Multivariate Age-Period-Cohort Models

Doctoral Examination

Andrea Riebler

Biostatistics Unit, Institute of Social and Preventive Medicine
University of Zurich

15 July 2010

University of Zurich

Main components of the doctoral thesis

Paper I: **The analysis of heterogeneous time trends in multivariate age-period-cohort models**
Andrea Riebler & Leonhard Held
Paper published in *Biostatistics (2010), 11, 1, pp. 57–69*.

Paper II: **A conditional approach for inference in multivariate age-period-cohort models**
Leonhard Held & Andrea Riebler
Paper accepted for publication in *Statistical Methods in Medical Research*.

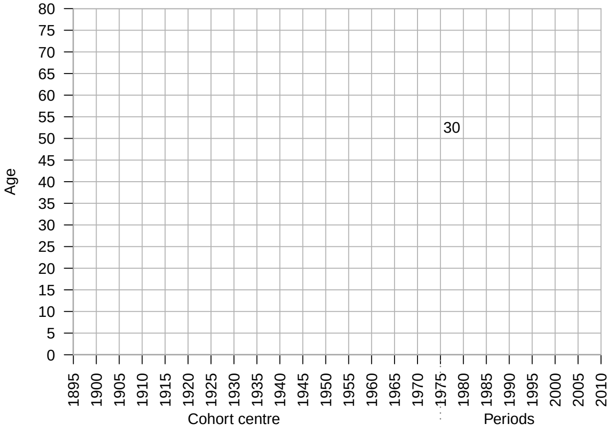
Paper III: **Correlated multivariate age-period-cohort models**
Andrea Riebler, Leonhard Held & Håvard Rue
Technical Report, University of Zurich.

Paper IV: **Suicide mortality in Switzerland: Gender-specific differences and the impact of family integration**
Andrea Riebler, Leonhard Held, Håvard Rue & Matthias Bopp
Technical Report, University of Zurich.

Overview

1. Introduction

- Incidence or mortality counts are typically stratified by age and period.
- Such data can be visualised in a Lexis-diagram.



Age-period-cohort analysis

- Goals:
  - Detecting temporal patterns in such data as they could provide important clues for disease etiology.
  - Extrapolation and forecasting.
- Age-period-cohort (APC) model is commonly used to describe vital rates using three time scales:
  - Age: age at diagnosis/death
  - Period: date of diagnosis/death
  - Cohort: date of birth
- Our extension: Multivariate APC models (Riebler & Held (2010), Biostatistics).

Age, period and cohort effects

On each time-scale operate causal factors potentially related to the disease.

Age effects

Related to factors that explain changes across different age groups.

Period effects

Related to factors that apply to all persons at a particular calendar time independent of their age, for instance advances in medical treatment.

Cohort effects

Related to factors that explain changes between different birth cohorts, for instance generation-specific habits or war.

2. The univariate age-period-cohort model

y_{ij}: Number of cases in age group i at period j
n_{ij}: Number of persons at risk in age group i at period j

$$\begin{aligned} y_{ij} \mid \eta_{ij} &\sim \text{Poisson}(n_{ij} \exp(\eta_{ij})) \\ \eta_{ij} &= \mu + \theta_i + \varphi_j + \psi_k \end{aligned}$$

with age effect θ_i, period effect φ_j and cohort effect ψ_k.

- To assure identifiability of the intercept μ we apply sum-to-zero constraints for each parameter vector $\boldsymbol{\theta}$, $\boldsymbol{\varphi}$, $\boldsymbol{\psi}$.
- The cohort index k is a linear function of i and j:
  - For equally spaced data: $k = I - i + j$.
  - For unequally spaced data: $k = M \cdot (I - i) + j$, where M is the ratio of age group to period interval length.

Identifiability problems

Time trends are not identifiable due to the linear dependence between age, period and cohort indices.

Equally spaced data

For any value of $a \in R$, the linear transformations

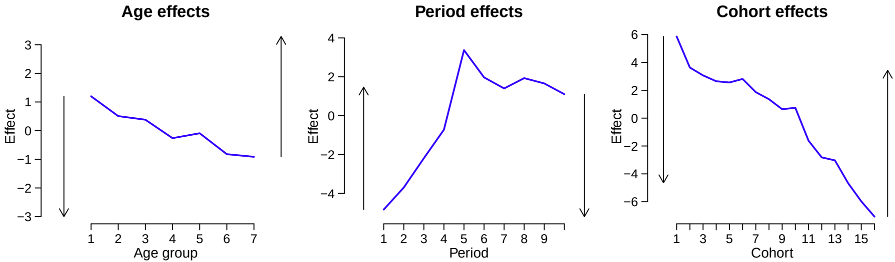
$$\begin{aligned}
\theta_i &\rightarrow \theta_i + a\left(i - \frac{(I+1)}{2}\right), \\
\varphi_j &\rightarrow \varphi_j - a\left(j - \frac{(J+1)}{2}\right), \\
\psi_k &\rightarrow \psi_k + a\left(k - \frac{(K+1)}{2}\right)
\end{aligned}$$

will leave η_{ij} unchanged and maintain the sum-to-zero constraints.

