NTNU – Trondheim
Norwegian University of
Science and Technology

The impact of varying time scales on the quality of cancer projections based on the Bayesian age-period-cohort model

James Korley Attuquaye

Master of Science in Mathematics (for international students)
Submission date: June 2015
Supervisor: Andrea Riebler, MATH

Norwegian University of Science and Technology
Department of Mathematical Sciences

NTNU – Trondheim
Norwegian University of
Science and Technology

James Korley Attuquaye

The impact of varying time scales on the quality of cancer projections based on the Bayesian age-period-cohort model

Master thesis for the degree of MSc.

Trondheim, August 2015

Norwegian University of Science and Technology

Faculty of Information Technology, Mathematics

and Electrical Engineering

Department of Mathematics

Supervisor: Andrea Riebler

Acknowledgments

I am exceedingly thankful to a number of people whose invaluable contribution, advice, love, and support enabled me to successfully complete the writing of this thesis.

First, I owe a debt of gratitude to my supervisor, Andrea Riebler, for her inspiring supervision, guidance and unflinching interest in my research. Her useful comments, invaluable suggestions, and constructive criticisms helped me to tremendously improve upon the overall quality of this thesis. By introducing me to the `BAPC` R-package, explicating its technicalities to me, and helping me out with any challenge on its usage whenever the need arose, she equipped me with a vital tool for cancer research âĂŞ and for that I am more grateful.

I extend my deep appreciation to Norwegian University of Science and Technology (NTNU) for offering me both admission and scholarship through the Quota Scheme Program to enable me to fund my masterâĂŹs program. Most especially, I would like to gratefully acknowledge the benevolence and geniality of the Norwegian government as well as the work of the NTNU Office of International Relations and Lanakassen in facilitating the scholarship.

For the personal support, motivation, and encouragement that kept me going and emboldened me to accomplish my goal, and for the inspiration of knowing that there was someone who always believed in me and âĂŸhad my backâĂŹ, I copiously thank my best friend Anna-Pearl Mante. Your constant words of reassurance and the extreme confidence you consistently reposed in me, Pearl, made me soldier on despite the difficulties âĂŞ we made it!

My profound gratitude also goes to the Ghanaian community in Trondheim and to all my colleagues and friends for the social support they offered me during my stay in Norway. Most especially, I would like to express my heartfelt thanks to the Ganaah

family for their hospitality, thoughtfulness, and selflessness. As well, I would like to acknowledge the friendship of my colleague Shipra Sachdeva. Still, I gratefully acknowledge the efforts of my good friend and brother Mark Nartey for proofreading various aspects of the thesis. Further, my special and unqualified thanks go to my parents (Joshua Attuquaye and Christiana Attuquaye), my uncle (Samuel Clottey), and my siblings for their incessant prayer and unwavering support throughout my study.

And finally Ebenezer! Thus far the Lord has brought me. Thee will I give all the thanks and praise. To the Almighty God who has been my strength, comforter, protector, and sustenance throughout this time of study, I exclusively render all the glory, praise, and honor.

Trondheim, June 2015
James Korley Attuquaye

Abstract

Projection of age-specific cancer incidence and mortality data play an integral role in planning and research. A wide range of methods used for these projections has been developed and put into practice. Many statistical software packages, such as Nordpred and the iterative Lee-Carter package, for projecting age-specific cancer incidence and mortality data implicitly assume that data are aggregated to five-year intervals on the time-scale (periods). However, data aggregation may not always be appropriate and information may get lost. In the field of spatial statistics, care is taken in choosing an appropriate spatial scale to analyse the data of interest; however, less care is taken in choosing an appropriate time-scale in temporal analyses.

In this thesis, the effect of varying temporal scales on the precision and accuracy of projections for selected female cancer mortality data sets from the World Health Organisation Mortality Database is investigated. Three temporal scales have been explored; five-year data aggregation, five-year model-specific aggregation and yearly data structure with no aggregation. Projections are obtained based on these three scales using the `BAPC` R-package, an R-package which implements Bayesian age-period-cohort (APC) models, where smoothing is applied to each time scale. Also, the hyperpriors for the precision parameter of the smoothing effects are carefully chosen to make them transferable between the different time aggregation. The models are estimated using integrated nested Laplace approximations (INLA). Calibration and sharpness of the projections are jointly assessed based on the absolute error and the continuous ranked probability score.

The study shows that annual to five-year data aggregation might not be ideal for projections since clues on yearly trends can not be monitored. However if any form of aggregation needs to be done, then model-specific aggregation might be useful, depending on the application.

Contents

Chapter 1

Introduction

Projections of cancer incidence and mortality with regard to the number of cases, deaths expected and rates relative to a population of interest play a vital role in effective planning and improving health services while serving as a baseline for assessing the impact of public health interventions (Bray and Møller, 2006). Cancer projections made on short term basis allow us to provide reliable estimates of current death rates based on rates observed in the past and also to identify significant changes in cancer trends before they are actually observed while long term projections come in handy when making decisive choices regarding public health interventions (Bray, 2002) like drawing national budgets and developing pragmatic scientific measures for cancer control and prevention. Projections of cancer data are therefore an essential part of cancer intervention programmes (Armstrong, 1992).

National health care registers are commonly used in developed countries for storing data on disease incidence and mortality. Cancer registries, like other disease registries, gather information on the prevalence of cancer in populations by registering cases by diagnosis, age and date of diagnosis, date of death, gender etc. (Carstensen, 2007). To visualise disease or mortality rates by age and time, the Lexis diagram, named after Wilhelm Lexis (Lexis, 1875), is probably the most com-

monly used (Carstensen, 2007). As far as mortality data is concerened, the Lexis diagram shows the number of deaths using the following demographic characteristics; the age of the deceased at the time of death, the time of death (period) and the time of birth of the deceased (cohort) (Vandeschrick et al., 2001) . The diagram provides a two-dimensional coordinate system representation, where age is represented on the vertical axis, while calendar time (period) is represented on the horizontal axis.

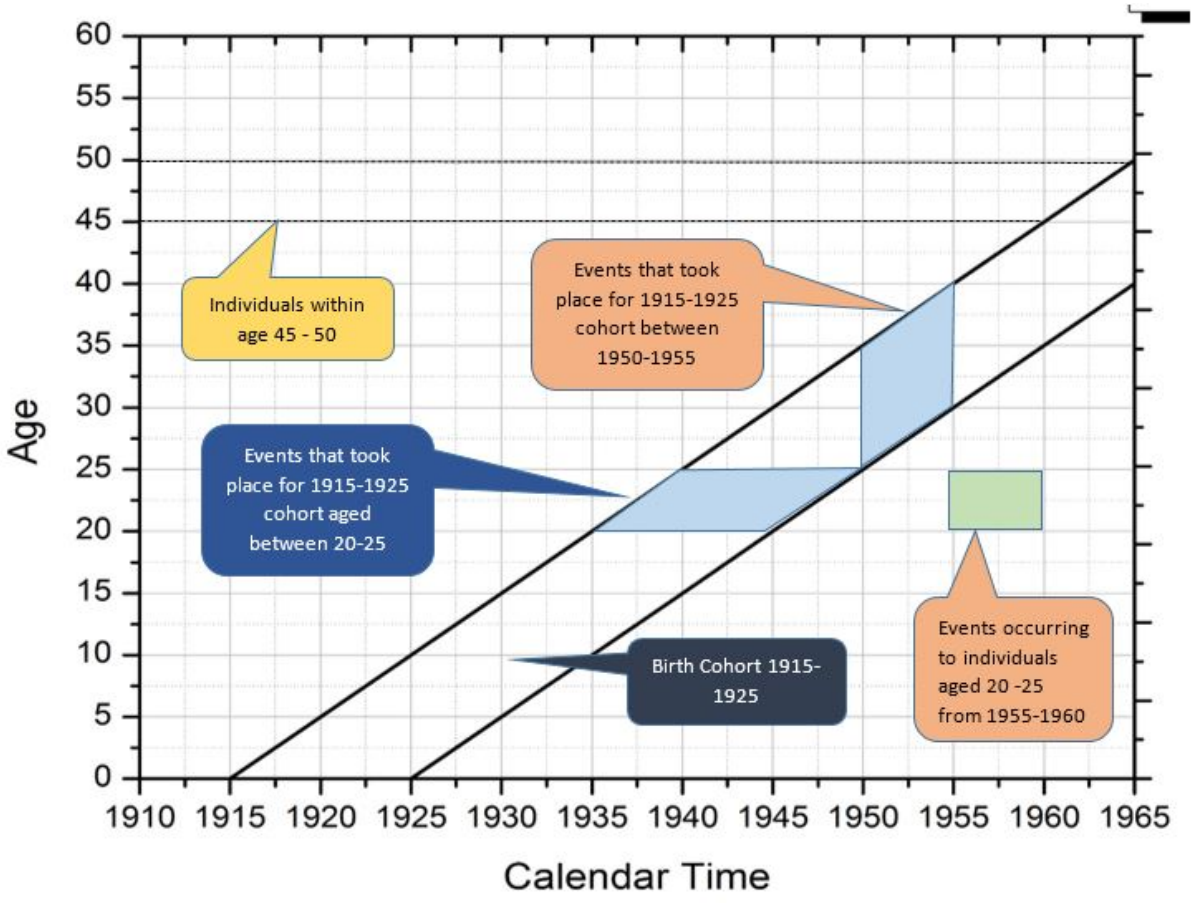


Figure 1: A Lexis diagram representation. Adapted from: `http://papp.iussp.org/sessions/papp101_s02/PAPP101_s02_100_010.html`, accessed 29.08.2014

The cohort is the third time-dimension featured on the diagram, which represents a group of people who were born at the same time and share a common demographic characteristic. It is represented on the Lexis diagram as a diagnonal band. The lives

of members of a particular cohort are displayed as a 45° diagonal line known as a life line (Vandeschrick et al., 2001). Figure 1 summarizes how events are represented on a Lexis diagram. Carrying out descriptive analysis on the diagram may give an indication on the general behaviour and qualitative opinions about future time trends can be formed (Riebler and Held, 2015).

Cancer mortality is commonly known to depend on the time scales age, period and birth cohort (Clayton and Schifflers, 1987a). Berzuini and Clayton (1994) have however drawn attention to the fact that disease events are not caused by time itself and that it only serves as a scale on which other causal factors, that are unknown or difficult to estimate, operate. For instance, the prevalence of chronic conditions such as hypertension, heart disease and arthritis increase with age (Anderson and Horvath, 2004).

Rate changes which are consequences of varying ages are known as age effect (Heuer, 1997). Increased cancer incidence with age, for example, resulting from accumulation of mutated genes (Kennedy et al., 2012) and tissue landscape changes with age (DeGregori, 2013), is an example of an age effect. Rate changes which occur at a particular calender time consistently for each age group are known as period effects; they are often due to new productive improvements in treatment or diagnosis, or a change in classification of disease at a particular date (Heuer, 1997). For instance, if a new chemotherapy is introduced for treatment of a certain cancer for all age groups, the effect it has on cancer mortality at this particular calendar time and beyond is considered as a period effect. Events which affect different birth cohorts (generations) are accounted for by cohort effect (Schmid and Held, 2007), e.g., the varying smoking habits of different generations.

In addition to the strictly descriptive approaches, mathematical alternatives have been developed and put into practice. For example, the Lee-Carter model, introduced in 1992 by Ronald D. Lee and Lawrence Carter in their article *Modelling and Forecasting US Mortality* (Lee and Carter, 1992), is popularly used for mortality

forecasts both in academic literature and practical applications (Girosi and King, 2007). Even though Lee and Carter developed the method mainly for U.S. mortality data ($1933 - 1987$), the approach is broadly applied to all-cause and cause-specific mortality data from several countries (Girosi and King, 2007). There exists also the Nordpred Software package, a classical approach (not Bayesian), written by Harald Fekjær and Bjørn Møller at the Cancer Registry of Norway used to fit an age-period-cohort regression model to observed data and to predict cancer incidence trends (Møller et al., 2003). In view of the colinearity that exists between age, period, and cohort (cohort = period − age), estimates of the linear effects of period and cohort cannot be found concurrently; Nordpred therefore estimates a common linear trend instead (Møller et al., 2002). Here, projected rates are based on the assumption that trends in the past will continue into the future. There exist various statistical models for predictions; see for example Bray and Møller (2006) for an overview.

The age-period-cohort (APC) model is used to analyze rates of diseases of interest in terms of age, period (which considers a group of individuals at a particular point in time), and cohort effects (influence that is considered to have an effect on individuals born around the same time) (Bray et al., 2001). It has its roots in fundamental generalised linear model theory (McCullagh, 1984) and is well known for analyses of age-specific cancer incidence or mortality data. The Bayesian age-period-cohort models, which we employ in this thesis, is used increasingly for cancer incidence and mortality projections (Besag et al., 1995; Bray, 2002); it takes into consideration prior knowledge about smoothness on each time scale to decrease the effect of random variation and has been shown to produce a precision gain in the projections (Bray, 2002). Moreover, unstructured variation can easily be accounted for.

Some statistical software, such as Nordpred (`http://www.kreftregisteret.no/software/nordpred`, accessed: 04.11.2014) and the iterative Lee-Carter (ilc) package (Butt and Haberman, 2009), for projecting age-specific cancer incidence and mortality rates implicitly assume that age and period groups are provided on same

interval lengths, often five-year intervals. This practice may lead to loss of information; for example Ocaña-Riola (2007) demonstrated that aggregation of count data over time may give misleading estimates of relative risks and subsequently introduce a compromise when making decisions. In some cases, yearly data are not readily available; hence, predictions can only be made using the data that are available. In spatial statistics, care is taken in choosing an appropriate spatial scale to analyse data of interest as seen for example in Kang et al. (2014). However, less attention is given to the choice of appropriate time-scale for temporal analyses.

This thesis seeks to investigate the impact of varying time scales on the quality of cancer projections based on the Bayesian age-period-cohort model. Using female lung and oesophagus cancer data in Norway and the United Kingdom, obtained from the World Health Organisation (WHO) mortality database, the study considers age-specific projections based on yearly data, five-year data aggregation and five-year model-specific aggregation, where we maintain the yearly data and aggregate the random effects. In each case, we exclude and predict observations in the last 10 years and are interested in the extent to which predictions agree with the observed counts from the data.

Structure of thesis

Chapter 2 begins by introducing some notation that has been used throughout the work and further reviews two alternative statistical models to the age-period-cohort (APC) model, namely the Lee-Carter and Power models, which are widely used for analyses and projections of age-specific cancer incidence or mortality data. In chapter 3, we present the Bayesian APC model, used in our analysis to obtain age-specific cancer projections under each of the time scales considered in this thesis. Chapter 4 discusses the assessment of quality of the predictions. We introduce a few proper scoring rules that can be used; of particular interest are the absolute

error and continuous ranked probabilty score which we have applied in this work. Chapter 5 begins by introducing the choice of data sets used in our analysis and how the data were obtained. Further, we introduce the different time scales investigated: five-year data aggregation, five-year model-specific aggregation and method of no aggegation and evaluate the predictions in each case based on age-specific plots and proper scoring rules discussed in chapter 4. We give a summary of our findings in chapter 6. Supplementary plots and some selected R-code are provided in the appendix.

Chapter 2

Common models for projection of age-specific cancer rates

Before discussing in detail the APC model, which forms the basis of this thesis, this chapter first reviews two alternative models to the APC model that are often used for cancer projection. The basic assumptions of the Lee-Carter Model and the Power model (also known as Nordpred model) are introduced and a description of how predictions are done using these models is outlined.

Throughout this chapter and also for the rest of the thesis, the following notation will be used:

- n_{ij} represents the number of persons at risk in the ith age group ($i = 1, \ldots, I$) during the jth period ($j = 1, \ldots, J$)
- y_{ij} represents the number of cases in the ith age group at the jth period.
- η_{ij} is the linear predictor for individuals in the ith age group during the jth period.
- λ_{ij} is the mortality rate for individuals in ith age group during the jth period.

- θ_i, φ_j and $\psi_k (k = 1, \ldots, K)$ represents age effects, period effects and cohort effects respectively with $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_I)^\top$, $\boldsymbol{\varphi} = (\varphi_1, \ldots, \varphi_J)^\top$, $\boldsymbol{\psi} = (\psi_1, \ldots, \psi_K)^\top$. Cohorts are written as

$$k = C \cdot (I - i) + j$$

and the total number of cohorts as $K = C \cdot (I - 1) + J$. Here the factor C is defined as the ratio of the width of age group and period intervals (Heuer, 1997).

2.1 Lee-Carter model

The classical Lee-Carter (LC) model originally proposed by Lee and Carter (1992), was first described as a log linear model for mortality rates with normally distributed additive error terms. Brouhns et al. (2002) later extended the classical linear model to a generalized linear model by replacing the additive error term on the logarithm of the mortality rates with a Poisson random variation for the number of deaths. It is important to mention that the Poisson distribution is well-suited for mortality analysis; see, e.g. Brillinger (1986) and Macdonald (1996a,b,c) for further details. We now consider the Poisson setting of the LC model. It assumes that y_{ij} are Poisson distributed with rate $n_{ij}\lambda_{ij}$ where

$$\eta_{ij} = \log(\lambda_{ij}) = \theta_i + \beta_i \kappa_j$$

Here, θ_i describes the average age-specific mortality, β_i represents the patterns of mortality change for ith age group- it indicates the sensitivity of the logarithm of the mortality rate at age group i to the variations in the time index κ_j, and κ_j represents the time trend; the shape of the β_i profile explains which rates decline speedily and which remain steady over time in response of change in κ_j. The model

lacks uniqueness, since one solution can be transformed by a linear combination to produce another set of solution. Lee and Carter proposed two constraints to curb this; ($\sum_j \kappa_j = 0$) and ($\sum_i \beta_i = 1$) (Lee and Carter, 1992). Parameters are determined by maximizing the log-likehood, $l(\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\kappa})$ based on the Poisson specification of the model.

Suppose parameter estimates ($\hat{\theta}_i$'s, $\hat{\beta}_i$'s, $\hat{\kappa}_j$'s) have been found. The time index, κ_j, is intrinsically viewed as a stochastic time series. In order to produce mortality forecasts for some $j = J + h$ beyond the last observed period J, the estimates $\hat{\kappa}_j$ are extrapolated into $\tilde{\kappa}_{J+h}$ using an appropriate ARIMA (p, d, q) time series model. An ARIMA (0, 1, 0), also known as the random walk with drift (RWD) model, was used by Lee and Carter (1992) for their data. They make clear the possibility of using other ARIMA models for different data sets, but in practice, the RWD has been used almost exclusively to model κ_j (Girosi and King, 2007). Then mortality forecasts for age group i at time point $J + h$ are found by inserting the estimates $\hat{\theta}_i$, $\hat{\beta}_i$ and the predicted values $\tilde{\kappa}_{J+h}$ into the model. Variance of the time index, κ_j, increases as projections go far into the future, and is used to estimate the uncertainty of a forecast.

The model assumes the absence of age×time interactions- that β_i is constant over time for all age groups and for all j, κ_j is constant over all age groups. Several studies have however shown that parameters are not constant over time, hence assuming such, increases the error associated with prediction, particularly for older age groups (Carter and Prskawetz, 2001). The assumption that the component of age is constant over time is the major challenge of the Lee-Carter method (Lee and Miller, 2001). Several modifications of the LC model have been proposed to solve this challenge of parameterisation; see (Tuljapurkar et al., 2000; Cairns et al., 2006; Carter and Prskawetz, 2001). Renshaw and Haberman (2006) also proposed an extension of the model which incorporates a cohort effect. Nevertheless, problems might be encountered in parameter estimation if a cohort effect is introduced (Yang et al., 2010). To implement methods that fit a generalised class of Lee-Carter models, the

iterative Lee-Carter (ilc) package, an R statistical software package (R Core Team, 2013), can be used (Butt and Haberman, 2009). The package assumes same lengths for age and period intervals, mostly five-year widths.

2.2 Power model

The power model fits a regression model to observed mortality data and is particularly used for long term prediction of trends in cancer mortality (Møller et al., 2003). It assumes that the number of cases, y_{ij} are Poisson distributed with rate $n_{ij}\lambda_{ij}$. Instead of a log-link, it adopts a power-link function with constant power of five. Motivation for the power link is that trends in the far future will not behave exponentially. Consequently in a bid to level off exponential growth, a power link is suggested (Møller et al., 2003). The model is of the form

$$\lambda_{ij} = (\theta_i + \varphi_j + \psi_k + D \cdot j)^5$$

where D is a common drift parameter; an overall trend term, θ, φ, and ψ are the non-linear components of age i, period j and cohort k, respectively. Jürgens et al. (2014) formulated this commonly used power model within the Bayesian framework, and further extended the model to allow for a random power parameter instead of the fixed one proposed by Møller (2004). In the power model, projected rates rely on the assumption that cohort-specific trends will carry on into the future. Birth cohort is calculated by subtracting age from calendar period (i.e $k = j - i$). In view of this colinearity between age, period and cohort, it is impossible to find estimates of the linear effect of period and cohort uniquely, but a common linear term known as the common drift parameter (D) can be estimated (Clayton and Schifflers, 1987b). However, cohort-specific deviations from D can be estimated and assumed to continue into the future.

Predictions are made by projecting forward the drift component, D, in order to obtain future estimates of cancer incidence (Møller et al., 2003). The age effects are assumed to be the same as the ones estimated from the existing data and the future non-linear components of period and cohort are set equal to the last estimated effect in the range of the data (Rutherford et al., 2012). In the implementation of the power model, a test is performed after fitting the model to determine if predictions should be based on the estimate of the time trend, $\hat{D}$ over the whole period of study or an estimated trend, $\tilde{D}$, which is based on the most recent ten years (Møller et al., 2002; Nowatzki et al., 2011). This test seeks to assess the presence of a significant curvature in the trend over recent periods. The detection of such a significant change suggests the use of $\tilde{D}$ as estimated trend coefficient in predictions. Møller et al. (2003) also suggested a gradual reduction in the trend coefficient (either D or $\tilde{D}$) for future predictions. For instance, instead of adding D to each new period, Møller et al. (2002) adds D, $0.75D$, 0.50, $0.25D$ and $0.25D$ to the latest five future period predictions. The reduction is based on the assumption that long term future trends will most likely not continue; current trends are of much interest to future predictions than those observed in the far past and that with time, factors affecting distant trends will gradually fade out (Møller et al., 2003). The model however performs poorly when used to predict sparse data (data in which a relatively high percentage of the entries do not have actual data, such "NA" or "empty") (Jürgens et al., 2014).

The power model is implemented in the `R` statistical software package (R Core Team, 2013), Nordpred, developed by the Cancer Registry of Norway; see (`www.kreftregisteret.no/software/nordpred`, accessed: 04.11.2014) (Møller et al., 2003). Of note, Nordpred usually uses 5-year aggregated data and provides no uncertainty estimates with predictions.

Chapter 3

The Bayesian age-period-cohort model

Age-period-cohort (APC) models are used for description and prediction of time series of age-specific mortality rates using three different time scales: age, period and birth cohort. In this chapter, an introduction to the main characteristics of APC analysis will be given and how it differs from the Lee-Carter and Nordpred model.

