Knorr-Held, Rainer:

Prognosis of Lung Cancer Mortality in West Germany: A Case Study in Bayesian Prediction. (REVISED, January 2000)

Sonderforschungsbereich 386, Paper 160 (1999)

Online unter: http://epub.ub.uni-muenchen.de/

Prognosis of Lung Cancer Mortality in West Germany A Case Study in Bayesian Prediction

Leonhard Knorr–Held and Evi Rainer

Institute of Statistics, Ludwig-Maximilians-University Munich

Ludwigstr. 33, 80539 Munich, Germany

Tel: +49 89/2180-3165, FAX: +49 89/2180-5040, Email: leo@stat.uni-muenchen.de

First version: July 1999

Revised: January 2000

Abstract

We apply a generalized Bayesian age-period-cohort (APC) model to a dataset on lung cancer mortality in West Germany, 1952-1996. Our goal is to predict future deaths rates until the year 2010, separately for males and females. Since age and period is not measured on the same grid, we propose a generalized APC model where consecutive cohort parameters represent strongly overlapping birth cohorts. This approach results in a rather large number of parameters, where standard algorithms for statistical inference by Markov chain Monte Carlo methods turn out to be computationally intensive. We propose a more efficient implementation based on ideas of block sampling from the time series literature. We entertain two different formulations, penalizing either first or second differences of age, period and cohort parameters. To assess the predictive quality of both formulations, we first forecast the rates for the period 1987-1996 based on data until 1986. A comparison with the actual observed rates is made based on a predictive deviance criterion. Predictions of lung cancer mortality until 2010 are then reported and a modification of the formulation in order to include information on cigarette consumption is finally described.

1 Introduction

Lung cancer is by far the leading cause of cancer death among males in West Germany with more than 22,000 cases each year. For females, the numbers of deaths are substantially lower, but constantly increasing, currently with more than 7,000 deaths per year. It is of strong public interest to predict the number of lung cancer deaths and the corresponding mortality rates for public health and demographical reasons.

In this paper we analyse a dataset on lung cancer mortality in West Germany, 1952-1996. Our objective is to provide predictions on the lung cancer epidemic for the following 14 years until 2010. Our database records the population size and the number of deaths from lung cancer in West Germany, stratified by age and gender for each year between 1952 and 1996. Age is categorized as under 5 years, 5–9, 10–14, ..., 80–84, and over 84 years, which makes a total of 18 age groups. Figure 1 shows perspective plots of the corresponding rates per 100,000 both for males and females. Note that age and period are not given on the same time grid.

Our modelling framework is a Bayesian version of the well-known age-period-cohort (APC) model as proposed in Berzuini, Clayton and Bernardinelli[2] and Besag, Green, Higdon and Mengersen[3], see also Berzuini and Clayton[1]. For non-Bayesian approaches see Holford[17] and Clayton and Schifflers[7, 8]. Bayesian formulations assume some sort of smoothness of age, period and cohort effects in order to improve estimation and facilitate prediction. In contrast to non-Bayesian approaches, the forecasts do not rely on strong parametric assumptions for future values of cohort and period effects and therefore seem to be particularly well suited for our objective. Furthermore, the models can take into account additional "unstructured" heterogeneity.

However, the usual APC model is not directly applicable to the case where age and

period are not given on the same time grid. The above authors indicate that a missing data formulation could be applied, if at least information on the number of persons at risk is available on the smaller grid for *both* age and period. For example, in our dataset, we would need information on the number of persons at risk, stratified by age groups of width of *one* year. However, such data were not available. We propose an alternative extension that only needs information on the number of persons at risk on the same grid as the number of cases are given. The idea is to introduce cohort parameters which represent strongly *overlapping* birth cohorts. A similar approach has been suggested in Heuer[15] for spline smoothing in APC models.

The extended version of the APC model involves a large number of parameters, especially for the cohort block. The implementation proposed in Besag *et al.*[3] turns out to be rather computationally intensive. We propose a more efficient algorithm, adopting ideas from the literature on Markov chain Monte Carlo simulation in state-space models[6, 13].

This paper is organized as follows. Section 2 reviews APC models and introduces our generalized Bayesian APC model. Section 3 gives some details on implementation and prediction issues. More details on the proposed implementation are given in the Appendix. Section 4 describes the analysis of the West Germany lung cancer data in three steps. The first step is an analysis of the complete data without any prediction. We then forecast the last ten years in order to assess the predictive power of the two model formulations. Finally predictions until the year 2010 are presented. Throughout we analyse males and females separately. We summarize our findings in Section 5 and describe a modification of the APC model in order to include information on cigarette consumption. However, we conclude that in our application the available smoking data are not able to improve our predictions.

2 The model

Let n_{ij} denote the number at risk in age group i ($i = 1, \ldots, I$) and year j ($j = 1, \ldots, J$); in our dataset $I = 18$ and $J = 45$. We assume that the number of deaths y_{ij} from lung cancer, in age group i during year j has a binomial distribution with parameters n_{ij} and π_{ij}, and that the likelihood for the entire data is the corresponding product of binomial terms. This formulation follows immediately if there is no variability in risk among all individuals in category (i, j) but an alternative derivation has been described in Besag *et al.*[3]. Suppose that the rth individual of the n_{ij} has probability π_{ijr} of death by lung cancer. We don't know what that probability is, so we therefore suppose that the π_{ijr} form a *random sample* from an arbitrary distribution indexed by i and j, having expectation π_{ij}. It follows that y_{ij} has a binomial distribution of the required form. The crucial assumption is that the π_{ijr} are independent, otherwise the binomial assumption will be violated and *overdispersion* will be introduced. For a more detailed discussion of the binomial model see also Knorr-Held and Besag[20].

If the age group bands would be of the same width as the period bands, for example years, then the classical APC model (e.g. Clayton and Schifflers[8]) could be adopted which decomposes the log odds $\eta_{ij} = \log\{\pi_{ij}/(1-\pi_{ij})\}$ additively into an overall level μ, age effects θ_i, period effects φ_j and cohort effects ψ_k,

$$\eta_{ij} = \mu + \theta_i + \varphi_j + \psi_k, \tag{1}$$

where the cohort index $k = 1, \ldots, K$ is defined through

$$k = I - i + j \tag{2}$$

with maximum value $K = I - 1 + J$. Note that the cohort parameters do not represent distinct birth cohorts, in fact, each parameter ψ_k corresponds to a birth cohort of the length

of two years, and the birth cohorts defined by two consecutive cohort parameters overlap by one year, see for example Holford[17].

Two types of additional constraints are necessary to assure identifiability. First, restrictions have to be imposed on each block of parameters θ, φ and ψ such as $\sum \theta_i = 0$, $\sum \phi_j = 0$ and $\sum \psi_k = 0$, say. The second redundancy is due to the linear dependence in (2), since, for any value of α, the linear transformations

$$\theta_i \to \theta_i + \alpha \left(i - \frac{I+1}{2} \right), \quad \varphi_j \to \varphi_j - \alpha \left(j - \frac{J+1}{2} \right), \text{ and } \quad \psi_k \to \psi_k + \alpha \left(k - \frac{K+1}{2} \right), \tag{3}$$

which still fulfill $\sum \theta_i = 0$, $\sum \phi_i = 0$ and $\sum \psi_i = 0$, leave η_{ij} (and π_{ij}) unchanged for all possible combinations of i and j. Hence, only nonlinear trends of the age, period or cohort blocks are interpretable, but linear trends are not. The problem is well known and thoroughly discussed in the literature[17, 7, 8, 2]. To assure identifiability in a classical framework, an additional constraint has to be imposed such as $\theta_1 = \theta_2$, say.

Bayesian versions of the APC model typically use a flat prior for μ (i.e. uniform on the whole real line) and assume that first or second differences of age, period or cohort parameters are independent Gaussian random variables. Such formulations fall into the class of dynamic models and have close connections to nonparametric smoothing methods, such as penalized likelihood or discrete spline smoothing, as reviewed in Fahrmeir and Knorr-Held[11]. For example, for the age effects θ, a smoothing prior based on first differences is given by

$$p(\theta|\kappa) \propto \exp\left(-\frac{\kappa}{2} \sum_{i=2}^{I} (\theta_i - \theta_{i-1})^2 \right),$$

where κ is a precision parameter which determines the smoothness of the age effects. This prior corresponds to the directed formulation as a *first-order random walk* (RW1) $\theta_i \sim N(\theta_{i-1}, \kappa^{-1})$, $i = 2, \ldots, I$ with a flat prior for θ_1. A highly dispersed but proper gamma distribution is assigned to the hyperparameter κ, say $\kappa \sim G(a, b)$, and κ is then also estimated

from the data.

Similarly, a prior based on second differences can be written as

$$p(\theta|\kappa) \propto \exp\left(-\frac{\kappa}{2}\sum_{i=3}^{I}(\theta_i - 2\theta_{i-1} + \theta_{i-2})^2\right).$$

The equivalent directed formulation is called a *second-order random walk* (RW2) defined by $\theta_i \sim N(2\theta_{i-1} - \theta_{i-2}, \kappa^{-1})$, $i = 3, \ldots, I$ with independent uniform priors both for θ_1 and θ_2. We will make use of the correspondence between directed and undirected definitions of the smoothing priors in order to ensure an efficient implementation by MCMC. Note that the RW2 model penalizes deviations from a linear trend $\theta_i = 2\theta_{i-1} - \theta_{i-2}$, $i = 3, \ldots, I$ while the RW1 model penalizes deviations from a model where all parameters are constant, i.e. $\theta_1 = \ldots = \theta_I$.