Note: Second differences are identifiable, but hard to interpret.

Holford (1983), Biometrics

Example: Equally spaced data



Identifiability problems (cont.)

Unequally spaced data

- A second identifiability problem emerges.
- Remember: M is the ratio of age group to period interval length.
- For any value of $b_1, \ldots, b_M \in R$ (subject to $b_1 + \ldots + b_M = 0$), the transformations

$$\begin{aligned} \varphi_j &= \varphi_j + b_{1+(j-1) \bmod M} \\ \psi_k &= \psi_k - b_{1+(k-1) \bmod M} \end{aligned}$$

will leave the linear predictor η_{ij} unchanged.

Holford (2006), Stat Med

Example: Unequally spaced data ($M = 5$)

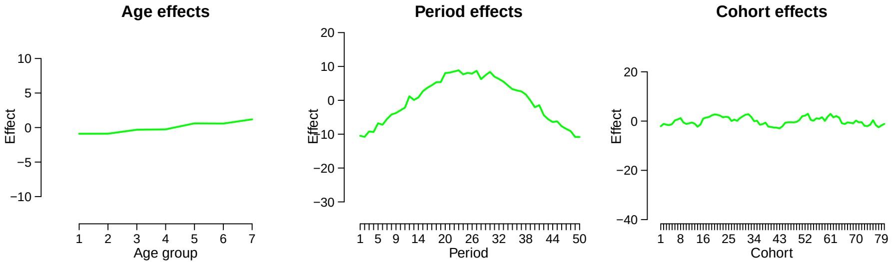


Table of contents

3. The multivariate age-period-cohort model

- Data now also depend on strata $r = 1, \ldots, R$.
- Most general formulation (apc model):

$$y_{ijr} | \eta_{ijr} \sim \text{Poisson}(n_{ijr} \exp(\eta_{ijr}))$$
$$\eta_{ijr} = \mu_r + \theta_{ir} + \varphi_{jr} + \psi_{kr},$$

  with stratum-specific intercept μ_r and stratum-specific age effect θ_{ir}, period effect φ_{jr} and cohort effect ψ_{kr}.

- Simpler models can be obtained. For example let us assume shared age effects (Apc model):

$$\eta_{ijr} = \mu_r + \theta_i + \varphi_{jr} + \psi_{kr}.$$

Identifiability of relative risks

- The multivariate APC model inherits all identifiability problems from the univariate APC model.
- However, differences

$$\begin{aligned}
\Delta_j^{(r)} &= \varphi_{jr} - \varphi_{jR} && \text{in the ApC model} \\
\Delta_k^{(r)} &= \psi_{kr} - \psi_{kR} && \text{in the APc model} \\
\Delta_{jk}^{(r)} &= \Delta_j + \Delta_k && \text{in the Apc model}
\end{aligned}$$

  are identifiable.
- Let $\Delta_\mu^{(r)} = \mu_r - \mu_R$. The adjusted differences

$$\begin{aligned}
\tilde{\Delta}_j^{(r)} &= \Delta_\mu^{(r)} + \Delta_j^{(r)} \\
\tilde{\Delta}_k^{(r)} &= \Delta_\mu^{(r)} + \Delta_k^{(r)}
\end{aligned}$$

  can be interpreted as (average) log relative risk.

Table of contents

3.1. Inference

Classical maximum likelihood (ML) approaches

- need at least one additional constraint for identifiability.
  Holford (1983), Biometrics
- overfit cohorts for which only a single observation exists.
  Besag *et al.* (1995), Stat Science
- may provide unstable estimates for unequally spaced intervals of age and period, resulting in an artificial saw-tooth pattern.
  Holford (1983), Biometrics; Holford (2006), Stat Med
- rely on strong parametric assumptions for future values of period and cohort effects. Bray (2002), JRSS-C

$\Rightarrow$ Here we use a Bayesian approach.

