Biometrical Journal **52** (2010) 61, zzz–zzz / DOI: 10.1002/bimj.200100000

Supplementary Material for

"Projecting the future burden of cancer: Bayesian age-period-cohort analysis with integrated nested Laplace approximations"

Andrea Riebler[*,1] and **Leonhard Held** [2]

[1] Department of Mathematical Sciences, Norwegian University of Science and Technology, Alfred Getz vei 1, 7th floor, 7491 Trondheim, Norway

[2] Department of Biostatistics, Epidemiology, Biostatistics and Prevention Institute, University of Zurich, Hirschengraben 84, 8001 Zurich, Switzerland

A.1 `BAPC` R-code

The R-package `BAPC` depends on the `INLA`-package, which is available from `www.r-inla.org` and can be installed in `R` using the command

```
install.packages("INLA",
  repos="http://www.math.ntnu.no/inla/R/testing")
```

In the following we will show how to apply the `BAPC`-package to the USA lung cancer mortality counts in females as used in this paper. The `BAPC` package itself is available from R-forge and can be installed using the command

```
install.packages("BAPC",
  repos="http://R-Forge.R-project.org")
```

Within the paper we used version INLA 0.0-1472635713, of the R-package `INLA` and version 0.0.33 of `BAPC`. The US count and population data can be downloaded from `http://www.math.ntnu.no/~andrerie/software.html`. The following R-code implements the retrospective analysis of Section 5 and shows how to reproduce Figure 1 and Figure 2.

```
# load the library BAPC
library(BAPC)
library(INLA)

# read the mortality and population counts.
# The rownames (i.e. periods 1950-2007) are given in the first column.
counts = read.table("us_data_2014.txt", row.names=1, header=F)
pop = read.table("us_pop_2014.txt", row.names=1, header=F)

# define the labels for the 12 age groups
agegroup = c("25-29", "30-34", "35-39", "40-44",
  "45-49", "50-54", "55-59", "60-64",
```

*Corresponding author: e-mail: andrea.riebler@math.ntnu.no, Phone: +47 735 93528, Fax: +47 7359 3524


```
13   "65-69", "70-74", "75-79", "80-84")
14
15 # define an object of class APCList were you specify
16 # the dataset together with the grid factor (here M=5)
17 # and the labels of the age groups.
18 us.APC = APCList(counts, pop, gf=5, agelab=agegroup)
19
20 # generate Figure 1
21 col <- c("grey20", "grey35", "grey50", "grey65", "grey75", "grey85")
22 ratesByAge(us.APC, scale=100000, age=seq(27,82,5),
23     per=1950:2007, col=col)
24
25 # perform retrospective projection for 10 years, see section 5
26 us.res = BAPC(us.APC, predict=list(npredict=10, retro=TRUE))
27
28 # to generate figure 2 in the paper use the following command
29 plotBAPC(us.res, scale=10^5, type="ageSpecProj", showdata=TRUE,
30          probs= seq(0.05, 0.95, by=0.05),
31          col.fan=sequential_hcl)
```

In line 2, the R-packages `BAPC` is loaded before the data (number of mortality counts and person-years of exposure) are read in from two separate text files (lines 7–8). Within the text files the rows represent periods in increasing order, and the columns age groups from young to old. That means, each row represents one period, and each column (besides the first, which shows the year label) represents one age group. Thus, the first five lines of the file "us_data_2014.txt" look as follows:

```
1950 16 43 69 138 205 302 394 513 519 488 371 212
1951 16 36 75 155 217 346 420 470 514 491 395 231
1952 12 38 89 144 236 327 433 543 572 574 433 221
1953 18 42 87 173 254 325 424 511 571 560 425 234
1954 14 43 91 158 283 328 424 518 585 555 449 274
```

The structure of the file "us_pop_2014.txt" is analogous. Next we define a string vector indicating the age groups used in the analysis (line 11–13). These labels will be used in consecutive plotting functions. Now, we can define the `APCList` object (line 18) where we assign the count and population data, as well as the grid factor M, termed `gf`, needed to define the birth cohort indices, see Section 3. These three arguments are obligatory and must be provided. The definition of the age group labels is optional (as they are used only in graphics). If they are not provided, the column names of the data matrix will be used. Using the function `ratesByAge` the data can be plotted as shown in Figure 1. This function is based on the `rateplot` function of the R-package `Epi` (Carstensen *et al.*, 2014), see documentation.

After the `APCList` object is declared the function `BAPC` is used to generate projections (line 26). The `BAPC` function takes as first argument the `APCList` object, here `us.APC`. The second argument `predict` requires a list consisting of two arguments. The first argument `npredict` specifies how many periods should be projected. The second argument `retro` is set to `TRUE` for retrospective projection. Setting `retro=TRUE`, the data for the last, i.e. most recent, `npredict` periods will be projected and can consequently be compared with the observed values. In contrast, setting `retro=FALSE`, all periods that are set to `NA` will be projected (irrespective of the argument `npredict`). The `BAPC` functions takes an additional optional argument `stdweight` to which a numeric vector, whose length is equal to the number of age groups, can be assigned. The elements of this vector are regarded as age-specific weights and used to generate age-standardised projections as described in A.2. If those weights do not sum to one, they will be internally divided through their overall sum. The exclusion of specific model components, such as the

overdispersion, can also be specified within the `BAPC` function, please see the package documentation for details.

Running the `BAPC` call in line 26 for the US data set takes about 5 seconds on a Dell laptop Intel(R) Core(TM) i7-3740QM CPU @ 2.70GHz with four CPUs. Age-specific projected rates and number of cases can be extracted using the functions `agespec.rate` and `agespec.proj`, respectively. Both functions return a list, where each list element correspond to one age group. For each group the mean and standard deviation are returned for all J periods. For example `agespec.proj(us.res)[[1]]` returns the expected number of counts for the first age group. Here, we show the result for the first five periods:

```
> agespec.proj(us.res)[[1]]
         mean       sd
1950 12.98705 3.650796
1951 13.08344 3.662114
1952 13.12273 3.667093
1953 13.21679 3.680450
1954 13.06686 3.658909
...
```

Equally, `agestd.rate` and `agestd.proj` can be used to extract age-standardized projected rates and counts, given that the age-specific weights were given in the `BAPC` call. The function `qapc` allows to derive quantiles from the predictive distribution:

```
> us.res = qapc(us.res, percentiles=c(0.025, 0.5, 0.975))
```

The desired quantiles will be added to the APCList object where mean and standard deviations are already saved:

```
> agespec.proj(us.res)[[1]]
         mean       sd    0.025Q     0.5Q   0.975Q
1950 12.98705 3.650796  5.831622 12.98705 20.14248
1951 13.08344 3.662114  5.905824 13.08344 20.26105
1952 13.12273 3.667093  5.935359 13.12273 20.31010
1953 13.21679 3.680450  6.003245 13.21679 20.43034
1954 13.06686 3.658909  5.895533 13.06686 20.23819
...
```

The built-in function `plotBAPC` can be used to plot the generated projections (line 29–31). Here, the argument `type` can take one of the values `"ageSpecRate"`, `"ageSpecProj"` to plot age-specific projected rates or counts, respectively. The analogous specification for plotting age-standardized projections are `"ageStdRate"` and `"ageStdProj"`. Figure 2 was generated using `type="ageSpecProj"`. Setting `showdata=TRUE` the data are shown in addition to the projections.

