University of Zurich[UZH]

Estimation and extrapolation of time trends in multivariate registry data using Bayesian age-period-cohort models

Andrea Riebler

Joint work with Leonhard Held and Håvard Rue

Research Seminar, Vienna, November 23, 2012

Registry data

- National health care registers typically collect incidence and/or mortality counts stratified by age and period.

TABLE 1

Observations: data from only the first seven time periods were used in fitting the model

Age group		Period									
		1935	1940	1945	1950	1955	1960	1965	1970	1975	1980
50–54	Cohort	7	8	9	10	11	12	13	14	15	16
	No. Deaths	177	271	312	382	321	305	308	304	274	278
	No. at Risk	301000	317000	353000	395000	426000	473000	498000	552000	598000	629000
55–59	Cohort	6	7	8	9	10	11	12	13	14	15
	No. Deaths	262	350	552	620	714	649	738	718	780	789
	No. at Risk	212000	248000	279000	301000	358000	411000	443000	435000	510000	583000
60–64	Cohort	5	6	7	8	9	10	11	12	13	14
	No. Deaths	360	479	644	949	932	1292	1327	1507	1602	1712
	No. at Risk	159000	194000	222000	222000	258000	304000	341000	404000	403000	482000
65–69	Cohort	4	5	6	7	8	9	10	11	12	13
	No. Deaths	409	544	812	1150	1668	1958	2153	2375	2742	2973
	No. at Risk	132000	144000	169000	210000	230000	264000	297000	322000	396000	401000
70–74	Cohort	3	4	5	6	7	8	9	10	11	12
	No. Deaths	328	509	763	1097	1593	2039	2433	3066	3432	3939
	No. at Risk	76000	94000	110000	125000	149000	180000	197000	213000	233000	293000
75–79	Cohort	2	3	4	5	6	7	8	9	10	11
	No. Deaths	222	359	584	845	1192	1638	2068	2671	3356	3928
	No. at Risk	37000	47000	59000	71000	91000	108000	118000	132000	141000	193000
80–84	Cohort	1	2	3	4	5	6	7	8	9	10
	No. Deaths	108	178	285	475	742	992	1374	1833	2353	3184
	No. at Risk	19000	22000	32000	39000	44000	56000	66000	77000	93000	94000

Prostate cancer data from Besag et al., 1995, Statistical Science

Age-period-cohort analysis

- Age-period-cohort (APC) model is commonly used to describe vital rates using three time scales:
  - **A**ge: age at diagnosis/death
  - **P**eriod: date of diagnosis/death
  - **C**ohort: date of birth
- Goals:
  - Detecting temporal patterns in such data as they could provide important clues for disease etiology.
  - Extrapolation and prediction.

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

$$\begin{aligned} y_{ij}|\eta_{ij} &\sim \text{Poisson}(n_{ij}\exp(\eta_{ij})) \\ \eta_{ij} &= \mu + \alpha_i + \beta_j + \gamma_k \end{aligned}$$

with age effect α_i, period effect β_j and cohort effect γ_k.

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

$$\begin{aligned} y_{ij}|\eta_{ij} &\sim \text{Poisson}(n_{ij}\exp(\eta_{ij})) \\ \eta_{ij} &= \mu + \alpha_i + \beta_j + \gamma_k \end{aligned}$$

with age effect α_i, period effect β_j and cohort effect γ_k.

- The cohort index $k = M \times (I - i) + j$ is a linear function of i and j, where M is the ratio of age group to period interval.
- To assure identifiability of the intercept μ we apply sum-to-zero constraints for each parameter vector $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$.

Identifiability problems

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

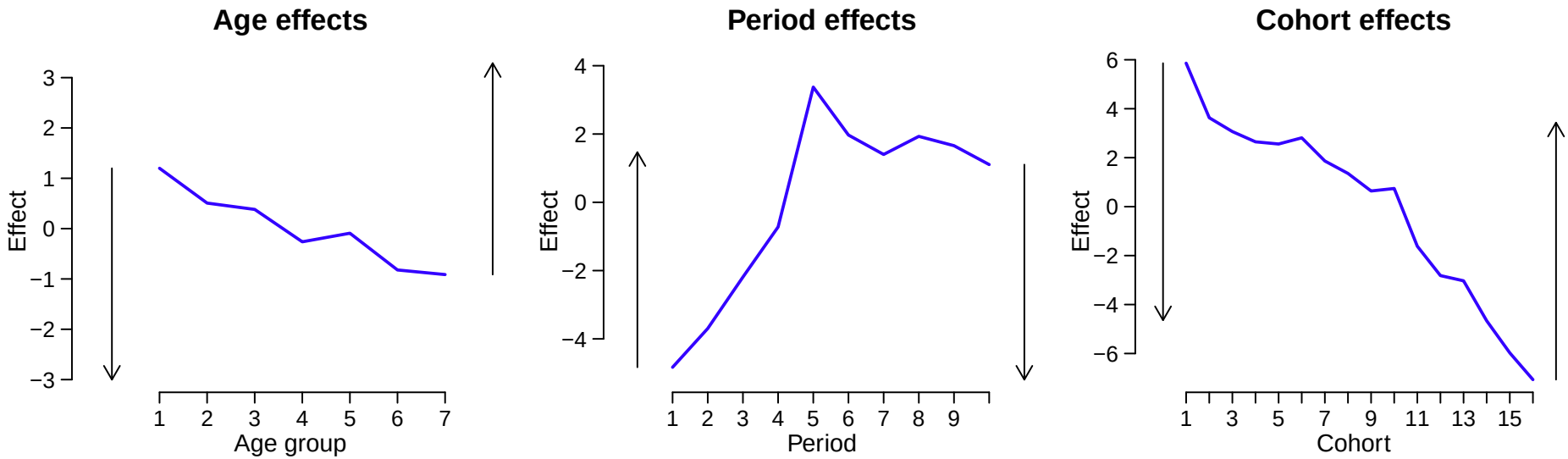
For any value of $a \in \mathbb{R}$, the linear transformations