Bayesian age-period-cohort model

- Non-parametric smoothing priors are used for the main effects.
- A natural choice is to penalise second differences.
- The second order random walk (RW2) prior for φ is

$$f(\boldsymbol{\theta}|\tau_{\boldsymbol{\theta}}) \propto \tau_{\boldsymbol{\theta}}^{(I-2)/2} \exp\left(-\frac{\tau_{\boldsymbol{\theta}}}{2}\sum_{i=3}^{I}(\theta_i - 2\theta_{i-1} + \theta_{i-2})^2\right) \quad = \tau_{\boldsymbol{\theta}}^{(I-2)/2} \exp\left(-\frac{1}{2}\boldsymbol{\theta}^\top \mathbf{R}\boldsymbol{\theta}\right),$$

where $\mathbf{R}$ is the precision matrix:

$$\mathbf{R} = \tau_{\boldsymbol{\theta}} \begin{pmatrix} 1 & -2 & 1 & & & & \\ -2 & 5 & -4 & 1 & & & \\ 1 & -4 & 6 & -4 & 1 & & \\ & \ddots & \ddots & \ddots & \ddots & \ddots & \\ & & 1 & -4 & 6 & -4 & 1 \\ & & & 1 & -4 & 5 & -2 \\ & & & & 1 & -2 & 1 \end{pmatrix}.$$

$\Rightarrow$ Latent Gaussian Markov random field (GMRF) model.

Bayesian age-period-cohort model (cont.)

- To account for overdispersion add $z_{ijr} \sim \mathcal{N}(0, \tau_z^{-1})$ to η_{ijr}:

$$\eta_{ijr} = \mu_r + \theta_i + \varphi_{jr} + \psi_{kr} + z_{ijr}.$$

- All precision parameters are treated as unknown and suitable gamma-hyperpriors are assigned.

Inference in the Bayesian APC model

- MCMC: Using a reparameterisation from z_{ijr} to η_{ijr} each block of the main effects can be updated using Gibbs sampling (Besag et al. (1995), Stat Science).
  However, the full-conditional distribution of the linear predictor ξ_{ijr} is a non-standard distribution.

  We implemented two options in the programming language C:

  - Second-order Taylor expansion of the (unnormalised) log-full-conditional around a suitable value

    $\Rightarrow$ normally distributed Metropolis-Hastings proposal.

  - Auxiliary mixture sampling (Frühwirth-Schnatter and Wagner (2006), Biometrika; Frühwirth-Schnatter et al. (2009), Stat Comput)

    $\Rightarrow$ Gibbs sampling possible.

  In our applications the Taylor expansion was roughly twice as efficient.

Model choice in MCMC

- Frequently used: deviance information criterion (DIC).
- In hierarchical models with many random effects, DIC tends to under-penalise complex models (Plummer (2008), Biostatistics).
- We additionally use leave-one-out cross-validatory predictive checks based on replicating the z_{ijr}'s and subsequently the y_{ijr}'s (Marshall & Spiegelhalter (2003), Stat Med).

  $\Rightarrow$

  - the mean Dawid-Sebastiani score $\overline{\text{DSS}}$.
  - the mean ranked probability score $\overline{\text{RPS}}$.

  Czado *et al.* (2009), Biometrics; Gneiting *et al.* (2008), Test

Inference in the Bayesian APC model (cont.)

- Integrated nested Laplace approximation (INLA)
  (Rue et al. (2009), JRSS-B)
  - Deterministic alternative to MCMC in general latent Gaussian Markov random field models.
  - Incorporates posterior uncertainty with respect to hyperparameters.
  - Can be used for model assessment and comparison based on leave-one-out cross-validation, DIC and log marginal likelihood.
  - R-Interface available (`www.r-inla.org`).

Chronic obstructive pulmonary disease (COPD)

COPD is a serious lung disease making it hard to breathe.

Known risk factors:

- Smoking
- Air pollution
- Exposures to dusts, gases and chemical fumes
- Fog, ...

Data: Mortality counts among males in England & Wales (Hansell et al., 2003)

- $I = 7$ age groups: 15–24, 25–34, ..., 75+
- $J = 50$ periods: 1950–1999
- $K = 110$ birth cohorts.
- $R = 3$ regions: conurbations excluding Greater London, Greater London and non-conurbations.

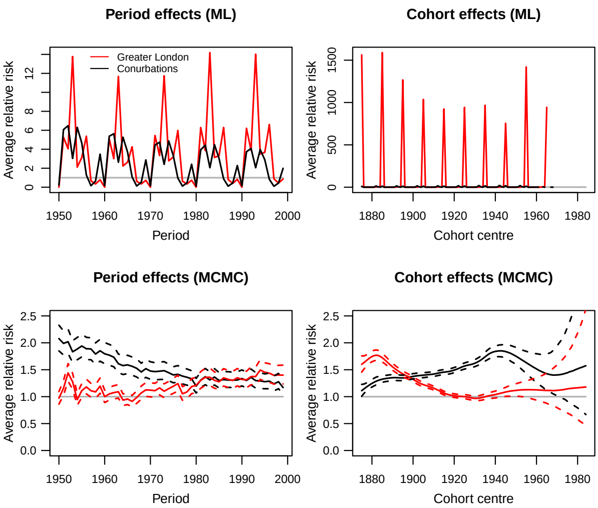
Note: Age groups and periods are unequally spaced.