As second example, we re-analyse mortality data on prostate cancer among the nonwhite male population of the United States from Holford (1983); Besag *et al.* (1995). The data are given in $J = 10$ five-year period intervals, 1935–1939, ..., 1980–1984, and $I = 7$ five-year age groups, 50–54, ..., 80–84, leading to $K = I - 1 + J = 16$ birth cohorts.

```
1 # load the BAPC package
2 library(BAPC)
3 library(INLA)
4
5 # define agegroup labels
```



```
agegroups = c("50-54", "55-59", "60-64",
    "65-69", "70-74", "75-79", "80-84")
# define period labels as the start-year of the five year interval
startperiod = c("1935", "1940", "1945", "1950", "1955",
              "1960", "1965", "1970", "1975", "1980")
# read the counts
counts <- read.table("counts.txt", header=F)
# read the population counts
pop <- read.table("pop.txt", header=F)

# define the APCList object. Note the gridfactor is equal to 1 here.
prostate = APCList(counts, pop, gf=1, agelab=agegroups)

# predict the last three periods
result = BAPC(prostate, predict=list(npredict=3, retro=TRUE),
   model=list(age=list(model="rw2",
                       prior = "loggamma", param = c(1, 0.005)),
         period= list(include=TRUE, model="rw2",
                       prior = "loggamma", param = c(1, 0.005)),
         cohort=list(include=TRUE, model="rw2",
                       prior = "loggamma", param = c(1, 0.005)),
         overdis=list(include=TRUE, model="iid",
                       prior = "loggamma", param = c(1, 0.005))))

# plot the predictive distribution together with the data
plotBAPC(result, scale=100000, type="ageSpecProj",
      probs = seq(0.1, 0.9, by=0.1))
```

In lines 21-28 details regarding the model can be specified. The age effects have to be included in the model, while other model components can be excluded by setting `include=FALSE`. As in the INLA package we specify hyperpriors for the precision parameters, which are internally represented on log-scale. Thus, a gamma distribution for the precision parameter, corresponds to a log-gamma distribution for the log precision parameter. However, as mentioned in Section 6.3 the gamma distribution can be replaced by any prior distribution, see INLA package. Here, we keep the gamma distribution and specify the value for the shape and rate parameter by setting `param = c(1, 0.005)`. Of note, these values were chosen corresponding to Besag *et al.* (1995), and may not be optimal. Figure 1 shows the resulting predictive distribution together with the observed counts.

A.2. Derivation of posterior age-standardized rates and addition of observation noise for the APC model

The derivation of age-standardized projected rates is complicated, since it requires the computation of weighted sums of correlated estimates. Hence, variance estimation using either linear approximations such as the delta method or resampling methods such as the bootstrap is required (Hakulinen and Dyba, 1994).

To compute approximate posterior variances, we use that

$$\boldsymbol{\eta} \,|\, \boldsymbol{y} \overset{a}{\sim} \mathcal{N}(\mathrm{E}(\boldsymbol{\eta} \,|\, \boldsymbol{y}), \mathrm{Cov}(\boldsymbol{\eta} \,|\, \boldsymbol{y})).$$

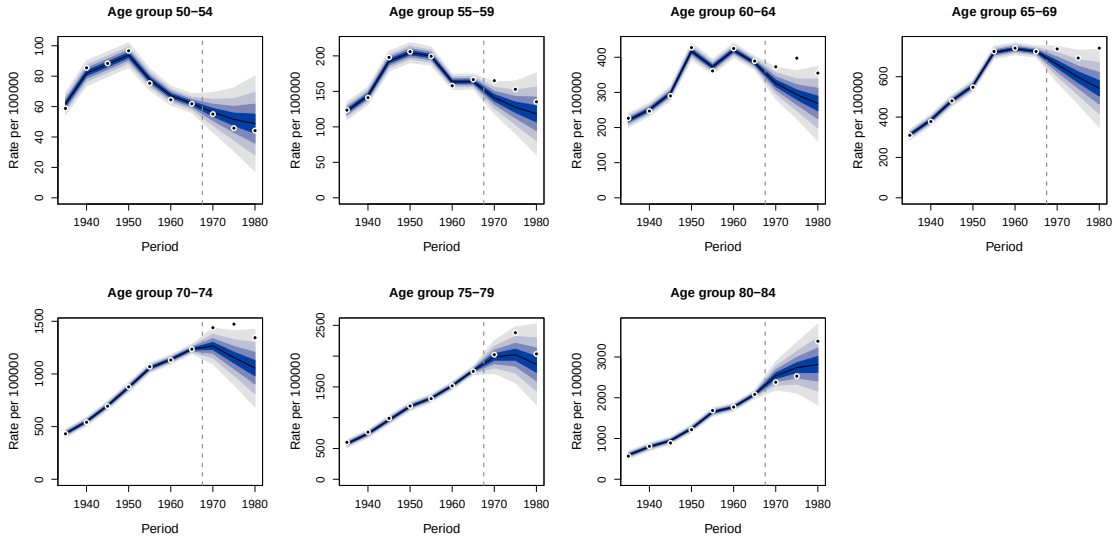

Figure 1 Prostate cancer mortality forecasts. The fan shows the predictive distribution between the 10% and 90% quantile, whereby the shaded bands show prediction intervals in increments of 20%. The predictive mean is shown as solid line. Observed number of cases are shown as a filled circle. The vertical dashed line indicates where prediction started.

Using the posterior correlations between all η_{ij}, $i = 1, \ldots, I$, $j = 1, \ldots, J$, and the posterior standard deviations of each η_{ij} we derive the corresponding posterior covariance matrix $\text{Cov}(\boldsymbol{\eta} \,|\, \boldsymbol{y})$ of dimension $(I \cdot J) \times (I \cdot J)$. To derive the covariance matrix of $\boldsymbol{\lambda} \,|\, \boldsymbol{y}$ we apply the multivariate delta rule:

$$\boldsymbol{\lambda} \,|\, \boldsymbol{y} \overset{a}{\sim} \mathcal{N}(\exp(\text{E}(\boldsymbol{\eta} \,|\, \boldsymbol{y})), \underbrace{\mathbf{D} \cdot \text{Cov}(\boldsymbol{\eta} \,|\, \boldsymbol{y}) \cdot \mathbf{D}}_{\boldsymbol{\Sigma}_\lambda}).$$

where the diagonal matrix $\mathbf{D}$ has $\exp(\text{E}(\boldsymbol{\eta} \,|\, \boldsymbol{y}))$ on the diagonal. Having the posterior covariance matrix between age-specific mortality rates λ_{ij}, we are able to compute posterior standard deviations of λ_j as follows

$$\text{SD}(\lambda_j) = \{\sqrt{\text{diag}(\mathbf{W}\boldsymbol{\Sigma}_\lambda\mathbf{W}^\top)}\}_j$$

with $\mathbf{W}_{J \times (I \cdot J)}$ containing the age-specific weight w_i at position $\mathbf{W}_{j,(i-1)\cdot I+j}$, $j = 1, \ldots, J$. As summary statistics BAPC returns the age-standardized expected value

$$\text{E}(\lambda_j) = \sum_{i=1}^{I} w_i \exp(\text{E}(\eta_{ij} \,|\, \boldsymbol{y})).$$

BAPC approximates a desired quantile to the given probability p using the corresponding quantile of a normal distribution with mean $\text{E}(\lambda_j)$ and variance $\text{SD}(\lambda_j)$. For age-standardisation we use the percentage of the population in each 5-year age group in the new WHO World Standard population as weights (Ahmad *et al.*, 2001).

Using the law of iterated expectations (Held and Sabanés Bové, 2014, Appendix A.3.4) the mean of the predictive distribution μ_{ij} can be derived. With $y_{ij} \,|\, \lambda_{ij} \sim \text{Po}(n_{ij} \cdot \lambda_{ij})$ it follows

$$\mu_{ij} = \text{E}(\text{E}(y_{ij} \,|\, \lambda_{ij})) = \text{E}(n_{ij} \cdot \lambda_{ij}) = n_{ij} \cdot \text{E}(\lambda_{ij}). \tag{1}$$

Analogously, the variance $\sigma_{ij}^2 = \text{Var}(y_{ij})$ follows from the law of total variance as

$$\begin{aligned}\sigma_{ij}^2 &= \text{E}(\text{Var}(y_{ij} \mid \lambda_{ij})) + \text{Var}(\text{E}(y_{ij} \mid \lambda_{ij})) \\ &= \text{E}(n_{ij} \cdot \lambda_{ij}) + \text{Var}(n_{ij} \cdot \lambda_{ij}) = n_{ij} \cdot \text{E}(\lambda_{ij}) + n_{ij}^2 \text{Var}(\lambda_{ij}). \end{aligned} \tag{2}$$

In the case of the age-standardized predictive distribution we obtain that

$$\begin{aligned}\mu_j &= \text{E}(\text{E}(y_j \mid \lambda_j)) \\ &= \text{E}(n_j \lambda_j) \\ &= n_j \text{E}(\lambda_j)\end{aligned}$$

with $n_j = \sum_{i=1}^{I} n_{ij}$. Analogously, the variance $\sigma_j^2 = \text{Var}(y_j)$ follows from the law of total variance as

$$\begin{aligned}\sigma_j^2 &= \text{E}(\text{Var}(y_j \mid \lambda_j)) + \text{Var}(\text{E}(y_j \mid \lambda_j)) \\ &= \text{E}(n_j \lambda_j) + \text{Var}(n_j \lambda_j) \\ &= n_j \text{E}(\lambda_j) + n_j^2 \cdot \text{SD}^2(\lambda_j).\end{aligned}$$