$$\begin{aligned}
\alpha_i &\rightarrow \alpha_i + M \times a\left(i - \frac{(I+1)}{2}\right), \\
\beta_j &\rightarrow \beta_j - a\left(j - \frac{(J+1)}{2}\right), \\
\gamma_k &\rightarrow \gamma_k + a\left(k - \frac{(K+1)}{2}\right)
\end{aligned}$$

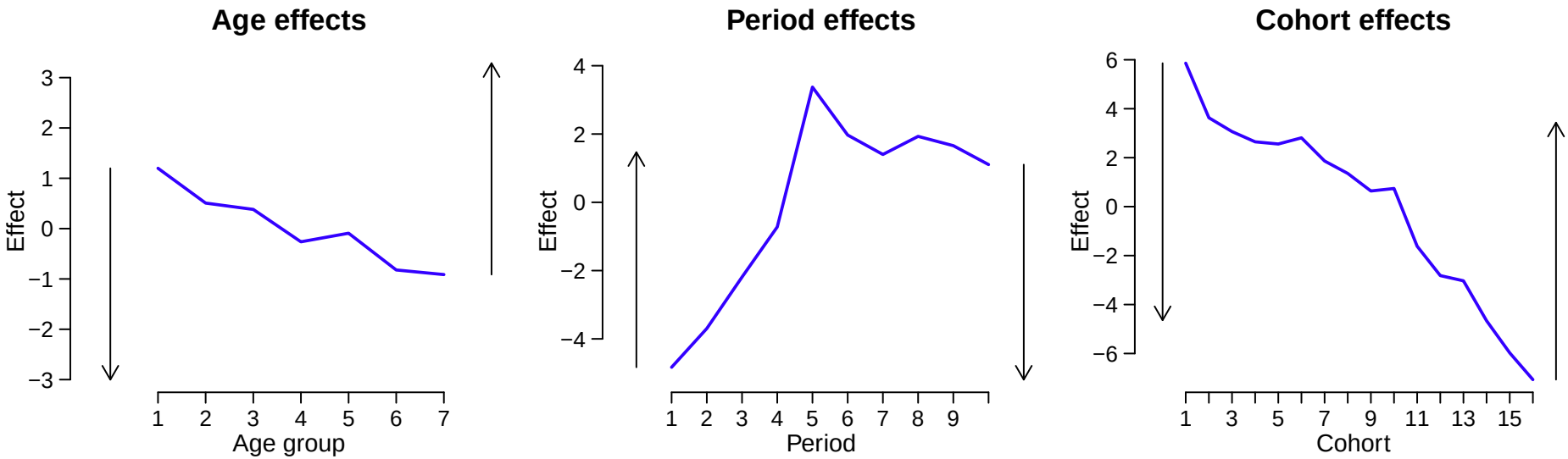
will leave η_{ij} unchanged and maintain the sum-to-zero constraints.

Holford, 1983, Biometrics

Identifiability problems: Illustration (M=1)



Identifiability problems: Illustration (M=1)