3.1 The APC model

Here we assume that the number of cases y_{ij} in age group i and at period j follows a Poisson distribution with rate $n_{ij} \times \lambda_{ij}$ and that the product of the corresponding Poisson terms gives the likelihood for the whole data. In the classical APC model

$$\eta_{ij} = \log(\lambda_{ij}) = \mu + \theta_i + \varphi_j + \psi_k.$$

The birth cohort index, k is directly determined by the age index i and the period index j. In the case of same interval widths for age group and period, $k = (I-i)+j$ and $K = (I-1)+J$. However, the definition of k has to be altered slightly if age group and period are defined on different time scales. For instance, Holford (1983) considered the case in which period intervals are C times wider than the age group intervals. In this thesis, we consider such a case where data are originally obtained with interval width of the age groups 5 times wider than the interval width of the periods. Suppose age is given in C-year intervals while period is given on yearly basis then, $k = C \times (I-i)+j$ and $K = C \times (I-1)+J$ (Heuer, 1997). Here, the factor C is defined as the ratio of the width of age group and period intervals.

To ensure identifiability of the intercept, so called sum-to-zero constraints ($\sum_i \theta_i = \sum_j \varphi_j = \sum_k \psi_k = 0$) are imposed (Holford, 2005). There is also a second identifiability problem due to the existence of the exact linear relationship between age, period and cohort effects (cohort can be expressed as a linear combination of the period and age) (Holford, 2005; Held and Riebler, 2013). Log-linear trends in rates cannot be accredited to the influences of age, period or cohort exclusively without introducing further assumptions that cannot be verified, due to the existence of colinearity between the three time scales (Holford, 1998). That is, there exist linear transformations of θ_i, φ_j and ψ_k that leads to same value of η_{ij}, i.e the linear predictor remains unaltered; see for example (Holford, 2005) and (Schmid and Held, 2007). Suppose age group interval is not equal to the period interval, then for transformations of the form

$$\theta_i \mapsto \theta_i + C\times\gamma\left(i - \frac{I+1}{2}\right);\ \varphi_j \mapsto \varphi_j - \gamma\left(j - \frac{J+1}{2}\right);\ \psi_k \mapsto \psi_k + \gamma\left(k - \frac{K+1}{2}\right),$$

for all i, j and k with $\gamma \in \mathbb{R}$, the linear predictor η_{ij} remains unchanged. This problem is seen as a distinct shortcoming of the APC model. Potential solutions to the identifiability problem have however been proposed- see for example Holford (1998); Clayton and Schifflers (1987b); Osmond and Gardner (1982). Irrespective

of this identifiability problem, non-linear trends, for instance, change points are interpretable because the non-identifiability only affects linear trends (Clayton and Schifflers, 1987b).

In this work, we are interested in predictions, which are only based on the values of η_{ij}. Hence, the identifiability problem is not an issue of concern in this case since η_{ij} is identifiable (Holford, 1985; Schmid and Held, 2007). We therefore employ the APC model without imposing any additional constraints for predictions in this work.

The main difference between the APC model and the Lee-Carter model is the bilinear effect ($\beta_i \kappa_j$) in the Lee-Carter model whereas the APC model considers only additive effects. Further, the Lee-Carter model commonly has no cohort effect. Softwares that implement the power and Lee-Carter models, Nordpred and the iterative Lee-Carter (ilc) package respectively implicitly assume same widths for age and period intervals, often five-year intervals. The power model, unlike the APC and Lee-Carter models, assumes a power link instead of a log link in the Poisson regression. Of note, the Nordpred software package (`http://www.kreftregisteret.no/software/nordpred`, accessed: 04.11.2014) does not report uncertainty estimates with its predictions.

3.2 Bayesian inference

The Bayesian age-period-cohort (BAPC) model imposes prior information about smoothness on each of the parameter vectors ($\boldsymbol{\theta}$, $\boldsymbol{\varphi}$ and $\boldsymbol{\psi}$) and also on hyperparameters (i.e precision parameters) to avoid overfitting and to improve precision of the projections (Knorr-Held and Rainer, 2001). It also minimizes random variation unexplained by age, period and cohort alone by introducing an additional parameter $z_{ij} \sim \mathcal{N}(0,\, \delta^{-1})$ to explain this random variation (Knorr-Held and Rainer, 2001). In this formulation, a random walk 2 prior, which is like an autoregressive process but non-stationary, that smooths effects on each time scale (age, period and cohort)

is specified, coercing parameter estimates not to differ unreasonably from those in neighbouring time intervals (Bray et al., 2000). Furthermore, difficulties resulting from the existence of colinearity in the APC models are avoided if interest lies in estimable functions (Besag et al., 1995).

Bayesian approaches are commonly used in the projection of mortality rates because they are not based on strict parametric assumptions for future values of period and cohort effects. Bray (2002) reviewed certain classical methods for prediction of incidence or mortality data, and provides a comparison to the classical and Bayesian versions of the APC model, and established that the Bayesian age-period-cohort model was the only approach to obtain reasonable predictions. Bayesian APC models have been applied and discussed in several articles; see for example Berzuini and Clayton (1994); Besag et al. (1995); Lunn et al. (2000); Knorr-Held and Rainer (2001); Bashir and Estève (2001); Bray et al. (2001); Bray (2002); Schmid and Held (2007); Riebler and Held (2015). Softwares such as **B**ayesian analysis software **U**sing **G**ibbs **S**ampling for **W**indows (WinBUGS) (Lunn et al., 2000) and **B**ayesian **A**ge-Period-Cohort **M**odeling and **P**rediction (BAMP) (Schmid and Held, 2007) are widely used softwares which implement the Bayesian approach of the age-period-cohort model, using Markov chain Monte Carlo (MCMC) algorithms. More recently, an R-package, `BAPC`, which uses integrated nested Laplace approximations (see section 3.2) for predicting future cancer rates and counts based on the Bayesian age-period-cohort model has been developed (Riebler and Held, 2015).

3.2.1 Smoothing priors for the time effects

We start by introducing Intrinsic Gaussian Markov random fields (IGMRFs). A vector $\boldsymbol{x} = (x_1, \ldots, x_n)^\top \in \mathbb{R}^n$ can be defined as an IGMRF if it has an improper Gaussian density with a sparse precision matrix which is not of full rank and has an order defined as the rank deficiency of its precision matrix, $\boldsymbol{Q}$ (Rue and Held, 2005). The precision matrix $\boldsymbol{Q}$ is sparse because it is endowed with the Markov

properties: $Q_{ij} = 0$ if and only if x_i and x_j are conditionally independent given $\boldsymbol{x}_{-ij} = (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_{j-1}, x_{j+1}, \ldots, x_n)^\top$. Thus, a zero-mean IGMRF of pth order has the form:

$$\pi(\boldsymbol{x}) = (2\pi)^{-(n-p)/2}(|\boldsymbol{Q}|^\star)^{1/2} \exp\left(-\frac{1}{2}\boldsymbol{x}^\top \boldsymbol{Q}\boldsymbol{x}\right).$$

Here, $\boldsymbol{Q}$ is a precision matrix assumed to be symmetric positive semi-definite with rank $n-p$ and $|\boldsymbol{Q}|^\star$ denotes the generalised determinant of $\boldsymbol{Q}$, defined as the product of the $n-p$ non-zero eigenvalues of $\boldsymbol{Q}$. An IGMRF can be described by expressing the precision matrix as $\boldsymbol{Q} = \kappa \boldsymbol{R}$, with marginal variances given by the diagonal elements of the generalized inverse matrix $\boldsymbol{\Sigma}^\star = \boldsymbol{Q}^{-1}$. Here, κ represents the random precision parameter while $\boldsymbol{R}$ is the so-called structure matrix which reflects the specific neighbourhood structure of the model. It can be represented as a labelled graph with nodes and edges; hence, the marginal variances of an IGMRF will depend on both the size and structure of $\boldsymbol{R}$ (Sørbye and Rue, 2014).

In APC models, it appears feasible that neighbouring effects in time are almost identical, so that Gaussian priors with zero mean are mostly used for pairwise differences (Besag et al., 1995). For instance, let $\mathbf{Q}^{(1)}$ be a precision matrix defined as

$$\mathbf{Q}^{(1)} = \kappa_\theta \begin{pmatrix} 1 & -1 & & & & & & \\ -1 & 2 & -1 & & & & & \\ & -1 & 2 & -1 & & & & \\ & & \ddots & \ddots & \ddots & & & \\ & & & -1 & 2 & -1 & & \\ & & & & -1 & 2 & -1 & \\ & & & & & -1 & 1 & \end{pmatrix}$$

where κ_θ is the precision parameter, which determines the degree of smoothness and will be estimated concurrently in the model. Higher precision yields smoother estimates of the parmeter vector. A prior based on the first-order differences for the age effects

$\boldsymbol{\theta} = (\theta_i, \ldots, \theta_I)^\top$ is given by

$$f(\boldsymbol{\theta} \mid \kappa_\theta) \propto \kappa_\theta^{(I-1)/2} \exp\left(-\frac{\kappa_\theta}{2}\sum_{i=2}^{I}(\theta_i - \theta_{i-1})^2\right)$$
$$= \kappa_\theta^{(I-1)/2} \exp\left(-\frac{1}{2}\boldsymbol{\theta}^\top \mathbf{Q}^{(1)}\boldsymbol{\theta}\right).$$

Note that unspecified entries in $\mathbf{Q}^{(1)}$ are zero. This prior corresponds with the directed formulation as a random walk of first order (RW1); $\theta_i \mid \theta_{i-1}, \ldots, \theta_1 \sim \mathcal{N}(\theta_{i-1}, \kappa_\theta^{-1})$, $i = 2, \ldots, I$ with a flat prior for θ_1 ($q(\theta_1) \propto constant$).

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

$$f(\boldsymbol{\theta} \mid \kappa_\theta) \propto \kappa_\theta^{(I-2)/2} \exp\left(-\frac{\kappa_\theta}{2}\sum_{i=3}^{I}((\theta_i - \theta_{i-1}) - (\theta_{i-1} - \theta_{i-2}))^2\right)$$
$$= \kappa_\theta^{(I-2)/2} \exp\left(-\frac{1}{2}\boldsymbol{\theta}^\top \mathbf{Q}^{(2)}\boldsymbol{\theta}\right).$$

here,

$$\mathbf{Q}^{(2)} = \kappa_\theta \begin{pmatrix} 1 & -2 & 1 & & & & & \\ -2 & 5 & -4 & 1 & & & & \\ 1 & -4 & 6 & -4 & 1 & & & \\ & 1 & -4 & 6 & -4 & 1 & & \\ & & \ddots & \ddots & \ddots & \ddots & \ddots & \\ & & & 1 & -4 & 6 & -4 & 1 \\ & & & & 1 & -4 & 5 & -2 \\ & & & & & 1 & -2 & 1 \end{pmatrix}$$

The corresponding directed formulation is a random walk of second order (RW2), $\theta_i \mid \theta_{i-1}, \ldots, \theta_1 \sim \mathcal{N}(2\theta_{i-1} - \theta_{i-2}, \kappa_\theta^{-1})$, $i = 3, \ldots, I$ with independent uniform priors both for θ_1 and θ_2. While first order random walk prior penalises deviations from a constant trend, second order random walk penalises deviations from a linear trend $2\theta_{i-1} - \theta_{i-2}$, $i = 3, \ldots, I$. Both RW1 and RW2 are well known examples of IGMRFs. The RW2 prior is particularly of interest in this work since it penalises the second order differences which are identifiable in APC models and also produces smoother estimates of the parameter vector

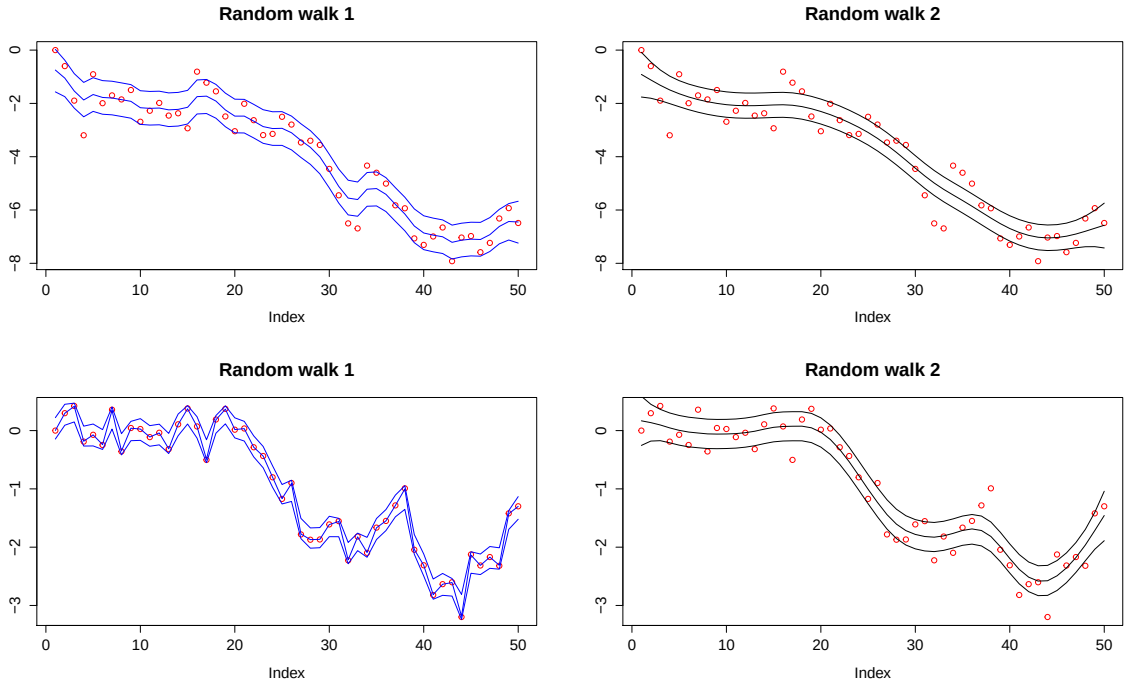


Figure 3.1: Simulated random numbers (points) together with posterior means and quantiles (95 % confidence intervals) using RW1-model (left) and RW2-model (right) as smoothing priors with fixed precision $\kappa = 1$. Shown are RW1 and RW2 smoothing for two different data sets (top and bottom).

than the RW1 prior. For example, we see in Figure 3.1, where we have used the RW1 and RW2 as smoothing priors for two simulated data sets (see Appendix B:1 for R-code), that the posterior means and quantiles (95 % credible intervals) of the RW2-model are smoother than in the case of the RW1 model and that the RW1 smoothing exhibits more dependence on the data. We prefer smoother estimates in order to reduce the dependence of projections on a local trend in data. The priors assumed for the age effects are the same used for the period and cohort parameters with κ_φ and κ_ψ denoting their precision parameters respectively.

A riveting characteristic of the RW1 is that it provides a solution to the implicit identifiability problem present in the APC model by imposing a restriction, not deterministic but stochastic. Considering all possible transformations of the parameters, the RW1 model will prefer the one that minimizes the quadratic first differences (i.e θ_i, φ_j and ψ_k are kept as constant as possible) (Knorr-Held and Rainer, 2001). This characteristic makes

it possible to study the existing trends in the APC model parameters under this implicit constraint. On the contrary, the identifiability problem remains unsolved in the RW2 model because the linear transformation of parameters neither alters the likelihood nor the prior (Knorr-Held and Rainer, 2001). In the Bayesian setting, however, it is not of utmost importance to establish identifiability of the parameters provided that quantities of interest (in our case, predictions) are identifiable (Besag et al., 1995).

One benefit of the Bayesian procedure is that uncertainty about the precision parameters (κ_θ, κ_φ and κ_ψ) is integrated in the estimation of the APC model parameters (θ_i, φ_j and ψ_k). To account for further "unstructured" heterogeneity which cannot be explained by age, period or cohort effects, Knorr-Held and Rainer (2001) proposed the inclusion of an additional parameter z_{ij} to capture this unexplained variation. The extended model will be

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

Here, a Gaussian distribution $z_{ij} \sim \mathcal{N}(0,\, \delta^{-1})$, with precision parameter δ is used as prior for the additional heterogeneity.

3.2.2 Hyperpriors

In a full Bayesian analysis, hyperprior distributions must be defined for the precision parameters (κ_θ, κ_φ, κ_ψ and δ). To avoid complications with improper hyperpriors, weakly informative but proper gamma distributions, $\texttt{Ga}(a,\, b)$ are assigned to all the precision parameters. In the same spirit as Knorr-Held and Rainer (2001), we use $a = 1$, $b = 0.00005$ for the precisions of the time effects (age, period and cohort effects) and $a = 1$, $b = 0.005$ for δ, precision of the overdispersion, as the overdispersion parameters are expected to have a slightly larger variation.

It is vital that appropriate hyperpriors are chosen for the precision parameters since this influences the degree of smoothness of the field and can have a strong influence on the posterior results (Sørbye and Rue, 2014). However, assigning a specific fixed hyperprior for the precisions of different IGMRF models (such as RW1, RW2 or spatial models) is

questionable since the structure matrices for each model differ and therefore the marginal variances of these models are unequal (Sørbye and Rue, 2014).

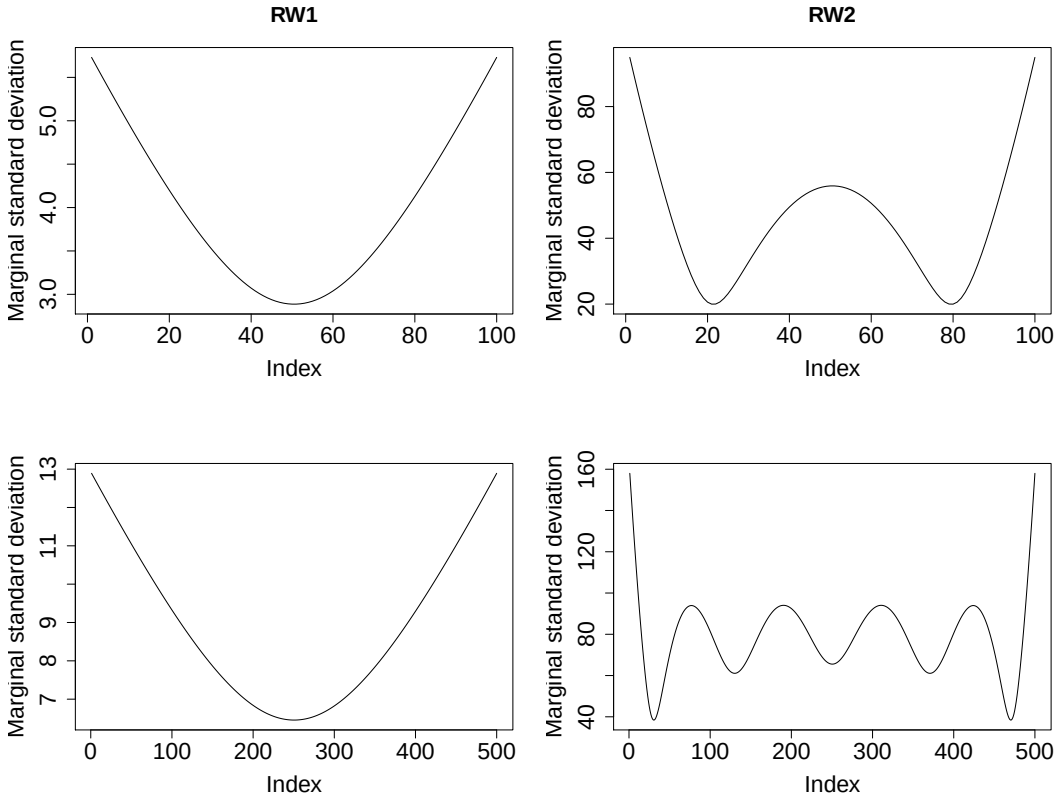


Figure 3.2: The marginal standard deviations for the RW1-model (left) and the RW2-model (right), computed using $n = 100$ points (top) and $n = 500$ points (bottom), with a fixed precision $\kappa = 1$.

Given a fixed precision κ, the marginal standard deviation of the components of a Gaussian vector $\boldsymbol{x}$ can be written as a function of κ as

$$\sigma_\kappa^2 = \frac{1}{\kappa} \exp\left(\frac{1}{n}\sum_{i=1}^{n} \log\left(\boldsymbol{Q}_{ii}^{-1}\right)\right), \; i = 1, \ldots, n.$$

For a fixed precision of one, $(\boldsymbol{Q}^{-1})_{ii}$ are the diagonal elements of the generalized inverse matrix $\boldsymbol{\Sigma}^\star = \boldsymbol{Q}^{-1}$ which represents the marginal variances of the Gaussian vector.

Figure 3.2 shows the marginal standard deviations for all components of the RW1 and RW2 models using fixed precision $\kappa = 1$, each defined on a graph with length $n = 100$ and also with $n = 500$ equidistant points. A consequence of the different lengths ($n = 100$ and $n = 500$) as well as different structures for the two models is that both the shape and level of these curves are different. It is evident from Figure 3.2 that the marginal standard deviation differs for the RW1 and RW2 priors and also for different lengths of the graph. This should therefore be taken into consideration when specifying hyperpriors to the precision parameters.

We follow Sørbye and Rue (2014), who recently proposed an approach for assigning hyperpriors specific for IGMRFs. The technique is to assign priors to scaled precision parameters based on their marginal standard deviations and hence ensure that the hyperpriors are not affected by the choice of IGMRF prior and the length of the graph. This ensures reasonable comparison between different IGMRF-models or same IGMRF-models with different lengths of the graph. Hence, for each IGMRF $\boldsymbol{x}$ with random precision κ, a reference standard deviation $\sigma_{\text{ref}}(\boldsymbol{x})$ for fixed $\kappa = 1$ is calculated as the geometric mean of the marginal standard deviations. An upper limit, U, for the marginal standard deviation is defined, indicating how large the marginal standard deviation, $\sigma(x_i)$ is allowed to be, such that

$$P(\sigma(x_i) > U) \approx P\left(\frac{\kappa}{\sigma^2_{\text{ref}}(x)} < \frac{1}{U^2}\right) = \alpha,$$

where α is a fixed small probability. This means that by asssigning a hyperprior to the scaled precision $\kappa/\sigma^2_{\text{ref}}(x)$, we establish the same interpretation for different models. Due to this, we can impose a common upper limit for the standard deviation of different models and redetermine hyperpriors between different IGMRF models.

If a Gamma distribution, $\texttt{Ga}(a, b)$, is assigned to the precision parameter, κ, the hyperprior assigned to the scaled precision, $\kappa/\sigma^2_{\text{ref}}(x)$, is given by $\texttt{Ga}(a, b\sigma^2_{\text{ref}}(x))$. Thus, it is easy to account for different marginal variances of IGMRFs with a common shape parameter a and an inverse scale parameter adjusted to the model. In our application, where we have

used $a = 1$ and $b = 0.00005$ for example, it implies that

$$U = \sqrt{\frac{b}{F^{-1}(\alpha, 1, 1)}} = \sqrt{\frac{0.00005}{F^{-1}(0.05, 1, 1)}} = 0.031 \quad \text{for} \quad \alpha = 0.05$$

and defines an upper limit for the marginal standard deviations, such that $\mathrm{P}(\sigma(x_i) > U) = 0.05$.

3.2.3 Model estimation using intergrated nested Laplace approximations (INLA)

Markov chain Monte Carlo (MCMC) sampling techniques have often been used for inference with respect to latent Gaussian models. Applying MCMC techniques to such models has been shown to perform poorly; one of the reasons fueling such poor performance is that the components of the latent field depend heavily on one another (Rue et al., 2009).

Rue et al. (2009) proposed a new and effective approach known as integrated nested Laplace approximations (INLA), for full Bayesian inference on the class of latent Gaussian models. INLA is a deterministic algorithm that employs nested Laplace approximations for its computations. It has been optimized for inference on the class of latent Gaussian models, and thus obtains approximations to posterior marginal distributions within very short computation time unlike long running times of commonly used MCMC sampling techniques for latent Gaussian models (Rue et al., 2009).