A.3 Derivation of the mean and standard deviation of the predictive distribution for the Lee-Carter model

The `demography` package returns the mortality rate $\hat{\lambda}_{ij}$ and a (symmetric) confidence interval $[\text{CI}_{\text{lower}}, \text{CI}_{\text{upper}}]$ for the log mortality rate at a predefined confidence level $l \cdot 100\%$. We derived the corresponding standard error of the linear predictor $\hat{\eta}_{ij} = \log \hat{\lambda}_{ij}$ on the log scale as

$$\text{SE}(\hat{\eta}_{ij}) = \frac{\log(\text{CI}_{\text{upper}}) - \log(\hat{\lambda}_{ij})}{q}$$

where q denotes the $(l+1)/2$ quantile of the standard normal distribution, e.g. if $l = 0.95$ it follows that $q \approx 1.96$. Using the delta rule the implied standard error for $\hat{\lambda}_{ij}$ is obtained as

$$\text{SE}(\hat{\lambda}_{ij}) = \text{SE}(\log \hat{\lambda}_{ij}) \cdot \hat{\lambda}_{ij}.$$

The Poisson observational noise has to be added to the distribution of $\hat{\lambda}_{ij}$ to get the mean and the variance of the predictive distribution for y_{ij}. To adjust for overdispersion we use the usual estimate of the GLM scale factor ϕ based on the deviance (McCullagh and Nelder, 1989) and then compute the predictive variance $\sigma_{ij}^2 = \text{Var}(y_{ij})$ using the law of total variance as:

$$\sigma_{ij}^2 = \phi \cdot n_{ij} \cdot \hat{\lambda}_{ij} + n_{ij}^2 \cdot \text{SE}(\hat{\lambda}_{ij})^2.$$

The predictive mean is computed as $\mu_{ij} = \text{E}(y_{ij}) = n_{ij} \hat{\lambda}_{ij}$.

A.4 Supplementary Figures

Figures 2, 3, 4, 5, show observed and predicted female lung cancer mortality rates in the UK, Australia, Sweden and New Zealand, respectively, for all age groups. The shading indicates various pointwise credible intervals of the predictive distribution.

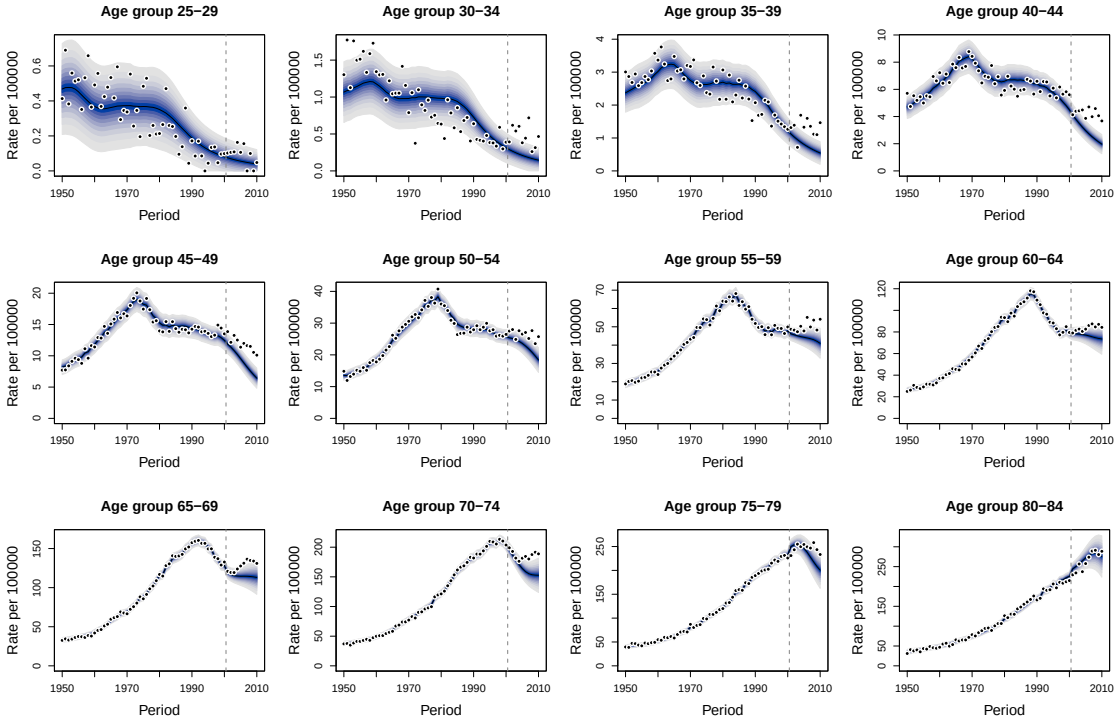

Figure 2 Female lung cancer mortality forecasts in the UK. The fan shows the predictive distribution between the 5% and 95% quantile, whereby the shaded bands show prediction intervals in increments of 10%. The predictive mean is shown as solid line. Observed number of cases are shown as a filled circle. The vertical dashed line indicates where prediction started.

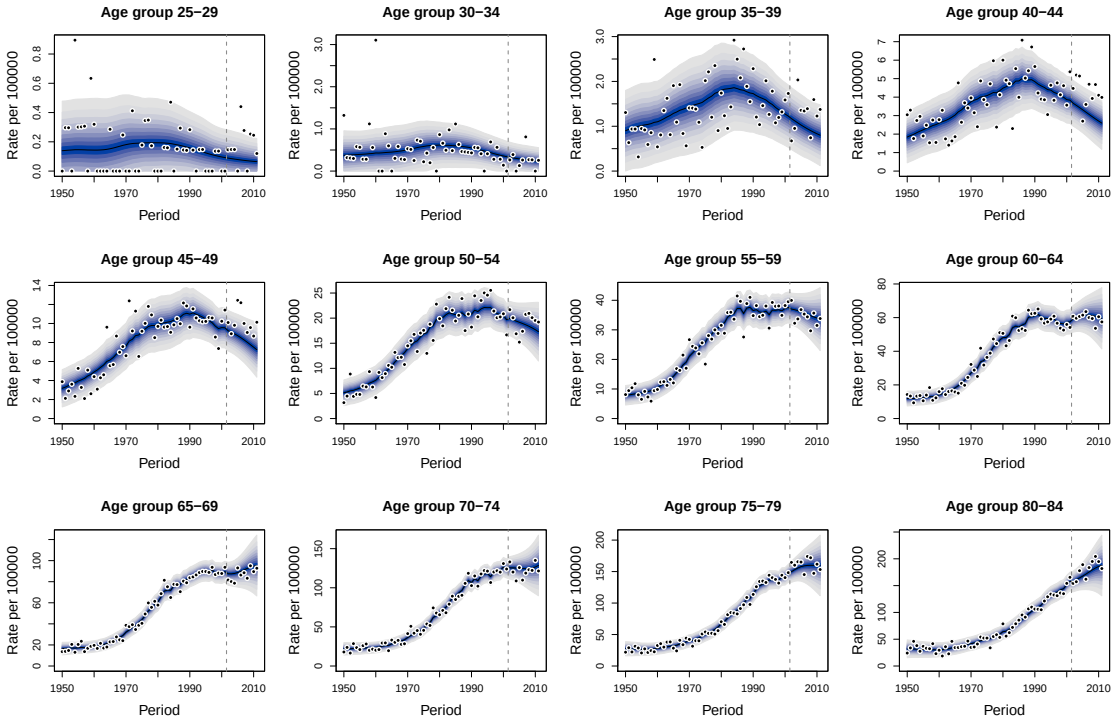

Figure 3 Female lung cancer mortality forecasts in Australia. The fan shows the predictive distribution between the 5% and 95% quantile, whereby the shaded bands show prediction intervals in increments of 10%. The predictive mean is shown as solid line. Observed number of cases are shown as a filled circle. The vertical dashed line indicates where prediction started.