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

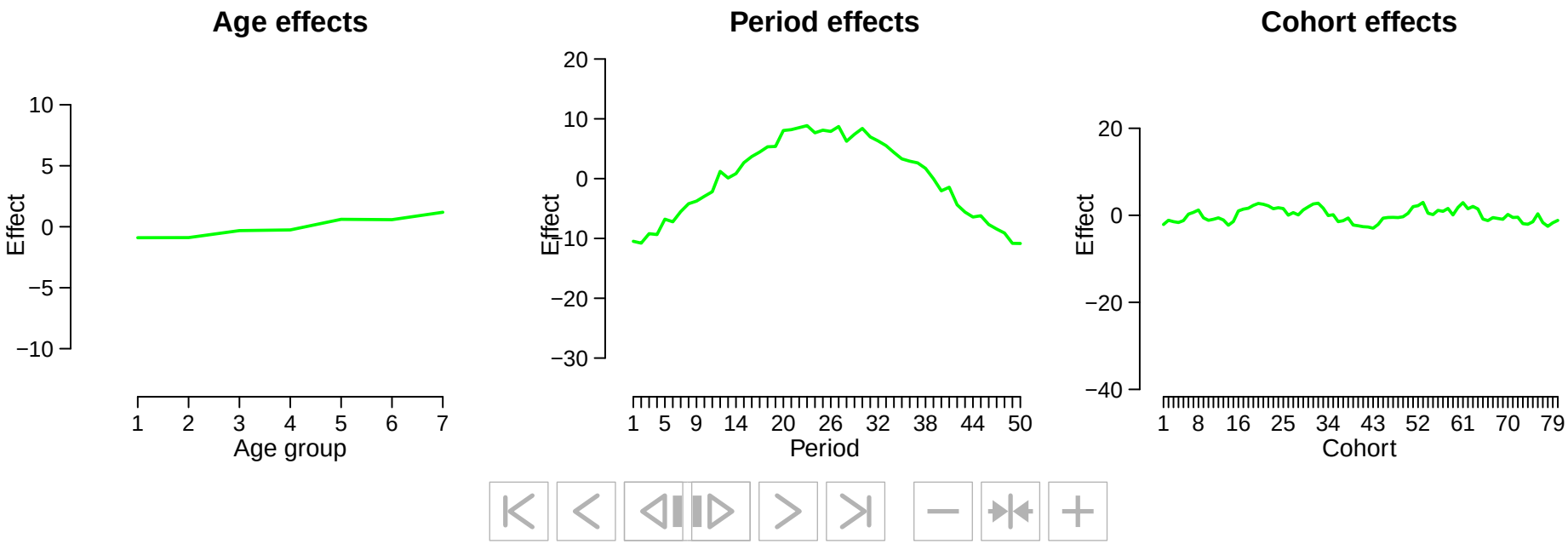
Identifiability problems cont.

- For unequally spaced data, a second identifiability problem induces artificial cyclical patterns (saw-tooth-pattern) in the period and cohort estimates. Holford, 2006, Stat Med
- Remember, M is the ratio of age group to period interval length.
- For any value of $b_1, \ldots, b_M \in \mathbb{R}$ (subject to $b_1 + \ldots + b_M = 0$), the transformations

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

will leave the linear predictor η_{ij} unchanged.

Identifiability problems: Illustration (M=5)



Likelihood Inference

Mainly classical maximum likelihood (ML) estimation has been used for APC models.

Disadvantages:

- Additional constraints are necessary for identifiability.
- The model overfits cohorts for which only a single observation exists.

  Besag et al., 1995, Stat Science
- For unequally spaced data, ML estimates become very unstable resulting in artificial saw-tooth pattern.

  Holford, 2006, Stat Med

$\Rightarrow$ We go for a Bayesian approach.

The Bayesian APC model

- A natural choice is to penalise second differences.

Berzuini and Clayton, 1994, Stat Med

- The second order random walk (RW2) prior for $\boldsymbol{\alpha}$ is

$$\pi(\boldsymbol{\alpha}|\tau_{\boldsymbol{\alpha}}) \propto \tau_{\boldsymbol{\alpha}}^{\frac{I-2}{2}} \exp\left(-\frac{\tau_{\boldsymbol{\alpha}}}{2}\sum_{i=3}^{I}(\alpha_i - 2\alpha_{i-1} + \alpha_{i-2})^2\right) = \tau_{\boldsymbol{\alpha}}^{\frac{I-2}{2}} \exp\left(-\frac{1}{2}\boldsymbol{\alpha}^\top \mathbf{R}\boldsymbol{\alpha}\right),$$

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

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

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

The Bayesian APC model (II)

- Use independent RW2 priors for α, β, γ.
- To account for overdispersion add $z_{ij} \sim \mathcal{N}(0, \tau_z^{-1})$ to η_{ij}:

$$\eta_{ij} = \mu + \alpha_i + \beta_j + \gamma_k + z_{ij}.$$

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

Inference in the Bayesian APC model

- Besag et al. (1995) propose a sophisticated MCMC algorithm using suitable reparameterisation and block sampling.
- However, they also note:

  "We anticipate that analytical approximations should work well on our model and on others similar to it, especially for the present data where there appears not to be any significant multimodality in the posterior distribution."

- Today, the INLA methodology can be used for routine application using an `R` interface (`www.r-inla.org`).

Integrated nested Laplace approximations (INLA)

(Rue et al, 2009, JRSS-B)

INLA is a fast alternative to inference via MCMC in latent Gaussian models. The methodology is particularly attractive if the latent Gaussian model is a GMRF.

The INLA approach

- incorporates posterior uncertainty with respect to hyperparameters,
- can be used for out-of-sample prediction,
- can be used for model assessment and comparison based on leave-one-out cross-validation.

INLA can be called in a modular way, just as `glm()` or `lme()`, say, using an R interface (`www.r-inla.org`).