Results

- Uppercase letters (A, P or C) denote a model where the corresponding effect is assumed to be the same over the different strata.
- Lowercase letters (a, p or c) denote a model where the effect is allowed to differ across strata.

	APC	aPC	ApC	APc	apC	aPc	Apc	apc
$\overline{\text{RPS}}$	51.52	39.37	52.38	41.4	24.23	27.64	**20.93**	**19.87**
$\overline{\text{DSS}}$	7.19	6.82	7.20	7.01	6.38	6.56	**6.29**	**6.17**

Relative risk estimates



Artificial cyclic patterns in period and cohort ML estimates!

Smoothing helps.

- 1952: “The Great Smog” in London.
- 1956, 1959, 1962: Air pollution episodes in Greater London.
- Differences in cohort effects might reflect lifetime smoking patterns.

INLA precision estimates

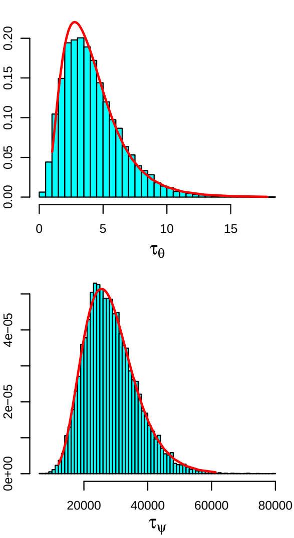


Table of contents

3.2. A conditional approach for inference

- Let $y_{ij\bullet} = y_{ij1} + \ldots + y_{ijR}$.
- It is easy to see that the Apc model for y_{ijr} implies that $\mathbf{y}_{ij}|y_{ij\bullet}$ is multinomial with individual success probability

$$\pi_{ijr} = \frac{\exp\left(\log\left(\frac{n_{ijr}}{n_{ijR}}\right) + \Delta_{\mu}^{(r)} + \Delta_{j}^{(r)} + \Delta_{k}^{(r)}\right)}{1 + \sum_{s=1}^{R-1} \exp\left(\log\left(\frac{n_{ijs}}{n_{ijR}}\right) + \Delta_{\mu}^{(s)} + \Delta_{j}^{(s)} + \Delta_{k}^{(s)}\right)}.$$

- Note that through conditioning, the original parameters are replaced by the differences $\Delta_j^{(r)}$ and $\Delta_k^{(r)}$.
- Age effects are no longer present in this formulation.
- It can be shown that the multinomial likelihood carries the same information about the relative risk parameters as the original Poisson likelihood.

Inference and model choice

- All parameters are identifiable and can be estimated with ML with suitable smoothing, if necessary (R-package `VGAM`).
- For overdispersion quasi-likelihood adjustments are possible
  $\Rightarrow$ However, if overdispersion is accounted for, different ML estimates as in a Poisson regression model are possible.
- Model choice is based on AIC or QAIC (Burnham & Anderson (2002), Springer, New York)

$$\text{QAIC} = -2 \log L/\phi + 2p,$$

  where ϕ is the estimated overdispersion parameter. If $\phi > 1$ one additional parameter must be added to p.
- Data stratified by more than one additional variable can be analysed.

COPD mortality: Extended model

Idea: Include also data on mortality among females.

- Potentially useful model:
  - different age effects across gender $g = ♀, ♂$.
  - different period and cohort effects across regions

$$\eta_{ijrg} = \mu_{rg} + \theta_{ig} + \varphi_{jr} + \psi_{kr}.$$

- Two separate multinomial logistic regression models can be fitted:
  - Conditioning $y_{ijr\bullet} \Rightarrow$ Estimation of age effects.
  - Conditioning $y_{ij\bullet g} \Rightarrow$ Estimation of period and cohort effects.

