. Soc. Ser. B* **28** 150–163. MR0196851

BAKER, A. and BRAY, I. (2005). Bayesian projections: What are the effects of excluding data from younger age groups? *Am. J. Epidemiol.* **162** 798–805.

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

BIATAT, V. D. and CURRIE, I. D. (2010). Joint models for classification and comparison of mortality in different countries. In 25*th International Workshop on Statistical Modelling* (A. W. Bowman, ed.) 89–94. Univ. Glasgow, UK.

BILLINGSLEY, P. (1986). *Probability and Measure*, 2nd ed. Wiley, New York. MR0830424

BOOTH, H. (2006). Demographic forecasting: 1980 to 2005 in review. *International Journal of Forecasting* **22** 547–581.

BOOTH, H., MAINDONALD, J. and SMITH, L. (2002). Applying Lee–Carter under conditions of variable mortality decline. *Popul. Stud.* (*Camb.*) **56** 325–336.

BOOTH, H., HYNDMAN, R. J., TICKLE, L. and DE JONG, P. (2006). Lee–Carter mortality forecasting: A multi-country comparison of variants and extensions. *Demographic Research* **15** 289–310.

BOUCHARDY, C., LUTZ, J.-M. and KÜHNI, C. (2011). *Krebs in der Schweiz*: *Stand und Entwicklung von* 1983 *bis* 2007. BFS, NICER, SKKR, Neuchâtel.

BRAY, I. (2002). Application of Markov chain Monte Carlo methods to projecting cancer incidence and mortality. *J. Roy. Statist. Soc. Ser. C* **51** 151–164. MR1900163

BRAY, I., BRENNAN, P. and BOFFETTA, P. (2001). Recent trends and future projections of lymphoid neoplasms—a Bayesian age-period-cohort analysis. *Cancer Causes and Control* **12** 813–820.

BRILLINGER, D. R. (1986). The natural variability of vital rates and associated statistics. *Biometrics* **42** 693–734. MR0872958

BROUHNS, N., DENUIT, M. and VERMUNT, J. K. (2002). A Poisson log-bilinear regression approach to the construction of projected lifetables. *Insurance Math. Econom.* **31** 373–393. MR1945540

BUTT, Z. and HABERMAN, S. (2009). ilc: A collection of R functions for fitting a class of Lee–Carter mortality models using iterative fitting algorithms. Technical report, Actuarial Research Paper No. 190, City Univ. London, UK.

BYERS, S. and BESAG, J. (2000). Inference on a collapsed margin in disease mapping. *Stat. Med.* **19** 2243–2249.

CARLIN, B. P. and BANERJEE, S. (2003). Hierarchical multivariate CAR models for spatio-temporally correlated survival data. In *Bayesian Statistics*, 7 (*Tenerife*, 2002) (J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman and A. F. M. Smith, eds.) 45–63. Oxford Univ. Press, New York. MR2003166

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

CURRIE, I. D., DURBAN, M. and EILERS, P. H. C. (2004). Smoothing and forecasting mortality rates. *Stat. Model.* **4** 279–298. MR2086492

CZADO, C., GNEITING, T. and HELD, L. (2009). Predictive model assessment for count data. *Biometrics* **65** 1254–1261. MR2756513

DOCKERY, D. W. and POPE, C. A. (1994). Acute respiratory effects of particulate air pollution. *Annu. Rev. Public Health* **15** 107–132.

ESS, S., SAVIDAN, A., FRICK, H., RAGETH, C., VLASTOS, G., LÜTOLF, U. and THÜRLI-MANN, B. (2010). Geographic variation in breast cancer care in Switzerland. *Cancer Epidemiol.* **34** 116–121.

FAHRMEIR, L. and TUTZ, G. (2001). *Multivariate Statistical Modelling Based on Generalized Linear Models*, 2nd ed. Springer, New York. MR1832899

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

FISHER, R. A. (1958). *Statistical Methods for Research Workers*, 13th (rev.) ed. Oliver and Boyd, Edinburgh.

FU, W. J. J. (2000). Ridge estimator in singular design with application to age-period-cohort analysis of disease rates. *Comm. Statist. Theory Methods* **29** 263–278.

GELFAND, A. E. and GHOSH, S. K. (1998). Model choice: A minimum posterior predictive loss approach. *Biometrika* **85** 1–11. MR1627258

GELFAND, A. E. and VOUNATSOU, P. (2003). Proper multivariate conditional autoregressive models for spatial data analysis. *Biostatistics* **4** 11–25.

GNEITING, T. and RAFTERY, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *J. Amer. Statist. Assoc.* **102** 359–378. MR2345548

GRECO, F. P. and TRIVISANO, C. (2009). A multivariate CAR model for improving the estimation of relative risks. *Stat. Med.* **28** 1707–1724. MR2675246

HANSELL, A. L. (2004). The epidemiology of chronic obstructive pulmonary disease in the UK: spatial and temporal variations. Ph.D. thesis, Faculty of Medicine, Univ. London, Imperial College, St Mary's Campus.

HANSELL, A., KNORR-HELD, L., BEST, N., SCHMID, V. and AYLIN, P. (2003). COPD mortality trends 1950–1999 in England and Wales—Did the 1956 Clean Air Act make a detectable difference? *Epidemiology* **14** S55.

HARVEY, A. (1990). *Forecasting*, *Structural Time Series Models and the Kalman Filter*, Reprinted ed. Cambridge Univ. Press, Cambridge.

HELD, L. and RIEBLER, A. (2011). A conditional approach for inference in multivariate age-period-cohort models. *Stat. Methods Med. Res.* To appear. DOI:10.1177/0962280210379761.

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

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

HOLFORD, T. R. (1992). Analysing the temporal effects of age, period and cohort. *Stat. Methods Med. Res.* **1** 317–337.