The general setting

Three-stage Bayesian hierarchical model:

- Observation model: $\pi(\boldsymbol{y}|\boldsymbol{x}) = \prod_u \pi(y_u|x_u, \boldsymbol{\theta})$.
- Parameter model: $\pi(\boldsymbol{x}|\boldsymbol{\theta})$, usually a GMRF.
- Hyperprior: $\pi(\boldsymbol{\theta})$.

The general setting

Three-stage Bayesian hierarchical model:

- Observation model: $\pi(\boldsymbol{y}|\boldsymbol{x}) = \prod_u \pi(y_u|x_u, \boldsymbol{\theta})$.
- Parameter model: $\pi(\boldsymbol{x}|\boldsymbol{\theta})$, usually a GMRF.
- Hyperprior: $\pi(\boldsymbol{\theta})$.

The posterior distribution is

$$\pi(\boldsymbol{x}, \boldsymbol{\theta}|\boldsymbol{y}) \propto \pi(\boldsymbol{\theta})\pi(\boldsymbol{x}|\boldsymbol{\theta}) \prod_u \pi(y_u|x_u, \boldsymbol{\theta}).$$

Dim($\boldsymbol{x}$) is large, while dim($\boldsymbol{\theta}$) is small.

INLA: non-Gaussian observations

Main goal: Compute the posterior marginals

$$\pi(x_u|\boldsymbol{y}) = \int_{\boldsymbol{\theta}} \underbrace{\int_{\boldsymbol{x}_{-u}} \pi(\boldsymbol{x},\boldsymbol{\theta}|\boldsymbol{y}) d\boldsymbol{x}_{-u}}_{\pi(x_u,\boldsymbol{\theta}|\boldsymbol{y})=\pi(x_u|\boldsymbol{\theta},\boldsymbol{y})\pi(\boldsymbol{\theta}|\boldsymbol{y})} d\boldsymbol{\theta},$$

$$\pi(\theta_v|\boldsymbol{y}) = \int_{\boldsymbol{\theta}_{-v}} \underbrace{\int_{\boldsymbol{x}} \pi(\boldsymbol{x},\boldsymbol{\theta}|\boldsymbol{y}) d\boldsymbol{x}}_{\pi(\boldsymbol{\theta}|\boldsymbol{y})} d\boldsymbol{\theta}_{-v}.$$

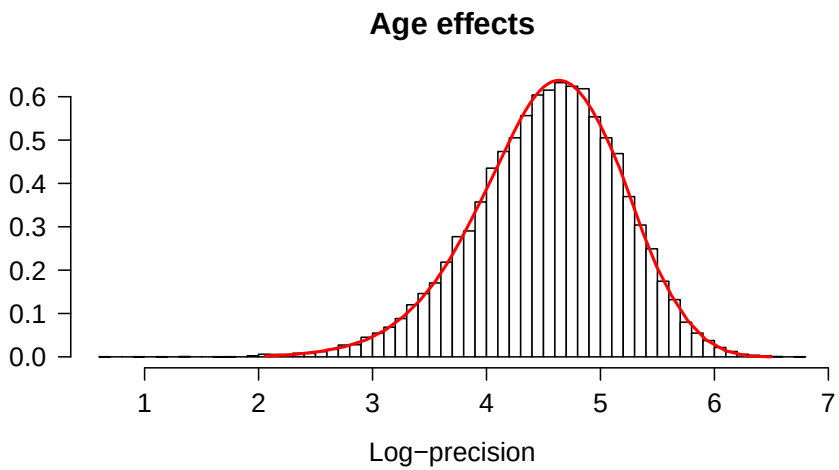
INLA uses nested Laplace approximations for this purpose.

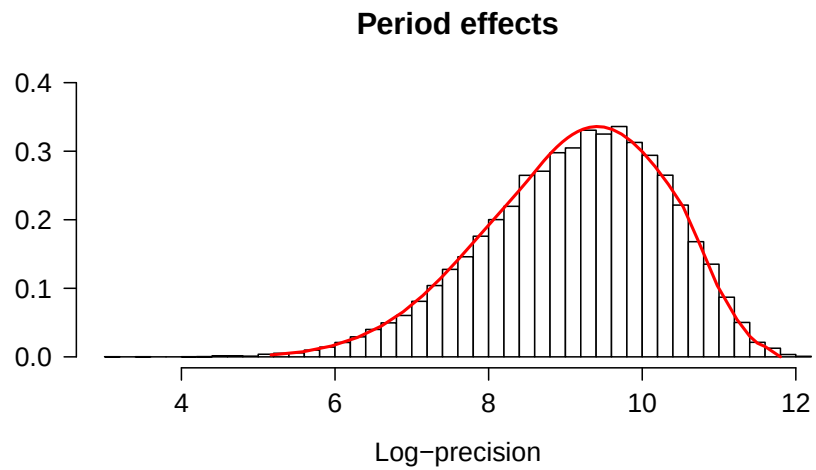
In our model: $\boldsymbol{x} = (\mu, \boldsymbol{\alpha}^\top, \boldsymbol{\beta}^\top, \boldsymbol{\gamma}^\top, \boldsymbol{z}^\top)^\top$, $\boldsymbol{\theta} = (\tau_{\boldsymbol{\alpha}}, \tau_{\boldsymbol{\beta}}, \tau_{\boldsymbol{\gamma}}, \tau_{\boldsymbol{z}})^\top$.