Results

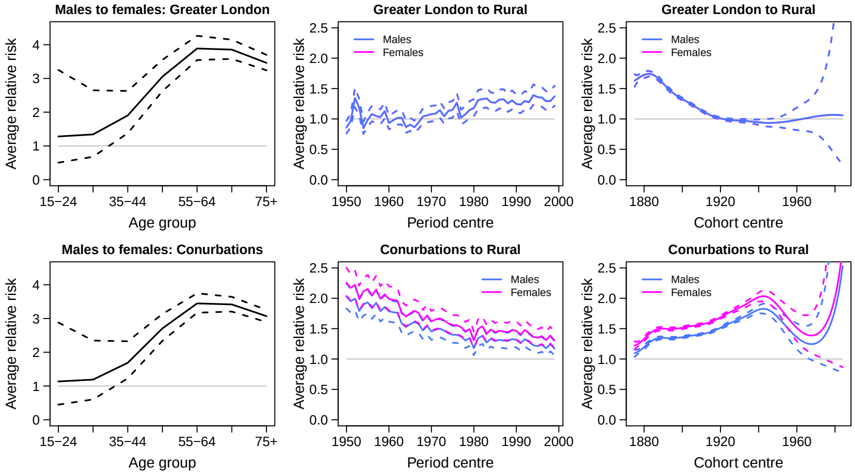

Males to females: Greater London
Average relative risk
Age group
15–24 35–44 55–64 75+
Greater London to Rural
Males
Females
Period centre
1950 1960 1970 1980 1990 2000
Greater London to Rural
Cohort centre
1880 1920 1960
Males to females: Conurbations
Conurbations to Rural
Conurbations to Rural

Table of contents

4. Correlated multivariate age-period-cohort models

When assuming separate time effects across strata it might nevertheless be plausible to assume some correlation. The inclusion of a correlation would actually represent a balance between separate and joint effects.

The benefits might be:

- Improvement of extrapolation of missing data for one region.
- Improvement of model choice criteria.

Correlated GMRF priors

- For simplicity: $R = 3$.
- Let $\mathbf{C} = \begin{pmatrix} 1 & \rho & \rho \\ \rho & 1 & \rho \\ \rho & \rho & 1 \end{pmatrix}$ denote a uniform correlation matrix.
- The random walks $\boldsymbol{\theta}_1$, $\boldsymbol{\theta}_2$, $\boldsymbol{\theta}_3$ can be correlated using the stacked vector $\tilde{\boldsymbol{\theta}} = (\boldsymbol{\theta}_1^\top, \boldsymbol{\theta}_2^\top, \boldsymbol{\theta}_3^\top)^\top$:

$$f(\tilde{\boldsymbol{\theta}}|\mathbf{C}_{\boldsymbol{\theta}}, \tau_{\boldsymbol{\theta}}) \propto |\tau_{\boldsymbol{\theta}} \mathbf{C}_{\boldsymbol{\theta}}^{-1}|^{(I-2)/2} \exp\left(-\frac{1}{2}\tilde{\boldsymbol{\theta}}^\top \{\mathbf{C}_{\boldsymbol{\theta}}^{-1} \otimes \mathbf{R}\} \tilde{\boldsymbol{\theta}}\right).$$

- Correlated overdispersion can also be incorporated: $\mathbf{z}_{ij} = (z_{ij1}, z_{ij2}, z_{ij3})^\top \sim \mathcal{N}(\mathbf{0}, \tau_{\mathbf{z}}^{-1}\mathbf{C}_{\mathbf{z}})$.
- All correlations are treated as unknown.
- Inference via MCMC or INLA

Prior on correlation parameters

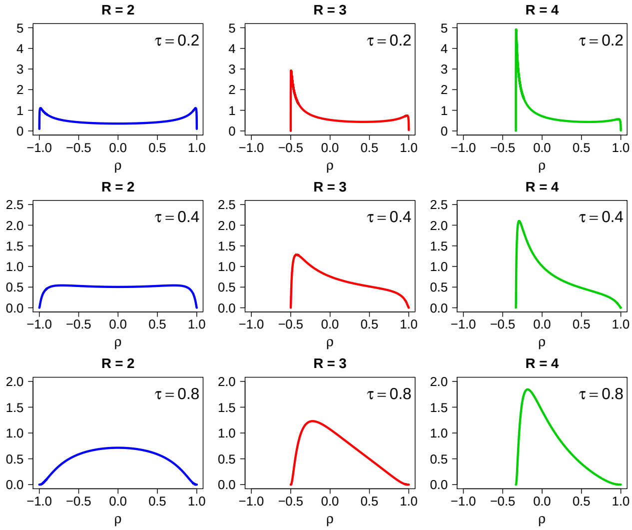
Reparameterize ρ using the general Fisher's z-transformation:

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

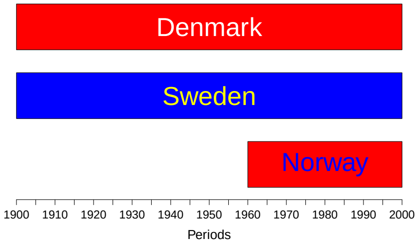
and assign a $\mathcal{N}(0, \tau^{-1})$ prior to $\rho^\star$.

- This prior automatically ensures that $\rho \in (-1/(R-1), 1)$, which is required to ensure positive definiteness of **C**, is fulfilled.
- In addition, $P(\rho > 0) = 0.5$, independent of R.
- Note: A flat prior on the variance-stabilizing transformation gives Jeffrey's (non-informative) prior (Lehmann (1999), pages 491-492).