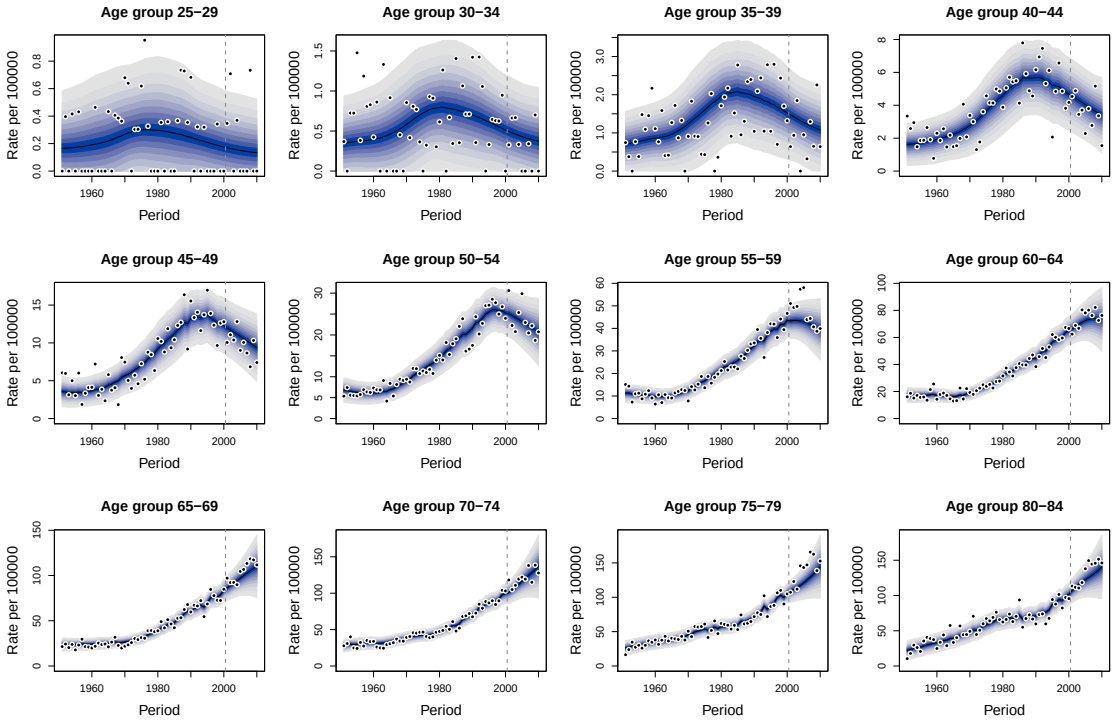

Figure 4 Female lung cancer mortality forecasts in Sweden. The fan shows the predictive distribution between the 5% and 95% quantile, whereby the shaded bands show prediction intervals in increments of 10%. The predictive mean is shown as solid line. Observed number of cases are shown as a filled circle. The vertical dashed line indicates where prediction started.

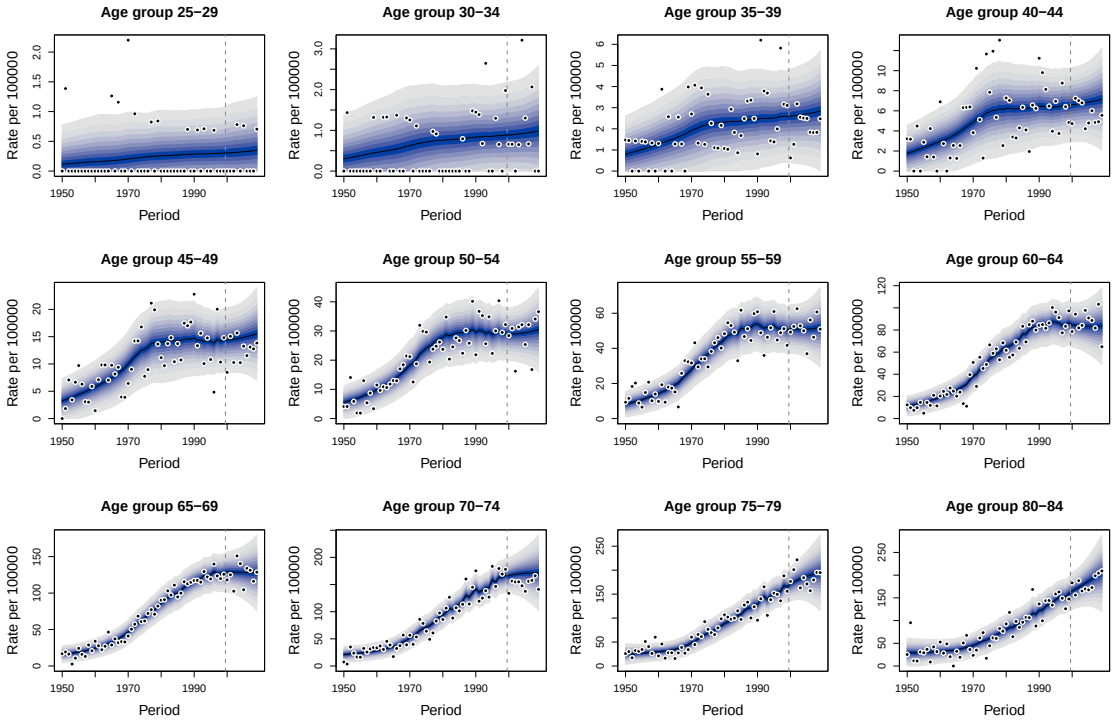

Figure 5 Female lung cancer mortality forecasts in New Zealand. The fan shows the predictive distribution between the 5% and 95% quantile, whereby the shaded bands show prediction intervals in increments of 10%. The predictive mean is shown as solid line. Observed number of cases are shown as a filled circle. The vertical dashed line indicates where prediction started.

A.5 Explanatory analysis of model components

To study how much age, period and cohort effects vary between countries Table 1 shows the posterior precision parameter estimates for the retrospective forecasting setting of the most recent 10 years, see Section 5 of the main text. The precision estimates for age, period and cohort seem comparable between countries, indicating more variation for age than for period and cohort effects. While for the USA we find lower precision estimates, that means more variation in the period effects than for the other countries, we find stronger cohort effects for the UK. Further differences are apparent in the precision parameter estimates for the overdispersion parameters. Here, higher estimates for the USA and UK are obtained compared to the three other countries. These estimates have to be set in relation to the population counts n_{ij} to quantify the increase of the variance (2) compared to the mean (1).

Sørbye and Rue (2014) showed that intrinsic Gaussian random fields, such as the random walks of second order, have to be suitably scaled to guarantee that a given prior has the same meaning when the length of the random walk changes. Of note, scaling according to Sørbye and Rue (2014) is readily available within the `BAPC` package see `?BAPC`. Since the random effects for age, period and cohort have varying lengths, we investigate the effect of scaling on the precision parameter estimates, see Table 2. We see that the precision estimates for age, period and cohort effects change. However, relative precision estimates between countries stay almost constant. Furthermore, the precision estimates for the overdispersion parameters are almost the same as in Table 1.

Although the original age, period and cohort effect are not identifiable, second differences, e.g. $(\alpha_{i+1} - \alpha_i) - (\alpha_i - \alpha_{i-1}) = \alpha_{i+1} - 2\alpha_i + \alpha_{i-1}$, representing measures of curvature can be inspected (Clayton and Schifflers, 1987). On the exponential scale, these contrasts represent ratios of two adjacent relative risks. A value of 1.1 for age group $i = 2$ would for example mean that the relative risk from age group 3 to age group 2 is 10% higher than from age group 2 to age group 1. Figure 6 shows the identifiable second differences parameter estimates on the exponential scale for female lung cancer mortality for all countries. Inspecting the results for the USA (first row) we see a change in the age trend around $35 - 40$ years and around $70 - 74$ years. For the period trend a change around the late 1960s is indicated and for the cohort trend a sudden change for persons born after the second world war is visible. Comparing the results between countries we see strong sudden changes in both the period trend and cohort trend of the USA, and the cohort trend of the UK and Sweden. This agrees well with the posterior precision estimates.

Figure 7 shows the corresponding results when the random effects are scaled. We see that the results are mostly the same, whereby some credible intervals get more narrow. Inspecting forecasts with or without scaling we only find minor differences (result not shown).