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

HUMAN MORTALITY DATABASE (2011). Univ. California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at www.mortality.org or www.humanmortality.de.

JACOBSEN, R., VON EULER, M., OSLER, M., LYNGE, E. and KEIDING, N. (2004). Women's death in Scandinavia—what makes Denmark different? *European Journal of Epidemiology* **19** 117–121.

KAZEROUNI, N., ALVERSON, C. J., REDD, S. C., MOTT, J. A. and MANNINO, D. M. (2004). Sex differences in COPD and lung cancer mortality trends—United States, 1968–1999. *J. Women's Health* **13** 17–23.

KNORR-HELD, L. (2000). Bayesian modelling of inseparable space–time variation in disease risk. *Stat. Med*. **19** 2555–2567.

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.

KOLTE, I. V., SKINHØJ, P., KEIDING, N. and LYNGE, E. (2008). The Spanish flu in Denmark. *Scand. J. Infect. Dis*. **40** 538–546.

KONISHI, S. (1985). Normalizing and variance stabilizing transformations for intraclass correlations. *Ann. Inst. Statist. Math*. **37** 87–94. MR0790377

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

LAGAZIO, C., BIGGERI, A. and DREASSI, E. (2003). Age-period-cohort models and disease mapping. *Environmetrics* **14** 475–490.

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

LEHMANN, E. L. (1999). *Elements of Large-Sample Theory*. Springer, New York. MR1663158

LEVI, F., RANDIMBISON, L., TE, V. C., ROLLAND-PORTAL, I., FRANCESCHI, S. and LA VECCHIA, C. (1993). Multiple primary cancers in the Vaud cancer registry, Switzerland, 1974-89. *Br. J. Cancer* **67** 391–395.

LEVI, F., LA VECCHIA, C., RANDIMBISON, L., ERLER, G., TE, V. C. and FRANCESCHI, S. (1998). Incidence, mortality and survival from prostate cancer in Vaud and Neuchâtel, Switzerland, 1974–1994. *Ann. Oncol*. **9** 31–35.

LEVI, F., RANDIMBISON, L., TE, V.-C. and LA VECCHIA, C. (2002). Thyroid cancer in Vaud, Switzerland: An update. *Thyroid* **12** 163–168.

LI, N. and LEE, R. (2005). Coherent mortality forecasts for a group of populations: An extension of the Lee–Carter method. *Demography* **42** 575–594.

LINDLEY, D. (1965). *Introduction to Probability and Statistics from a Bayesian Viewpoint*, *Part* 2, *Inference*. Cambridge Univ. Press, Cambridge.

MARDIA, K. (1988). Multi-dimensional multivariate Gaussian Markov random fields with application to image processing. *J. Multivariate Anal*. **24** 265–284.

NAKAMURA, T. (1986). Bayesian cohort models for general cohort table analyses. *Ann. Inst. Statist. Math*. **38** 353–370.

OGATA, Y., KATSURA, K., KEIDING, N., HOLST, C. and GREEN, A. (2000). Empirical Bayes age-period-cohort analysis of retrospective incidence data. *Scand. J. Stat*. **27** 415–432.

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

PAUL, M., RIEBLER, A., BACHMANN, L. M., RUE, H. and HELD, L. (2010). Bayesian bivariate meta-analysis of diagnostic test studies using integrated nested Laplace approximations. *Stat. Med*. **29** 1325–1339. MR2757228

R DEVELOPMENT CORE TEAM (2010). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

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

RIEBLER, A., HELD, L. and RUE, H. (2011). Supplement to "Estimation and extrapolation of time trends in registry data—Borrowing strength from related populations." DOI:10.1214/11-AOAS498SUPP.

RIEBLER, A., HELD, L., RUE, H. and BOPP, M. (2011). Gender-specific differences and the impact of family integration on time trends in age-stratified Swiss suicide rates. *J. Roy. Statist. Soc. Ser. A*. To appear.

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

RUE, H. and HELD, L. (2005). *Gaussian Markov Random Fields*: *Theory and Applications*. *Monographs on Statistics and Applied Probability* **104**. Chapman and 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

SCHMID, V. and HELD, L. (2004). Bayesian extrapolation of space-time trends in cancer registry data. *Biometrics* **60** 1034–1042. MR2133556

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

SCHRÖDLE, B., HELD, L., RIEBLER, A. and DANUSER, J. (2011). Using integrated nested Laplace approximations for the evaluation of veterinary surveillance data from Switzerland: A case study. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **60** 261–279.

SUNYER, J. (2001). Urban air pollution and chronic obstructive pulmonary disease: A review. *Eur. Respir. J.* **17** 1024–1033.

VERKOOIJHEN, H. M., FIORETTA, G., VLASTOS, G., MORABIA, A., SCHUBERT, H., SAPPINO, A., PELTE, M., SCHAFER, P., KURTZ, J. and BOUCHARDY, C. (2003). Important increase of invasive lobular breast cancer incidence in Geneva, Switzerland. *International Journal of Cancer* **104** 778–781.

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. *Sociological Methodology* **34** 75–110.

A. RIEBLER
L. HELD
DIVISION OF BIOSTATISTICS
ISPM, UNIVERSITY OF ZURICH
HIRSCHENGRABEN 84
CH-8001 ZURICH
SWITZERLAND
E-MAIL: andrea.riebler@ifspm.uzh.ch
          leonhard.held@ifspm.uzh.ch

H. RUE
DEPARTMENT OF MATHEMATICAL SCIENCES
NORWEGIAN UNIVERSITY OF SCIENCE
    AND TECHNOLOGY
N-7491 TRONDHEIM
NORWAY
E-MAIL: havard.rue@math.ntnu.no