We always use the same type of prior for the period and cohort parameters as for the age effects and denote the corresponding precision parameters by λ (for ϕ) and ν (for ψ). An interesting feature of the RW1 model is that it solves the identifiability problem caused by (3) by imposing a stochastic constraint: the RW1 model will prefer, among all possible transformations of the form (3), the one where the quadratic first differences (weighted by the precision parameters) are minimal. In other words, age, period and cohort effects are kept "as constant as possible". This feature allows us to examine the trends present in the model parameters θ, ϕ and ψ. In contrast, in the RW2 model parameters are fully unidentifiable, because the linear transformation (3) will neither change the likelihood, nor the prior. We do not impose an additional constraint here, because, in a Bayesian setting, it is not crucial to ensure identifiability of latent parameters as long as the quantities we are interested (the π's) are identifiable.

One advantage of the Bayesian approach is that the uncertainty about κ, λ and ν is incorporated in the estimation of θ, ϕ and ψ. Also, the procedure does not require any arbitrary knot selection as in spline smoothing[15]. Furthermore, model (1) can easily be

extended, if necessary. For example, to account for additional "unstructured" heterogeneity which cannot be explained by the age, period or cohort effects, random effect type parameters can be added in (1). The model will then be

$$\eta_{ij} = \mu + \theta_i + \varphi_j + \psi_k + z_{ij} \tag{4}$$

with independent Gaussian variables

$$z_{ij} \quad \sim \quad N(0, \delta^{-1}). \tag{5}$$

In a frequentist setting, this extension would be referred to as overdispersion or extrabinomial variation, but here we prefer to think of the z_{ij}'s as surrogates for unobserved observation-specific covariates. A gamma prior will be assigned to the hyperparameter δ, similarly as to the other hyperparameters.

An inherent problem with the definition of cohorts through formula (2) is that it is not applicable when age and period are not on the same grid. In a classical framework, interpolation methods have been proposed in Schifflers, Smans and Muir[23]. For a Bayesian analysis by MCMC, Berzuini *et al.*[2] as well as Besag *et al.*[3] suggest to use a missing data formulation. However, this requires the knowledge of the number at risk at an annual basis both for age and calendar time. Such data are typically not available. A way out of this difficulty is to use the same model as in (1) or (4), but with a different definition of the cohorts as proposed in Heuer[15]. Suppose age is given in five years intervals whereas period is given on an annual basis, as in our dataset. Cohorts can then be defined by

$$k = 5 \cdot (I - i) + j \tag{6}$$

with maximum value $K = 5 \cdot (I - 1) + J$. In our application, there are $K = 130$ cohort parameters. With the new definition of cohorts, the log-odds η_{ij} in model (1) and (4) are

now invariant with respect to linear transformations of the type

$$\theta_i \to \theta_i + 5 \cdot \alpha \left(i - \frac{I+1}{2} \right), \quad \varphi_j \to \varphi_j - \alpha \left(j - \frac{J+1}{2} \right), \text{ and } \quad \psi_k \to \psi_k + \alpha \left(k - \frac{K+1}{2} \right). \tag{7}$$

The birth cohorts, defined through (6), do now overlap to a larger extent than in the usual APC model. For example, if ψ_1 represents the cohort born between 1862 and 1867 (as it does in our dataset) then ψ_2 will represent the birth cohort between 1863 and 1868. This is not any problem in our Bayesian approach, in fact, it justifies the assumption that consecutive cohort parameters are similar a priori, as they represent to a large extent the same birth cohorts. The implicit assumption of prior independence in an unrestricted Maximum Likelihood approach seems to be not appropriate here and this holds also in the APC model on a regular grid, where birth cohorts still overlap to some extent.

3 Implementation and Prediction

The model described above implies the posterior distribution for all unknown parameters μ, θ, ϕ, ψ, z and hyperparameters κ, λ, ν and δ. To sample from such a high dimensional distribution by MCMC, parametrization and implementation issues are very important in order to ensure a reliable and efficient algorithm. We follow Besag *et al.*[3] and reparametrize the model from z_{ij} to η_{ij} for all values of i and j. As a consequence, full conditional distributions for each block θ, ϕ and ψ are now multivariate Gaussian. Because of the completely confounded model, it is not to be recommended to use univariate update steps for each component of each block. Instead, updates of large blocks should be preferred. Besag *et al.* suggest to use Cholesky decomposition to sample from each block θ, ϕ and ψ. However, this turned out to be computationally intensive as the dimension of some blocks, especially of the cohort block, is quite high. As a valuable alternative, we suggest

to use ideas from the literature on MCMC sampling in state space models (Carter and Kohn[6], Frühwirth-Schnatter[13]) to sample θ, ϕ and ψ. Details are given in the Appendix. Note that our algorithm is of linear computational order, which is in contrast to Cholesky decomposition. In fact, each of the analyses presented here took around ten minutes on a SUN Ultra, while the computing time for the alternative implementation by Cholesky was two orders of magnitude higher.

Updating of all other parameters is similar as described in Besag *et al.*[3] with univariate Metropolis steps for η_{ij} and Gibbs steps for all hyperparameters. Tuning of the Metropolis steps was done in an automatic fashion (prior to the burn-in period) to achieve acceptance rates around 40%, following recommendations of Gelman, Roberts and Gilks[14]. In all analyses presented here, we have chosen highly uninformative values for the hyperpriors of precision parameters κ, λ, ν, namely $G(1, 0.00005)$. For δ we choose a $G(1, 0.005)$ prior, as we expect a bit more variation for the random effects parameters. However, both priors are highly dispersed and do not impose any severe constraints on the hyperparameters. All results are based on samples of 1,980 realizations, collected by saving the current state after every 50th iteration after a burn–in period of 1,000.

Given samples from the posterior distribution, prediction of future death probabilities $\pi_{ij} = 1/(1+\exp(-\eta_{ij}))$, $j = J+1, \ldots, J+T$, is straightforward to implement. Samples from the predictive distributions of ϕ_j, $j = J+1, \ldots, J+T$ and ψ_k, $k = K+1, \ldots, K+T$ are obtained by repeated application of the directed RW1 or RW2 model definitions. Similarly, samples from z_{ij}, $i = 1, \ldots, I$, $j = J+1, \ldots, J+T$ are generated independently from (5). Samples from the predictive distribution of π_{ij} are finally given through (4).

It has recently been suggested[10, 24] to consider the posterior distribution of the (binomial) log-likelihood of the data as a measure of fit, equivalent to examining the posterior

distribution of the *saturated deviance*

$$D = \sum_{i=1}^{I} \sum_{j=1}^{J} d_{ij}^2 \tag{8}$$

with the squared deviance residual

$$d_{ij}^2 = y_{ij} \log\left(\frac{y_{ij}}{n_{ij}\pi_{ij}}\right) + (n_{ij} - y_{ij}) \log\left(\frac{n_{ij} - y_{ij}}{n_{ij}(1 - \pi_{ij})}\right).$$

Note that D is a functional of the unknown model parameters θ, φ, ψ, and z but not of the second-stage hyperparameters. Hence D can be seen as a measure of fit, comparing observed cases y_{ij} with fitted cases $n_{ij}\pi_{ij}$. Spiegelhalter, Best and Carlin[24] show by asymptotic arguments that, if a model is true, the expected posterior saturated deviance is approximately equal to the number of observations (here $I \cdot J = 810$). If not, then D will be larger. For a recent application of the posterior deviance for screening different formulations in the context of time-space modelling of disease rates see Knorr-Held[19].

For a retrospective analysis, the *predictive deviance* D^P, which is a slight modification of (8) with the sum over j from $J+1$ to $J+T$, seems to be a reasonable quantity to assess the predictive ability of a model if data are already available for the periods that are predicted. Furthermore, for a refined analysis we use

$$\begin{aligned} D_i^P &= \sum_{j=J+1}^{J+T} d_{ij}^2 \quad (i = 1, \ldots, I) \text{ and} \\ D_j^P &= \sum_{i=1}^{I} d_{ij}^2 \quad (j = J+1, \ldots, J+T), \end{aligned}$$

the predicted deviance stratified by age group and by period, respectively. We will compare the predictive quality of the RW1 and RW2 model on the basis of the predictive deviance.

We also simulate from the predictive distribution for the number of cases y_{ij} by generating a sample from the binomial observation model (for known n_{ij}) for each sample from the predictive distribution of π_{ij}. This allows us to calculate credible intervals for future number

of cases. In a retrospective analysis, those can be compared with the actual observed number of cases. We use this as a method for model validation, as the percentage of the number of cases that lie within the corresponding credible interval should approximately match the credibility level for a perfectly appropriate model.

4 Application

4.1 First analysis

We have first analysed the whole dataset to make inference about model parameters describing age, period, and cohort effects in order to discern what trends might be present. As outlined in Section 2, this is possible only for the RW1 model, where parameters are identifiable through a stochastic constraint in the prior. Figure 2 displays posterior median estimates of age, period, and cohort parameters (within 80% pointwise credible intervals) from separate analyses of males and females. One should keep in mind that the patterns can be transformed through linear transformations of type (7) without any change in the likelihood, so only nonlinear trends are useful quantities to interpret.

The age effects exhibit a similar pattern for males and females with an increasing slope for younger and a decreasing slope for older age groups. The period effects have less variability with a decreasing slope around 1965, more distinct for females. This drop might be related to a shortage of cigarettes within the second world war, which apparently had a positive effect on lung cancer mortality rates later. Another interesting explanation is that anti-tobacco campaigns in Nazi Germany may have had an effect on smoking prevalence and consumption, see Procter[22]. A lag of 25 years between exposure and lethal development of the disease can be considered as quite realistic. Similarly, there is an increasing slope around 1970, which might correspond to the end of the second world war. After 1975 the

period effects for females are linear while for males they show a steadily decreasing trend. The cohort effects, finally, have more variation than the period effects. For females there is a distinct peak with a sharply decreasing slope for birth cohorts born around 1950. For males there are three peaks, one around 1905, one around 1930, and the last one around 1950. Note that, for both sexes, the uncertainty of the estimates increases a lot for younger birth cohorts born after 1965. There is not much information to estimate those cohort effects, so the prior starts to dominate the pattern with more or less time-constant effects.