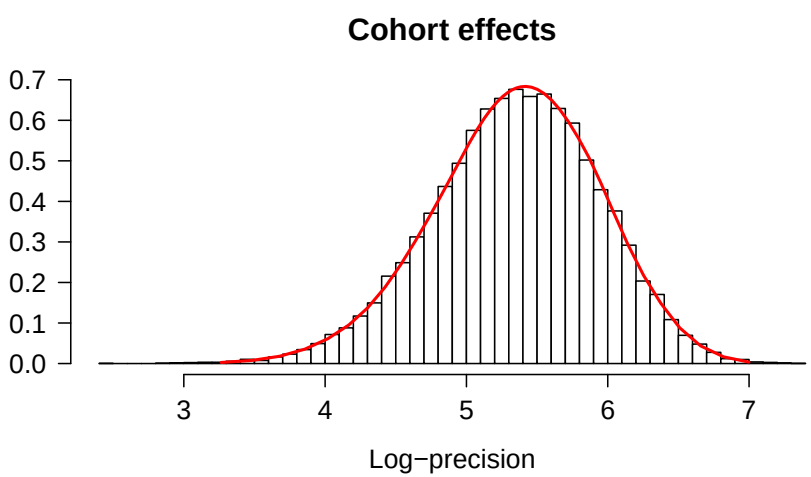
Using INLA

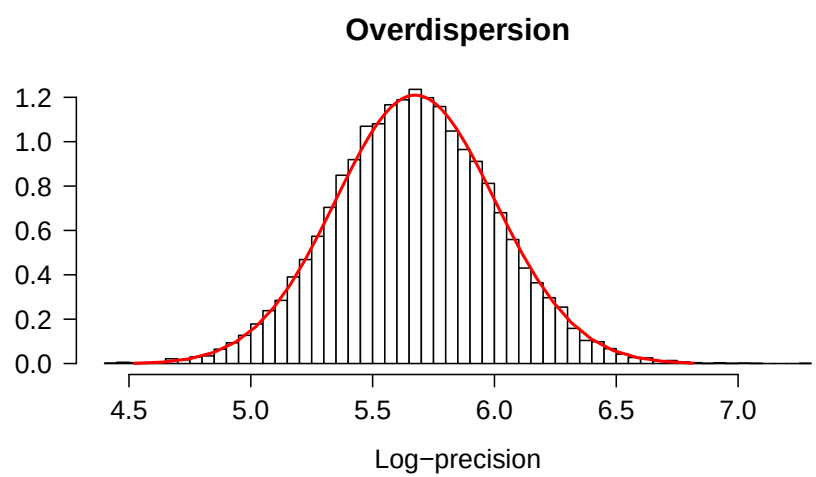
```
> head(ProstateCancer, 4)

  deaths    pop age.group period cohort index
1    177 301000         1      1      7     1
2    262 212000         2      1      6     2
3    360 159000         3      1      5     3
4    409 132000         4      1      4     4

> library(INLA)
> my.hyper <- list(prec=list(param=c(1, 0.005)))
> model <- deaths ~ f(age.group, model="rw2", hyper=my.hyper) +
                    f(period, model="rw2", hyper=my.hyper) +
                    f(cohort, model="rw2", hyper=my.hyper) +
                    f(index, model="iid", hyper=my.hyper)
> results <- inla(model, family="poisson", data=ProstateCancer,
                  E=pop, quantiles=c(0.1, 0.5, 0.9))
```