Illustration of prior on correlation parameters



Mortality among females in Denmark, Sweden and Norway

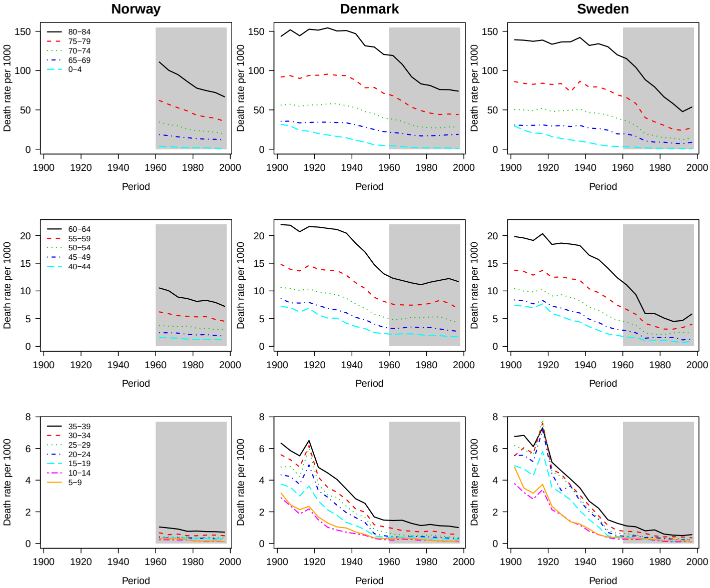


- $R = 3$ regions: Denmark, Sweden and Norway.
- $I = 17$ age groups: 0–4, 5–9, . . . , 75–79, 80–84.
- $J = 20$ periods from 1900–1999 for Denmark and Sweden (only $J = 8$ periods for Norway)

$\rightarrow$ $K = 36$ birth cohorts for Denmark and Sweden (only $K = 24$ birth cohorts for Norway)

(Jacobsen et al. (2004), Eur J Epidemiol)

Mortality among females in Denmark, Sweden and Norway



Mortality extrapolation for Norway from 1900-1959

We consider the model with separate age and cohort, but joint period effects (aPc model), which was classified as the best model in an analysis with data of Denmark and Norway from 1960 -1999.

Results from INLA:

- Posterior median correlation estimates:
  Age effects: 0.97, 80% CI (0.95, 0.99)
  Cohort effects: 0.74, 80% CI (0.50, 0.88)
  Overdispersion: 0.67, 80% CI (0.62, 0.71)
- Running time & model choice

Model	Time	$\bar{D}$	p_D	DIC	$\log(p(\mathbf{y}))$
ordinary	0:20	817.66	789.19	1606.85	-6357.15
correlated	14:11	830.68	771.32	**1602.01**	**−6228.45**

We assess the quality of the predictions with data available from the Human Mortality Database (HMD) (`wwww.mortality.org`).

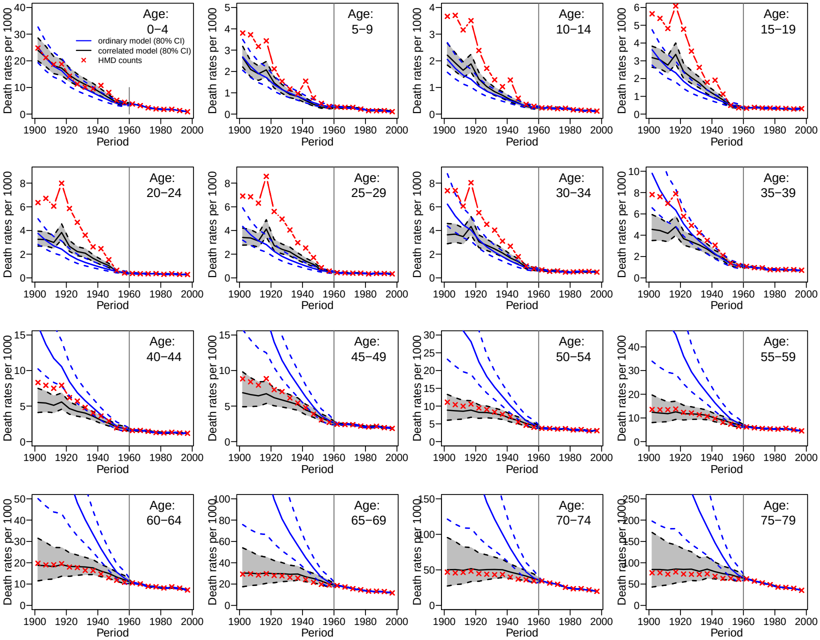

Age: 0–4
ordinary model (80% CI)
correlated model (80% CI)
HMD counts
Age: 5–9
Age: 10–14
Age: 15–19
Age: 20–24
Age: 25–29
Age: 30–34
Age: 35–39
Age: 40–44
Age: 45–49
Age: 50–54
Age: 55–59
Age: 60–64
Age: 65–69
Age: 70–74
Age: 75–79
Death rates per 1000
Period