The estimates of the hyperparameters κ, λ and ν differ very much in the expected order. For males, the posterior median of κ is 2.58 (4.07), while the estimates of λ and ν are 3,864 (1,382) and 367.3 (353.3), respectively (values for females in brackets). Hence the most important factor for both sexes is age, followed by cohort and period. This is in accordance with the range of the corresponding parameters estimates in Figure 2.

We have also examined convergence and mixing diagnostics. In particular, we have calculated autocorrelations for the 1,980 samples of each of the 193 model parameters, and each of the 810 binomial probabilities π_{ij}. The results are very similar for males and females. For some of the model parameters, there is dependence between the samples with larger autocorrelations for parameters on the edge of the age, period, or cohort range. This is due to the unidentifiability (7) in the likelihood. However, the real interest lies in the probabilities π_{ij}, which are *all* virtually independent for age group 30-34 and larger. Only for the first six age groups (age below 30) there is some dependence, with autocorrelations of lag 10 in the range of 0.05 to 0.75. For the RW2 model, the parameters θ, ϕ and ψ are now fully unidentifiable, so there is a stronger drift among individual model parameters with high autocorrelations. However, samples from π_{ij} remain stable and are again very close to independence for age above 30 with some dependencies only for the very first age groups.

We were then interested if, for both models, the inclusion of the parameters z_{ij} is really

necessary. Based on the samples from the corresponding posterior distribution we have estimated $p_{ij} = P(z_{ij} > 0)$, the posterior probability that each of the z_{ij}'s is greater than zero. In Figure 3 a value of $p_{ij} < 0.05$ is denoted by "$-$", while $p_{ij} > 0.95$ is denoted by "$+$". It can be seen that the random effect parameters z_{ij} appear to be necessary, both for males and females. Extreme values of p_{ij} can be found mostly for the older age groups and birth cohorts which are present over the whole period range. Interestingly, there are more extreme values of p_{ij} in the RW2 model. A possible explanation is that, for a given observation (i, j), the RW2 model implies a stronger dependence structure on the parameters θ_i, ϕ_j and ψ_k. Thus it is more likely that the unstructured term z_{ij} will be substantially different from zero as it has to compensate for an otherwise bad fit of model (4) without z_{ij}.

We have also looked at the model fit, based on samples from the posterior deviance (8). There was a slightly better fit for males than for females (median posterior deviance equal to 925 and 943, respectively, RW1 model) but differences between the two models have been negligible and no preference for a specific model could be made based on the posterior deviance. The reason is probably the inclusion of the z_{ij}'s, which seem to prevent both models from a bad fit. Incidentally, most posterior samples from D are (for both models and both datasets) in the range of 850 to 1,000, so not too far from the asymptotically optimal value of 810.

4.2 Retrospective prediction

To assess the predictive quality of the two models, we have deleted all observations of the last ten years 1987-1996 in both datasets. Based on the data until 1986, we have then predicted those last ten years. This allows us to compare our predictions with the actual observed rates.

Figures 4 gives median predictive deviances D_i^P and D_j^P for 1987-1996. The top rows show

the predictive deviance stratified by period with the expected increase as time progresses. The predictions are nearly everywhere better for the RW2 model, both for males and females. The bottom rows give the predictive deviance stratified by age group. For males, the worst fit can be seen for the age groups between 50 and 69, while for age over 69, the values of the predictive deviance are surprisingly small. A similar pattern can be seen for females. The reason might be, that, for higher age groups, there is already substantial information about the corresponding cohorts in the data so the predictions become better. Note that, both for males and females, the predictions are more accurate for the RW2 model except for females in the last two age groups. Overall the RW2 model gives far better predictions than the RW1 model. This can be seen from the median total predictive deviance D_p which is 3537 (3040) for RW1 and 2114 (2098) for the RW2 model (values for females in brackets).

We have also calculated 50, 80 and 95% pointwise credible intervals for the number of cases y_{ij}, $i = 1, \ldots, 18$, $j = 36, \ldots, 45$, given n_{ij}. The percentages of observed values that lie within the corresponding intervals are reported in Table 1. The values for the RW2 model seem to come reasonably close to the actual credibility levels, both for males and females. For example, 74% of the observed number of cases for females are within our 80% credible regions. The percentages for the RW1 model are always considerably smaller.

For a more detailed comparison, Figure 5 shows the observed and predicted number of male lung cancer deaths (per 100,000) within 80% pointwise credible regions for the four age groups 50-54, 55-59, 60-64 and 65-69, where the prognosis was relatively bad in terms of predictive deviance. From these plots it becomes rather obvious that the RW2 model gives better predictions than the RW1 model. Note, incidentally, that also for the RW1 model, the predictions are not merely constant rates as they incorporate the variation of the cohort parameters.

Figure 6 gives the corresponding plot for females in age groups 70-74, 75-79, 80-84 and

85+. In the last two age groups the RW1 model gives better forecasts than the RW2 model, in accordance with the values of the predictive deviance D_i^P. The reason seems to be that the temporal trend in the rates suddenly flattens out. The RW1 model, which totally ignores any trends in the period and cohort forecasts, is therefore better in prediction. We conclude, that it is not clear a priori that the more ambitious RW2 model is always better in forecasting. Overall, however, we certainly prefer now the RW2 model for our objective.

4.3 Prediction of lung cancer mortality until 2010

We finally present predictions of lung cancer mortality until 2010 based on the RW2 model. Figures 7 and 8 give observed and median predicted number of cases y_{ij} (per 100,000) together with 80% (pointwise) credible intervals for all age groups above 50 years. The observed cases are marked by crosses. The predictions require the knowledge of the population at risk, which is not known in advance. We have used demographical forecasts by the Federal Statistical Office of Germany, Wiesbaden. Note that we do ignore the uncertainty involved in these forecasts, but, since we report the number of cases per 100,000, the impact will be minimal. As an alternative, one could calculate credible intervals for $\pi_{ij} \times 100,000$ which does not require information on the population at risk. Those estimates are very similar for older age groups, for younger age the credible intervals are slightly narrower due to the missing binomial variation.

It is interesting to see that the predictions are not merely straight lines but all have different temporal patterns due to the strong influence of the cohort effects. Furthermore, it can be seen how additional variability is produced in the forecasts as time progresses. For males, we mostly predict decreasing rates, in particular for younger age. However, for females, rates will increase. This is in accordance with known increases in cigarette consumption by women[4].

Finally, Figure 9 give predictions of the absolute number of cases by lung cancer in West Germany. Of course, these predictions now really depend on the accuracy of the demographical projections. It can be seen that although the age-specific rates for males are mostly decreasing, the total number of cases will stay roughly constant because of changes in the demographical structure with more and more older people. For females one can see the expected dramatic increase. Note that the estimates involve a lot of uncertainty, adding up the uncertainty involved in the estimates of the age-specific numbers. However, the temporal trends are obvious and it seems not unlikely that by 2010, every third death by lung cancer in West Germany will be female.

5 Extensions and Conclusions

This paper has demonstrated how Bayesian APC models can be used to predict future disease rates. Our formulation is applicable to data where age and period do not have to be available on the same time grid. We have proposed a specific algorithm for a highly efficient implementation of the necessary MCMC simulation.

In a retrospective analysis we have shown by means of the predictive deviance that the RW2 model gives considerably better forecasts than the alternative RW1 model. Interval predictions for the actual number of observed cases turned out to be sufficiently accurate. We note that other prior formulations for the model parameters seem plausible. For example, it seems reasonable to impose a constraint that the age parameters be monotonic increasing. Also, the local linear trend model (e.g. Fahrmeir and Tutz[12]) might be a valuable, but more complex, alternative to the RW1 or RW2 model. Finally, the structure apparent in Figure 3 seems to suggest that there is some heterogeneity in the variation of the z parameters, with substantial departures from zero only for higher age. Hence, another interesting extension

might be to let the precision parameter δ depend on age group i in a smooth way.

There is of course no doubt that appropriate information on smoking prevalence and consumption would improve our predictions. However, only few detailed data on smoking habits in Germany have so far been published[16]. Yearly estimates of prevalence of cigarette smoking are available from surveys conducted by the "Institute of Demoscopy Allensbach", but only from 1974 onwards. Heuer and Becker[16] use this gender-, age- and calendar-year specific data to extrapolate prevalence rates to previous calendar time periods. This was done only for birth cohorts, on which there is prevalence data in 1974. Hence, more and more back-extrapolated prevalences are missing. Heuer and Becker also report the number of cigarettes manufactured per adult in West Germany, but only for 10 years between 1948 and 1995. There was no information on sex-specific cigarette consumption, nor on the average tar content per cigarette available.

To include smoking data in the APC model, Brown and Kessler[5] replace the period parameters with a regression variable related to the number of cigarettes sold and to the average tar content per cigarette, following a suggestion by Day and Charnay[9]. This approach seems to be justifiable as long as other relevant factors, which could be attributed to the period effects, are time-constant. For example, improvements in therapy might affect the period effects as well, but are probably not so relevant for lung cancer. Also it assumes that the age group-specific proportions of consumption do not change over time, which might be questionable.

Thus model (4) could be modified to

$$\eta_{ij} = \mu + \theta_i + \beta \cdot x_{j-L} + \psi_k + z_{ij} \tag{9}$$

where the period parameter φ_j is replaced by a regression on x_j which denotes some measure of the (male or female) population's exposure to cigarettes during the jth calendar period. Of course, changes in carcinogenic exposure are not reflected immediately in changing cancer

mortality, so the covariate x in model (9) is lagged by L years. Peace[21] found a 21-year lag when correlating overall lung cancer mortality with cigarette tobacco sales in England and Wales during 1880-1983. Brown and Kessler[5] analyse data from the US and suggest an even larger lag of 24 years based on minimizing a deviance criterion, obtained from a Maximum Likelihood analysis of model (9) (without the z_{ij}'s) for different values of L.