For a wide range of INLA applications, Rue et al. (2009) have shown that the approximations are of very high quality, yielding more precise posterior estimates than MCMC methods. For example, it has been demonstrated that INLA performs well in the estimation of stochastic volatility models (Martino et al., 2011) in the analysis of animal models (Holand et al., 2013), in meta-analysis (Paul et al., 2010), and in space-time modelling (Schrödle and Held, 2011a,b), just to mention a few. The software can be freely downloaded from www.r-inla.org and run under Linux, Windows and Macintosh via an R-Interface (R Core Team, 2013).

When assigning priors in INLA, we can scale them to have a generalized variance of 1 (as described in the preceding subsection) by specifying "scale.model=TRUE" (Sørbye and Rue, 2014). See for example, line 18 – 24 of Appendix B:3, where the priors for the RW2 variance parameters of the time effects (age, period and cohort) have not been scaled and hence "scale.model=FALSE". For a detailed description of the INLA methods, we refer to Rue et al. (2009).

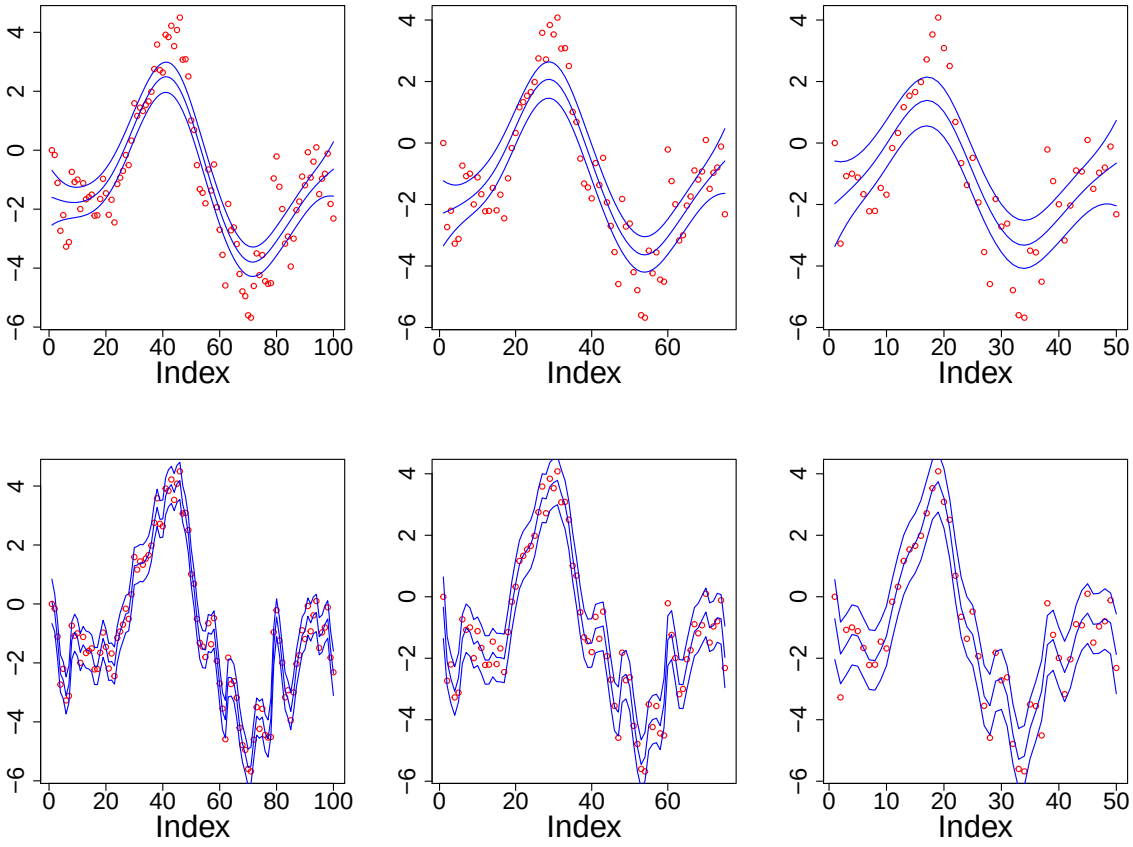


Figure 3.3: Simulated random numbers (points) together with posterior means and quantiles (95% confidence intervals) fitted with RW2-model as smoothing prior and fixed precision $\kappa = 1$. For the top three plots, we have assigned priors to scaled precision parameters but for bottom three, precision parameters have not been scaled.

Chapter 4

Assessment of predictions

Forecasts play a vital role in a wide range of diverse disciplines, including meteorology, agriculture, finance, medicine and betting in sports - see for example, Pulwarty and Redmond (1997) and Jones et al. (2000). Even laypersons often rely on forecasts to make decisions; for instance, making a choice of clothing based on a weather forecast. Projections of cancer mortality or incidence can have a great impact on public health planning (Bray, 2002); projections give prior knowledge as to how limited resources can be distributed and how lives can be saved by introducing preventive measures. It also puts researchers in a better position to earn grants to facilitate research into specific cancer predicted to have increasing mortality rates. In view of the value of predictions, its accuracy must not be overlooked.

Forecasts are nowadays often probabilistic, i.e they take the form of a probability distribution over future happenings (Dawid, 1984) so that their expected accuracy can be assessed (Keilman et al., 2002). Single-valued or point forecasts can also be used for predictions. However, probabilistic forecasts are considered to be more effective than single-valued forecasts, which provide no uncertainty with the forecasts. In recent times, more attention has been given to probabilistic forecasts in a wide range of diverse disciplines (Gneiting, 2008).

Calibration is an important concept when it comes to probabilistic forecasts (Gneiting and Raftery, 2007). It reflects the statistical closeness between the probabilistic forecasts and the actual observations (Gneiting and Raftery, 2007). In simple terms, it explains how reliable a probabilistic forecast is by comparing it with the actual behaviour of the variable predicted. Ideally, observations can be seen as random realisations of the predictive distribution. Forecast sharpness is another vital characteristic of the probabilistic forecasts that cannot be overlooked. It refers to the concentration of the predictive distribution, reflecting the degree of confidence in the forecast (Gneiting and Raftery, 2007).

To illustrate this characteristic, we consider the prediction interval corresponding to a predictive distribution. If, for instance, lung cancer rate for females aged $25-29$ in a country is projected to surely be within the range $0.002-0.003$ per $100,000$ person years, it would be considered a very sharp prediction as compared to a rather vague prediction interval of $0.002-0.900$ per $100,000$ person years. In as much as a forecaster would love to make sharp predictions (i.e obtain a narrow prediction interval), the calibration of the forecast cannot be ignored (i.e the forecast must be statistically consistent with the true observation). Gneiting and Raftery (2007) therefore hypothesized that, the challenge of obtaining quality probabilistic forecasts can be seen as the issue of "maximizing the sharpness of the predictive distributions, subject to calibration".

Following the frequent use of probabilistic forecasts, methods that measure their quality have been suggested. In this chapter, we present some of the methods used for assessment of probability forecasts or predictive distributions. In this work, we follow (Riebler and Held, 2015), assuming a normal predictive distribution for the forecasts.

4.1 Proper scoring rules

Scoring rules are methods used to evaluate the quality of probabilistic forecasts. They are a function of the predictive distribution and the value that is observed (Gneiting and Raftery, 2007). A scoring rule is considered to be proper if it is maximized by forecasting the true distribution. Scoring rules are required to be proper in order to promote truthful and

cautious forecasting (Gneiting and Katzfuss, 2014). Using proper scoring rules, both the sharpness and the calibration of predictive distributions can jointly be evaluated (Gneiting and Raftery, 2007). In what follows, we will present some proper scoring rules. The scores presented here are negatively oriented, i.e smaller scores suggest that predictions are of better quality.

4.1.1 Logarithmic score

The logarithmic score proposed by Good (1952) is a proper scoring rule that depends on the predictive distribution $F(.)$ only through the probabilty mass $f(y_{ij})$ at the actually observed count, y_{ij} in age group i and period j. The logarithmic score is defined as

$$\mathrm{LS}(F, y_{ij}) = -\log f(y_{ij}),$$

Consider a probabilistic forecast Y_{ij} which assumes a normal density $\mathcal{N}(\mu_{ij}, \sigma_{ij}^2)$, then the logarithmic score becomes

$$\mathrm{LS}(y_{ij}) = \frac{1}{2}\left[\log(\sigma_{ij}^2) + \tilde{y}_{ij}^2\right],$$

where $\tilde{y}_{ij} = (y_{ij} - \mu_{ij})/\sigma_{ij}$ are standardized observed number of cases with respect to their predictive distributions and the constant $\log(2\pi)/2$ has been omitted.

4.1.2 Continuous ranked probability score

This proper scoring rule is specified directly in terms of the cumulative distribution function (CDF), $F(.)$, of the forecast distribution (Gneiting and Raftery, 2007). It evaluates the closeness of the predictive distribution and the actual realization. The continuous ranked probability score (CRPS) is defined as

$$\mathrm{CRPS}(F, y_{ij}) = \mathbb{E}_F\{|Y_1 - y_{ij}|\} - \frac{1}{2}\mathbb{E}_F\{|Y_1 - Y_2|\},$$

where Y_1 and Y_2 are independent random variables with CDF F with finite first moment and $\mathbb{E}_F$ is the expectation operator. Assuming normality of the probabilistic forecast with density $\varphi(.)$ and CDF $\Phi(.)$, the CRPS is obtained as

$$\text{CRPS}(F, y_{ij}) = \sigma_{ij}\left[\tilde{y}_{ij}\{2\Phi(\tilde{y}_{ij}) - 1\} + 2\varphi(\tilde{y}_{ij}) - \frac{1}{\sqrt{\pi}}\right]$$

The CRPS is a generalised form of the absolute error for the ij-th observation defined as $|y_{ij} - \mu_{ij}|$. That is, if F is a point forecast, then the CRPS reduces to the absolute error; hence, it enables a forecaster to make comparison between point and deterministic forecasts (Gneiting and Raftery, 2007). The absolute error (AE) measures how far the mean of the predictive distribution is, in absolute value, from the observations and is mostly used to evaluate the quality of point predictions.

The average CRPS in all age groups and in all predicted periods up to and including the jth predicted period, is called the cumulative average $\overline{\text{CRPS}}_j$. It can be used to compare the change in predictive quality from short-term to long-term forecasts (Riebler et al., 2012). As an overall criterion to assess the quality of different probabilistic prediction models, the mean CRPS, $\overline{\text{CRPS}} = \overline{\text{CRPS}}_J$ can be used (Gneiting and Raftery, 2007). The cumulative average of the absolute error, $\overline{\text{AE}}_j$ and the corresponding mean AE, $\overline{\text{AE}}$ are defined analogously.

4.1.3 Dawid-Sebastiani score

In situations where the forecast distribution is complex, it can be problematic to apply the expectation operator in the CRPS fomula, and can therefore be difficult to compute the CRPS (Gneiting and Katzfuss, 2014). The proper Dawid-Sebastiani score (DSS) is a feasible alternative since it depends on the probabilistic forecast only through its first two central moments, μ_{ij} and σ_{ij} (Dawid and Sebastiani, 1999). It is computed by the equation,

$$\text{DSS}(F, y_{ij}) = \tilde{y}_{ij}^2 + 2\log\sigma_{ij}$$

For Gaussian predictive distributions, the logarithmic score can be seen as a special case of the Dawid-Sebastiani score (Dawid and Sebastiani, 1999; Gneiting and Raftery, 2007) and the Dawid-Sebastiani sccore has (up to a multiplicative constant) the same form as the logarithmic score (Held et al., 2010; Gneiting and Katzfuss, 2014):

$$\mathrm{DSS}(F, y_{ij}) = \frac{1}{2}\left[\log(\sigma_{ij}^2) + \tilde{y}_{ij}^2\right],$$

In our application, the predictive distribution is assumed to be normal (Riebler and Held, 2015) and thus CRPS can easily be computed. In order to assess the quality of predictions obtained under the different time scales, the cumulative average $\overline{\mathrm{CRPS}}_j$ and $\overline{\mathrm{AE}}_j$ have been computed to compare the change in predictive quality from short-term to long-term projections (Riebler et al., 2012). The mean CRPS and AE have also been used as an overall measure (Gneiting and Raftery, 2007) to assess the predictions.

Chapter 5

Data analysis

Cancer is a leading public health problem in most countries across the world, responsible for 8.2 million deaths in 2012 (Stewart and Wild, 2014). As a motivation for this thesis, lung cancer is acknowledged as the most frequent cancer worldwide both in terms of cases and deaths (Ferlay et al., 2010). In the United Kingdom, over 50% of all new cancer cases are attributed to breast, lung, prostate and bowel cancers combined; lung cancer is the second most common of these, the first being breast cancer in women and prostate cancer in men (`cruk.org/cancerstats`, accessed: 20.04.2015). The 2012 annual report by the Cancer Registry of Norway reveals that 50% of cancer deaths can be attributed to cancer of the lung, colon, rectum, prostate and female breast; lung cancer being responsible for most cancer deaths both in men and women.

Cancer of the oesophagus is recorded as the sixth most frequent cause of cancer deaths worldwide (Ferlay et al., 2013), and is responsible for close to 5% of all cancer deaths in the UK (`cruk.org/cancerstats`, accessed: 20.04.2015). Cancer deaths caused by oesophagus cancer is also quite rare in Norway; see Figure 5.1 for age-specific oesophagus cancer death rates per 100,000 for females in Norway.

Given the health burden imposed by cancer on most population groups, it is crucial for forecasters to make cancer predictions using appropriate time scales to ensure that quality projections are obtained for decision making. The purpose of this thesis is to model and

assess the quality of predicted age-specific cancer rates and counts using three different time scales. Using lung and oesophagus cancer data in Norway and the United Kingdom, we consider predictions based on five-year data aggregation, model-specific aggregation (aggregating the random effects based on five-year intervals) and the case where we use the yearly data witout an imposed aggregation.

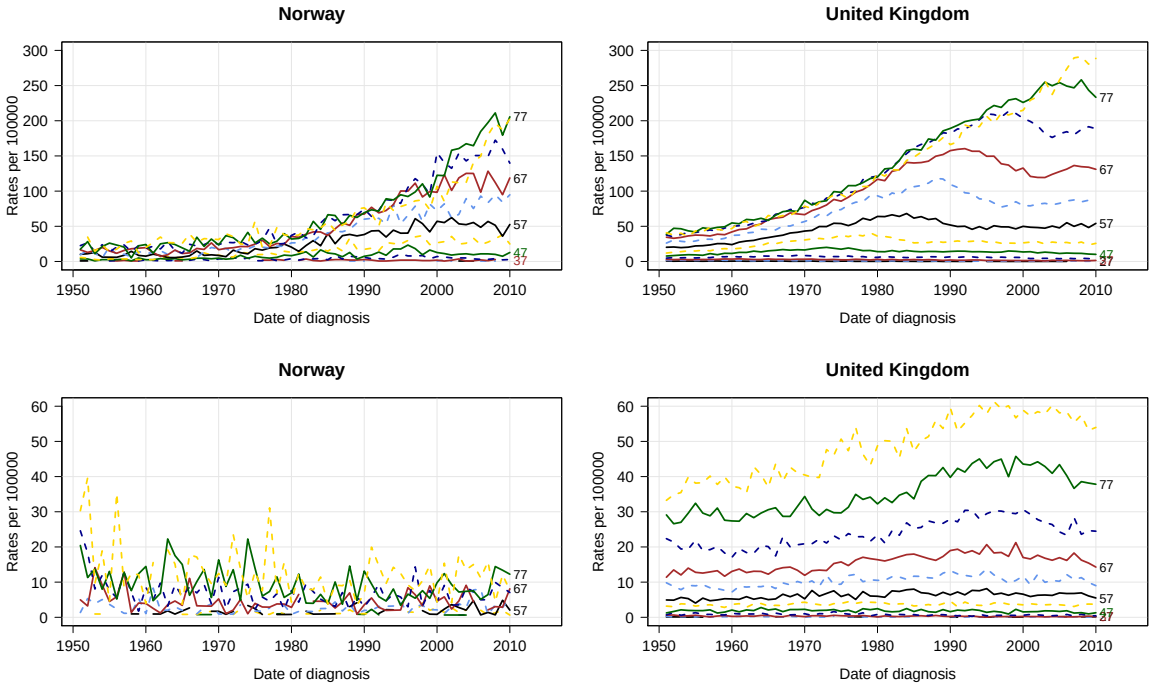


Figure 5.1: Age-specific lung cancer (top) and oesophagus cancer (bottom) death rates per 100,000 for females in Norway and the United Kingdom.

The choice of lung and oesophagus cancer data is informed by the curiosity to investigate how the different time scales influence the quality of prediction of cancers with high mortality rate (lung cancer) and cancers with very low mortality rate (oesophagus cancer), thereby making the deployment of lung and oesophagus cancer data suitable for the study. Similarly, by using both densely and sparsely populated countries like the United Kingdom and Norway respectively, we indirectly and tacitly explore the quality of cancer predictions under the different time scales, for diverse types of population (for instance, populations that may neither be densely nor sparsely populated) without necessarily examining data from these countries in the current study. Consequently, the idea of representativeness

suggested by the use of lung and oesophagus cancer data from the United Kingdom and Norway informs the choice of data for the study.

Age-specific mortality rates from 1951 to 2010 are displayed in Figure 5.1. Lung cancer mortality rates for females with the age 50 and above in Norway shows a steady rise over the past decades, in contrast to UK which appears to have witnessed a decrease after 1990 and subsequently levels off after year 2000. Rates in females below the age of 50 remain fairly constant for both UK and Norway. In the United Kingdom, we see fairly constant oesophagus cancer death rates with slight increases after 1980 for females aged 70 and above. It is difficult to identify trends in oesophagus cancer rates for females in Norway due to very low counts. The noise in the data suggests the use of model-based prediction to project data properly.

In this chapter, we describe how we obtain the data and give an overview of the `BAPC`-package, an R-package which implements Bayesian age-period-cohort models (see section 3.2) with focus on predictions (Riebler and Held, 2015). Further, we obtain cancer rate projections for different time scales based on the Bayesian age-period-cohort model using the `BAPC`-package and some modification of it to suit this thesis. The impact of these varying time scales on the quality of cancer projections will be assessed using proper scoring rules, here the continuous ranked probability score and absolute error, discussed in chapter 4.

5.1 Cancer data sets

The principal data source for this thesis was the World Health Organisation (WHO) mortality database (`http://www.who.int/healthinfo/statistics/mortality_rawdata/en/`, accessed: 26.02.2015), which contains number of deaths by country, year, sex, age group and cause of death, dating as far back as 1950. Only data of countries that have been coded appropriately using the International Classification of Diseases (ICD) are available in the database. ICD is recognised in epidemiology, health management and medicine as a benchmark tool used to keep incidence and prevalence of diseases in population

Cause	ICD revision	Detailed list
Malignant neoplasm of trachea, bronchus and lung	ICD-7	A050
	ICD-8	A051
	ICD-9	B101
	ICD-10	1034
code 1034 *was combined from* $C33$*,* $C340$*,* $C341$*,* $C342$*,* $C343$*,* $C348$ *and* $C349$		
Malignant neoplasm of oesophagus	ICD-7	A045
	ICD-8	A046
	ICD-9	B090
	ICD-10	1028
code 1028 *was combined from* $C150$*,* $C151$*,* $C153$*,* $C154$*,* $C155$*,* $C158$ *and* $C159$		

Table 5.1: Codes used for the data extraction for the different revisions of the ICD

groups under surveillance (`http://www.who.int/classifications/icd/en/`, accessed: 20.04.2015). The database contains mortality data recorded in four separate files according to different revisions of the ICD (ICD-7, ICD-8, ICD-9, ICD-10). Also contained in the database is a file which reports population mid-year figures. Each country in the database is identified uniquely by a country code- country codes and names are recorded in one data file. A list of countries and the years for which mortality and population data included in the database is available. The last data file included in the database is a notes document with notes related to data for some countries.

Data were retrieved for female population estimates (mid-year population figures), female cancer deaths for oesophaus and combined female cancer deaths for lung, brunchus and trachea according to the ICD using the appropriate country codes for United Kingdom (4308), England and Wales (4310), Northern Ireland (4320), Scotland (4330) and Norway (4220). For the different revisions of the ICD, the codes we used for the data extraction are tabulated in Table 5.1. Mortality data were available for both countries, by five-year age groups and yearly intervals. We retrieved data for females between the ages of 25 and 84 for a period spanning $1951 - 2010$ for both Norway and United Kingdom.

For the year 2000, mortality counts are unavailable for the United Kingdom due to the fact that ICD-10 was not introduced the same year in the countries of the UK. In view of this, the sum of mortality counts from England & Wales (ICD-9), Northern Ireland (ICD-9) and Scotland (ICD-10) is used as suggested by the WHO in the notes document-

see http://www.who.int/entity/healthinfo/statistics/notes.zip?ua=1 (updated: 03.11.2014, accessed: 26.02.2015).

5.2 Statistical software and package

All statistical analysis in this work were carried out using the R statistical software (R Core Team, 2013) using the BAPC-package (Riebler and Held, 2015) and some modifications of it to suit the purposes of this thesis. Some selected R-codes are documented in Appendix B.

The BAPC-package

The BAPC-package (Riebler and Held, 2015) is an R-package based on the Bayesian APC model for predicting future cancer rates and counts within a full Bayesian inference setting (see section 3.2). Contrary to some softwares which are mostly used for cancer rate projections based on the Bayesian version of the age-period-cohort model such as the independent software, BAMP (Schmid and Held, 2007) and WinBugs (Lunn et al., 2000) which employ Markov chain Monte Carlo (MCMC) algorithms, the BAPC-package uses integrated nested Laplace approximations (INLA) (See subsection 3.2.3) for full Bayesian inference (Riebler and Held, 2015). Since INLA has been optimized for inference on the class of latent Gaussian models (Rue et al., 2009), it avoids long running times which are characteristic of MCMC techniques and its corresponding needed convergence checks (Bray, 2002). Riebler and Held (2015) obtained probabilistic forecasts based on the Bayesian APC model implemented in the BAPC-package and demonstrated that the forecasts were well calibrated and the prediction intervals not too wide. The BAPC-package allows very straightforward model specification and readily generates desired outputs, such as age-standardized and age-specific projected rates and counts. For a detailed description of the methodological details of the BAPC-package, we refer to Riebler and Held (2015).

In this thesis we include all three time effects (age, period and cohort effects) and also incorporate an overdispersion component (see section 3.2). We include the time effects as smooth functions based on the RW2 model (see subsection 3.2.1) and hence sum-to-zero constraints are incorporated automatically (Riebler and Held, 2015). To account for model-specific aggregation, we make alterations to the standard model specification within the `BAPC`-packaage.