Table of contents

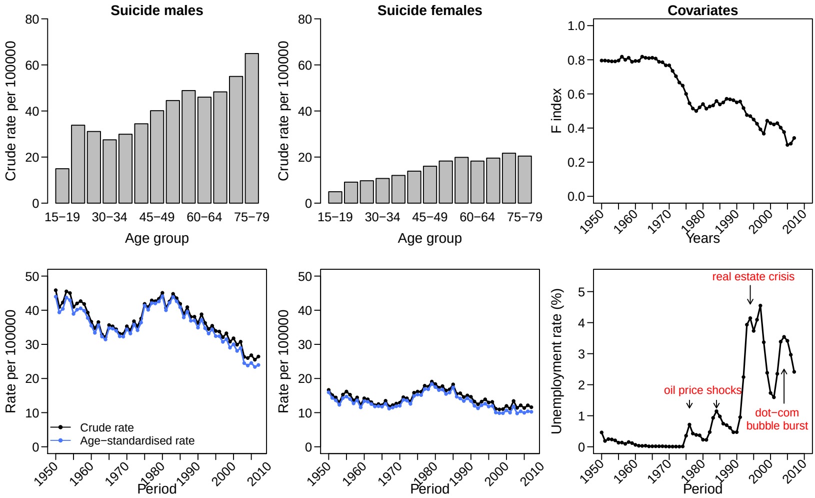
5. Suicide mortality in Switzerland

- Suicide has become one of the leading courses of death of Swiss males aged between 15 and 44 years.
- Goal:
  - Investigation of gender-specific differences over the last 58 years.
  - Investigation of the impact of family integration.
- Data:
  - $I = 13$ age groups: 15–19, ..., 75–79.
  - $J = 58$ periods: 1950–2007.
  - $\Rightarrow K = 118$ brith cohorts.
  - Covariate information:

    F-index = (marriage rates - divorce rates)/(marriage + divorce).

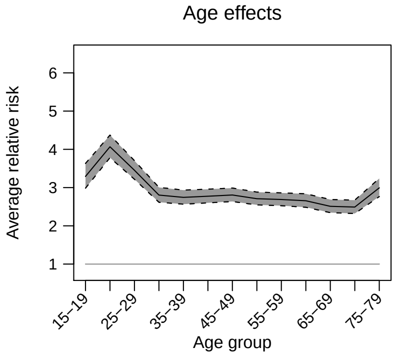
    The rate of unemployment to account for confounding.

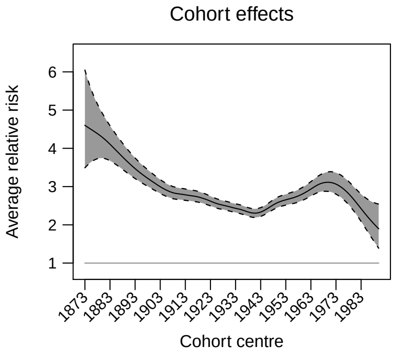
Mortality rates and covariates



Gender-specific differences: Males to females

The fully correlated "joint period effects model" was regarded as the best.





Analysis of the impact of family integration

In contrast to standard time-series models, we want to keep the age-specific data structure when including covariate information.

- Model formulation:
  - Keep correlated age and cohort effects, and overdispersion.
  - Replace the joint period effects by correlated covariate effects related to family integration.
  - To adjust for confounding we include a time-constant effect of the unemployment rate.

    $$\eta_{ijg} = \mu_g + \theta_{ig} + \psi_{kg} + f_g(\text{F\_index}_j) + \beta \times \text{Unempl}_j + z_{ijg},$$

    where $f_g(\text{F\_index}_j)$ is the unknown effect of the F-index on sex g, which is modelled as a correlated RW2.
- Model estimation using INLA.