Based on the (interpolated) number of cigarettes manufactured per adult in Germany, we have calculated a sex-specific proxy x_j for smoking consumption in year j, the number of cigarettes sold per capita (age ≥ 15) from 1955 to 1994. To do so, we have assumed that, on average, a male smoker consumes as many cigarettes as a female smoker. This assumption is not so crucial as long as at least the male-female proportion does not vary over time. If it does vary, however, then our estimates, displayed in Figure 10, are certainly questionable. The calculations are based on the backcalculated (age-adjusted) smoking prevalence estimates by Heuer and Becker, ignoring the missing values in higher age groups. For example, for 1955, there was missing information on smoking prevalence for age above 60 years. It was not attempted to calculate the numbers of cigarettes before 1955 because even more prevalence information would be missing and, in addition, the available (backcalculated) prevalence estimates are less reliable.

We have then tried to use the covariate x_j (divided by 1,000) in model (9). However, the assumption of a time-constant effect β is certainly unrealistic, because our estimates ignore the decreasing tar-content of cigarettes (for which no data were available). We therefore use the model

$$\eta_{ij} = \mu + \theta_i + \beta_j \cdot x_{j-L} + \psi_k + z_{ij} \tag{10}$$

where β is replaced by a smoothly *time-varying effect* β_j to which a RW2 smoothing prior is assigned, just as the priors used for the age and cohort parameters. The original program was modified to include the covariate effects β_j instead of the period parameters φ_j and was

also used to predict future number of cases.

However, even by assuming a moderate time lag of $L = 20$ years, only half of the original dataset (from 1975 onwards) could enter in the extended analysis. As a consequence, credible intervals and predictive intervals are considerably wider than in the analysis without information on cigarette consumption. The point predictions are similar to those obtained without incorporation of smoking consumption, but the large predictive intervals did not allow to draw any useful conclusions about future temporal trends.

An alternative approach might be to replace the cohort parameters ψ by smoking prevalence when young. This is a suggestion by Day and Charnay[9], who interpret the cohort parameters as reflecting the number and type of cigarettes a cohort becomes habituated to smoking when young. Figure 11 gives estimated smoking prevalence in age group 20-24 by birth cohort, based on the back-projections by Heuer and Becker[16]. Interestingly, the interpretation by Day and Charnay is supported by comparing Figure 11 to the estimated cohort effects in our RW1 model (Figure 2). For example, the peak prevalences for males at 1905, 1930 and 1955 match those of the cohort effects quite accurately. The patterns for females are also quite similar. This indicates that the cohort effects in the original APC model reconstruct the proposed surrogate for cohort effects (smoking prevalence when young) quite precisely. We therefore do not expect that a replacement of the cohort parameters by this surrogate will improve our predictions, especially because, similar to the analysis with model (10), only a fraction of the original dataset could be used in such an analysis.

If more precise information on smoking behaviour is available, then more accurate projections are to be expected. In particular age- and period-specific data on cigarette, or even better, tar *consumption* (not just prevalence), would be extremely valuable. However, it is not immediately clear how the APC model should be modified to incorporate such information properly. One approach might be to calculate both period- and cohort-specific

surrogates for smoking consumption from such data and to replace not only the period effects, as in model (10), but also the cohort effects by regression terms with smooth effects. As an alternative, Holford, Zhang, Zheng and McKay[18] try to incorporate more refined information on the proportion of never, current and ex-smokers in an analysis of lung cancer incidence in Connecticut. They use a formulation for the incidence rate where an APC model is combined multiplicatively with a term depending on the proportions of the three subgroups and on (age- and period-specific) mean duration of smoking exposure among current and ex-smokers. We plan to consider such extensions once we have more accurate information on smoking consumption.

Nevertheless, we think that, even without incorporating the relevant smoking data, our model is realistic and is able to produce quite reliable estimates of future lung cancer mortality. The structured model parameters in combination with smoothing priors determine the direction of the projections, while additional variability is added through the unstructured terms z_{ij}. Note that the implementation of such a complex formulation would have been impossible without the computational advances during the past decade, particularly the adoption of Markov chain Monte Carlo as a standard Bayesian inference machine.

Acknowledgements

This project was supported by the German Science Foundation (DFG), SFB 386. The authors express thanks to Nikolaus Becker for access to the dataset, to Carsten Heuer for helpful comments on the initial project, to the Federal Statistical Office of Germany in Wiesbaden for the demographical forecasts, and to Volker Schmid for implementing model (10) with the smoking covariate. Insightful comments by the editor and two referees have helped us to improve the manuscript.

References

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

[2] Berzuini, C., Clayton, D. and Bernardinelli, L. 'Bayesian inference on the Lexis diagram', *Bulletin of the International Statistical Institute*, **50**, 149–164 (1993).

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

[4] Brenner, H. 'A birth cohort analysis of the smoking epidemic in West Germany', *Journal of Epidemiology and Community Health*, **47**, 54–58 (1993).

[5] Brown, C. C. and Kessler, L. G. 'Projections of lung cancer mortality in the United States: 1985-2025', *Journal of the National Cancer Institute*, **80**, 43–51 (1988).

[6] Carter, C. K. and Kohn, R. 'On Gibbs sampling for state space models', *Biometrika* **81**, 541–553 (1994).

[7] Clayton, D. G. and Schifflers, E. 'Models for temporal variation in cancer rates. I: Age-period and age-cohort models', *Statistics in Medicine*, **6**, 449–467 (1987).

[8] Clayton, D. G. and Schifflers, E. 'Models for temporal variation in cancer rates. II: Age-period-cohort models', *Statistics in Medicine*, **6**, 469–481 (1987).

[9] Day, N. E. and Charnay, B. 'Time trends, cohort effects, and aging as influence on cancer incidence'. In: K. Magnus (ed.), *Trends in Cancer Incidence (Causes and Practical Implications)*, Washington, DC: Hemisphere, 1982, 51-65.

[10] Dempster, A. P. 'The direct use of likelihood for significance testing', *Statistics and Computing*, **7**, 247-252 (1997). Reprinted from a Proceedings contribution from 1974.

[11] Fahrmeir, L. and Knorr-Held, L. 'Dynamic and semiparametric models', In: M. Schimek (ed.), *Smoothing and Regression: Approaches, Computation and Application*, New York: John Wiley & Sons, to appear, 2000.

[12] Fahrmeir, L. and Tutz, G. *Multivariate statistical modelling based on generalized linear models*. Springer–Verlag, New York, 1994.

[13] Frühwirth–Schnatter, S. 'Data augmentation and dynamic linear models', *J. Time Ser. Anal.*, **15**, 183–202 (1994).

[14] Gelman, A., Roberts, G. O. and Gilks, W. R. 'Efficient Metropolis jumping rules', In: J. Bernardo, J. Berger, A. P. Dawid and A. F. M. Smith (eds.), *Bayesian Statistics 5*, Oxford University Press, 599–607 (1995).

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

[16] Heuer, C. and Becker, N. 'Smoking prevalence and lung cancer mortality in Germany', *Journal of Epidemiology and Biostatistics* **4**, 45–52 (1999).

[17] Holford, T. R. 'The estimation of age, period and cohort effects for vital rates', *Biometrics* **39**, 311–324 (1983).

[18] Holford, T. R., Zhong, Z., Zheng, T. and McKay, L. A. 'A model for the effect of cigarette smoking on lung cancer incidence in Connecticut', *Statistics in Medicine* **15**, 565–580 (1996).

[19] Knorr-Held, L. 'Bayesian modelling of inseparable space-time variation in disease risk', *Statistics in Medicine*, forthcoming.

[20] Knorr-Held, L. and Besag, J. 'Modelling risk from a disease in time and space', *Statistics in Medicine* **17**, 2045–2060 (1998).

[21] Peace, L. R. 'A time correlation between cigarette smoking and lung cancer', *The Statistician* **34**, 371–381 (1985).

[22] 'The anti-tobacco campaign of the Nazis: A little known aspect of public health in Germany, 1933-45.' *British Medical Journal*, **313**, 1450-1453, 1996.

[23] Schifflers, E., Smans, M. and Muir, C. S. 'Birth cohort analysis using irregular cross-sectional data: a technical note', *Statistics in Medicine* **4**, 63–75 (1985).

[24] Spiegelhalter, D. J., Best, N. G. and Carlin, B. P. 'Bayesian deviance, the effective number of parameters, and the comparison of arbitrarily complex models', Technical Report, MRC Biostatistics Unit, Cambridge, UK (1998).

6 Appendix: Efficient sampling from age, period and cohort blocks

The full conditionals for the age, period and cohort effects all have a multivariate Gaussian distribution. These distributions can be rewritten in a linear Gaussian state space model as will be shown below. This allows the implementation of an efficient "forward filtering-backward sampling" algorithm by Carter and Kohn[6] for sampling from each of the full conditionals. We outline the procedure only for the full conditional for the age effects θ. Similarly, the method has been applied to period and cohort parameters.