5.3 Methods

Based on the Bayesian APC model (see chapter 3), we obtained predictions using three time scales (data aggregation, model-specific aggregation and yearly data structure). In each case, we included an overdispersion parameter (Knorr-Held and Rainer, 2001) to capture the variation unexplained by age, period or cohort effects. The overall model is of the form

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

Here, η_{ij}, μ, θ_i, φ_j, ψ_k, and z_{ij} follow the same definition as in chapter 3. We now describe the different time scales explored. In all cases considered, the age index, $i = 1, 2, \ldots, 12$, remains unaltered and is included in all the models. With the help of age-specific plots, we compare projected mortality counts and rates under each of the methods to the true observations. In addition, we make use of proper scoring rules to assess how the quality of the predictions vary from short-term to long-term forecasts (Riebler and Held, 2015), here the continuous ranked probability score and absolute error have been used. A motivating basis for this work is Kang et al. (2014), where the impact of varying definitions of geographical areas on the estimation of individual disease risks was investigated. They explored several spatial scales that explain geographical variation effectively. Here, the impact of three temporal scales on the quality of cancer projections based on the Bayesian APC model is investigated.

Data aggregation

Each cancer data set is given by the WHO based on yearly intervals (here, 60 years) and 12 five-year age groups. For example, a preview of the first few lines of the lung cancer mortality counts data extracted for Norway looks like this:

	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80-84
1951	1	0	0	1	4	6	9	7	10	11	6	0
1952	1	0	0	0	4	5	10	12	8	13	10	7
1953	0	0	1	0	1	3	12	9	10	7	4	3
1954	0	0	0	0	3	2	6	4	13	12	8	5
1955	0	2	2	1	3	4	6	17	10	7	10	3

The structure of the data for "person-years of exposure" is analogous. We begin by considering cancer projections based on five-year period intervals. This was motivated by the fact that some software packages such as Nordpred and the iterative Lee-Carter package implicitly assume same lengths for age and period intervals, mostly five-year widths, thus requiring that yearly data is artificially aggregated in order to make cancer rate projections. This practice was also employed by Clements et al. (2005), where they aggregated the yearly periods to five-year intervals and made predictions based on the Bayesian APC model. In comparison with predictions obtained by fitting frequentist generalised additive models (GAMs) based on yearly data structure, Clements et al. (2005) found fault with the large credible intervals given by the Bayesian APC model. Riebler and Held (2015) suspected that besides others, data aggregation might have contributed to this unsatisfactory performance of the Bayesian APC model predictions. We investigate retrospective projections of mortality counts based on the Bayesian APC model implemented in the `BAPC`-package using the five-year aggregated data.

For each cancer mortality data set, we aggregate the data by summing the counts within a period of five-years to obtain same lengths for both age and period intervals (five-year lengths). We do the same for corresponding person-years data sets. The data for lung cancer mortality in Norway now looks like this:

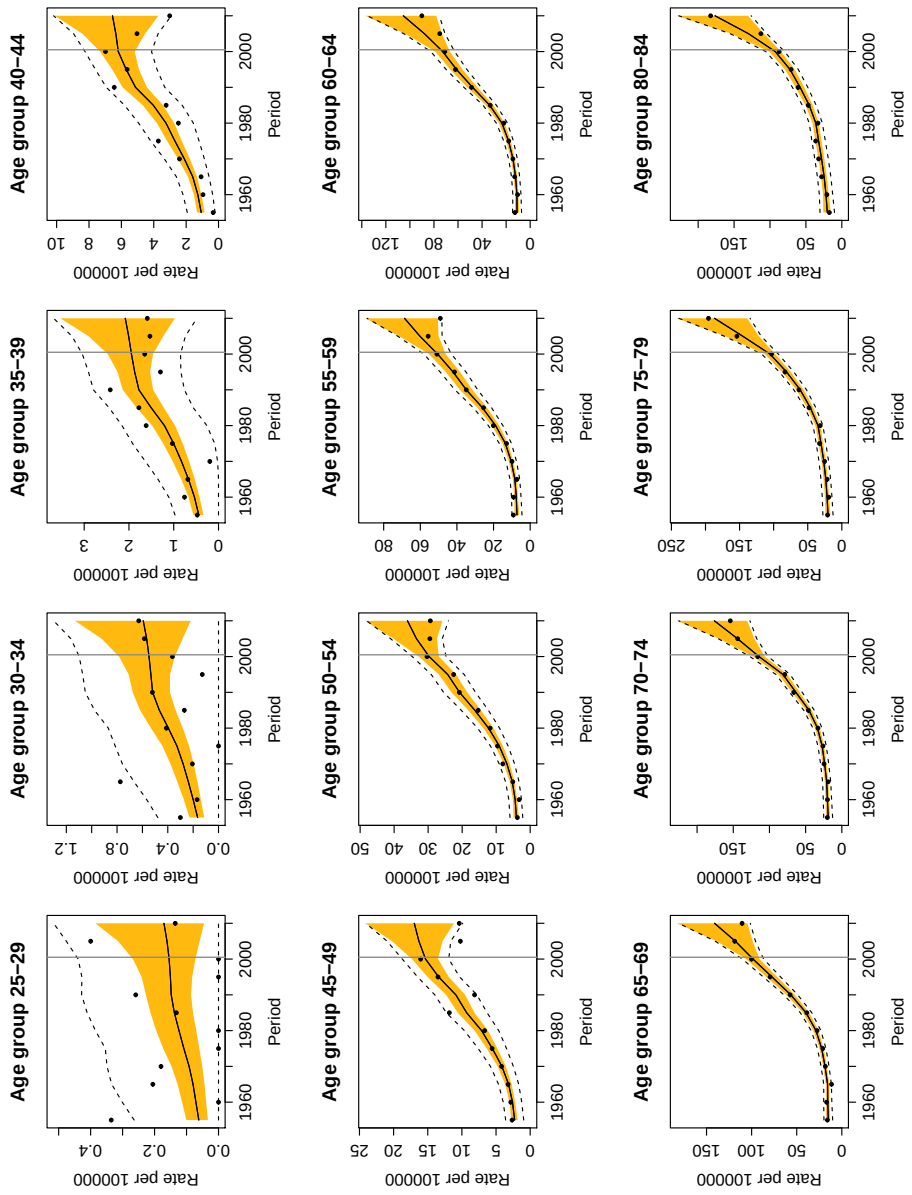


Figure 5.2: Observed number of cases aggregated over five-year intervals (dots) together with predicted mean rate within 95%-pointwise credible intervals (gold shaded) and predicted number of cases aggregated over five-year intervals within 95%-pointwise credible intervals (dashed) for all age groups for female lung cancer mortality in Norway. The vertical line shows where prediction started. The points represent the sum of counts within the five-year period and is shown at the last time point in very five-year period. For example, the sum of counts between $1951 - 55$ is shown at year 1955.

	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65-69	70-74	75-79	80-84
1951-55	2	2	3	2	15	20	43	49	51	50	38	18
1956-60	0	1	5	6	17	18	47	47	61	56	38	25
1961-65	1	4	4	7	20	30	39	64	48	60	48	37
1966-70	1	1	1	14	27	49	57	75	85	92	65	49
1971-75	0	0	5	19	32	61	77	99	104	109	100	65
1975-80	0	3	9	12	34	67	125	126	144	147	110	72

Here there are 12 periods, hence the period index runs from $j = 1, 2, \ldots, 12$. The age group and period intervals have the same lengths; hence, the cohort index is computed as $(I - i) + j$ (Holford, 1983). For both lung and oesophagus cancer in both countries, we make projections by excluding the observed number of mortality cases across all age groups in the respective last 2 periods, representing a 10-year period since data has been aggregated over five-year intervals. Observed and predicted number of mortality cases for female lung cancer in Norway within 95%-pointwise credible intervals obtained using data aggregation are displayed in Figure 5.2. Similar figures for the United Kingdom and oesophagus cancer are shown in Appendix A.

From the plots, it appears that projections are fairly good in most age groups, especially for younger age groups and for both cancers. For instance, in Figure 5.2, lung cancer predictions in age group $30 - 34$ and $35 - 39$ appears reasonable in comparison with the aggregated counts (points) plotted on the same graph. However, in age groups $40 - 44$, $45 - 49$ and $50 - 54$, predictions suggest an increasing trend whereas one would expect a decrease in the trends given the observed counts. Moreover, the five-year data aggregation is not informative for yearly dynamics. This does not mean that it does not exist but then we lose it due to the aggregation, thus it is not possible to obtain clues on yearly trends using aggregated data. On the contrary, we can aggregate predictions made based on yearly intervals to obtain five-year interval predictions. It is difficult to compare the predictions obtained based on five-year data aggregation to the true observations. This is because the predictions are for five-year intervals while the true obsevations follow a yearly data structure. Moreover, the aggregated predicted counts hide possible variations in cancer trends and hence counts in individual years cannot be tracked over time. We now shift attention to a method of aggregation which is specific to the model and not the data.

Period	Index	Period	Index	Period	Index	...	Period	Index
1951	1	1956	2	1961	3	...	2006	12
1952	1	1957	2	1962	3	...	2007	12
1953	1	1958	2	1963	3	...	2008	12
1954	1	1959	2	1964	3	...	2009	12
1955	1	1960	2	1965	3	...	2010	12

Table 5.2: Indexing of cases and person-years according to period partitions

Here, we maintain the yearly data structure and thus we can compare the retrospective projections with the observed data.

Model-specific aggregation

In this approach, we preserve the yearly data structure and aggregate the random effects instead. Here, we partition periods into 12 groups and assign the same period index to neighbouring periods before modelling the random effects. We partition the periods into five-year intervals and assign to each partition one period index ranging from $j = 1, \ldots, 12$. The indexing of cases and person-years based on the period partitions is shown in Table 5.2. The cohort index k is computed here as $(I - i) + j$ since we assume same intervals for age groups and period during model specification (Holford, 1983). Observed and predicted number of mortality cases for female lung cancer in Norway within 95%-pointwise credible intervals, obtained using model-specific aggegation are displayed in Figure 5.3. Similar figures for the United Kingdom and oesophagus cancer are shown in the Appendix.

The figures suggest that the projections seem sensible for most age groups and countries. Projections are seen as steps due to the fact that the random effects have been aggregated and thus the predicted rates are fairly constant for periods which were assigned the same period index. However, projections fairly depict the trend in the data in most age groups. In comparison to Figure 5.2, it is observed that information is lost during data aggregation and that, by maintaining the yearly data structure and imposing a model-specific aggregation, clues on yearly trend can be monitored and predictions depict more of this trend than when the data is aggregated.

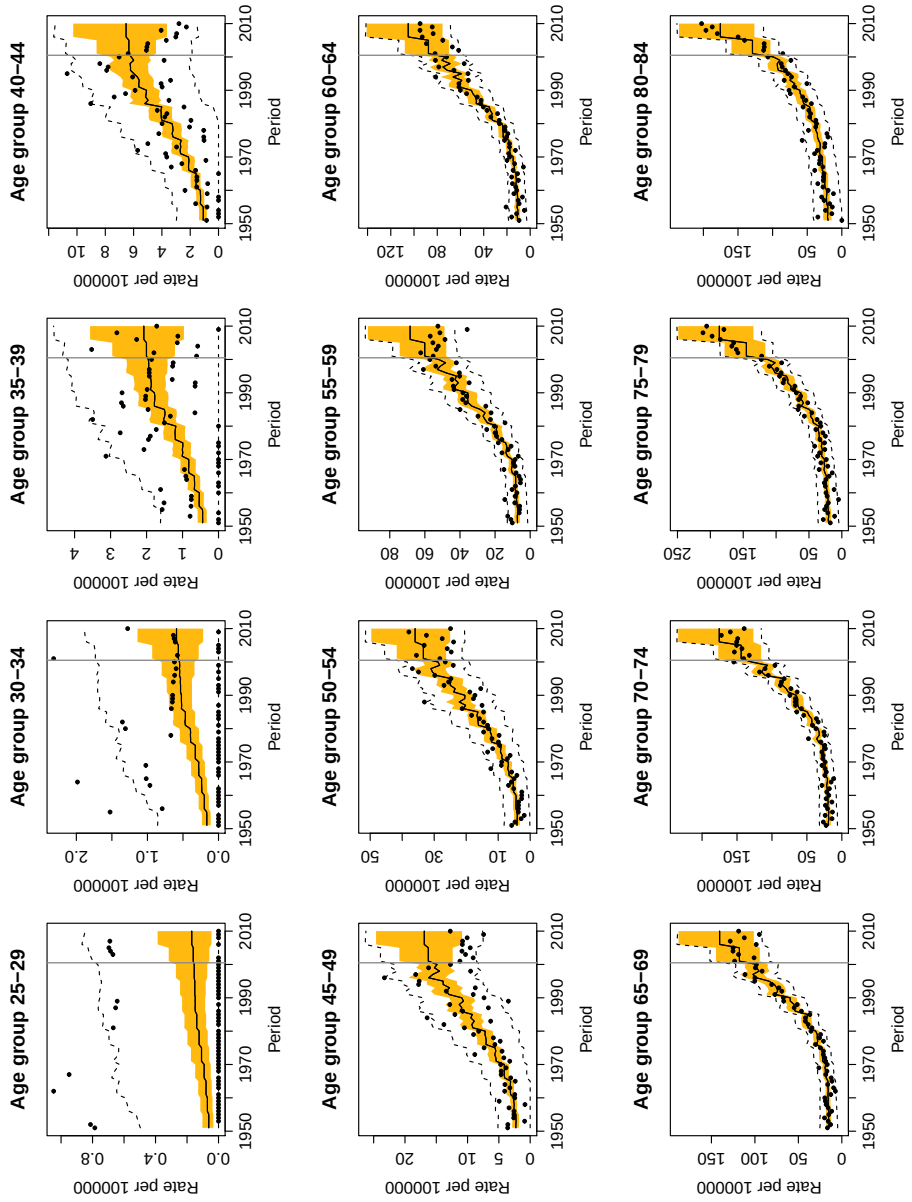


Figure 5.3: Observed number of cases (dots) together with predicted mean rate within 95%-pointwise credible intervals (gold shaded) and predicted number of cases within 95%-pointwise credible intervals (dashed) for all age groups for female lung cancer mortality in Norway obtained using model-specific aggregation (5-year). The vertical line shows where prediction started.

No aggregation

Here, we follow Riebler and Held (2015), keeping the yearly data structure and fit the Bayesian APC model. The cohort index k does not depend only on the age group and period index as in the two previous formulations, but also on the length of the age group and period intervals (Holford, 1983). It is computed as $5 \cdot (I - i) + j$ since the age group intervals are 5 times wider than the period interval (Heuer, 1997). Riebler and Held (2015) showed that the predictions obtained with the Bayesian APC model was of better quality than the generalized Lee-Carter model. We will make comparison of the results obtained here to the predictions obtained using model-specific aggregation. Observed and predicted number of mortality cases for female lung cancer in Norway within 95%-pointwise credible intervals obtained with no aggregation induced is displayed in Figure 5.4. Similar figures for the United Kingdom and oesophagus cancer are shown in the Appendix. The plots suggest that the predictions appear sensible for almost all age groups, for both cancers and both countries. It is observed that the retrospective projections here are smoother than that of the model-specific aggregation (see Figure 5.3 for example), seen as steps, and closer to the observed counts. The smoother prediction curves seen in the method of no aggregation is due to the fact that smoothing has been carried out over a longer period (60-year period), whereas there are only 12 period effects in the model-specific aggregation.

To assess the change in predictive performance from short-term to long-term projections, Figure 5.5 shows the cumulative average $\overline{\text{AE}}_j$ (dashed) and $\overline{\text{CRPS}}_j$ (solid lines). For lung cancer in the UK and Norway and also for oesophagus cancer in the UK, the curves $\overline{\text{AE}}_j$ and $\overline{\text{CRPS}}_j$ of the method with no aggregation (black) lie below those of the model-specific aggregation (gold) suggesting better predictive quality. For oesophagus cancer in Norway, we observe similar performance between the two methods. This is attributed to the fact that oesophagus cancer counts are very low in Norway; hence, the predictions based on the yearly data structure are very close to that of the model-specific aggregation. This indicates that predictive quality is almost the same for both methods when counts in the data are very small. It is also seen from Figure 5.5 that the cumulative average $\overline{\text{AE}}_j$ (dashed) and $\overline{\text{CRPS}}_j$ (solid lines) for both methods stay fairly constant for both cancers in Norway, indicating that predictive quality is preserved if more periods are predicted

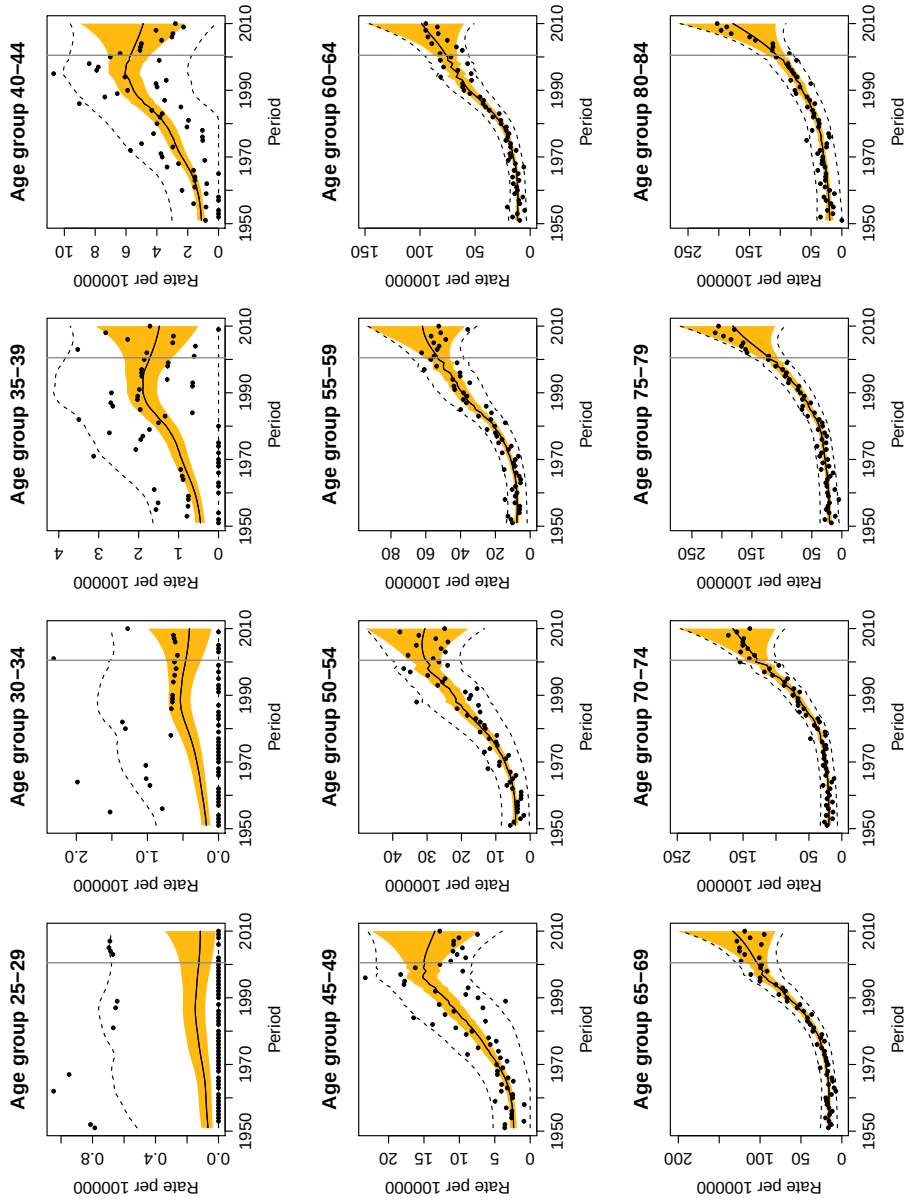


Figure 5.4: Observed number of cases (dots) together with predicted mean rate within 95%-pointwise credible intervals (gold shaded) and predicted number of cases within 95%-pointwise credible intervals (dashed) for all age groups for female lung cancer mortality in Norway obtained without any form of aggregation. The vertical line shows where prediction started.

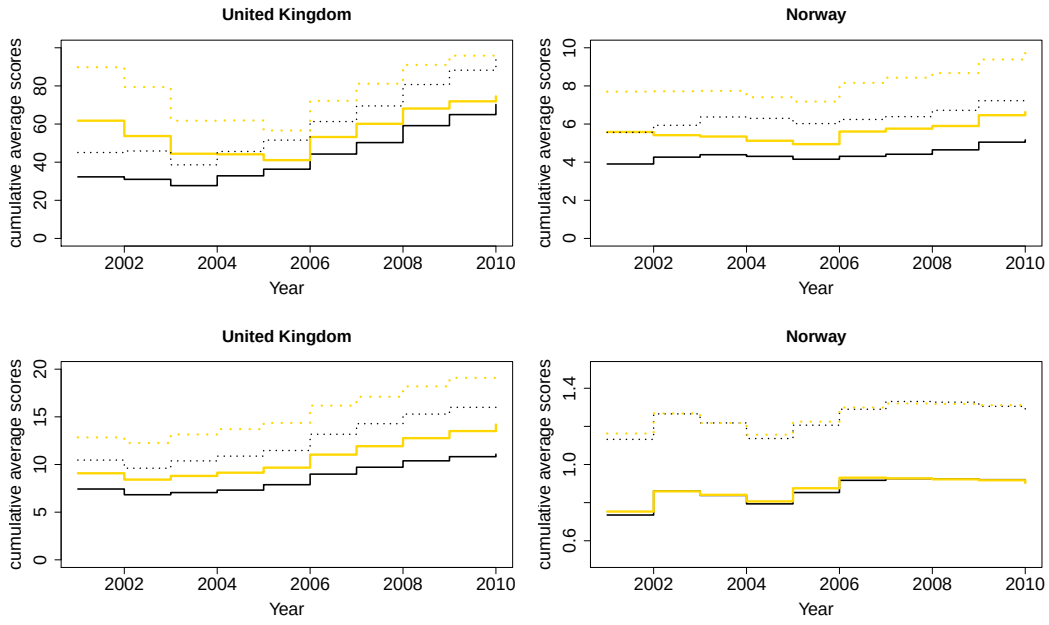


Figure 5.5: Cumulative average of mean absolute errors (dotted) and continuous ranked probability scores (lines) across age groups for lung cancer (top) and oesophagus cancer (bottom) in Norway and the United Kingdom obtained by the Bayesian APC model with no aggregation imposed (black) and model-specific aggregation (gold).

irrespective of whether model-specific aggregation is used or method of no aggregation. There is a steady rise in the scores observerd for both cancers in the UK, indicating a decrease in predictive quality over time when the data is large. The cumulative average $\overline{\text{AE}}_j$ mostly lies above that of the CRPS.

Table 5.3 displays the mean CRPS and AE, which can also be seen in Figure 5.5 as the values at the last time points in the curves. It is observed from the table that the method with no aggregation always gives better predictions than the model-specific aggregation with relatively smaller mean scores apart from oesophagus cancer in Norway where scores are the same in both methods due to low counts observed in the data. Considering lung cancer in Norway, for instance, the mean CRPS is 70.27 cases for the method with no aggregation, but 74.54 cases for the model-specific aggregation. In view of the fact that Norway is sparsely poplulated, the CRPS and AE score differences are smaller compared to the score difference in the UK. For example, the mean CRPS score for Norway lung cancer records 6.63 cases (model-specific) against 5.17 cases (method of no aggregation).

		Model-specific		No aggregation	
		$\overline{\text{AE}}$	$\overline{\text{CRPS}}$	$\overline{\text{AE}}$	$\overline{\text{CRPS}}$
Lung Cancer	United Kingdom	99.13	74.54	95.89	70.27
	Norway	9.76	6.63	7.40	5.17
Oesophagus Cancer	United Kingdom	19.94	14.19	16.45	11.09
	Norway	1.29	0.90	1.29	0.91

Table 5.3: Mean absolute error $\overline{\text{AE}}$ and mean continuous ranked probability score $\overline{\text{CRPS}}$. Shown are results obtained with the Bayesian APC model using model specific data aggregation and model with no aggregation imposed. Results for both lung and oesophagus cancer in the UK and Norway are displayed. Of note, the priors for the RW2 precision parameters of the time effects under both methods have not been scaled to have a generalised variance = 1.