Precision	Mean	SD	0.025Q	0.5Q	0.975Q
		USA			
Age	81.36	33.64	31.85	76.24	161.52
Period	21942.60	8760.31	9220.05	20521.09	43066.73
Cohort	17205.42	5706.28	8482.00	16402.39	30605.14
Overdispersion	2042.31	351.55	1445.95	2009.65	2820.93
		UK			
Age	128.56	53.94	49.68	120.19	257.46
Period	56612.64	24467.31	21846.97	52430.67	116166.01
Cohort	10540.11	3407.96	5295.32	10071.18	18508.57
Overdispersion	1614.00	295.97	1118.29	1584.20	2276.69
		Australia			
Age	97.83	44.67	34.81	90.19	206.06
Period	30516.28	17942.73	8034.57	26635.70	75905.83
Cohort	19718.87	10230.04	6264.05	17665.28	45324.36
Overdispersion	829.27	288.23	410.32	780.50	1525.68
		Sweden			
Age	82.58	41.48	26.66	74.68	184.95
Period	29558.86	16762.75	8203.19	26025.62	71634.04
Cohort	15775.27	7709.23	5432.77	14278.37	34965.17
Overdispersion	590.07	237.20	268.01	542.94	1182.15
		New Zealand			
Age	64.31	32.73	20.29	58.08	145.04
Period	23830.06	17571.62	4315.26	19418.14	69521.56
Cohort	29230.94	16383.89	8418.27	25746.02	70658.54
Overdispersion	483.93	245.98	172.14	430.09	1110.22

Table 1 Summary estimates (mean, standard deviation, 2.5% quantile, median and 97.5% quantile) of all precision (inverse variance) parameters in the APC model with overdispersion provided for all countries.

Precision	Mean	SD	0.025Q	0.5Q	0.975Q
		USA			
Age	30.72	12.74	12.10	28.73	61.14
Period	83.32	35.04	33.70	77.22	168.64
Cohort	7.84	2.69	3.78	7.44	14.20
Overdispersion	2035.81	350.24	1440.69	2003.78	2810.27
		UK			
Age	48.15	20.25	18.64	44.97	96.60
Period	426.33	263.62	117.44	362.91	1108.82
Cohort	4.26	1.40	2.11	4.06	7.54
Overdispersion	1603.52	293.97	1110.28	1574.31	2260.29
		Australia			
Age	36.50	16.68	12.98	33.65	76.93
Period	194.83	150.49	37.18	154.94	590.15
Cohort	10.02	5.68	2.86	8.80	24.41
Overdispersion	815.90	283.84	403.96	767.66	1501.78
		Sweden			
Age	31.16	15.65	10.11	28.16	69.90
Period	189.06	142.62	38.66	151.37	564.16
Cohort	7.77	4.10	2.48	6.91	18.09
Overdispersion	584.78	235.47	265.49	537.85	1172.83
		New Zealand			
Age	24.27	12.37	7.68	21.89	54.81
Period	235.78	237.99	24.90	166.30	866.31
Cohort	22.86	16.31	4.94	18.70	65.46
Overdispersion	481.52	245.12	171.08	427.81	1105.74

Table 2 Summary estimates (mean, standard deviation, 2.5% quantile, median and 97.5% quantile) of all precision (inverse variance) parameters in the APC model with overdispersion provided for all countries. Here, we assume scaled random effects.

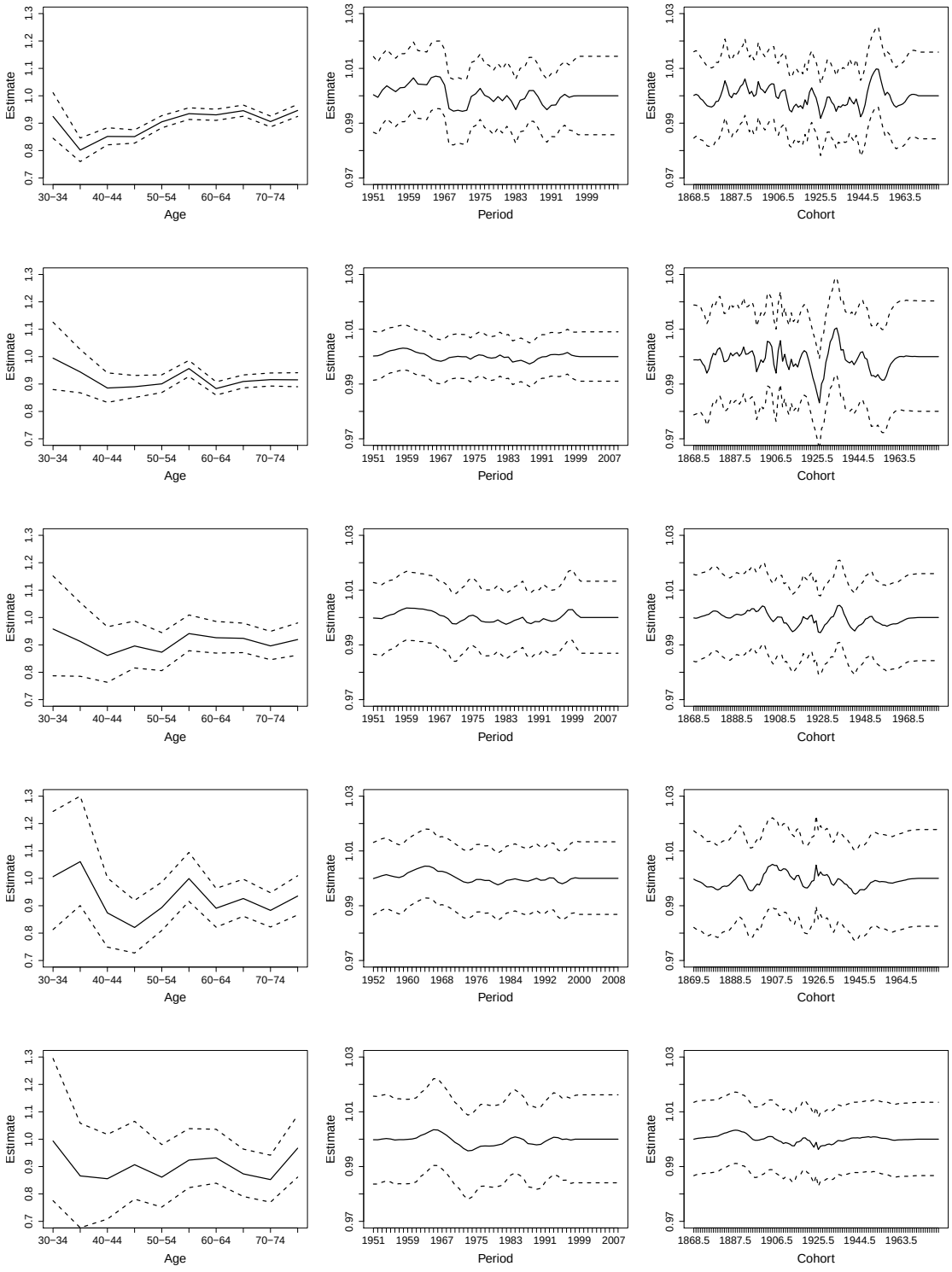

Figure 6 Female lung cancer mortality: Summary estimates (2.5% quantile, median, 97.5% quantile) of the identifiable second differences on exponential scale for all countries (row-wise from top to bottom: USA, UK, Australia, Sweden, New Zealand).