For a prior based on first differences (RW1 model), the full conditional for θ is given by

$$p(\theta|\cdot) \propto \exp\left[-\frac{1}{2}\left(\kappa\sum_{i=2}^{I}(\theta_i-\theta_{i-1})^2+\delta J\sum_{i}^{I}(\bar{\xi}_{i\cdot}-\theta_i)^2\right)\right]$$

with $\bar{\xi}_{i\cdot} = \frac{1}{J}\sum_{j=1}^{J}\xi_{ij}$ where $\xi_{ij} = \eta_{ij} - \phi_j - \psi_k$ and k is given by (6). This distribution is equivalent to the distribution of latent parameters θ in a state space model (see for example Fahrmeir and Tutz[12])

$$\theta_i \sim N(\theta_{i-1}, \kappa^{-1}) \qquad (i = 2, \ldots, I)$$

$$\bar{\xi}_{i\cdot} \sim N(\theta_i, (\delta J)^{-1}) \qquad (i = 1, \ldots, I)$$

with a prior for θ_1 uniform on the whole real line. Note that $\bar{\xi}_{i\cdot}$ takes over the role of the observations in the usual state space model. A sample θ^* from $p(\theta|\cdot)$ can now be generated as follows:

1. Run the Kalman Filter to calculate mean θ_i and variance v_i of filtered distributions $p(\theta_i|\bar{\xi}_{1\cdot}, \ldots, \bar{\xi}_{i\cdot})$:

   $v_1 := (\delta J)^{-1}$

   $\theta_1 := \bar{\xi}_{1\cdot}$

   `for` $i := 2$ `to` I `do`

   $v_i = ((v_{i-1} + \kappa^{-1})^{-1} + \delta J)^{-1}$

   $\theta_i = \theta_{i-1} + v_i \cdot \delta J(\bar{\xi}_{i\cdot} - \theta_{i-1})$

2. Generate θ_I^* from $N(\theta_I, V_I)$.

3. Use backward sampling to generate $\theta_{I-1}^*, \ldots, \theta_1^*$:

for $i := I-1$ to 1 do

$$\tilde{v}_i = (v_i^{-1} + \kappa)^{-1}$$

$$\tilde{\theta}_i = \theta_i + \tilde{v}_i \cdot \kappa(\theta^*_{i+1} - \theta_i)$$

$$\theta^*_i \sim N(\tilde{\theta}_i, \tilde{v}_i)$$

For a prior based on second differences (RW2 model), the full conditional for θ is

$$p(\theta|\cdot) \propto \exp\left[-\frac{1}{2}\left(\kappa\sum_{i=3}^{I}(\theta_i - 2\theta_{i-1} + \theta_{i-2})^2 + \delta J\sum_{i}^{I}(\bar{\xi}_{i\cdot} - \theta_i)^2\right)\right]$$

with $\bar{\xi}_{i\cdot}$ defined as above. This distribution is equivalent to the distribution of latent parameters θ in a dynamic model

$$\theta_i \sim N(2\theta_{i-1} - \theta_{i-2}, \kappa^{-1}) \qquad (i = 3, \ldots, I)$$

$$\bar{\xi}_{i\cdot} \sim N(\theta_i, (\delta J)^{-1}) \qquad (i = 1, \ldots, I)$$

with priors for θ_1 and θ_2 uniform on the whole real line. A sample θ^* from $p(\theta|\cdot)$ can be generated in similar lines as above. However, to apply the Kalman Filer and the backward sampling algorithm, the dynamic model has to be written in a state space form:

$$\begin{pmatrix} \theta_i \\ \theta_{i-1} \end{pmatrix} = \begin{pmatrix} 2 & -1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} \theta_{i-1} \\ \theta_{i-2} \end{pmatrix} + u_i \qquad u_i \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \kappa^{-1} & 0 \\ 0 & 0 \end{pmatrix}\right)$$

$$\bar{\xi}_{i\cdot} = \begin{pmatrix} 1 & 0 \end{pmatrix}\begin{pmatrix} \theta_i \\ \theta_{i-1} \end{pmatrix} + v_i \qquad v_i \sim N\left(0, (\delta J)^{-1}\right) \qquad (i = 3, \ldots, I).$$

We have calculated a scalar form of the filter and backward sampling algorithm by means of the computer package MAPLE, as to avoid time consuming matrix manipulations. However, the update steps are more complicated than for the prior based on first differences, so we do not report details here.

credibility level	males		females	
	RW1	RW2	RW1	RW2
50%	41%	48%	39%	51%
80%	59%	68%	53%	74%
95%	78%	85%	70%	90%

Table 1: The percentage of observed cases that lie within predictive credible intervals.

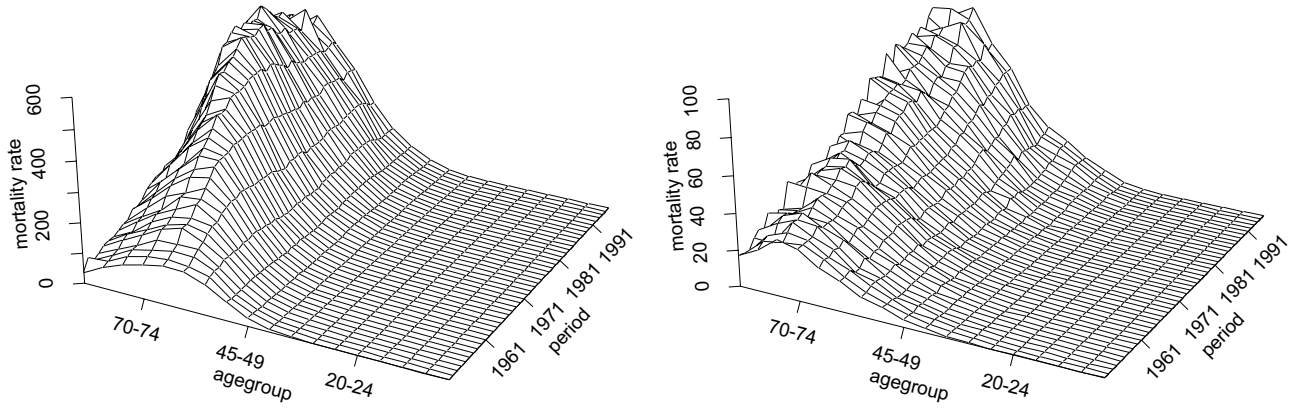

Figure 1: Perspective plots of lung cancer mortality rates per 100,000 in West Germany for males (left) and females (right).

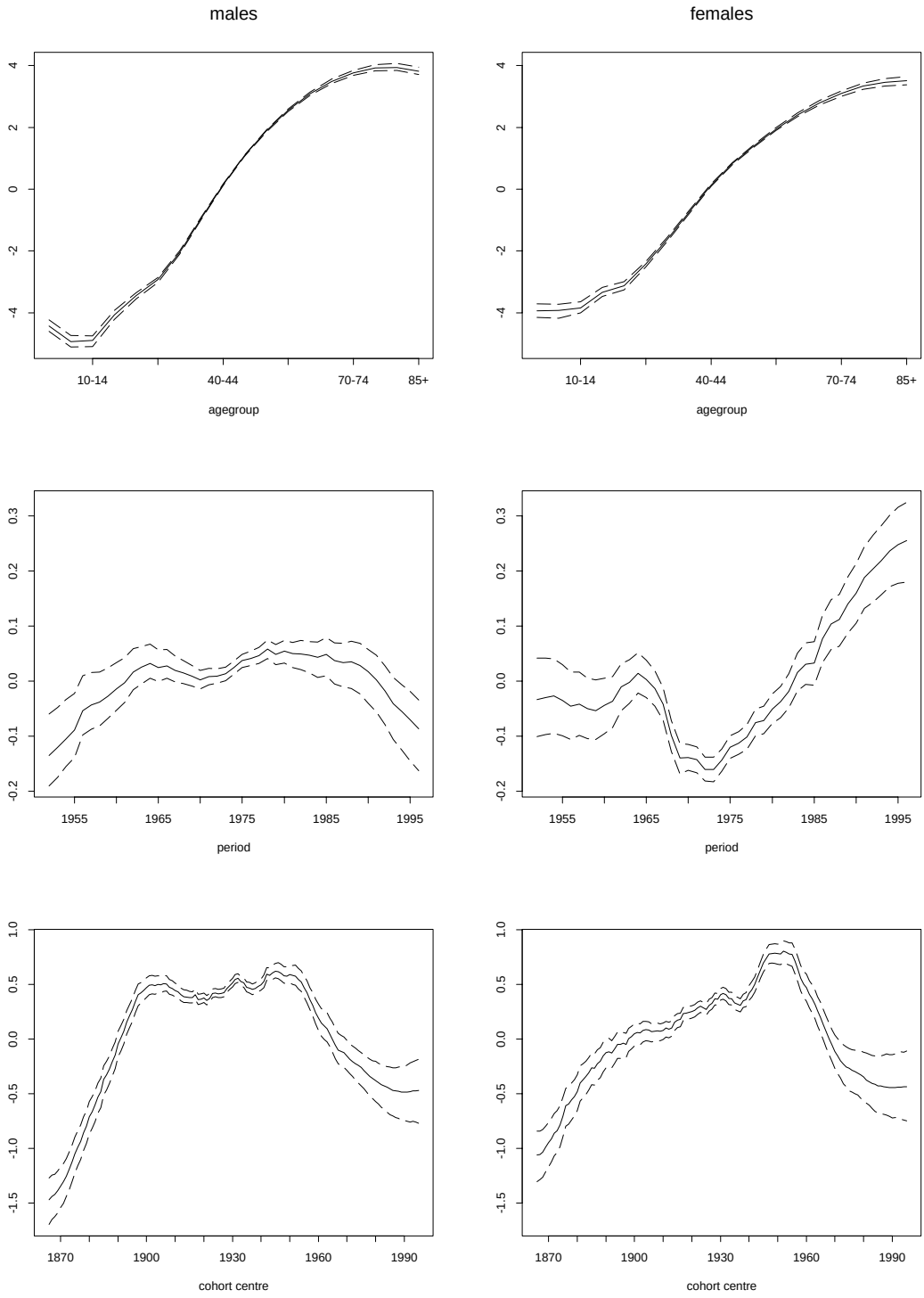

Figure 2: Posterior median estimates within 80% pointwise credible regions of age, period, and cohort parameters. RW1 model.