Comparing INLA with MCMC









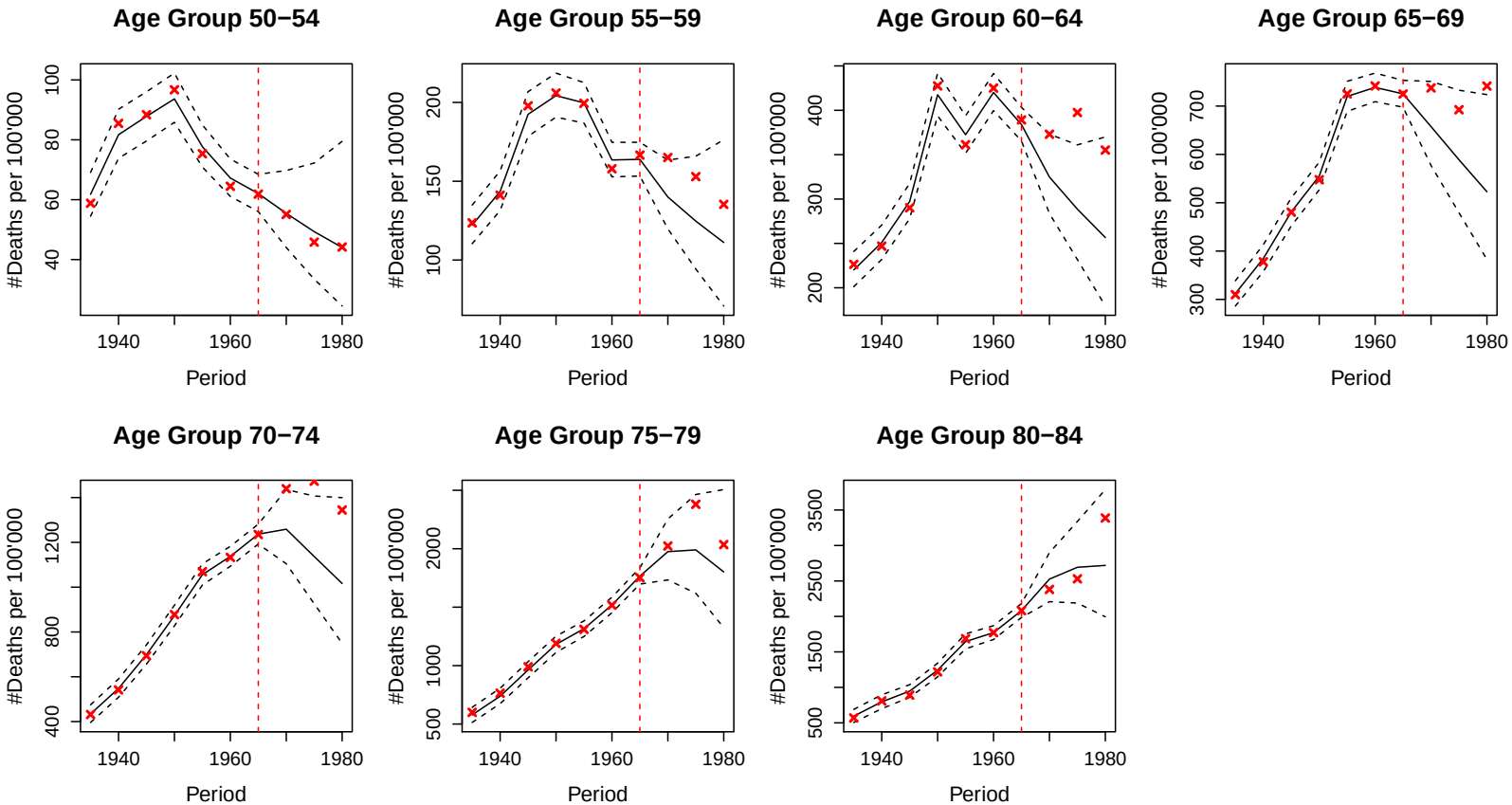
Predictions in INLA

- Prediction of future deaths rates was one of the major goals in Besag et al. (1995).
- This can be also done in INLA by setting the observations to be predicted to NA.
- Post-processing of the posterior predictive distribution of the linear predictor η_{ij} gives the predictive distribution of y_{ij}.
- Even simultaneous credible bands can be computed.

Sørbye and Rue, 2010

Prediction: Prostate cancer

- Assume, we would like to predict the last three five-year periods 1970–1974, 1975–1979, 1980–1984.



Observed and predicted number of cases within 80% point-wise credible bands.

Multivariate APC models

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

$$\begin{aligned} y_{ijr}|\eta_{ijr} &\sim \text{Poisson}(n_{ijr} \exp(\eta_{ijr})) \\ \eta_{ijr} &= \mu_r + \alpha_{i,r} + \beta_{j,r} + \gamma_{k,r} + z_{ijr} \end{aligned}$$

with independent $z_{ijr} \sim \mathcal{N}(0, \kappa_{\boldsymbol{z}}^{-1})$, say.

Multivariate APC models

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

$$\begin{aligned} y_{ijr}|\eta_{ijr} &\sim \text{Poisson}(n_{ijr} \exp(\eta_{ijr})) \\ \eta_{ijr} &= \mu_r + \alpha_{i,r} + \beta_{j,r} + \gamma_{k,r} + z_{ijr} \end{aligned}$$

  with independent $z_{ijr} \sim \mathcal{N}(0, \kappa_z^{-1})$, say.
- Simpler models can be obtained, e.g. assuming shared age effects (Apc model):

$$\begin{aligned} y_{ijr}|\eta_{ijr} &\sim \text{Poisson}(n_{ijr} \exp(\eta_{ijr})) \\ \eta_{ijr} &= \mu_r + \alpha_i + \beta_{j,r} + \gamma_{k,r} + z_{ijr} \end{aligned}$$

- As a start: independent RW2 priors for α, β_r, γ_r, $r = 1, \ldots, R$.

Identifiability of relative risks

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

$$\begin{aligned}
\Delta_j^{(r)} &= \beta_{j,r} - \beta_{j,R} && \text{in the ApC model} \\
\Delta_k^{(r)} &= \gamma_{k,r} - \gamma_{k,R} && \text{in the APc model} \\
\Delta_{jk}^{(r)} &= \Delta_j + \Delta_k && \text{in the Apc model}
\end{aligned}$$

  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.