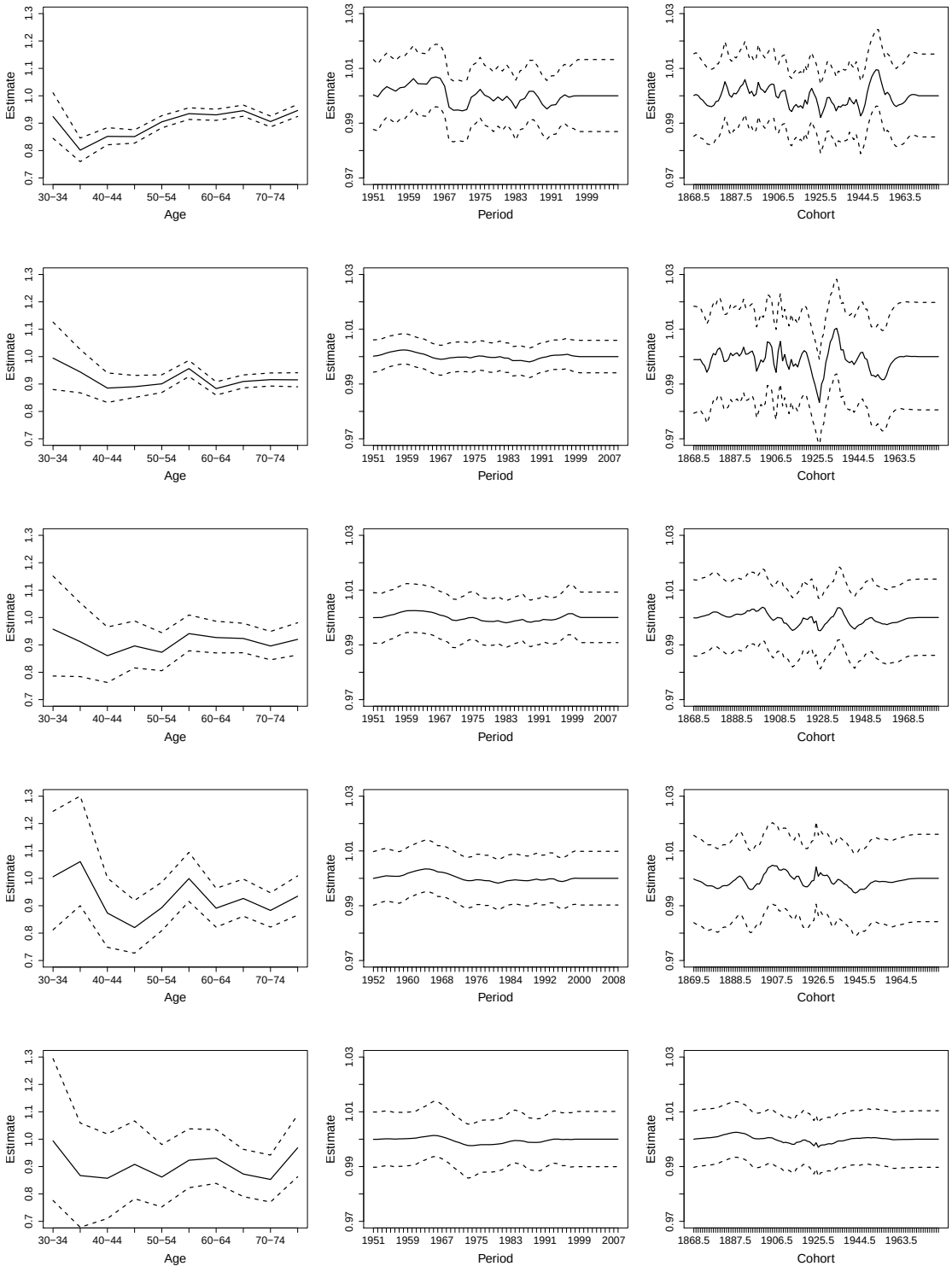

Figure 7 Female lung cancer mortality: Summary estimates (2.5% quantile, median, 97.5% quantile) of the identifiable second differences on exponential scale for all countries (row-wise from top to bottom: USA, UK, Australia, Sweden, New Zealand). Here, we assume scaled random effects.

A.6 Effect of excluding data from the youngest age group

In this section, we investigate the effect of excluding data from the youngest age group, i.e. those aged between 25 to 29 years, on the predictive quality of retrospective projections for the last 10 years and the calibration of one-step-ahead forecasts. The observed number of cases per 100 000 person years can be seen in Figure 2 of the main text, and Figures 2,3,4 and 5 of the supplementary material. Table 3 shows the amount of zero or lower than two cases for the youngest age group. Data are sparsest for New Zealand where there are a maximum of two cases observed over all years. In contrast data seem not sparse for lung cancer mortality in the USA where we observe for all years at least three deaths. We therefore expect that excluding data from the youngest age group for forecasts in the USA has a different effect than for New Zealand.

Baker and Bray (2005) found that using all available data improves Bayesian APC projections when there are strong cohort effects and when etiology is similar across all ages. For the age groups 25 to 84 years we would expect a similar etiology for lung cancer for all countries. However, the strength of cohort effects is different for the different countries as shown in the supplementary material A.5.

Table 3 Explanatory analysis of the youngest age group (25-29 years) over all years and countries. Shown are the total number of cases, number of one or two cases and the number of more than two cases over all years.

Country	# cases	# zeros	#(1-2)	#(> 2)
USA	1053	0	0	58
United Kingdom	316	3	14	44
Australia	51	26	33	3
Sweden	39	31	28	1
New Zealand	15	46	14	0

We start by repeating the analysis presented in Section 6 of the main text omitting data for the youngest age group. Table 4 shows the empirical coverage of the one-step-ahead Bayesian APC forecasts for people aged 30-84 years when including data for persons aged 25-84 years and obtained when using data for persons aged 30-84 years. The empirical coverage for the USA and UK is with a maximum of 3 percentage points closer to the nominal level when excluding data from the youngest age group. For the other three countries results are almost the same. The results of the CRPS calibration test do not change when

Table 4 Empirical coverage of the one-step-ahead predictive credible bands for three different credible levels 50%, 80% and 95%. Shown are results for all age groups except the youngest (25-29 years) obtained with the Bayesian APC model using data for all age groups and the Bayesian APC model omitting data from the youngest age group.

	All age groups			Omitting the youngest		
Credible level	50%	80%	95%	50%	80%	95%
USA	35%	65%	91%	36%	68%	92%
United Kingdom	43%	67%	85%	44%	69%	88%
Australia	51%	82%	97%	49%	82%	97%
Sweden	50%	82%	95%	51%	82%	95%
New Zealand	51%	81%	93%	52%	81%	94%

excluding data for the youngest age group, see Table 5. For the USA we see a light improvement in the mean CRPS score of 3 cases when omitting data from the youngest age group. For smaller countries the

Table 5 Mean absolute error $\overline{\text{AE}}$, mean predictive standard deviation $\overline{\text{SD}}$ and mean continuous ranked probability score $\overline{\text{CRPS}}$ with z-statistic and p-value from the corresponding calibration test. Shown are results for the one-step-ahead projections for all age groups except the youngest (25-29 years) obtained with the Bayesian APC model using data for all age groups and the Bayesian APC model omitting data from the youngest age group.

	Country	$\overline{\text{AE}}$	$\overline{\text{SD}}$	$\overline{\text{CRPS}}$	z	p-value
All age groups	USA	169.39	162.93	118.01	5.26	< 0.0001
	United Kingdom	43.34	44.74	31.22	4.40	< 0.0001
	Australia	11.54	15.92	8.11	-1.93	0.054
	Sweden	8.61	10.97	6.04	-0.48	0.63
	New Zealand	5.27	6.91	3.78	-0.63	0.53
Omitting the youngest	USA	165.45	162.66	115.35	4.76	< 0.0001
	United Kingdom	43.16	44.75	31.03	4.26	< 0.0001
	Australia	11.58	15.93	8.12	-1.91	0.057
	Sweden	8.63	10.98	6.06	-0.45	0.65
	New Zealand	5.27	6.93	3.78	-0.66	0.51

score differences are negligible. Figure 8 shows the PIT histograms based on all projections for people aged 30-84 years obtained using either data from all agegroups, i.e. 24-84 years, or a reduced data set, i.e. 30-84 years. Differences are again minor. A slight improvement toward uniformity is visible for the USA and UK. The results can be summarized as follows, predicting only one period at a time we see no

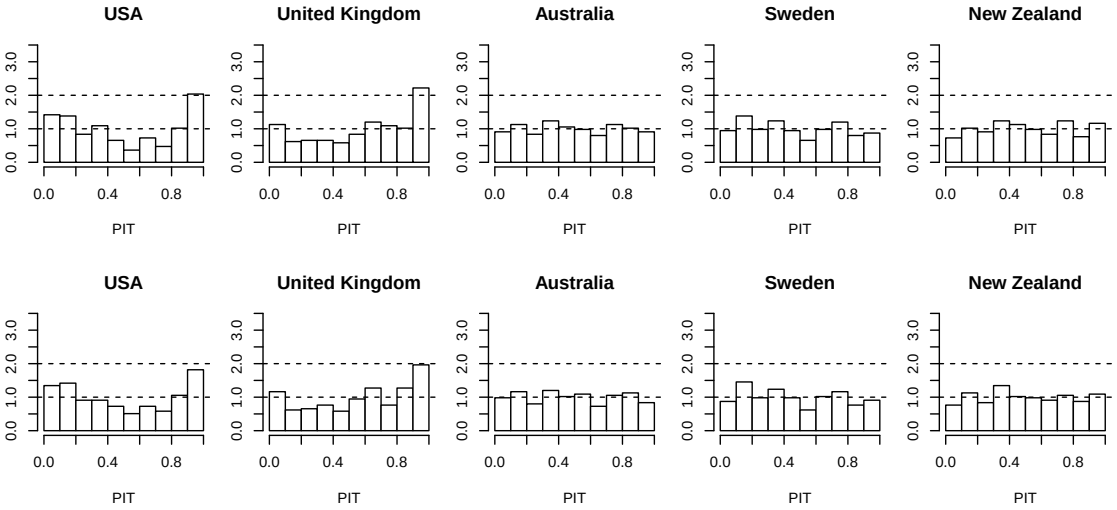

Figure 8 PIT histograms. Shown are results for all age groups except the youngest (25-29 years) obtained with the Bayesian APC model using data for all age groups (top) and the Bayesian APC model omitting data from the youngest age group (bottom).

differences when omitting data for the youngest age groups when predicting data for all age groups except the youngest for the three smaller countries. For those countries sparsity was strongest. Looking at the USA and UK we observe slight improvement when omitting the youngest age group. The reason is not clear, but might be due to stronger varying cohort effects as indicated in supplementary Figure 6.