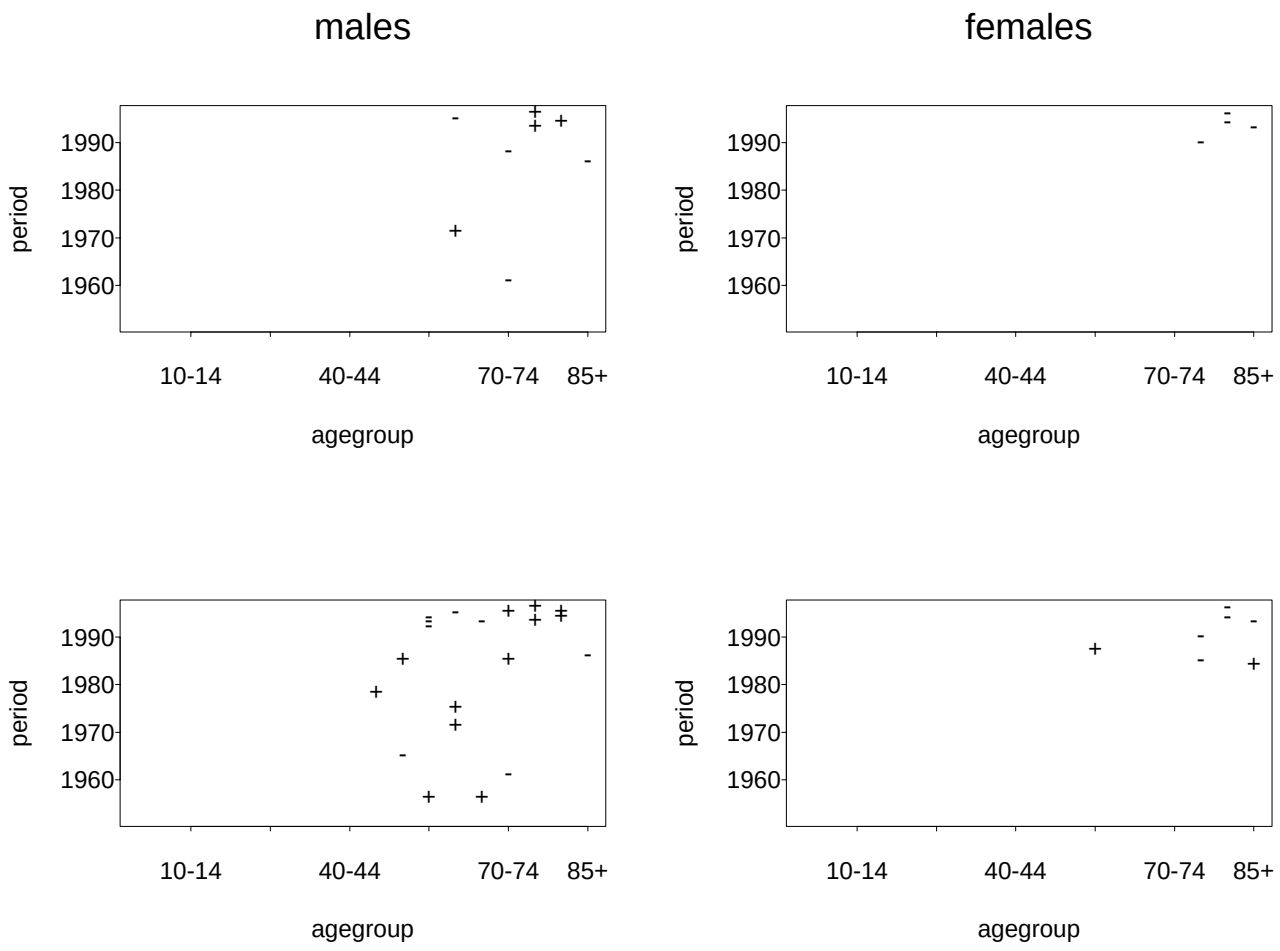

Figure 3: Classification of posterior probabilities. ("+" corresponds to $P(z_{ij} > 0) > 0.95$, "−" to $P(z_{ij} > 0) < 0.05$). Top row: RW1 model. Bottom row: RW2 model.

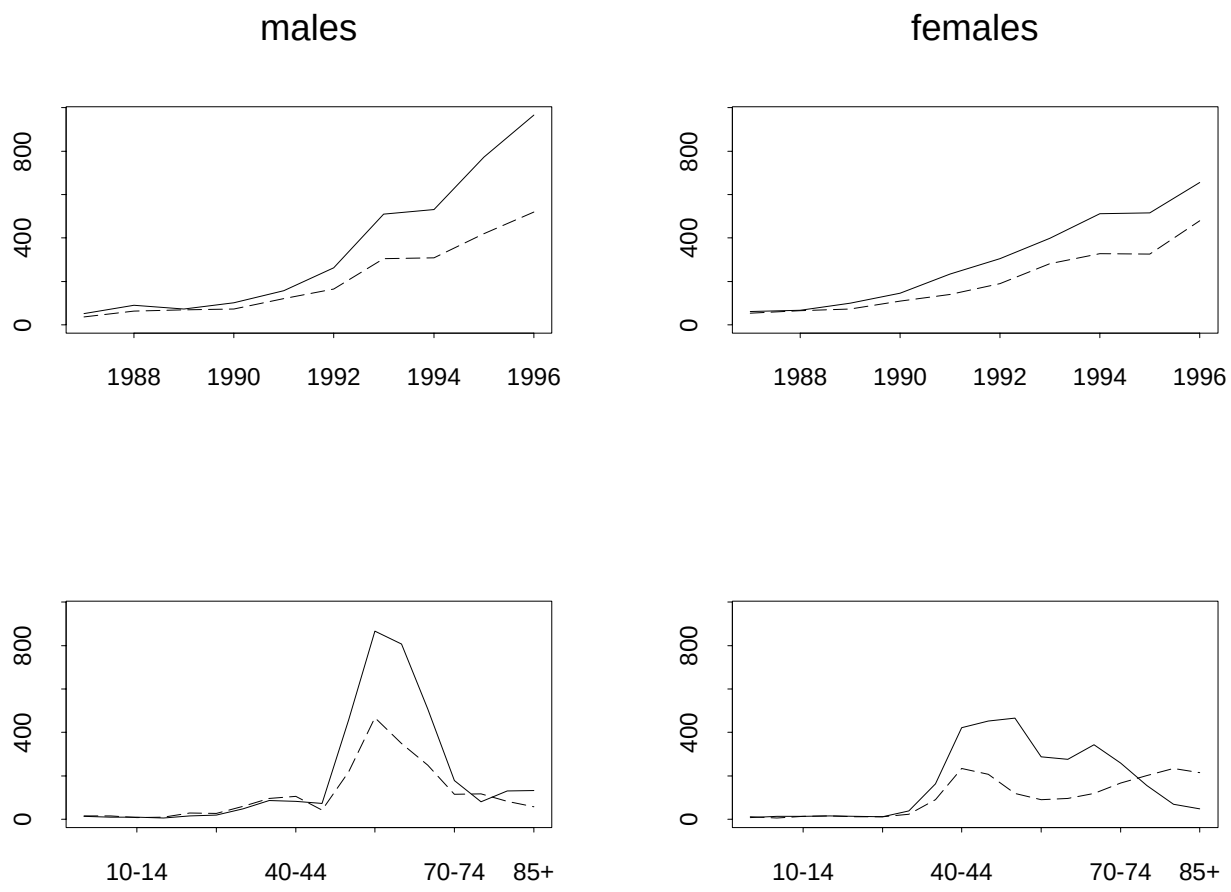

Figure 4: Median predictive deviance by period (top row) and by age group (bottom row). Solid lines: RW1 model. Dashed lines: RW2 model.

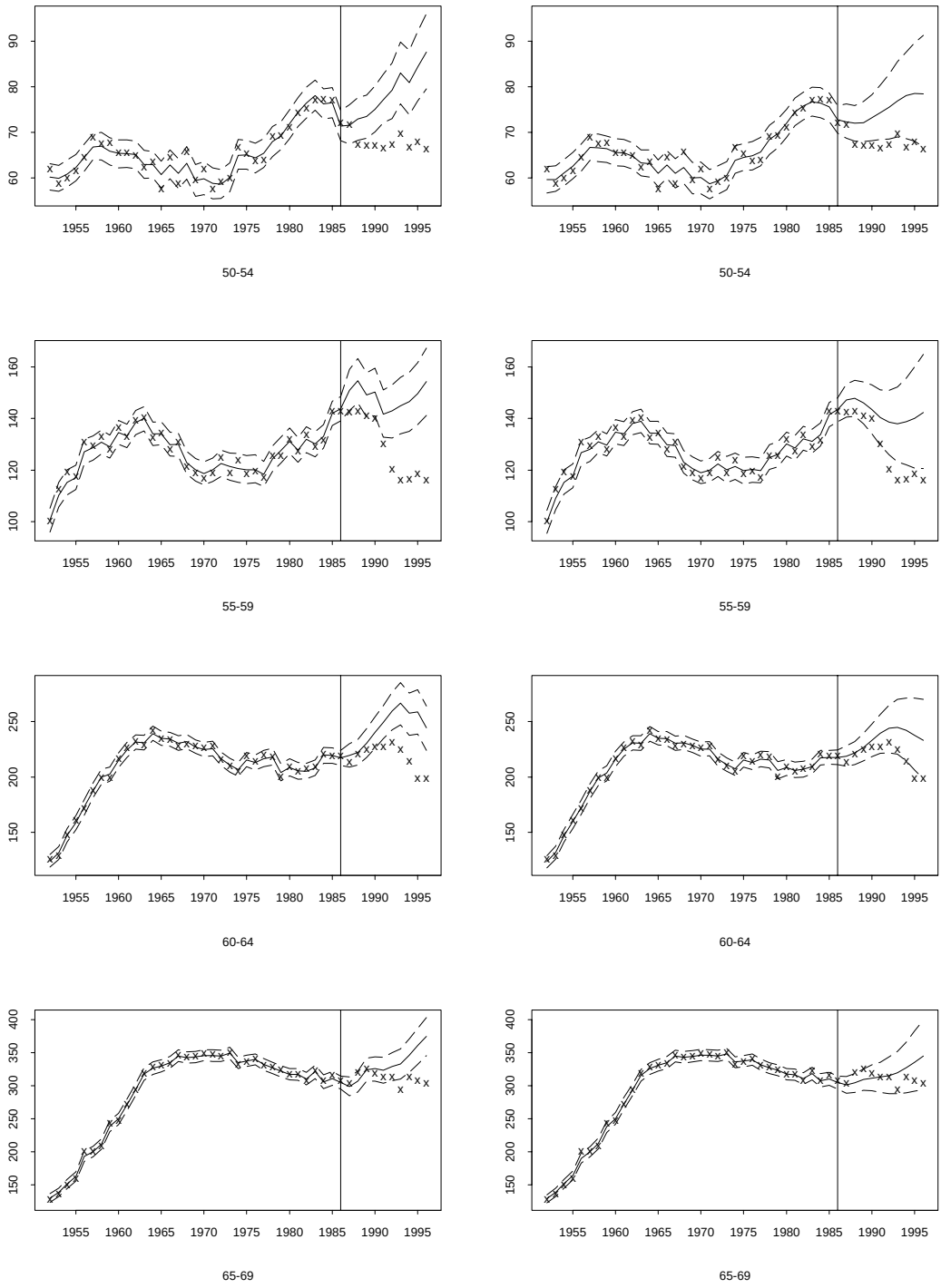

Figure 5: Observed and predicted number of cases within 80% credible regions for males in age groups 50-54, 55-59, 60-64 and 65-69. Left panel: RW1 model. Right panel: RW2 model.