Of note, the score differences between both methods are quite small for both cancers and in both countries, suggesting similar performance between the two methods in our application based on the scores investigated, with the method of no aggregation having slightly better predictive quality and smoother predictive mean rate. Of note, scores have not been computed for the five-year data aggregation method because predictions obtained here are for five-year intervals while the true observations are on an annual scale and so it is not clear how to compare them using AE or CRPS.

Prior specification

In the analysis, priors for the RW2 precision parameters of the time effects (age, period and cohort) were not scaled to have a generalised variance of 1 (see section 3.2). We will now explore how scaling the priors for the precision parameters of the time effects influences our results. This is interesting for us since the random effects in the two methods are of different lengths. Recall from section 3.2 that the marginal standard deviations for the RW2-model varied for different lengths of the graph. Hence, we will scale the RW2 priors in each method (method of no aggregation and model-specific aggregation) to have a similar interpretaion for the precision parameter. To evaluate how scaling the priors affects our results, we fit the Bayesian age-period-cohort model again under the two time scales, model-specific aggregation and the method of no aggregation, but this

time we scale the priors of the RW2 precision parameters of the time effects to have a generalised variance = 1. Table 5.4 displays the mean CRPS and AE scores obtained by fitting the models with scaled priors. Scaling the priors had a minor influence on the mean scores; most mean scores experienced slight changes, with the mean scores under method of no aggregation still seen to be slightly lower than those for model-specific aggregation, apart from UK lung cancer, where the latter performs slightly better. Before scaling the priors (see Table 5.3), the method of no aggregation performed slightly better in UK lung cancer, considering that the mean AE and CRPS were a bit smaller. However, the random effects were of different lengths in the two methods and might have influenced this finding. Scaling the priors to have a similar interpretation for the precision parameters in both methods (see Table 5.4, we see that model-specific aggregation is seen to perform slightly better when the data is large, and might be a preferred choice when making predictions for large data sets.

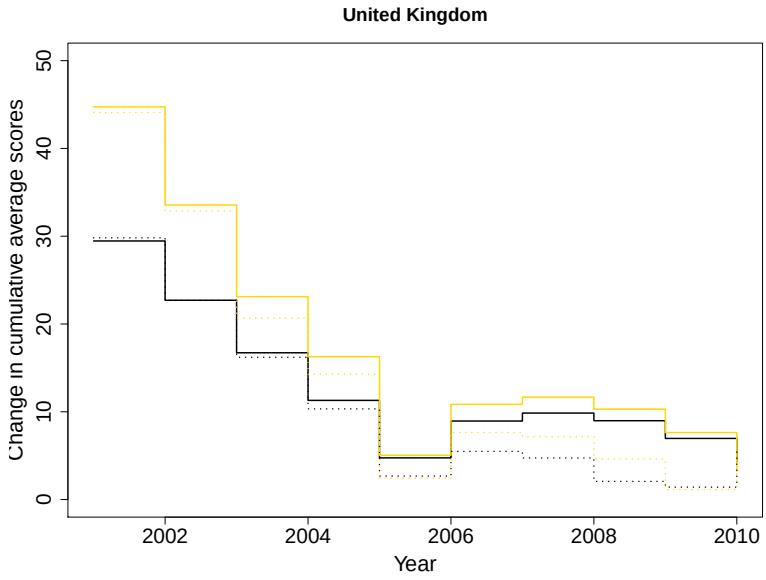


Figure 5.6: Absolute difference in cumulative average of mean absolute errors (gold) and continuous ranked probability scores (black) of model with no aggregation and model with data specific aggregation across age groups for lung cancer in the United Kingdom obtained by the Bayesian APC model with scaled priors for the precision parameters of time effects (dotted) and priors not scaled (lines).

		Model-specific		No aggregation	
		$\overline{\text{AE}}$	$\overline{\text{CRPS}}$	$\overline{\text{AE}}$	$\overline{\text{CRPS}}$
Lung Cancer	United Kingdom	98.91	74.41	103.24	80.12
	Norway	9.78	6.65	7.65	5.28
Oesophagus Cancer	United Kingdom	20.05	14.32	18.71	13.10
	Norway	1.30	0.90	1.29	0.90

Table 5.4: Mean absolute error $\overline{\text{AE}}$ and mean continuous ranked probability score $\overline{\text{CRPS}}$. Shown are results obtained with the Bayesian APC model using data aggregation, model specific data aggregation and no aggregation imposed. Results for both lung and oesophagus cancer in the UK and Norway are displayed. Here, priors for the RW2 precision parameters of the time effects have been scaled to have a generalised variance = 1.

Figure 5.6 shows how far (in absolute value) the cumulative average of mean scores for model-specific aggregation is from that of method with no aggregation. The curves were obtained by scaling the priors of the RW2 precision parameters and computing the absolute difference in the cumulative average of the mean CRPS obtained for model-specific aggregation and that of method of no aggregation (dotted). The absolute difference was computed again when the priors are not scaled (solid lines). The difference in the AE scores are displayed in gold and that of CRPS in black. The difference in the cumulative average of the mean scores between the two methods appear to get smaller when the priors for the RW2 precision parameters of the time effects are scaled since the dotted curves mostly lies below the solid lines. Scaling the priors of the RW2 precision parameters to have a generalised variance = 1, reveals the closeness in predictive quality of both methods since the mean CRPS and AE scores get closer.

Evaluation of random effects

The precision of estimation of random effects for model-specific aggregation and method with no aggregation are displayed in Figure 5.7 for Norway lung cancer. Of note, the priors for the precision parameters of the time effects (age, period and cohort) have not been scaled. The model-specific aggregation seemed to influence the precision of estimation of

the overdispersion since the standard deviation of the overdispersion effect varies when we compare no aggregation method to the model-specific aggregation. The box plot of the standard deviation of the overdispersion effects under model-specific aggregation (5-year) has a larger length compared to the method with no aggregation (1-year) imposed. This indicates that we obtain larger variation and smaller precision in the overdispersion when model-specific aggregation is imposed. Here, the standard deviations of the time effects (age, period and cohort) are too large and thus not interpretable.

Further, we investigate box plots of the standard deviation of the random effects for the other cancer data sets and also the effect of scaling the RW2 precision parameters of the time effects (age, period and cohort). Similar to Figure 5.7, the scale of the standard deviation of the time effects (age, period and cohort) is very large irrespective of the data set and the model used, and so cannot be interpreted (results are not shown). We suspect this may be due to the identifiability problem discussed in section 3.1. Figure A.11 illustrates this problem- where one observes that there is no clear difference in the effects of age, period and cohort. The time effects therefore need to be made identifiable in order to investigate the standard deviation of the time effects (age, period and cohort effects). Here, we use two-factor models (age and cohort effects) to avoid the issue of collinearity between the effects of age, period and cohort (Clayton and Schifflers, 1987a). By fitting the models with only age and cohort effects, while adjusting for overdispersion, the scale of the standard deviation of the age and cohort effects is now reasonable (see, for example, Figure A.12 and similar plots in appendix) and thus can be interpreted. The plots suggest very little differences in the precision of estimation of the age effects since the standard deviations of the age effects stay almost constant from the method of no aggregation (1-year) to model-specific aggregation (5-year), whether the priors are scaled or not. The box plots of the standard deviation of the cohort effects under model-specific aggregation have a slightly larger length compared to that of the method with no aggregation imposed. This indicates that we obtain slightly larger variation and smaller precision in the cohort effects when model-specific aggregation is imposed. The results also indicate that the frequency of cancer influences the precision of estimation of the overdispersion effects in the two methods. For oesophagus cancer, which is rare, the standard deviation of the overdispersion effect experiences slight changes when we move

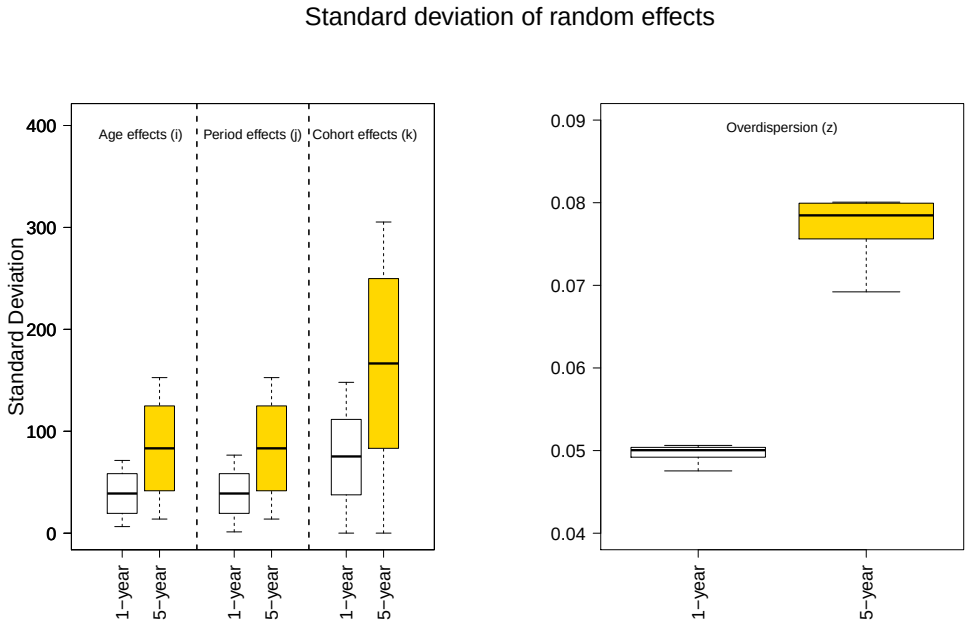


Figure 5.7: Box plots of the standard deviation of the random effects for both the model with no aggregation (1-year) and that of model specific data aggregation (5-year)

from method of no aggregation to model-specific aggregation. However for lung cancer, which is more frequent, the box plots of standard deviation of the overdispersion effects have wider length compared to those of method of no aggregation, indicating a larger variance and smaller precision.

Chapter 6

Summary

There is a wide variety of models used for analysis and projection of age-specific cancer incidence and mortality data. Softwares that implement these models have been developed and put in routine use. The Bayesian age-period-cohort (APC) model is well known for analysis and prediction of age-specific cancer incidence and mortality data. Riebler and Held (2015) developed an easy to use R-package `BAPC`, which uses intergrated nested Laplace approximations for complete Bayesian inference and subsequently showed that uncertainty of predictions obtained based on the Bayesian APC model seemed reasonable and outperformed the generalised Lee-Carter model. Cancer data are commonly provided by the World Health Organisation on yearly time-scale and five-year age groups. Some softwares such as Nordpred (`http://www.kreftregisteret.no/software/nordpred` accessed: 04.11.2014) and the iterative Lee-Carter package (Butt and Haberman, 2009) however require the data to have the same lengths for the time-scale and age group intervals, usually five-year intervals, which may not always be ideal for projections since information may get lost and projections might be misleading when data are aggregated.

This work is motivated by Kang et al. (2014), where the effect of varying definitions of geographical areas on the estimation of individual disease risks was investigated. They explored various spatial scales in effectively explaining geographical variation. Here, the impact of three temporal scales on the quality of cancer projections based on the Bayesian

APC model is investigated. Based on retrospective projections, with the help of age-specific plots and the use of proper scoring rules, the quality of predictions obtained under each time-scale is assessed. First, we obtained projections by aggregation of yearly to five-year data. The question arises whether annual aggregation of data to five-year intervals preserves every information in the data and if it provides unquestionable predictions. Instead of aggregating the data, we also investigate in this work aggregation of the random effects where the yearly data is maintained and a five-year model-specific aggregation via the random effects is carried out. We combined the yearly data structure and the effect of neighbouring periods by assigning period indexes based on five-year intervals while preserving the annual data. Finally, we obtained predictions based on yearly data without any form of aggregation and compared the predictions with the model-specific aggregation approach. In each of the three methods, we included all time effects (age, period and cohort) and also adjust for overdispersion (Knorr-Held and Rainer, 2001)

Riebler and Held (2015) suspected that annual data aggregation to five-year intervals and the exclusion of an overdispersion parameter may have contributed to the imprecise predictions of the Bayesian age-period-cohort model as shown in Clements et al. (2005). By obtaining retrospective projections of mortality counts and rates, one-step ahead forecasts and projected age-standardized rates based on yearly data with an adjustment for overdispersion, Riebler and Held (2015) showed that the Bayesian APC model outperformed the generalized Lee-Carter model and that the prediction intervals were not too wide. The question therefore arises which aggregation is appropriate in analyses and projection of age-specific cancer incidence or mortality data and how this influences projections.

In this thesis, we analysed lung and oesophagus cancer data for females in the United Kingdom and Norway obtained from the World Health Organisation (WHO) mortality database (`http://www.who.int/healthinfo/statistics/mortality_rawdata/en/` accessed: 26.02.2015). Based on age-specific plots, we found that as expected, clues about yearly trends are lost when annual data is aggregated to five-year intervals. Retrospective predictions, however, seemed reasonable in most age groups. For oesophagus cancer, especially in Norway, the retrospective projections obtained with five-year aggregated data are close to trends in the data since the counts are very low and hence aggregating the data

five-year intervals had very little effect on the trends in the yearly data. For Norway and UK lung cancer, predictions seemed reasonable in some age groups while slight deviations in trend were observed for other age-groups in comparison with the true observations. Furthermore, the annual counts are lost by aggregation of yearly to five-year data and hence it is not clear how to compare the predictions with the true observations based on the scores used in our application since predictions are on a five-year scale and the true observations are on yearly intervals.

In contrast with the five-year data aggregation, model-specific aggregation preserved the yearly data structure. Here, yearly trends can still be monitored since the annual data has not been lost. Predictions are however seen as steps, with rates staying fairly constant across periods which were assigned the same period index (five-year intervals). This is a consequence of the aggregation of random effects since the smoothing effect takes into account the five-year model-specific aggregation to make the projections. For Norway lung and oesophagus cancers, projection quality mostly remained constant when moving from short-term to long-term projections, while minimal decline was observed for both cancers in the UK where data are larger. Thus when the country is not large or cancer is rare, e.g oesophagus cancer in Norway, predictive quality is not lost when more years are predicted using five-year model-specific aggregation.

In comparison with the method with no aggregation, we observed similar predictive quality between the two methods based on the scores investigated. For oesophagus cancer in the UK and lung cancer in Norway, the method of no aggregation performed slightly better than the model-specific aggregation. The reverse was, however, seen for UK lung cancer, where the model-specific aggregation performed slightly better. Based on retrospective predictions which we have investigated in this work, we found that the differences between the two methods were very small. The age-specific plots however favoured the method with no aggregation, where predictions are smoother than in model-specific aggregation where predictions appear as steps and predicted rates are fairly constant for each five-year period predicted.

We reckon that when yearly data is readily available, data aggregation to five-year intervals may not be an ideal practice for projecting age-specific cancer incidence or mortality

data. Specific to our application, yearly data without any form of aggregation gave the best predictions. However, if any form of aggregation is desired, it seems beneficial and more natural to preserve yearly data structure and aggregate the random effects instead, depending on the application. In future work, it will be interesting to investigate the impact of data and model-specific aggregation on the quality of cancer projections where data are more detailed, having information on exact dates of death. We have investigated aggregation of yearly data (both data and model specific), however having individual characteristics such as exact dates of death, one will want to consider some sort of aggregation since there will be a lot of zero counts in the data. Future studies have the potential of revealing interesting and illuminating findings in such a case.

Appendix A

Additional figures

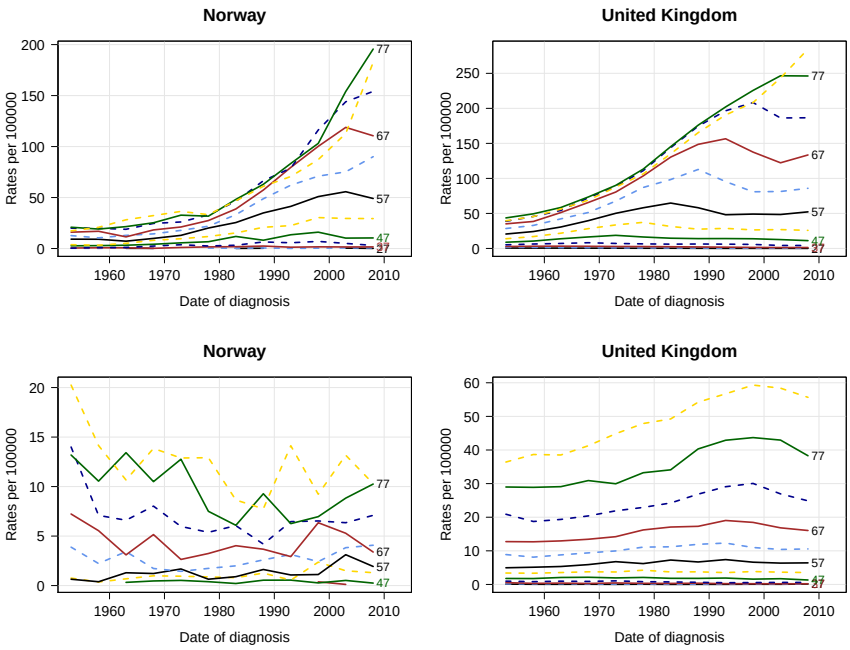


Figure A.1: Age-specific lung cancer (top) and oesophagus cancer (bottom) death rates per 100000 for females in Norway and the United Kingdom. Shown are rates obtained based on data aggregated over five-year intervals

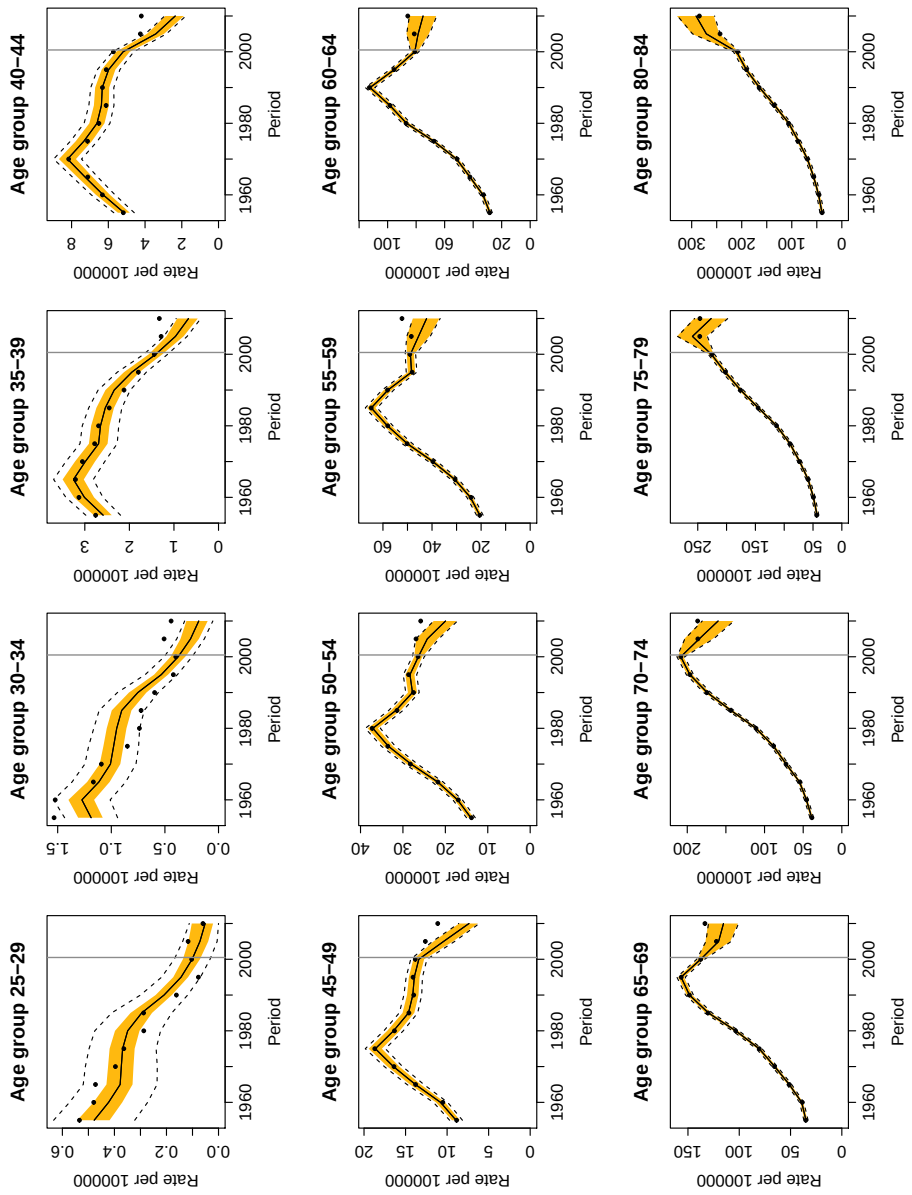


Figure A.2: Observed number of cases aggregated over five-year intervals (dots) together with predicted mean rate within 95%-pointwise credible intervals (gold shaded) and predicted number of cases aggregated over five-year intervals within 95%-pointwise credible intervals (dashed) for all age groups for female lung cancer mortality in the UK. The vertical line shows where prediction started.

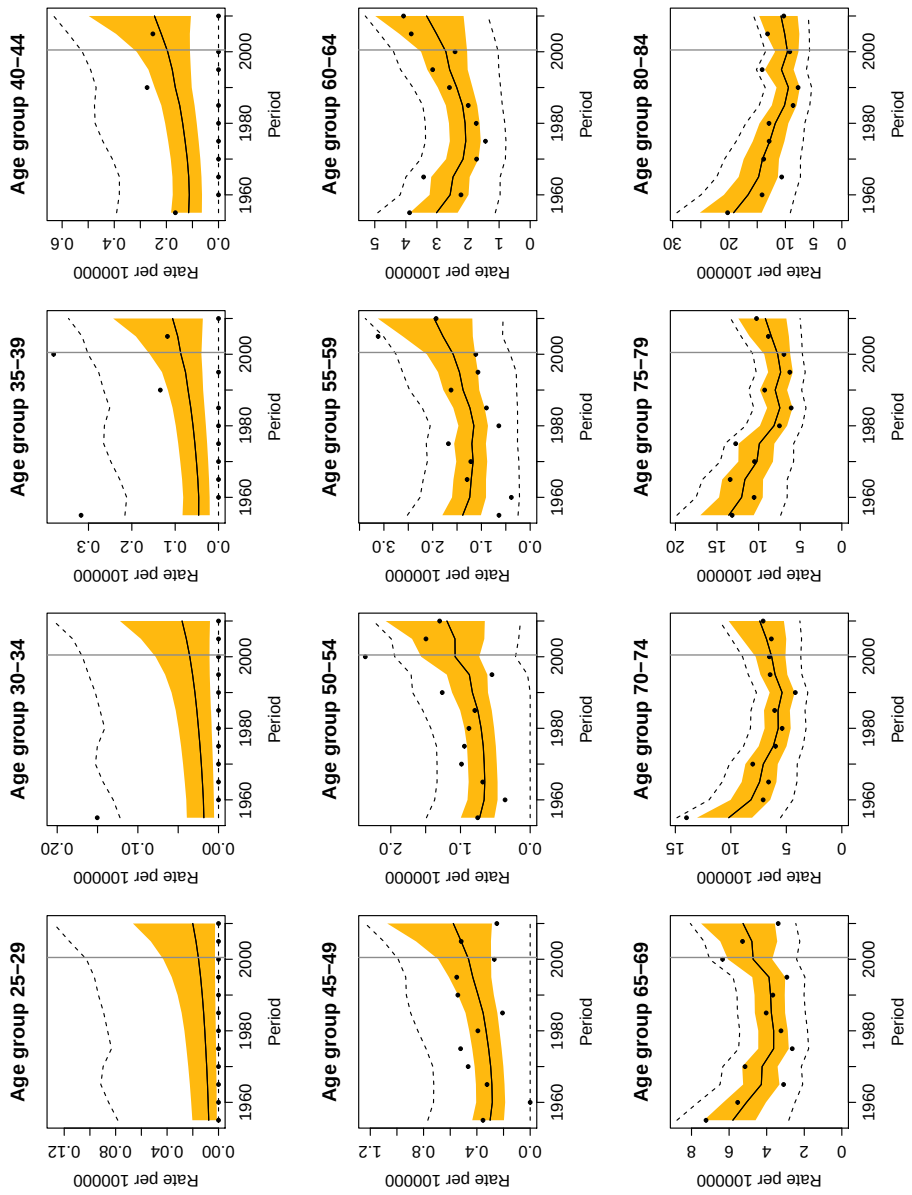


Figure A.3: Observed number of cases aggregated over five-year intervals (dots) together with predicted mean rate within 95%-pointwise credible intervals (gold shaded) and predicted number of cases aggregated over five-year intervals within 95%-pointwise credible intervals (dashed) for all age groups for female oesophagus cancer mortality in the Norway. The vertical line shows where prediction started.