Analysing heterogeneous time trends: Apc model

Data: COPD mortality counts among males in England & Wales

Hansell et al., 2003, Epidemiology

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

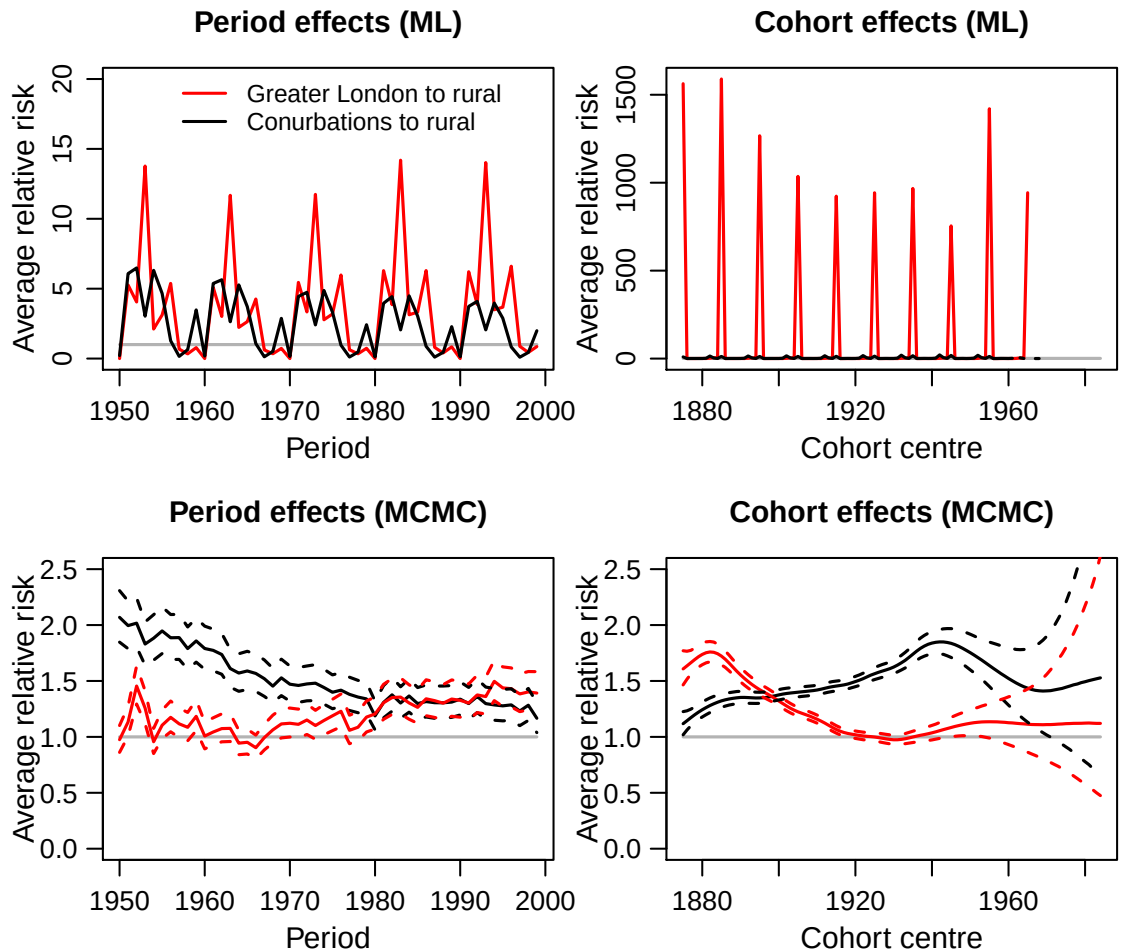
Riebler & Held, 2010, Biostatistics.

Analysing heterogeneous time trends: Apc model

Data: COPD mortality counts among males in England & Wales

Hansell et al., 2003, Epidemiology

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



Riebler & Held, 2010, Biostatistics.

A side comment: A conditional approach

- Let $y_{ij\bullet} = y_{ij1} + \ldots + y_{ijR}$.
- It is easy to see that the Apc model for y_{ijr} implies that $\boldsymbol{y}_{ij}|y_{ij\bullet}$ 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.
- All parameters are identifiable and can be estimated with ML with suitable smoothing, if necessary (R-package `VGAM`).