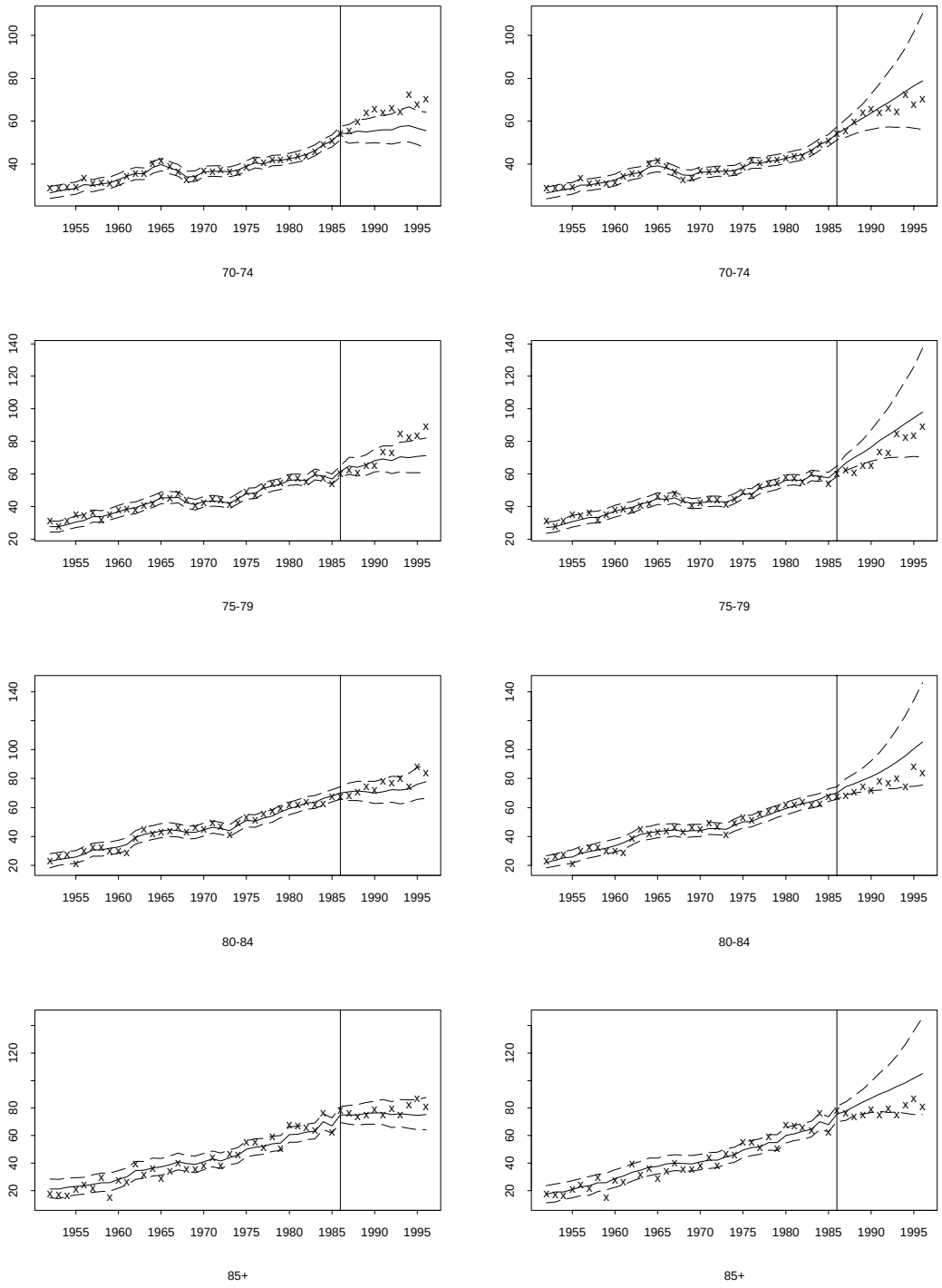

Figure 6: Observed and predicted number of cases within 80% credible regions for females in in age groups 70-74, 75-79, 80-84 and 85+. Left panel: RW1 model. Right panel: RW2 model.

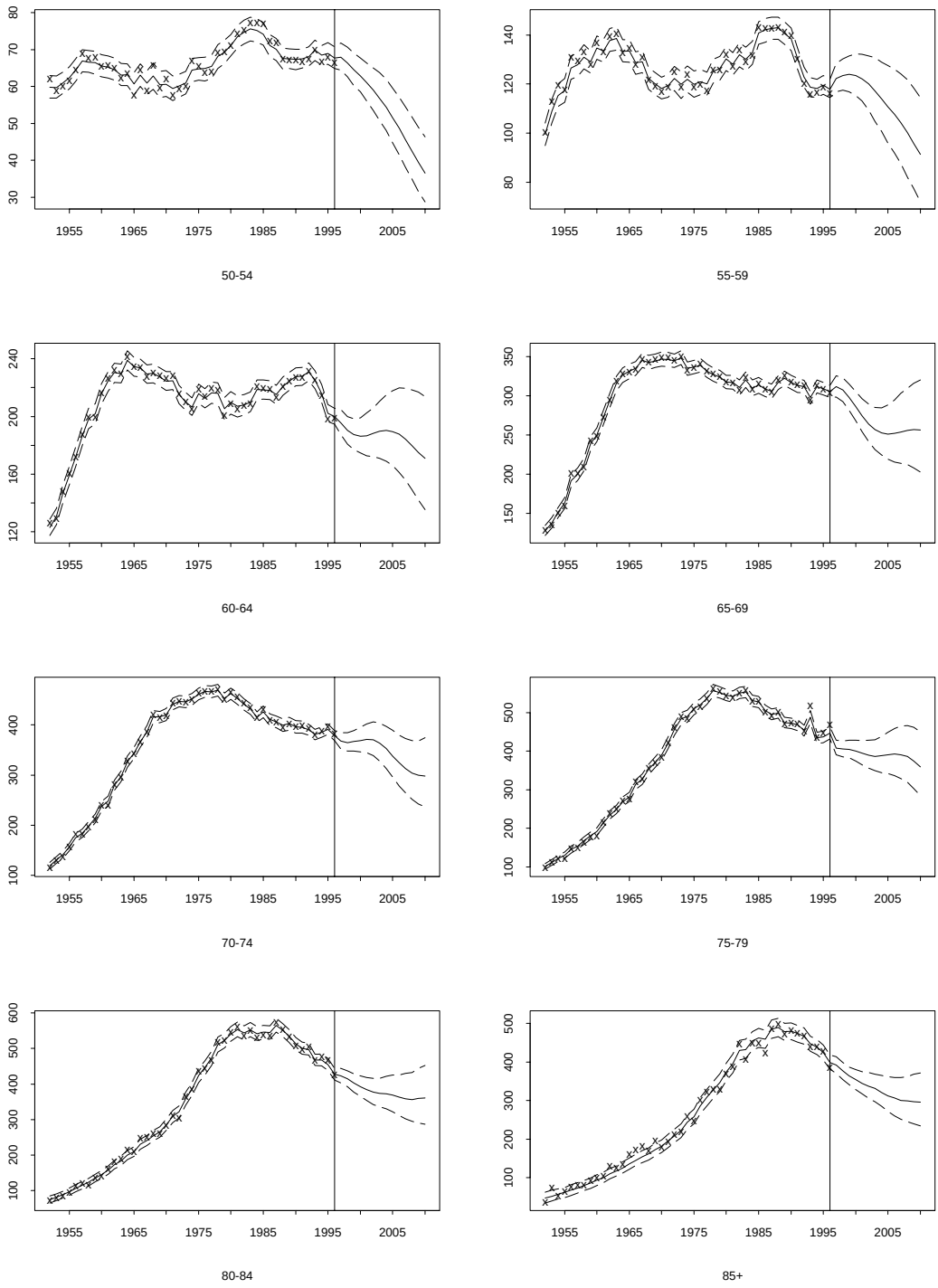

Figure 7: Observed and predicted number of cases per 100,000 within 80% pointwise credible intervals based on the RW2 model, males. Crosses represent the observed rates.