Investigating the effect when the length of prediction increases we repeat the analysis shown in Section 5 of the main text. Figure 9 shows the difference in cumulative average $\overline{\text{CRPS}}_j$ and cumulative average $\overline{\text{AE}}_j$ including projections for people age 30-84 years between the model using data for all age groups and when omitting the youngest age group. For the USA we see that projection quality further improves over time when excluding data for the youngest age group. The contrary is true for Sweden and New Zealand where the predictive quality is always better when including all data and improves further when prediction time increases. For Australia the short-term predictive quality is better when including all data but gets worse with increasing prediction time where omitting the first age group seems beneficial. For the UK differences start favoring the excluding of the youngest age group but go towards zero with increasing prediction time.

We conclude that there cannot be made an overall statement regarding the effect of excluding data for the youngest age group. For the USA where data are not sparse, using a reduced data set might be beneficial, whereas for the other countries using all data might be in particular recommended when doing long-term predictions.

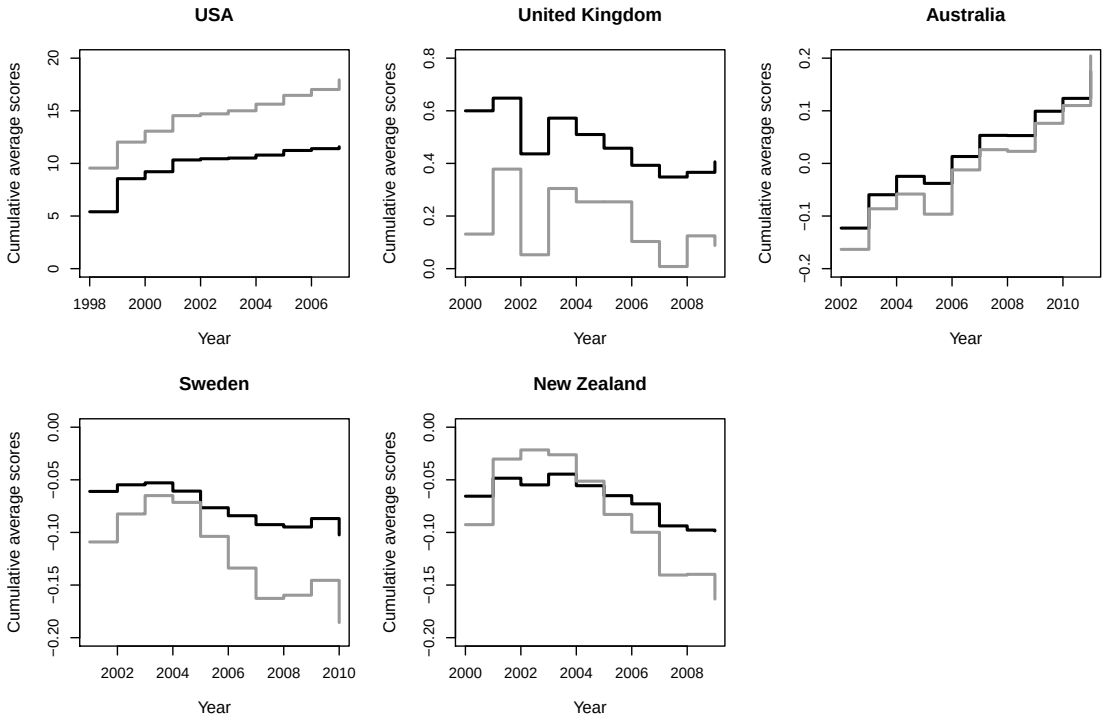

Figure 9 Difference in cumulative average of mean absolute errors (grey) and continuous ranked probability scores (black) across all age groups except the youngest (25-29 years) for all five countries obtained between the Bayesian APC model when including data from all age groups and the Bayesian APC model omitting data from the first age group.

A.7 Prior sensitivity when changing parameters in the inverse gamma priors

To study to what extent results change when changing the parameters of the inverse gamma priors for the variance parameters, we compare the results to those obtained when using the prior distributions proposed by Smith and Wakefield (2016, Section 6.3), in the following denoted as SW priors. The shape parameter of the inverse gamma distributions is assumed to be 1 for all variance parameters. The rate parameter for the variance of the age effects is consequently derived by assuming that the relative risk lies within (0.83, 1.2) of $\exp(2\alpha_{i-1} + \alpha_{i-2})$. For the period and cohort parameters smaller changes are expected and the range (0.91, 1.1) is assumed. Due to less information about the overdispersion a slightly larger interval, i.e. (0.67, 1.5), is used for the overdispersion parameters. This leads to the following inverse gamma priors: IG(1, 0.0008977873) for κ_α^{-1}, IG(1, 0.0002453443) for κ_β^{-1} and κ_γ^{-1}, and IG(1, 0.008880438) for κ_z^{-1}.

Figure 10 and 11 compare the cumulative average of mean absolute errors and mean CRPS scores, respectively, across age groups using the original priors, and changing the priors for the RW2 and/or overdispersion to the SW priors. The scores seem very stable when changing the hyperpriors. Small

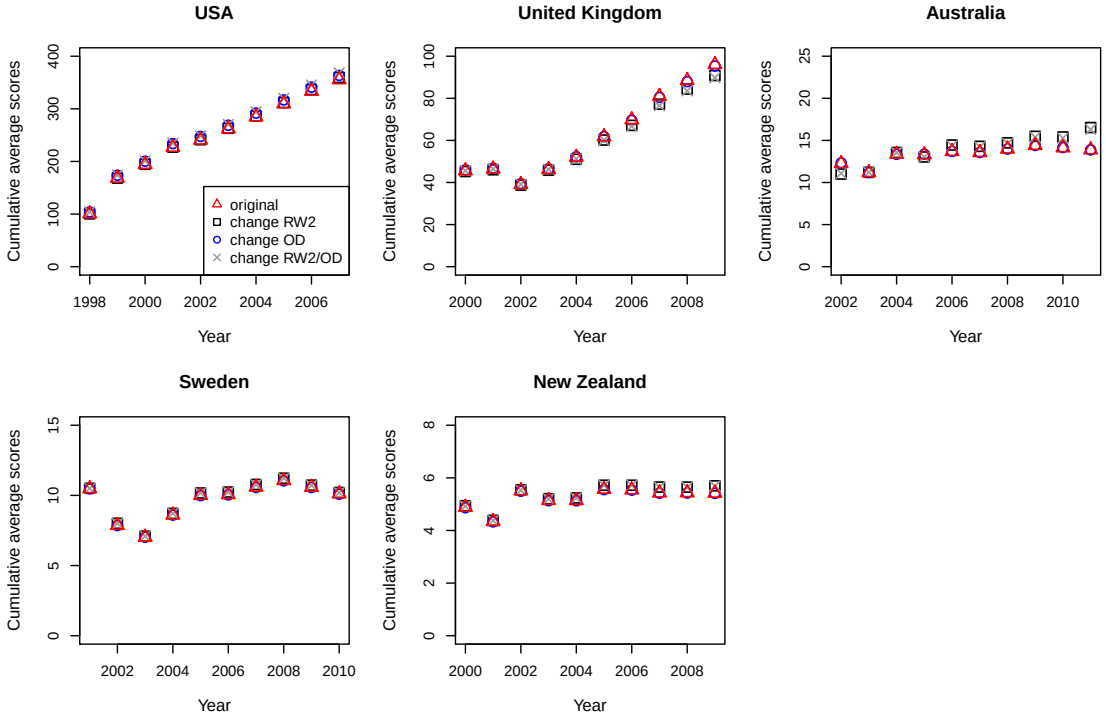

Figure 10 Cumulative average of mean absolute errors across age groups for all five countries obtained by the Bayesian APC model using four different hyperprior sets (Original: Ga(1, 5e-5) for RW2, Ga(1, 0.005) for OD; change RW2: SW priors for all RW2; change OD: SW prior for OD; change RW2/OD: SW priors for all RW2 and OD).

deviations can be seen when predicting further into the future for UK, Australia and New Zealand.