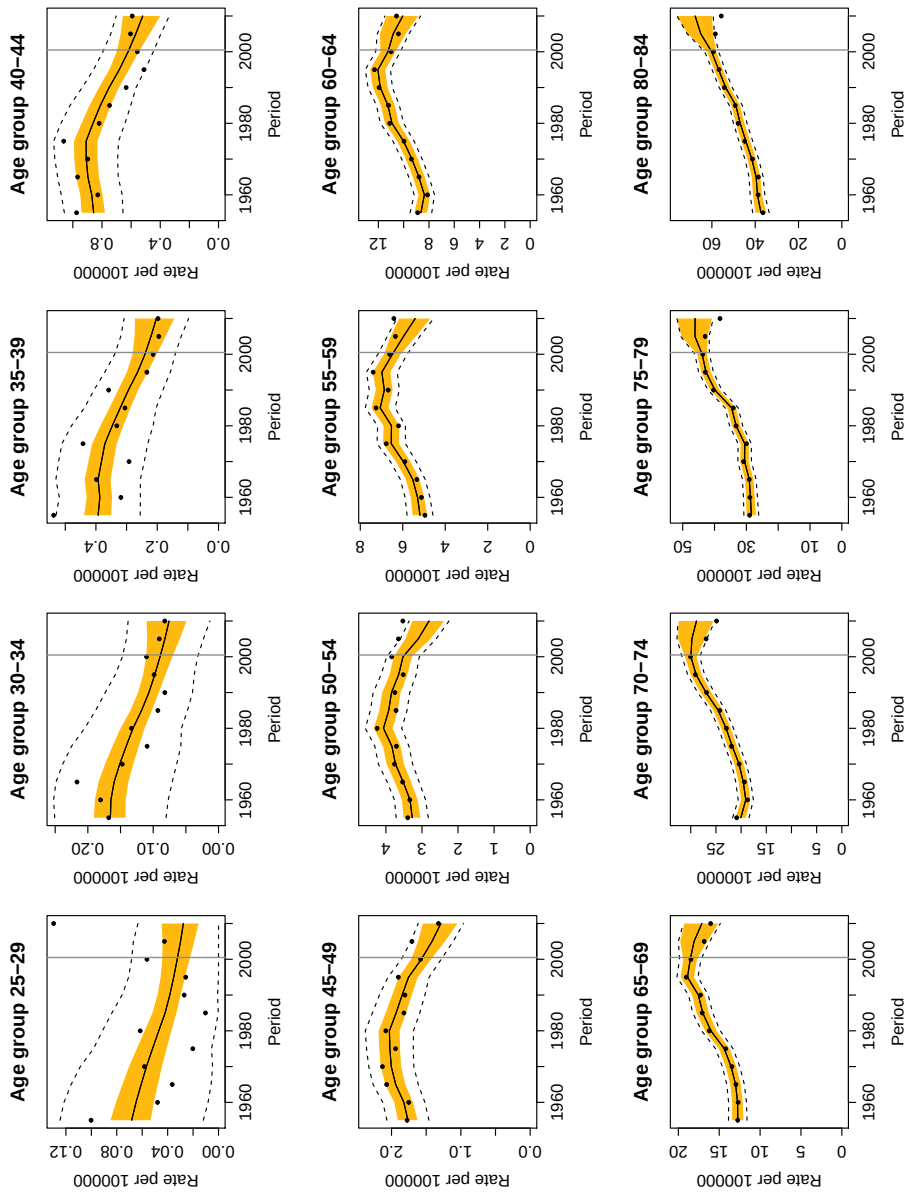


Figure A.4: Observed number of cases aggregated over five-year intervals (dots) together with predicted mean rate within 95%-pointwise credible intervals (gold shaded) and predicted number of cases aggregated over five-year intervals within 95%-pointwise credible intervals (dashed) for all age groups for female oesophagus cancer mortality in the UK. The vertical line shows where prediction started.

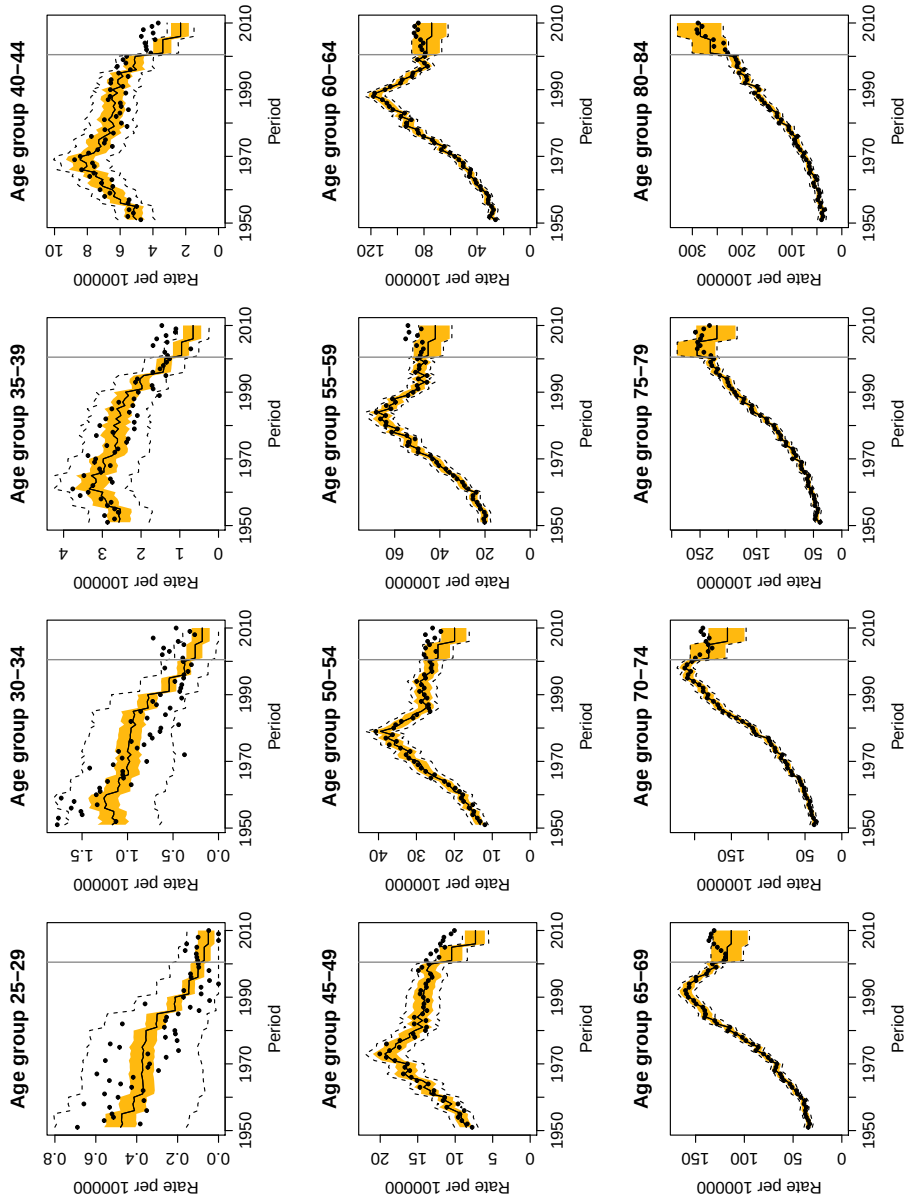


Figure A.5: Observed number of cases (dots) together with predicted mean rate within 95%-pointwise credible intervals (gold shaded) and predicted number of cases within 95%-pointwise credible intervals (dashed) for all age groups for female lung cancer mortality in the UK obtained using model-specific aggregation. The vertical line shows where prediction started.

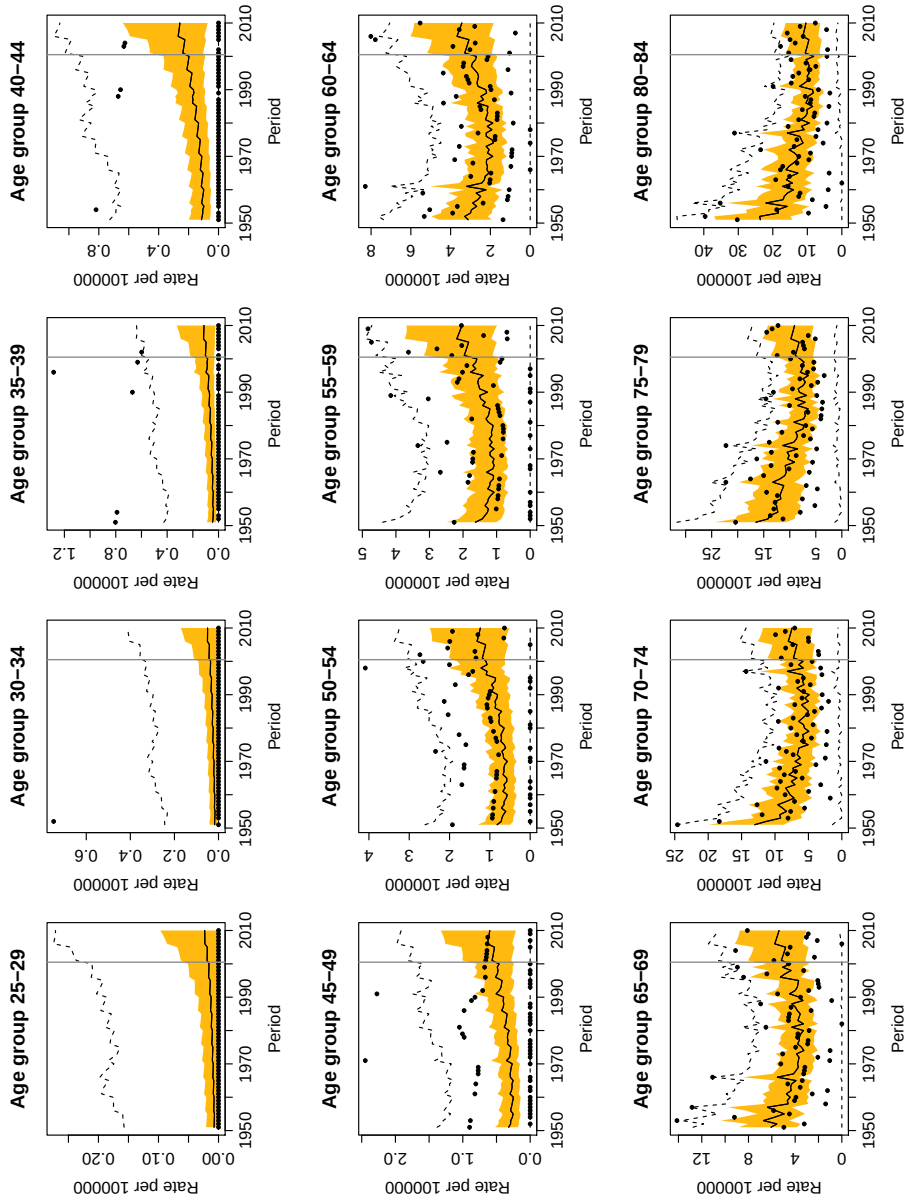


Figure A.6: Observed number of cases (dots) together with predicted mean rate within 95%-pointwise credible intervals (gold shaded) and predicted number of cases within 95%-pointwise credible intervals (dashed) for all age groups for female oesophagus cancer mortality in Norway obtained using model-specific aggregation. The vertical line shows where prediction started.

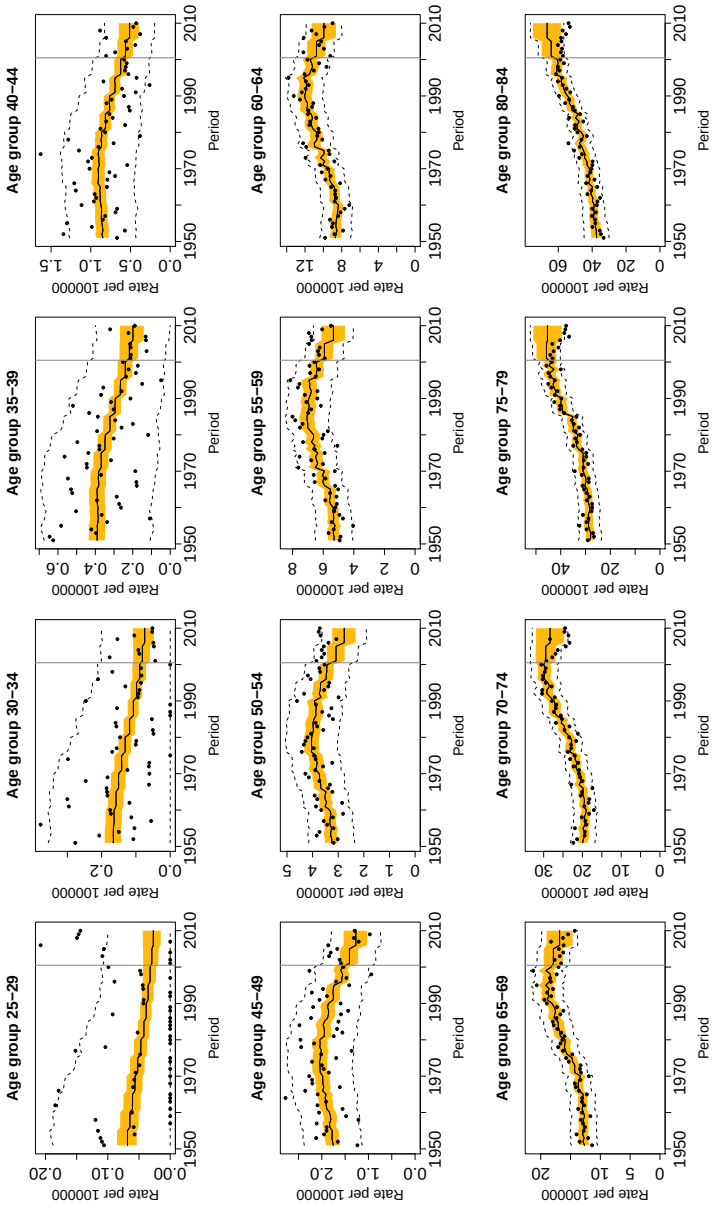


Figure A.7: Observed number of cases (dots) together with predicted mean rate within 95%-pointwise credible intervals (gold shaded) and predicted number of cases within 95%-pointwise credible intervals (dashed) for all age groups for female oesophagus cancer mortality in the UK obtained using model-specific aggregation. The vertical line shows where prediction started.

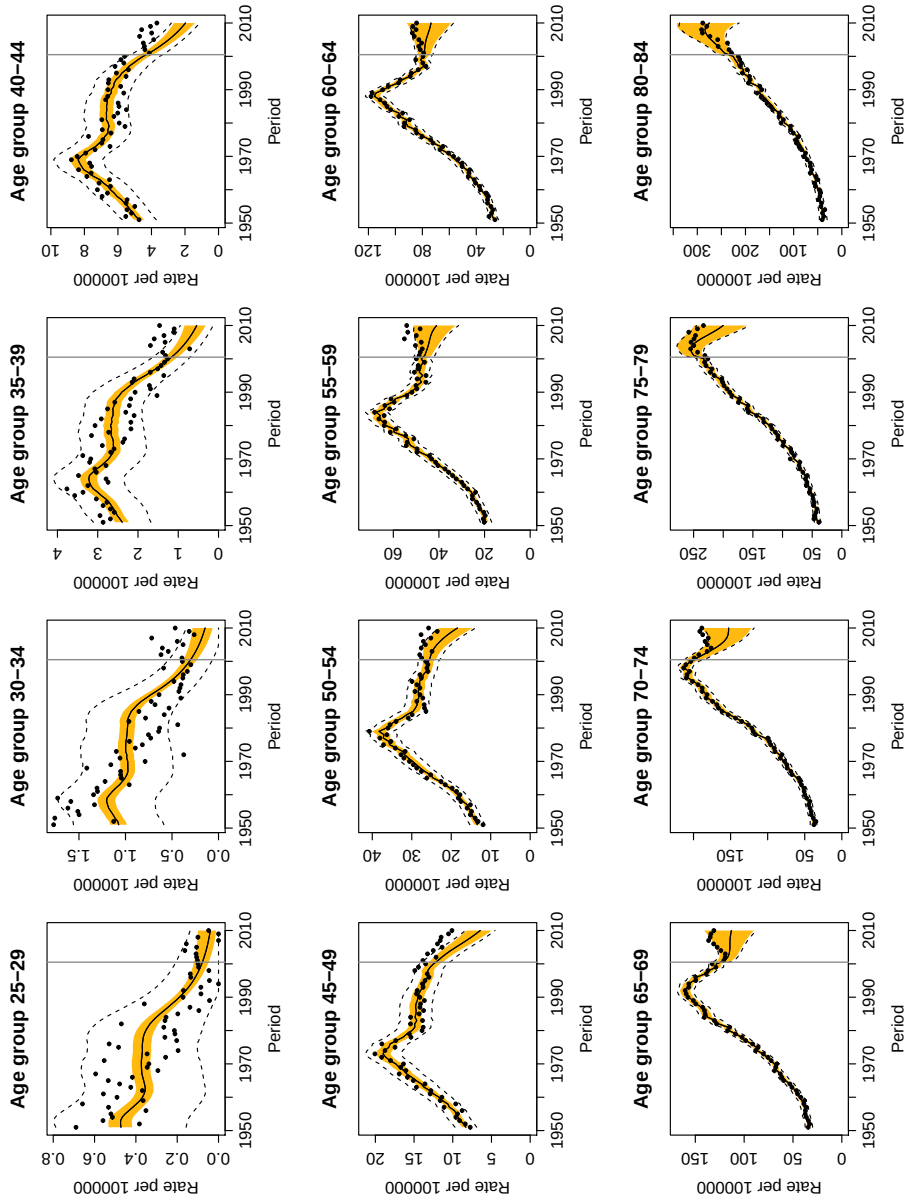


Figure A.8: Observed number of cases (dots) together with predicted mean rate within 95%-pointwise credible intervals (gold shaded) and predicted number of cases within 95%-pointwise credible intervals (dashed) for all age groups for female lung cancer mortality in the UK obtained without any form of aggregation. The vertical line shows where prediction started

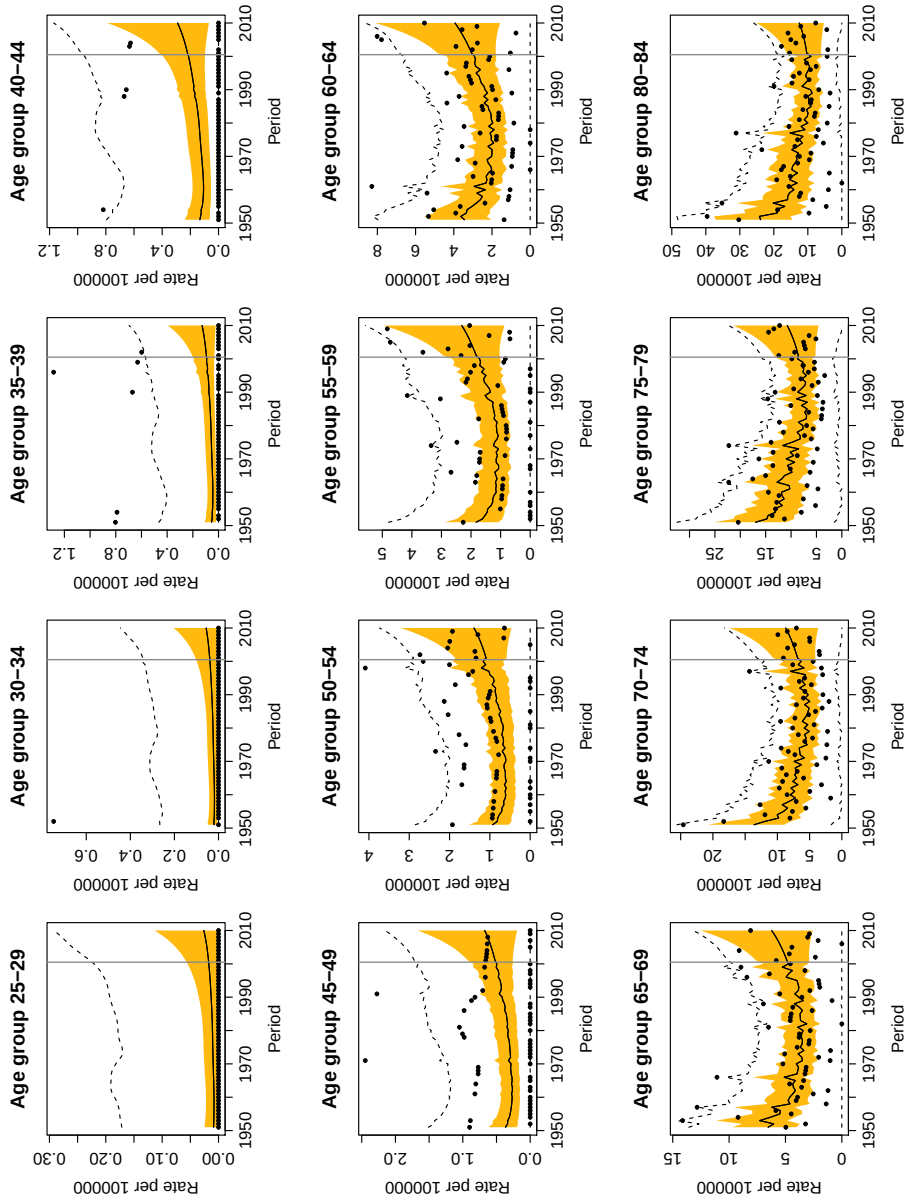


Figure A.9: Observed number of cases (dots) together with predicted mean rate within 95%-pointwise credible intervals (gold shaded) and predicted number of cases within 95%-pointwise credible intervals (dashed) for all age groups for female oesophagus cancer mortality in Norway obtained without any form of aggregation. The vertical line shows where prediction started