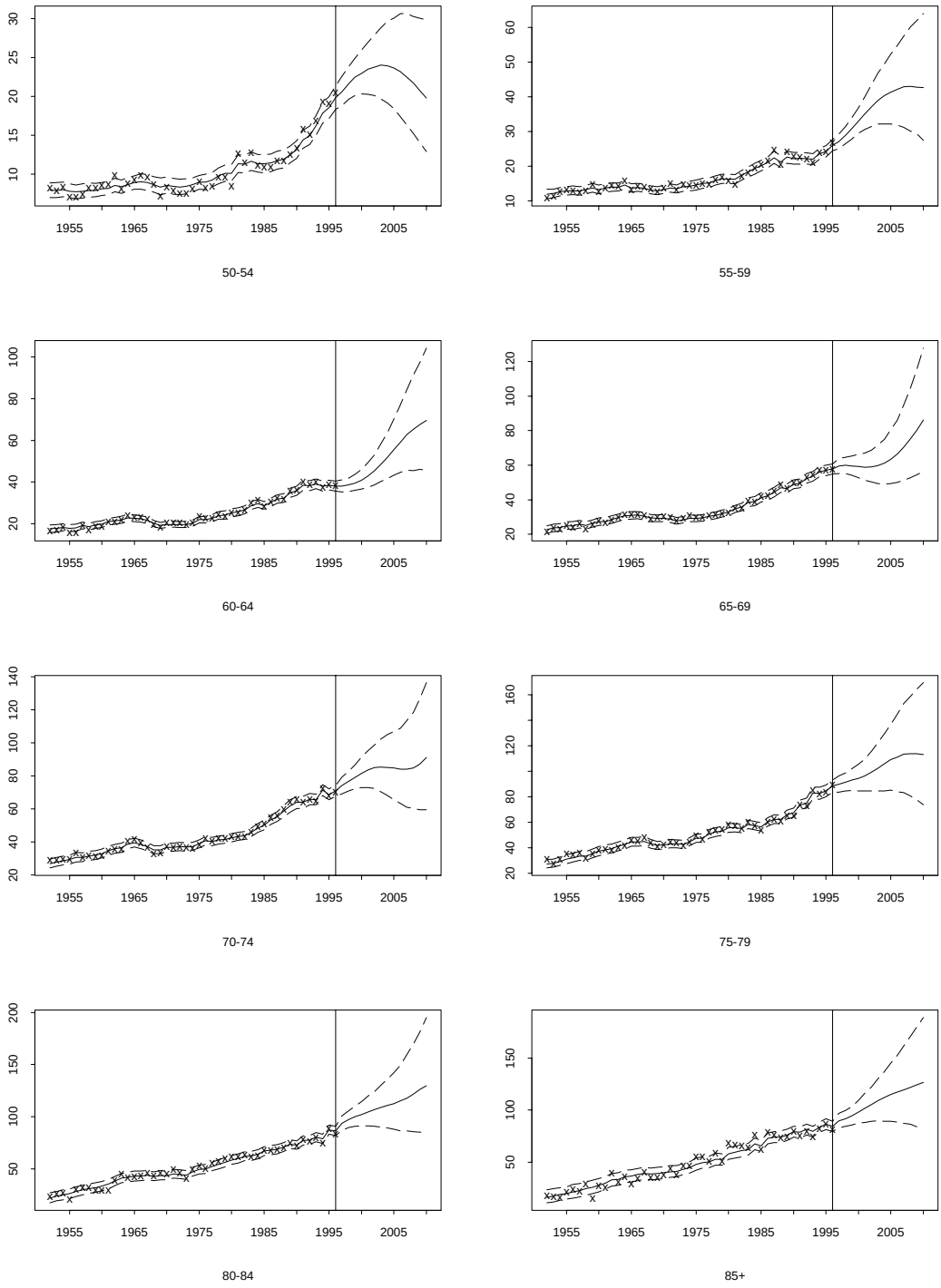

Figure 8: Observed and predicted number of cases per 100,000 within 80% pointwise credible intervals based on the RW2 model, females. Crosses represent the observed rates.

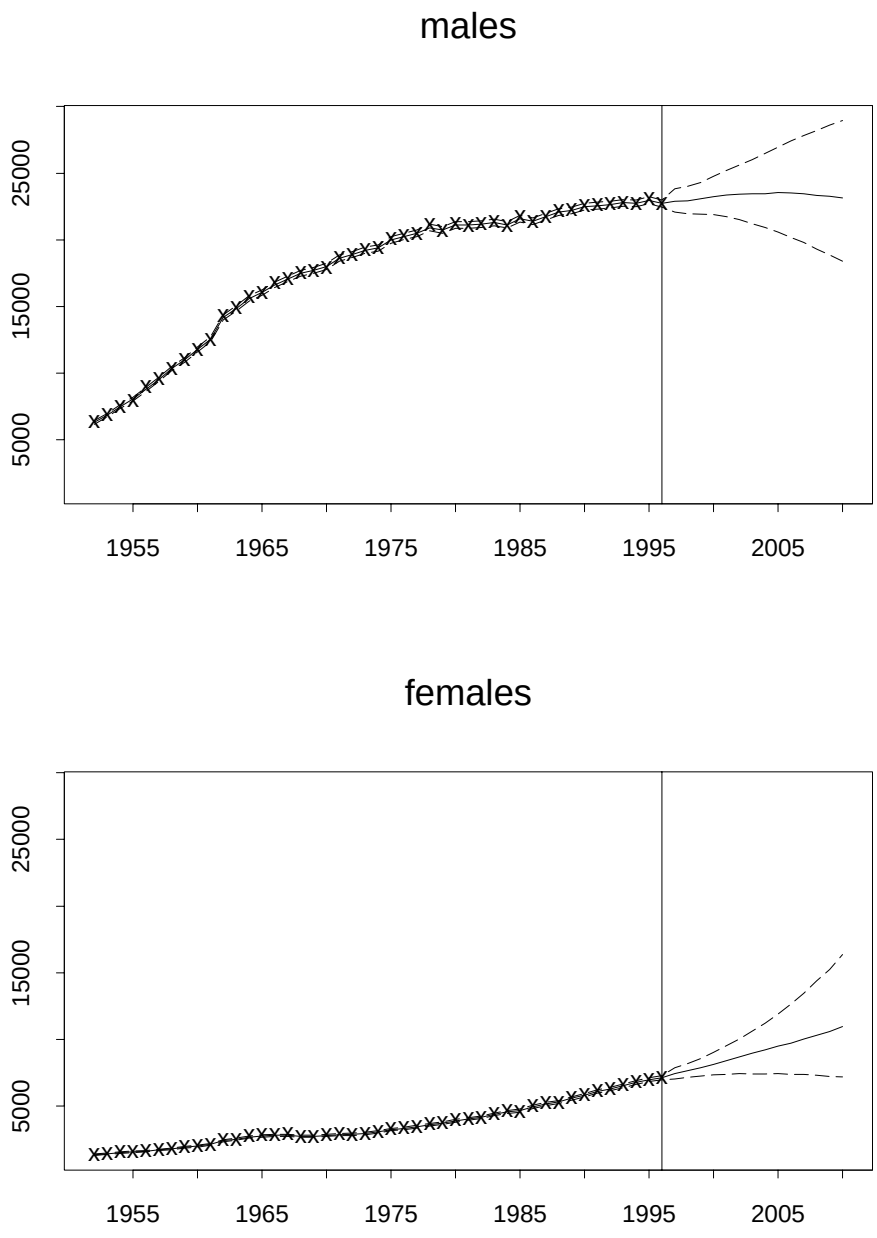

Figure 9: Predicted absolute number of lung cancer deaths in West Germany within 80% credible intervals based on the RW2 model.

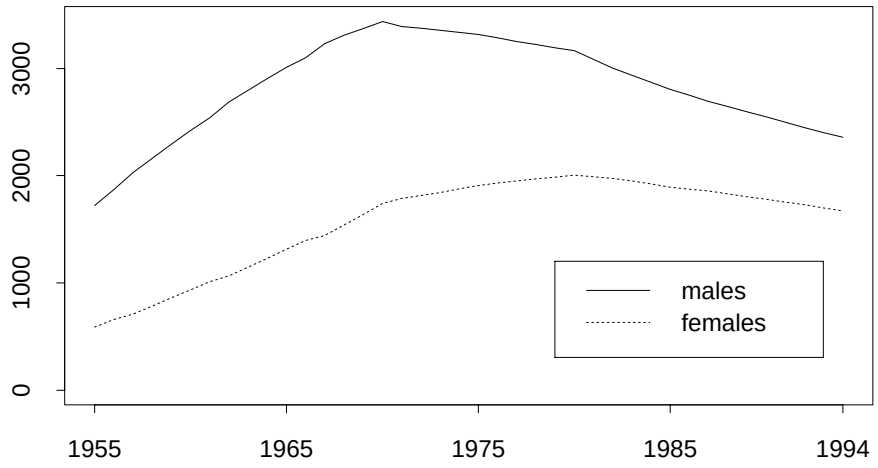

Figure 10: Number of cigarettes sold per capita (age $\geq$ 15), 1955-1994.

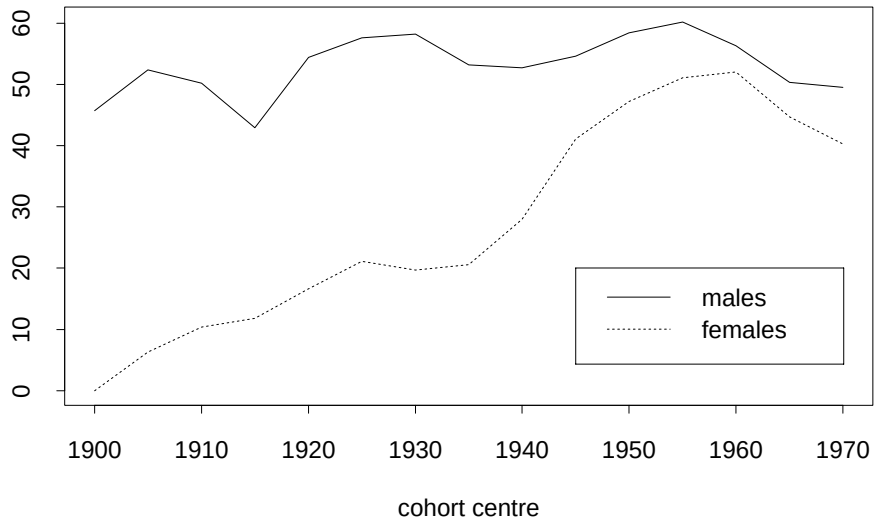

Figure 11: Smoking prevalence (in %) in age group 20-24 by birth cohort.