Inspecting the sensitivity of one-step-ahead predictions in more detail, Table 6 shows coverage probabilities when changing the hyperpriors. Compared to Table 1 in the main paper, we see that there are only minor changes. A similar conclusion can be drawn by comparing Table 2 in the main paper and Table 7. The mean predictive scores are fairly stable, whereas we observe changes in the mean predictive standard deviation and z-statistics. The conclusions from the miscalibration test are, however, almost unchanged.

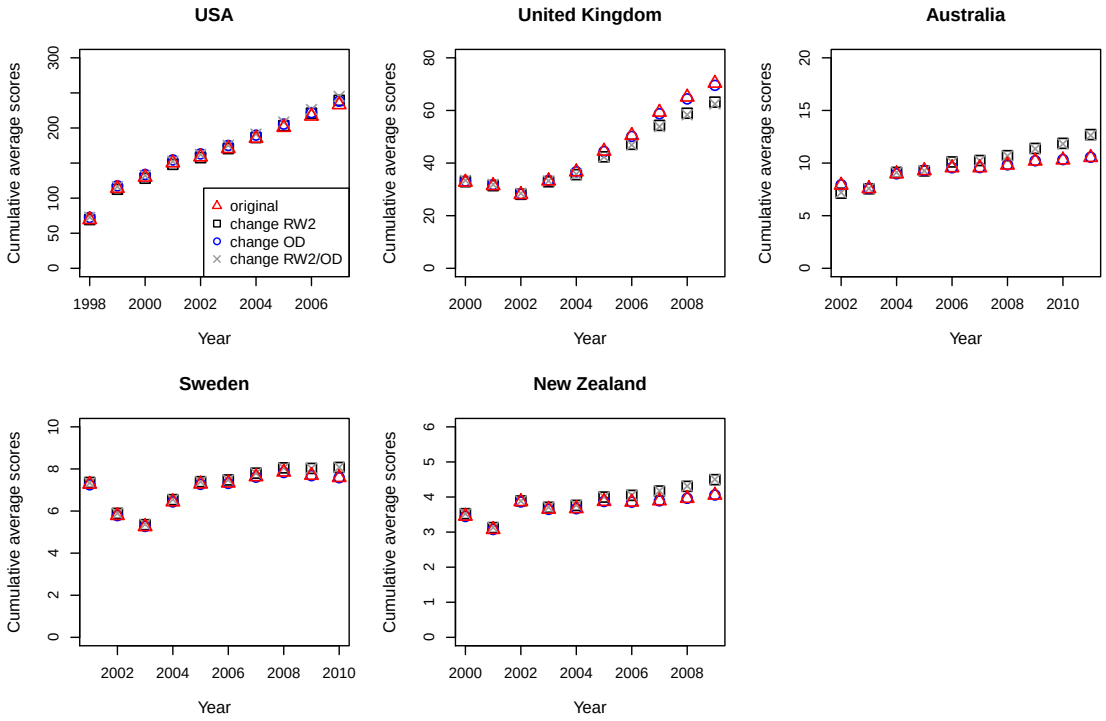

Figure 11 Cumulative average of mean continuous ranked probability scores across age groups for all five countries obtained by the Bayesian APC model using four different hyper prior sets (Original: Ga(1, 5e-5) for RW2, Ga(1, 0.005) for OD; change RW2: SW priors for all RW2; change OD: SW prior for OD; change RW2/OD: SW priors for all RW2 and OD).

Table 6 Empirical coverage of the one-step-ahead predictive credible bands for three different credible levels 50%, 80% and 95%. Shown are results obtained with the Bayesian APC model using three different hyperprior sets compared to the one shown in Table 1 in the main paper where we used a Ga(1, 5e-5) for all RW2s and a Ga(1, 0.005) for OD. (Prior change RW2: SW priors for all RW2; Prior change OD: SW prior for OD; Prior change RW2/OD: SW priors for all RW2 and OD).

	Prior change RW2			Prior change OD			Prior change RW2/OD		
Credible level	50%	80%	95%	50%	80%	95%	50%	80%	95%
USA	37%	67%	93%	36%	67%	92%	37%	69%	93%
United Kingdom	43%	69%	87%	45%	68%	87%	44%	70%	88%
Australia	51%	83%	98%	52%	84%	97%	52%	85%	98%
Sweden	52%	82%	95%	51%	83%	95%	52%	83%	95%
New Zealand	53%	81%	93%	52%	82%	95%	53%	81%	94%

Table 7 Mean absolute error $\overline{\text{AE}}$, mean predictive standard deviation $\overline{\text{SD}}$ and mean continuous ranked probability score $\overline{\text{CRPS}}$ with z-statistic and p-value from the corresponding calibration test. Shown are the results for the one-step-ahead projections obtained with the Bayesian APC model using three different hyperprior sets compared to the one shown in Table 2 in the main paper where we used a Ga(1, 5e-5) for all RW2s and a Ga(1, 0.005) for OD. (Prior change RW2: SW priors for all RW2; Prior change OD: SW prior for OD; Prior change RW2/OD: SW priors for all RW2 and OD).

Priors	Country	$\overline{\text{AE}}$	$\overline{\text{SD}}$	$\overline{\text{CRPS}}$	z	p-value
Prior change RW2	USA	155.05	149.76	108.12	4.59	< 0.0001
	United Kingdom	39.73	41.16	28.73	3.64	0.0003
	Australia	10.82	14.67	7.57	-2.05	0.041
	Sweden	7.96	10.13	5.59	-0.76	0.45
	New Zealand	4.95	6.38	3.55	-0.68	0.49
Prior change OD	USA	159.24	83.40	110.29	4.59	< 0.0001
	United Kingdom	40.31	31.18	28.95	3.61	0.0003
	Australia	10.63	12.45	7.48	-2.44	0.015
	Sweden	7.95	9.01	5.58	-0.91	0.36
	New Zealand	4.86	5.94	3.49	-1.03	0.30
Prior change RW2/OD	USA	158.67	508.12	110.00	3.96	< 0.0001
	United Kingdom	40.15	100.90	28.98	2.91	0.0036
	Australia	10.80	23.91	7.58	-2.54	0.011
	Sweden	7.96	11.99	5.60	-1.18	0.24
	New Zealand	4.95	8.70	3.55	-1.05	0.29

References

Ahmad, O., Boschi-Pinto, C., Lopez, A.D., Murray, C.J.L., Lozano, R., and Inoue, M. (2001) Age standardization of rates: a new WHO standard, in *GPE Discussion Paper Series*, Geneva. (GPE Discussion Paper No. 31).

Baker, A. and Bray, I. (2005) Bayesian projections: What are the effects of excluding data from younger age groups? *American Journal of Epidemiology*, **162** (8), 798–805.

Besag, J., Green, P., Higdon, D., and Mengersen, K. (1995) Bayesian computation and stochastic systems (with discussion). *Statistical Science*, **10** (1), 3–66.

Carstensen, B., Plummer, M., Laara, E., and Hills, M. (2014) Epi: A package for statistical analysis in epidemiology, R package version 1.1.67. `http://CRAN.R-project.org/package=Epi`.

Clayton, D. and Schifflers, E. (1987) Models for temporal variation in cancer rates. II: Age-period-cohort models. *Statistics in Medicine*, **6** (4), 469–481.

Hakulinen, T. and Dyba, T. (1994) Precision of incidence predictions based on Poisson distributed observations. *Statistics in Medicine*, **13** (15), 1513–1523.

Held, L. and Sabanés Bové, D. (2014) *Applied Statistical Inference - Likelihood and Bayes*, Springer, Heidelberg.

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

McCullagh, P. and Nelder, J.A. (1989) *Generalized Linear Models*, no. 37 in Monographs on Statistics and Applied Probability, Chapman and Hall, New York, 2nd edn..

Smith, T.R. and Wakefield, J. (2016) A review and comparison of age-period-cohort models for cancer incidence. *Statistical Science*. To appear.

Sørbye, S.H. and Rue, H. (2014) Scaling intrinsic Gaussian Markov random field priors in spatial modelling. *Spatial Statistics*, **8**, 39–51.