Model choice

		Period effects replaced by F-index	
	A,C,O	no correlation	A,F-index,C,O
Model choice			
Log Score	**3.386**	3.402	3.393
DIC	**1854.60**	1890.05	1870.26
Correlation coefficients			
Correlation age	${}_{0.50}0.82_{0.95}$	-	${}_{0.50}0.82_{0.95}$
Correlation F-index	-	-	${}_{0.50}0.94_{1.00}$
Correlation cohort	${}_{-0.61}{+0.20}_{+0.79}$	-	${}_{-0.61}{+0.23}_{+0.81}$
Correlation overdispersion	${}_{0.18}0.56_{0.84}$	-	${}_{0.26}0.63_{0.85}$

(Note: "A" stands for age, "C" for cohort and "O" for overdispersion)

Posterior estimates

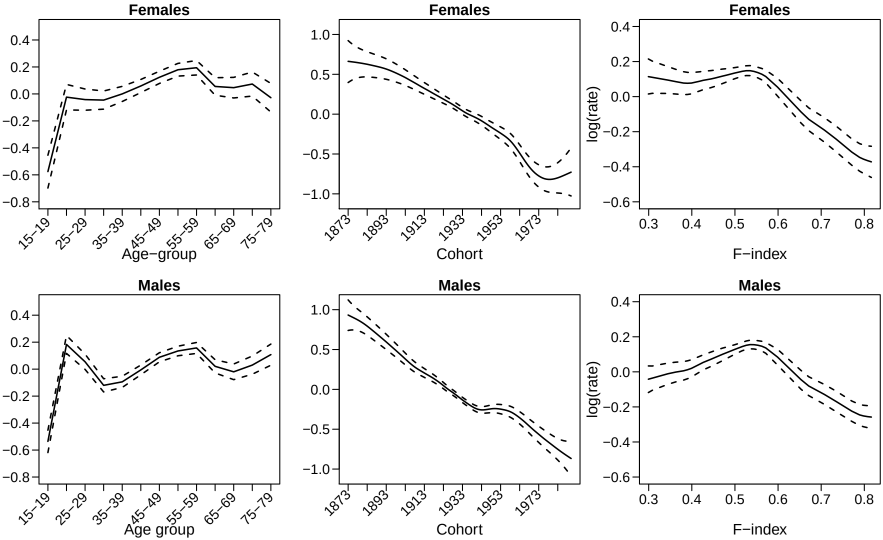


Table of contents

Summary

This thesis aimed at improving the methodology for statistical inference in multivariate APC models.

Multivariate APC models

- A multivariate analysis of APC models makes it possible to identify relative time trends.
- The Bayesian approach provides useful result even in the case of unequal time intervals.

A conditional approach for inference

- A conditional formulation simplifies the analysis as nuisance parameters are removed.
- In addition data with multiple stratification factors can be analysed.
- For model estimation standard software can be used.

Summary and future work

Correlated multivariate APC models

- Useful to improve precision of relative risk estimates and extrapolation of missing data.
- The possibility to correlate a wide range of other GMRF models is available in INLA, e.g. CAR, seasonal models, ...

Application: Suicide mortality in Switzerland

- Strong gender-specific differences.
- Social integration cannot fully explain period effects.

Future work will focus on correlated GMRF priors:

- Allow for other correlation structures.
- Application to space-temporal modelling.

Further publications during my PhD

Correlated GMRF priors for multivariate age-period-cohort models

Riebler, A., Held, L. & Rue, H.

In: Proceedings of the 25th International Workshop on Statistical Modelling, A. Bowman, Editor. 455–460, 2010.

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

Held, L. & Riebler, A.

Journal of the Royal Statistical Society, Series B (2009), 71, 357–358.

Bayesian bivariate meta-analysis of diagnostic test studies using integrated nested Laplace approximations

Paul, M., Riebler, A., Bachmann, L., Rue, H. & Held, L.

Statistics in Medicine (2010), 29, 12, pp. 1325–1339.

Bayesian variable selection for detecting adaptive genomic differences among populations

Riebler A., Held L. & Stephan W.

Genetics (2008), 178, pp. 1817–1829.

A whole genome Bayesian scan for adaptive genetic divergence in West African cattle

Gautier, M., Flori, L., Riebler, A., Jaffrézic, F., Laloé, D., Gut, I., Moazami-Goudarzi, K. & Foulley, J. L.

BMC Genomics (2009), 10, pp. 550+.

Thank you for your attention!

Acknowledgements:

Rupali Akerkar - Lucas Bachmann - Andrew Barbour - Matthias Bopp - Julia Braun - Peter Bühlmann - Sylvia Frühwirth-Schnatter - Reinhard Furrer - Hans-Peter Jauss - Sarah Haile - Andrew Hector - Leonhard Held - Ulrike Held - Sinikka Kohler - Anne Kajander - Christian Lang - Sara Martino - Alessandro Ottavi - Michaela Paul - Margitta Riebler - Malgorzata Roos - Håvard Rue - Kaspar Rufibach - Daniel Sabanés Bové - Burkhardt Seifert - Sigrunn Sørbye - Birgit Schrödle - Alois Tschopp - Christian von Mering - Andreas Wagner.

Financial support by the Swiss National Science Foundation (SNF) and partly the Research Council of Norway is gratefully acknowledged.