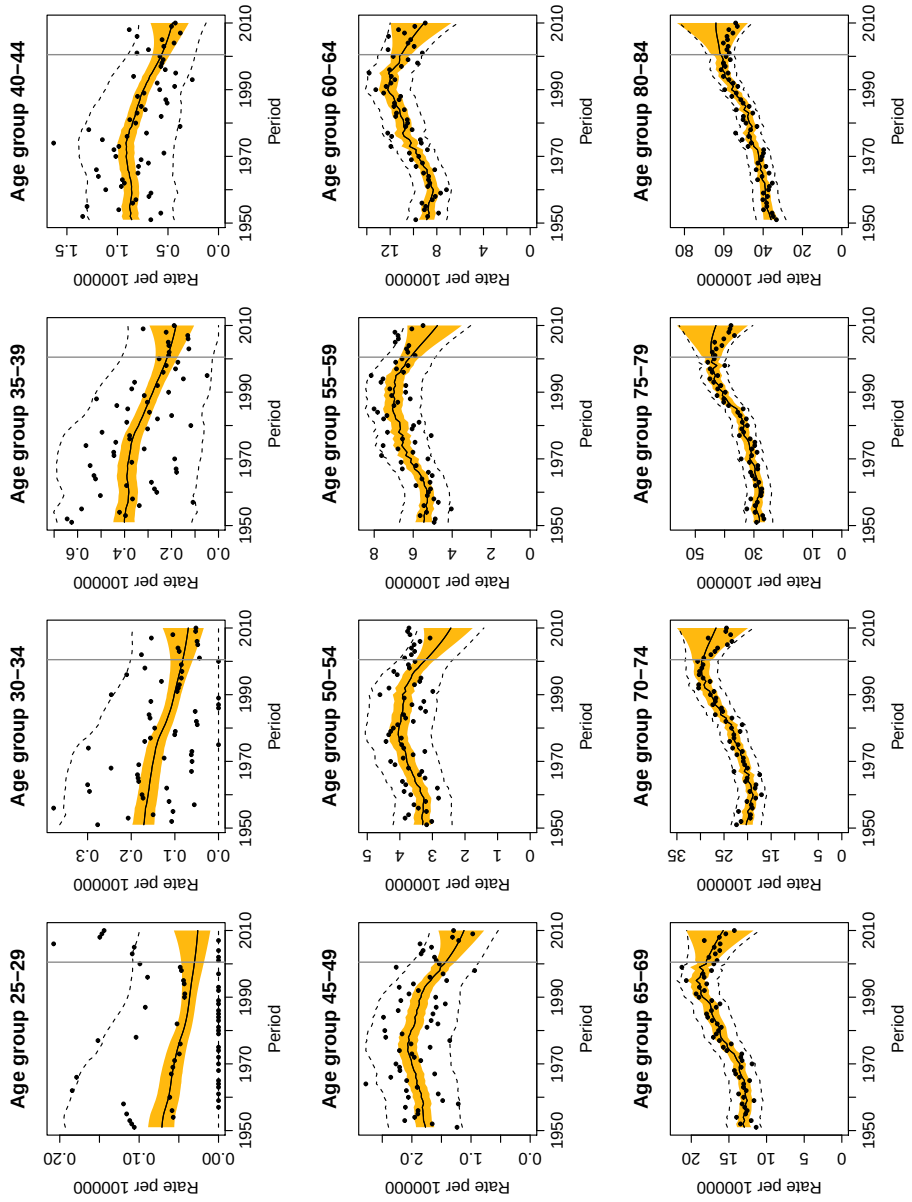
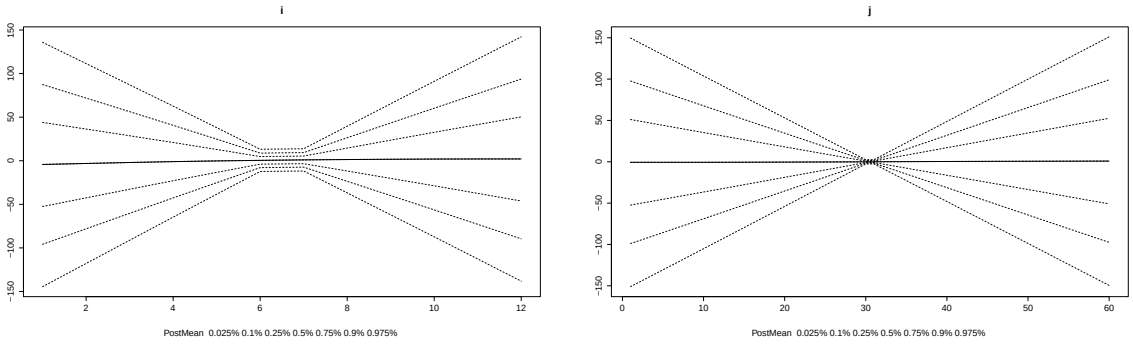
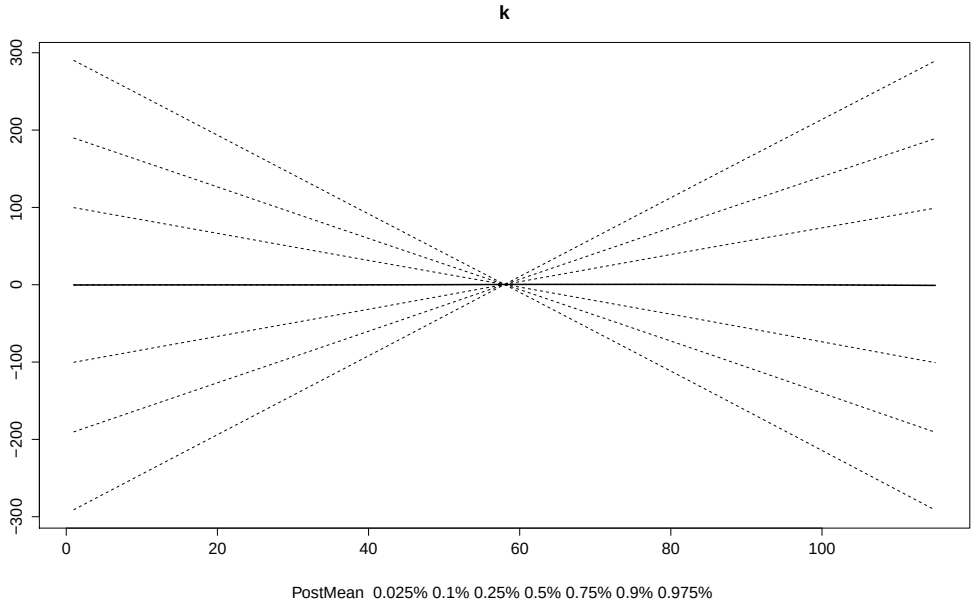


Figure A.10: Observed number of cases (dots) together with predicted mean rate within 95%-pointwise credible intervals (gold shaded) and predicted number of cases within 95%-pointwise credible intervals (dashed) for all age groups for female oesophagus cancer mortality in the UK obtained without any form of aggregation. The vertical line shows where prediction started



(a) Age effects

(b) Period effects



(c) Cohort effects

Figure A.11: Age, period and cohort effects for lung cancer Norway under method of no aggregation.

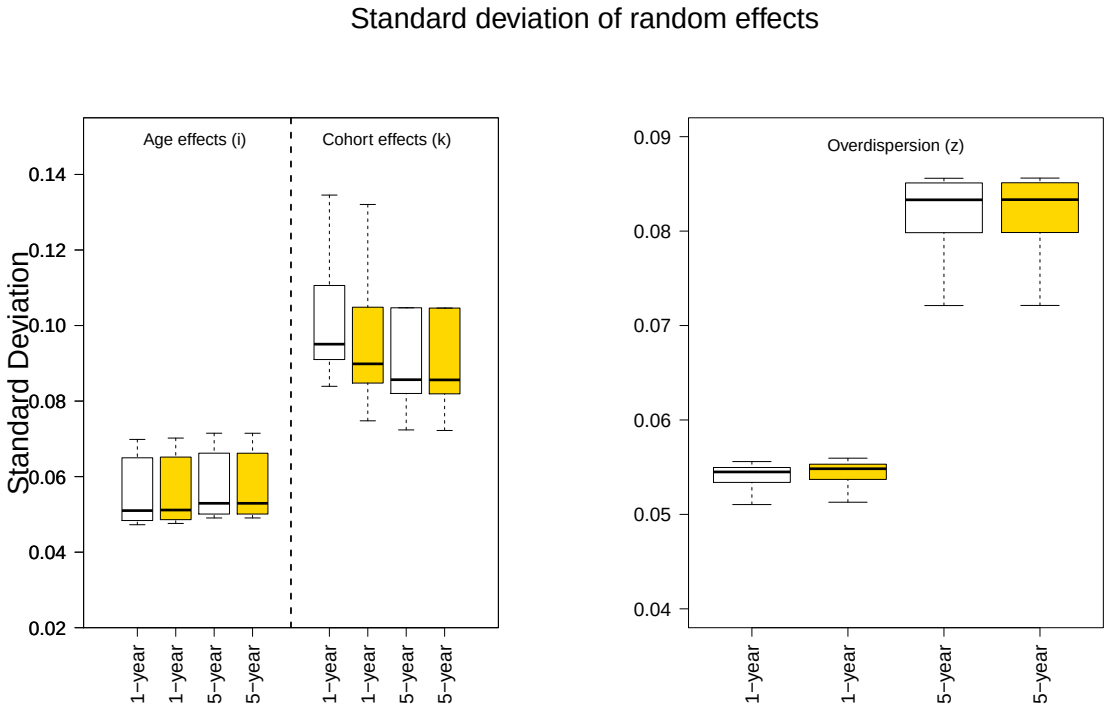


Figure A.12: Box plots of the standard deviation of the random effects (period effects have been excluded from the models for identifiabililty) for both the model with no aggregation (1-year) and that of model-specific aggregation (5-year) for Norway lung cancer data. Shown are results obtained when the priors for the RW2 precision parameters of the time effects (age and cohort) are scaled to have a marginal variance = 1 (gold) and when they are not scaled (plain).

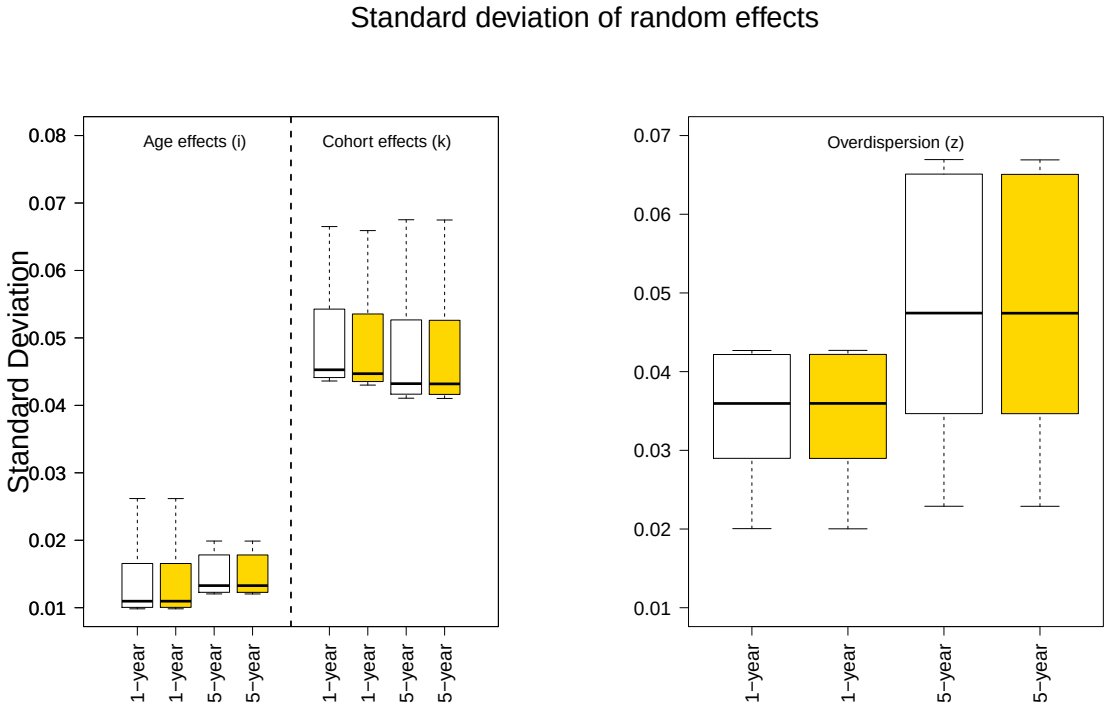


Figure A.13: Box plots of the standard deviation of the random effects (period effects have been excluded from the models for identifiabililty) for both the model with no aggregation (1-year) and that of model-specific aggregation (5-year) for UK lung cancer data. Shown are results obtained when the priors for the RW2 precision parameters of the time effects (age and cohort) are scaled to have a marginal variance = 1 (gold) and when they are not scaled (plain).

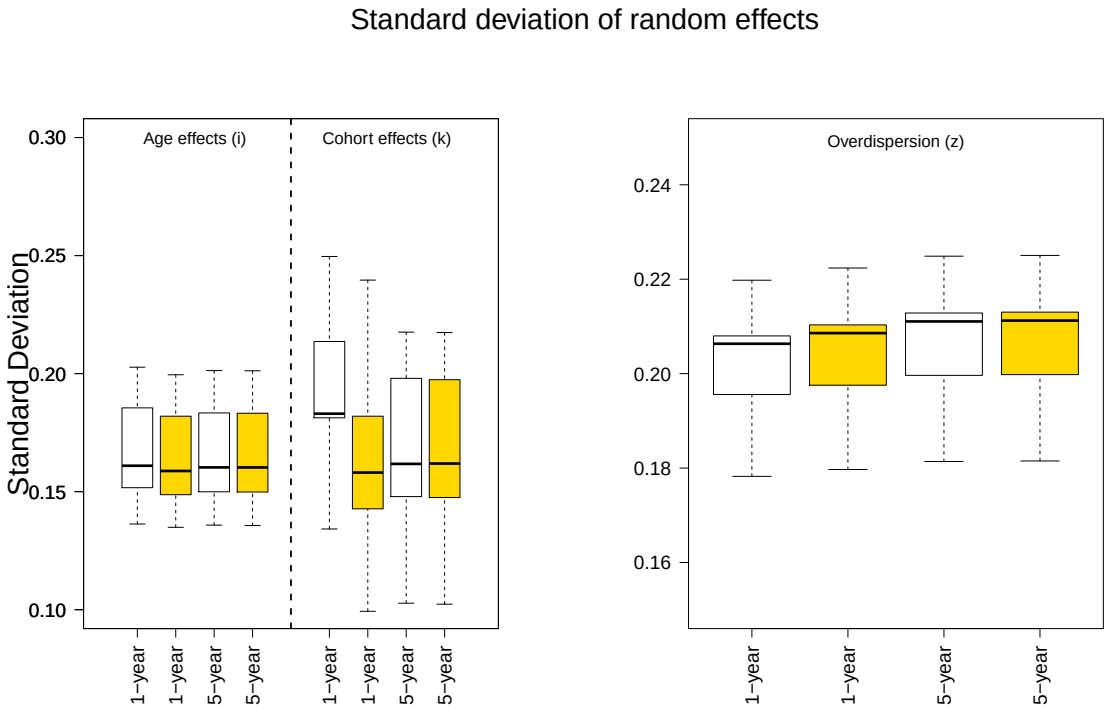


Figure A.14: Box plots of the standard deviation of the random effects (period effects have been excluded from the models for identifiabililty) for both the model with no aggregation (1-year) and that of model-specific aggregation (5-year) for Norway oesophagus cancer data. Shown are results obtained when the priors for the RW2 precision parameters of the time effects (age and cohort) are scaled to have a marginal variance = 1 (gold) and when they are not scaled (plain).

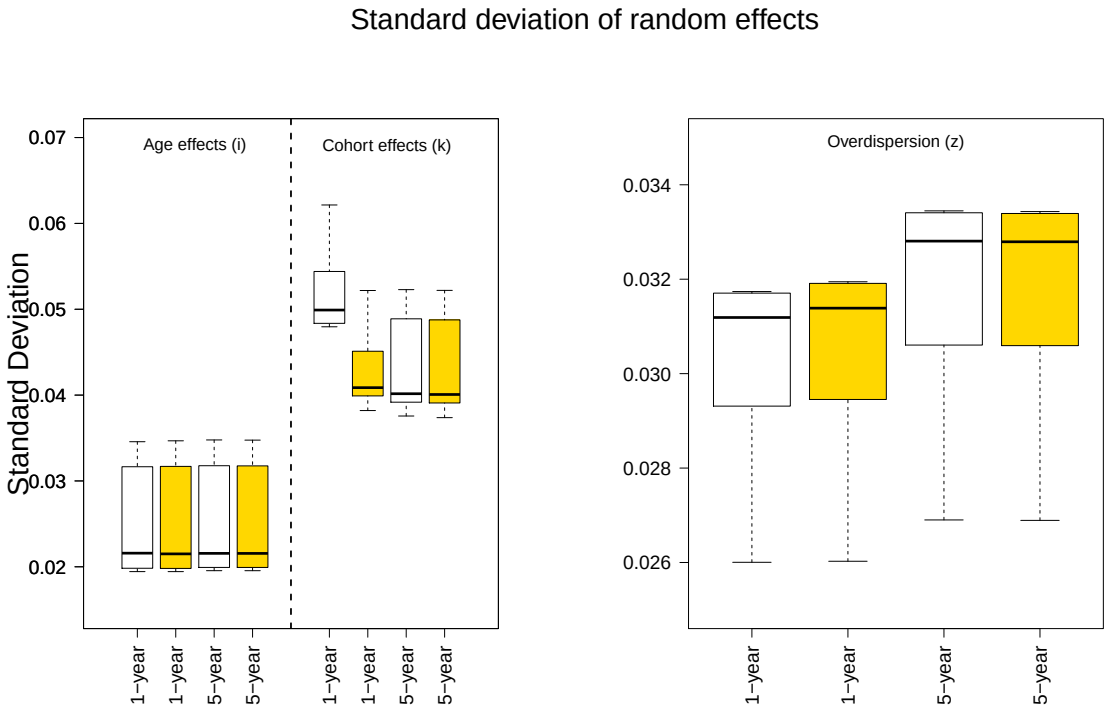


Figure A.15: Box plots of the standard deviation of the random effects (period effects have been excluded from the models for identifiabililty) for both the model with no aggregation (1-year) and that of model-specific aggregation (5-year) for UK oesophagus cancer data. Shown are results obtained when the priors for the RW2 precision parameters of the time effects (age and cohort) are scaled to have a marginal variance = 1 (gold) and when they are not scaled (plain).

Appendix B

Selected R-codes

1: Simulation study

The following simulates random numbers and plots them together with posterior means and quantiles (95 % credible intervals) obtained using RW1-model and RW2-model as smoothing priors with fixed precision $\kappa = 1$. Generates Figure 3.1

```
# load the INLA library
library(INLA)
# empty vector to store random numbers
x = c()
x[1] = 0
# length of the vector
n = 50
# simulate the numbers
for(i in 2:n){
x[i] = x[i-1] + rnorm(1, 0, sd=1)
}

par(mfrow=c(2,2))
# smoothing using the RW1
y = 1:n
data = data.frame(x=x, y=y)
# plot the points
plot(x, col = "red", xlab = "Index", ylab = "", main = "Random walk 1")
# define the model, specify the hyperpriors
```

```
formula1 =x~1+f(y, model="rw1", scale.model=TRUE, hyper = list(prec=list
           (fixed=TRUE, initial=0, prior="loggamma", param=c(1,0.00005))))
result1 =inla(formula1, data=data, family="gaussian", control.predictor=list(
    compute=T))
# add a plot of the mean and 95% credible intervals based on RW1 smoothing
matplot(result1$summary.fitted.values[,3:5], add=TRUE, type="l", lty=1, col="blue")

# smoothing using the RW2
plot(x, col="red", xlab ="Index", ylab="", main= "Random walk 2")
# define the model, specify the hyperpriors
formula2= x~1+f(y, model="rw2", scale.model= TRUE, hyper = list(prec=list
           (fixed=TRUE, initial=0, prior="loggamma", param=c(1,0.00005))))
result2= inla(formula2, data=data, family="gaussian", control.predictor=list(
    compute=T))
# add a plot of the mean and 95% credible intervals based on RW2 smoothing
matplot(result2$summary.fitted.values[ ,3:5], add=TRUE, type="l", lty=1, col="black
    ")

# we repeat the same for another simulated data set
x = c()
x[1] = 0
n = 50
for(i in 2:n){
x[i] = x[i-1] + rnorm(1, 0, sd=0.5)
}

y = 1:n
# smoothing using the RW1
plot(x, col="red", xlab="Index", ylab="", main= "Random walk 1")
data = data.frame(x=x, y=y)
formula1= x~1+f(y, model="rw1", scale.model=TRUE, hyper=list(prec=list
           (fixed=TRUE, initial=0, prior="loggamma", param=c(1,0.00005))))
result1= inla(formula1, data=data, family="gaussian", control.predictor=list(
    compute=T))
matplot(result1$summary.fitted.values[ ,3:5], add=TRUE, type="l", lty=1, col="blue"
    )

# smoothing using the RW2
plot(x, col="red", xlab ="Index", ylab="", main= "Random walk 2")
formula2 = x~1+f(y, model="rw2", scale.model=TRUE, hyper=list(prec=list
           (fixed=TRUE, initial=0, prior="loggamma", param=c(1, 0.00005))))
result2 =inla(formula2, data=data, family="gaussian", control.predictor=list(
    compute=T))
matplot(result2$summary.fitted.values[ ,3:5], add=TRUE, type="l", lty=1, col="black
    ")
```

2: Marginal standard deviation

This generates Figure 3.2, the marginal standard deviations of the RW1-model and the RW2-model.

```
# load Matrix and MASS libraries
library(Matrix)
library(MASS)

# Function that calculates the standard marginal deviationns
marginal_std_dev = function(n) {
  # Define the structure matrix
  # for Random walk 1
  RW1 = toeplitz(c(2, -1, rep(0, n-3), -1))
  RW1[1, 1] = RW1[n, n] = 1
  RW1[n, 1] = RW1[1, n] = 0

  # for Random walk 2
  RW2 = toeplitz(c(6, -4, 1, rep(0, n-4), 1))
  RW2[1, 1] = RW2[n, n] = RW2[n-1, n-3] = 1
  RW2[n, 1] = RW2[1, n] = 0
  RW2[1, 3] = RW2[n, n-2]= 1
  RW2[1, 2] = RW2[2, 1] = RW2[n, n-1] = RW2[n-1, n] = -2
  RW2[2, 2] = RW2[n-1, n-1] = 5
  RW2[2, 3] = -4

  # Since structure matrices are not of full rank,
  # compute the generalised inverse instead
  inverse_RW1 = ginv(RW1)
  inverse_RW2 = ginv(RW2)
  # Extract diagonal elements in each case (these are the marginal variances)
  # and take the square root to obtain marginal standard deviations
  sd_RW1 = sqrt(diag(inverse_RW1))
  sd_RW2 = sqrt(diag(inverse_RW2))
  return (list(sd_RW1=sd_RW1, sd_RW2=sd_RW2))
}

par(mfrow=c(2,2), cex.lab= 1.3, cex.axis=1.3)
```

```
34
35 # compute marginal standard deviation for length n = 100
36 n = 100
37 marginal_std = marginal_std_dev(n)
38 # plot the marginal standard deviation for RW1
39 plot(marginal_std$sd_RW1, main="RW1", xlab="Index",
40      ylab="Marginal standard deviation", type="l")
41 # plot the marginal standard deviation for RW2
42 plot(marginal_std$sd_RW2, main= "RW2", xlab="Index",
43      ylab="Marginal standard deviation", type="l")
44
45 # compute marginal standard deviation for length n = 500
46 n = 500
47 marginal_std = marginal_std_dev(n)
48 # plot the marginal standard deviation for RW1
49 plot(marginal_std$sd_RW1, xlab="Index", ylab="Marginal standard deviation", type="l
    ")
50 # plot the marginal standard deviation for RW2
51 plot(marginal_std$sd_RW2, xlab="Index", ylab="Marginal standard deviation", type="l
    ")
```

3: BAPC R-code

This code shows how to obtain the retrospective predictions under the three temporal scales, method of no aggregation, five-year data aggregation and model-specific aggregation. It describes how to obtain Figure 5.2, 5.3 and 5.4. Plots for UK and oesophagus cancer are obtained analogously.

```
1 # load the BAPC and INLA libraries
2 library(BAPC)
3 library(INLA)
4
5 # load annual female lung cancer mortality and population counts for Norway
6 # 60 rows and 12 columns in both data sets
7 FemLC_NW <- read.table ("nw_counts.txt", row.names=1, header=T)
8 FemPY_NW <- read.table ("nw_pop.txt", row.names=1, header=T)
9
10 # generate APCList object based on annual data with no aggregation
11 # we specify the count and population data and also
```

```
# the grid factor (C=5), indicating that age-groups are 5 times
# wider than period intervals
lc_nw <- APCList(FemLC_NW, FemPY_NW, gf=5)

# obtain retrospective projection for 10 years
# based on yearly data with no aggregation
lc_nw <- BAPC(lc_nw, predict=list(npredict=10, retro=TRUE), model=list(age=
         list(model="rw2", prior="loggamma", param=c(1,0.00005), initial=4,
         scale.model=FALSE), period=list(include=TRUE, model="rw2", prior =
         "loggamma", param = c(1, 0.00005), initial=4, scale.model = FALSE),
         cohort=list(include=TRUE, model="rw2", prior="loggamma", param =
         c(1, 0.00005), initial=4, scale.model=FALSE), overdis=list(include=
         TRUE, model="iid", prior="loggamma", param=c(1, 0.005), initial=4)))

# this generates figure 5.4
plotBAPC(lc_nw, scale=10^5, type="ageSpecBoth", showdata=T, coladd="darkgoldenrod1"
    )

# analysis based on five year data aggregation

# load five-year aggregated data for lung cancer mortality and population counts
# 12 row and 12 columns in both data sets
# each row represents data aggregated over a five-year interval
aggFemLC_NW <- read.table("agg_nw_counts.txt", row.names=1, header=T)
aggFemPY_NW <- read.table("agg_nw_pop.txt", row.names=1, header=T)

# generate APCList object based on five-year aggregated data
# here the grid factor (C=1)
# since age groups and period intervals have the same width
agglc_nw<- APCList(aggFemLC_NW, aggFemPY_NW, gf=1)

# obtain retrospective projection for 10 years
# based on five-year aggregated data
agglc_nw <- BAPC(agglc_nw, predict=list(npredict=2, retro=TRUE), model=list(age=
            list(model="rw2", prior="loggamma", param=c(1,0.00005), initial = 4,
            scale.model=FALSE), period=list(include = TRUE, model="rw2", prior =
            "loggamma", param = c(1, 0.00005), initial=4,  scale.model = FALSE),
            cohort = list(include=TRUE, model="rw2", prior = "loggamma", param =
            c(1,0.00005), initial=4, scale.model=FALSE), overdis=list(include =
            TRUE, model="iid", prior="loggamma", param=c(1,0.005), initial=4)))

# this generates Figure 5.2
plotBAPC(agglc_nw, scale=10^5, type="ageSpecBoth", showdata=T, coladd="
    darkgoldenrod1")

```

```
55
56 # analysis based on model-specific aggregation
57
58 # change the period index of the BAPC object obtained based on annual data
59 # length=720 is the length of the data
60 # here we partition the periods into 12 groups
61 # each group made up of five-year periods
62 # and assign the same period index to all periods in the same group
63 periodindex(lc_nw) <- rep(1:12, each=5, length=720)
64
65 # based on the new period indexes,
66 # compute the corresponding cohort index k=(I-i)+j
67 cohortindex(lc_nw)<- (nage(lc_nw)-ageindex(lc_nw))+periodindex(lc_nw)
68
69 # Of note, the age index remains unchanged
70
71 # create new BAPC objects under model-specific aggregation
72 lc_nw_idx <- BAPC(lc_nw, predict=list(npredict=10, retro=TRUE), model=list(age=
73              list(model="rw2", prior="loggamma", param=c(1,0.00005), initial=4,
74              scale.model=FALSE), period=list(include=TRUE, model="rw2", prior=
75              "loggamma", param = c(1, 0.00005), initial=4, scale.model = FALSE),
76              cohort = list(include=TRUE, model="rw2", prior = "loggamma", param=
77              c(1,0.00005), initial=4, scale.model=FALSE), overdis=list(include=
78              TRUE, model="iid", prior="loggamma", param=c(1,0.005), initial=4)))
79
80 # obtain plots. This generates Figure 5.3
81 plotBAPC(lc_nw_idx, scale=10^5, type="ageSpecBoth", showdata=T, coladd="
       darkgoldenrod1")
```

4: Computing Scores

The following shows how we compute the absolute error and continuous ranked probability mean scores. Subsequently demonstrates how values in Table 5.3 and plots in Figure 5.5 are obtained. Norway lung cancer has been used as an illustration. Scores for UK and oesophagus cancer are obtained analogously.

```
1 # In the BAPC object (lc_nw) you find after running BAPC,
2 # all the predictions can be extracted using agespec.proj(lc_nw)
3
```

```
# For method of no aggregation,

# extract the mean and standard dev of the predictive distributions
pred_lc_nw<- agespec.proj(lc_nw)
# mean of the predictive distribution
mu_lc_nw<- sapply(pred_lc_nw, function(x){x[51:60, 2]})
# standard dev of the predictive distribution
sigma_lc_nw<- sapply(pred_lc_nw, function(x){x[51:60, 4]})

# Similarly we extract mean and standard dev
# under model-specific aggregation
pred_lc_nw_idx  <- agespec.proj(lc_nw_idx)
# mean of the predictive distribution
mu_lc_nw_idx <- sapply(pred_lc_nw_idx, function(x){x[51:60, 2]})
# standard dev of the predictive distribution
sigma_lc_nw_idx <- sapply(pred_lc_nw_idx, function(x){x[51:60, 4]})

# true observations, which we compare the predictions to
# mortality counts data for female lung cancer
fem_lc_nw <- as.matrix(FemLC_NW[51:60, ])

# function that computes the scores, CRPS and AE
compute_score <- function (y, mu, sigma){
     CRPS= matrix(data=NA, nrow=nrow(y), ncol=ncol(y))
     AE= matrix(data=NA, nrow=nrow(y), ncol=ncol(y))
     ybar= matrix(data=NA, nrow=nrow(y), ncol=ncol(y))
     for(j in 1:nrow(y)){
     for(i in 1:ncol(y)){
          AE[j, i]   = abs(y[j, i] - mu[j, i])
          ybar[j, i] = (y[j, i] - mu[j, i]) / sigma[j, i]
          CRPS[j, i] = sigma[j, i]*(ybar[j, i]*(2*pnorm(ybar[j, i],0,1)-1)
                     + 2*dnorm(ybar[j, i], 0, 1) - (1/sqrt(pi)))
         }
      }
     return(list(CRPS=CRPS, AE=AE))
}

# now compute scores
# under method of no aggregation
score_lc_nw <- compute_score(fem_lc_nw, mu_lc_nw, sigma_lc_nw)

# under model-specific aggregation
score_lc_nw_idx <- compute_score(fem_lc_nw, mu_lc_nw_idx, sigma_lc_nw_idx)

# function that computes the cumulative average CRPS and AE
```

```
cumulative_average <- function (CRPS, AE){
      CRPS_j = matrix(data=NA, nrow=nrow(CRPS), ncol=1)
      AE_j = matrix(data=NA, nrow=nrow(CRPS), ncol=1)
      CRPS_j[1, ] = mean(CRPS[1, 1:ncol(CRPS)])
      AE_j[1, ] = mean(AE[1, 1:ncol(AE)])
      for(j in 2:nrow(CRPS)){
         CRPS_j[j, 1] = mean(c(CRPS[1:j, ]))
         AE_j[j ,1] = mean(c(AE[1:j, ]))
       }
      return(list(CRPS_j = CRPS_j, AE_j = AE_j))
}

# compute cumulative average under method of no aggregation
score_lc_nw_j <- cumulative_average(score_lc_nw$CRPS, score_lc_nw$AE)

# extract cumulative average CRPS and AE
# a plot of these values are seen in Figure 5.5
AE_lc_nw_j <- score_lc_nw_j$AE_j
CRPS_lc_nw_j <- score_lc_nw_j$CRPS_j

# the mean CRPS and AE
# which correspond to the last time points of the cumulative average
meanAE_lc_nw<- AE_lc_nw_j[nrow(AE_lc_nw_j), 1]
meanCRPS_lc_nw<- CRPS_lc_nw_j[nrow(CRPS_lc_nw_j), 1]

# compute cumulative average for model-specific aggregation
score_lc_nw_idx_j<- cumulative_average(score_lc_nw_idx$CRPS, score_lc_nw_idx$AE)

# extract cumulative average CRPS and AE
AE_lc_nw_idx_j<- score_lc_nw_idx_j$AE_j
CRPS_lc_nw_idx_j<- score_lc_nw_idx_j$CRPS_j

# the mean CRPS and AE
# which correspond to the last time points of the cumulative average
meanAE_lc_nw_idx <- AE_lc_nw_idx_j[nrow(AE_lc_nw_idx_j), 1]
meanCRPS_lc_nw_idx <- CRPS_lc_nw_idx_j[nrow(CRPS_lc_nw_idx_j), 1]

# these values are the entries of Table 5.3 and 5.4
meanAE_lc_nw
meanAE_lc_nw_idx
meanCRPS_lc_nw
meanCRPS_lc_nw_idx

# plot the cumulative average of mean AE (dotted) and CRPS (lines)
# based on yearly data without aggregation (black)
```

```
# and that of five-year model-specific aggregation (gold)
# generates Figure 5.5
year <- 2001:2010
plot(year, CRPS_lc_nw_j, type="s", main ="Norway", ylab = "cumulative
    average scores", xlab ="Year", xlim=c(2001,2010), ylim=c(0,10), lwd=2)
lines (year, CRPS_lc_nw_idx_j, type= "s", lwd=2, col="gold")
lines (year, AE_lc_nw_j, type="s", lty=3, lwd=2)
lines (year, AE_lc_nw_idx_j, type= "s", lty=3, lwd=2, col="gold")
```

Bibliography

Anderson, G. and Horvath, J. (2004). The growing burden of chronic disease in America. *Public Health Reports*, 119(3):263–270.

Armstrong, B. K. (1992). The role of the cancer registry in cancer control. *Cancer Causes & Control*, 3(6):569–579.

Bashir, S. and Estève, J. (2001). Projecting cancer incidence and mortality using Bayesian age–period–cohort models. *Journal of Epidemiology and Biostatistics*, 6(3):287–296.

Berzuini, C. and Clayton, D. (1994). Bayesian analysis of survival on multiple time scales. *Statistics in Medicine*, 13(8):823–838.

Besag, J., Green, P., Higdon, D., and Mengersen, K. (1995). Bayesian computation and stochastic systems. *Statistical Science*, 10(1):3–41.

Bray, F. and Møller, B. (2006). Predicting the future burden of cancer. *Nature Reviews Cancer*, 6(1):63–74.

Bray, I. (2002). Application of Markov chain Monte Carlo methods to projecting cancer incidence and mortality. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 51(2):151–164.

Bray, I., Brennan, P., and Boffetta, P. (2000). Projections of alcohol-and tobacco-related cancer mortality in central Europe. *International Journal of Cancer*, 87(1):122–128.

Bray, I., Brennan, P., and Boffetta, P. (2001). Recent trends and future projections of lymphoid neoplasms–a Bayesian age–period–cohort analysis. *Cancer Causes & Control*, 12(9):813–820.

Brillinger, D. R. (1986). A biometrics invited paper with discussion: the natural variability of vital rates and associated statistics. *Biometrics*, 42(4):693–734.

Brouhns, N., Denuit, M., and Vermunt, J. K. (2002). A Poisson log-bilinear regression approach to the construction of projected lifetables. *Insurance: Mathematics and Economics*, 31(3):373–393.

Butt, Z. and Haberman, S. (2009). Ilc: A collection of R functions for fitting a class of Lee-Carter mortality models using iterative fitting algorithms. *Actuarial Research Paper*, (190).

Cairns, A. J., Blake, D., and Dowd, K. (2006). A two-factor model for stochastic mortality with parameter uncertainty: Theory and calibration. *Journal of Risk and Insurance*, 73(4):687–718.

Carstensen, B. (2007). Age–period–cohort models for the Lexis diagram. *Statistics in Medicine*, 26(15):3018–3045.

Carter, L. R. and Prskawetz, A. (2001). Examining structural shifts in mortality using the Lee-Carter method. pages 1–17.

Clayton, D. and Schifflers, E. (1987a). Models for temporal variation in cancer rates. i: age–period and age–cohort models. *Statistics in medicine*, 6(4):449–467.

Clayton, D. and Schifflers, E. (1987b). Models for temporal variation in cancer rates. ii: age–period–cohort models. *Statistics in medicine*, 6(4):469–481.

Clements, M. S., Armstrong, B. K., and Moolgavkar, S. H. (2005). Lung cancer rate predictions using generalized additive models. *Biostatistics*, 6(4):576–589.

Dawid, A. P. (1984). Present position and potential developments: Some personal views: Statistical theory: The prequential approach. *Journal of the Royal Statistical Society. Series A (General)*, 147(2):278–292.

Dawid, A. P. and Sebastiani, P. (1999). Coherent dispersion criteria for optimal experimental design. 27(1):65–81.

DeGregori, J. (2013). Challenging the axiom: does the occurrence of oncogenic mutations truly limit cancer development with age? *Oncogene*, 32(15):1869–1875.

Ferlay, J., Shin, H.-R., Bray, F., Forman, D., Mathers, C., and Parkin, D. M. (2010). Estimates of worldwide burden of cancer in 2008: Globocan 2008. *International Journal of Cancer*, 127(12):2893–2917.

Ferlay, J., Steliarova-Foucher, E., Lortet-Tieulent, J., Rosso, S., Coebergh, J., Comber, H., Forman, D., and Bray, F. (2013). Cancer incidence and mortality patterns in Europe: estimates for 40 countries in 2012. *European Journal of Cancer*, 49(6):1374–1403.

Girosi, F. and King, G. (2007). Understanding the Lee-Carter mortality forecasting method. `http://gking.harvard.edu/files/lc.pdf`. [Online; Accessed 20-November-2014].

Gneiting, T. (2008). Editorial: probabilistic forecasting. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 171(2):319–321.

Gneiting, T. and Katzfuss, M. (2014). Probabilistic Forecasting. *Annual Review of Statistics and Its Application*, 1(1):125–151.

Gneiting, T. and Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378.

Good, I. J. (1952). Rational decisions. *Journal of the Royal Statistical Society. Series B (Methodological)*, 14(1):107–114.

Held, L. and Riebler, A. (2013). Comment on âĂIJassessing validity and application scope of the intrinsic estimator approach to the age-period-cohort (apc) problemâĂİ. *Demography*, 50(6):1977–1979.

Held, L., Rufibach, K., and Balabdaoui, F. (2010). A score regression approach to assess calibration of continuous probabilistic predictions. *Biometrics*, 66(4):1295–1305.

Heuer, C. (1997). Modeling of time trends and interactions in vital rates using restricted regression splines. *Biometrics*, 53(1):161–177.

Holand, A. M., Steinsland, I., Martino, S., and Jensen, H. (2013). Animal models and integrated nested Laplace approximations. *G3: Genes| Genomes| Genetics*, 3(8):1241–1251.

Holford, T. (2005). Age-period-cohort analysis. In Armitage, P. and Colton, T., editors, *Encyclopaedia of Biostatistics*, pages 105–123. John Wiley and Sons, West Sussex, 2nd edition.

Holford, T. R. (1983). The estimation of age, period and cohort effects for vital rates. *Biometrics*, 39(2):311–324.

Holford, T. R. (1985). An alternative approach to statistical age-period-cohort analysis. *Journal of Chronic Diseases*, 38(10):831–836.

Holford, T. R. (1998). Age–period–cohort analysis. *Encyclopedia of biostatistics*.

Jones, J. W., Hansen, J. W., Royce, F. S., and Messina, C. D. (2000). Potential benefits of climate forecasting to agriculture. *Agriculture, ecosystems & environment*, 82(1):169–184.

Jürgens, V., Ess, S., Cerny, T., and Vounatsou, P. (2014). A Bayesian generalized age–period–cohort power model for cancer projections. *Statistics in Medicine*, 33(26):4627–4636.

Kang, S. Y., McGree, J., Baade, P., and Mengersen, K. (2014). An investigation of the impact of various geographical scales for the specification of spatial dependence. *Journal of Applied Statistics*, 41(11):2515–2538.

Keilman, N., Pham, D. Q., and Hetland, A. (2002). Why population forecasts should be probabilistic–illustrated by the case of Norway. *Demographic Research*, 6(15):409–454.

Kennedy, S. R., Loeb, L. A., and Herr, A. J. (2012). Somatic mutations in aging, cancer and neurodegeneration. *Mechanisms of Ageing and Development*, 133(4):118–126.

Knorr-Held, L. and Rainer, E. (2001). Projections of lung cancer mortality in West Germany: a case study in bayesian prediction. *Biostatistics*, 2(1):109–129.

Lee, R. and Miller, T. (2001). Evaluating the performance of the Lee-Carter method for forecasting mortality. *Demography*, 38(4):537–549.

Lee, R. D. and Carter, L. R. (1992). Modeling and forecasting US mortality. *Journal of the American Statistical Association*, 87(419):659–671.

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

Lunn, D. J., Thomas, A., Best, N., and Spiegelhalter, D. (2000). Winbugs-a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing*, 10(4):325–337.

Macdonald, A. (1996a). An actuarial survey of statistical models for decrement and transition data-i: Multiple state, poisson and binomial models. *British Actuarial Journal*, 2(01):129–155.

Macdonald, A. (1996b). An actuarial survey of statistical models for decrement and transition data, ii. competing risks, non-parametric and regression models. *British Actuarial Journal*, 2(02):429–448.

Macdonald, A. (1996c). An actuarial survey of statistical models for decrement and transition data, iii. counting process models. *British Actuarial Journal*, 2(03):703–726.

Martino, S., Aas, K., Lindqvist, O., Neef, L. R., and Rue, H. (2011). Estimating stochastic volatility models using integrated nested Laplace approximations. *The European Journal of Finance*, 17(7):487–503.

McCullagh, P. (1984). Generalized linear models. *European Journal of Operational Research*, 16(3):285–292.

Møller, B. (2004). Prediction of cancer incidence–methodological considerations and trends in the Nordic countries 1958–2022. *Faculty of Medicine, University of Oslo, Unipuc AS, Oslo.*

Møller, B., Fekjær, H., Hakulinen, T., Sigvaldason, H., Storm, H. H., Talbäck, M., and Haldorsen, T. (2003). Prediction of cancer incidence in the Nordic countries: empirical comparison of different approaches. *Statistics in Medicine*, 22(17):2751–2766.

Møller, B., Fekjær, H., Hakulinen, T., Tryggvadóttir, L., Storm, H. H., Talbäck, M., and Haldorsen, T. (2002). Prediction of cancer incidence in the Nordic countries up to the year 2020. *European Journal of Cancer Prevention*, 11(1):S1–S96.

Nowatzki, J., Moller, B., Demers, A., et al. (2011). Projection of future cancer incidence and new cancer cases in Manitoba, 2006–2025. *Chronic Diseases in Canada*, 31(2):71–78.

Ocaña-Riola, R. (2007). The misuse of count data aggregated over time for disease mapping. *Statistics in Medicine*, 26(24):4489–4504.

Osmond, C. and Gardner, M. (1982). Age, period and cohort models applied to cancer mortality rates. *Statistics in Medicine*, 1(3):245–259.

Paul, M., Riebler, A., Bachmann, L., Rue, H., and Held, L. (2010). Bayesian bivariate meta-analysis of diagnostic test studies using integrated nested Laplace approximations. *Statistics in Medicine*, 29(12):1325–1339.

Pulwarty, R. S. and Redmond, K. T. (1997). Climate and salmon restoration in the Columbia river basin: The role and usability of seasonal forecasts. *Bulletin of the American Meteorological Society*, 78(3):381–397.

R Core Team (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Renshaw, A. E. and Haberman, S. (2006). A cohort-based extension to the Lee–Carter model for mortality reduction factors. *Insurance: Mathematics and Economics*, 38(3):556–570.

Riebler, A. and Held, L. (2015). Projecting the future burden of cancer - Bayesian age-period-cohort analysis ready for routine use. Technical Report (Unpublished), Department of Mathematics, NTNU, Norway & Department of Biostatistics, University of Zurich, Switzerland.

Riebler, A., Held, L., Rue, H., et al. (2012). Estimation and extrapolation of time trends in registry dataÃČÂćÃćâĂŽÂňÃćâĆňÂİborrowing strength from related populations. *The Annals of Applied Statistics*, 6(1):304–333.

Rue, H. and Held, L. (2005). *Gaussian Markov random fields: theory and applications*. CRC Press.

Rue, H., Martino, S., and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the Royal Statistical Society: Series B (statistical methodology)*, 71(2):319–392.

Rutherford, M. J., Thompson, J. R., and Lambert, P. C. (2012). Projecting cancer incidence using age-period-cohort models incorporating restricted cubic splines. *The International Journal of Biostatistics*, 8(1).

Schmid, V. J. and Held, L. (2007). Bayesian age-period-cohort modeling and prediction-BAMP. *Journal of Statistical Software*, 21(8):1–15.

Schrödle, B. and Held, L. (2011a). A primer on disease mapping and ecological regression using INLA. *Computational Statistics*, 26(2):241–258.

Schrödle, B. and Held, L. (2011b). Spatio-temporal disease mapping using INLA. *Environmetrics*, 22(6):725–734.

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

Stewart, B. and Wild, C. P. (2014). International Agency for Research on Cancer, WHO. (2014) *World Cancer Report* 2014 [Online]. Available From: `http://www.thehealthwell.info/node/725845`. [Accessed: 20th May 2015].

Tuljapurkar, S., Li, N., and Boe, C. (2000). A universal pattern of mortality decline in the g7 countries. *Nature*, 405(6788):789–792.

Vandeschrick, C. et al. (2001). The Lexis diagram, a misnomer. *Demographic Research*, 4(3):97–124.

Yang, S. S., Yue, J. C., and Huang, H.-C. (2010). Modeling longevity risks using a principal component approach: A comparison with existing stochastic mortality models. *Insurance: Mathematics and Economics*, 46(1):254–270.