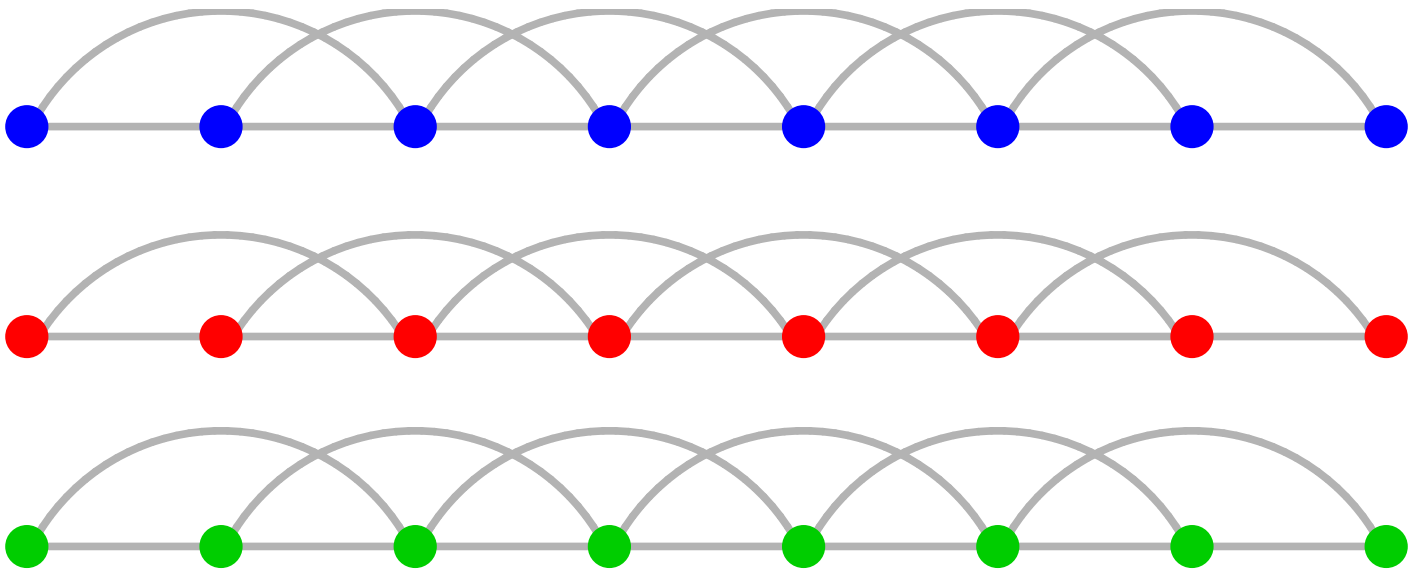
Held & Riebler, 2012, Stat Methods Med Res

Correlate separate random walks

When assuming separate time effects across strata it might nevertheless be plausible to assume some correlation.

$\Rightarrow$ Use of correlated random walks

Illustration for $R = 3$:

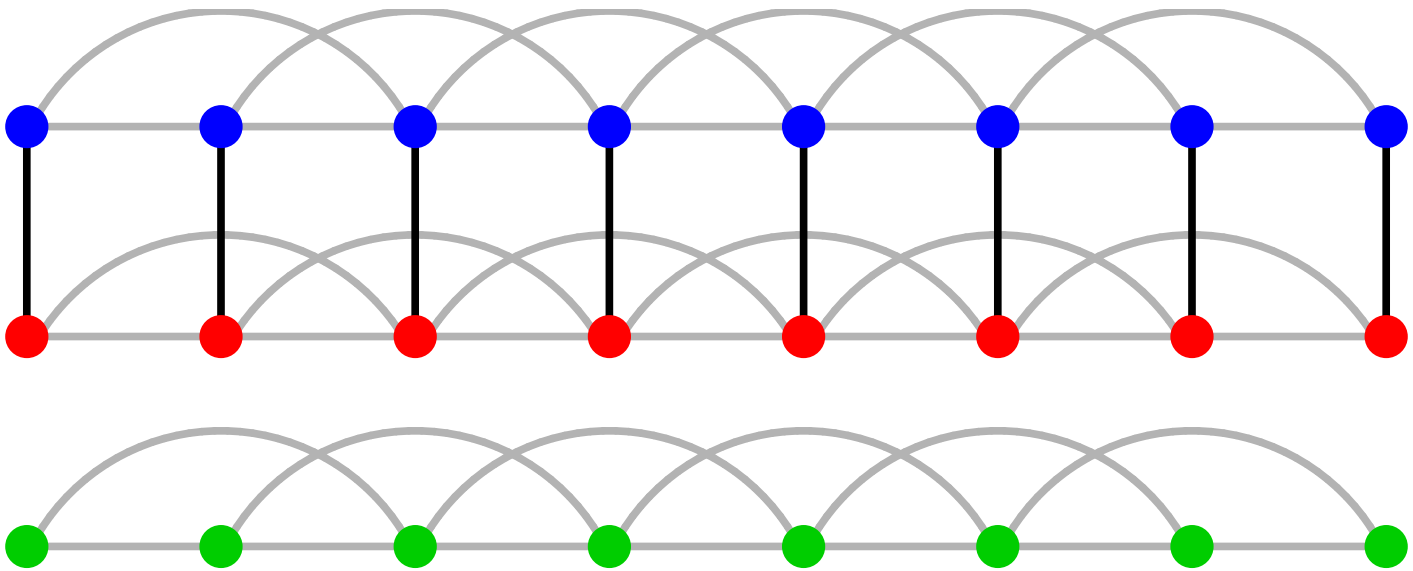


Correlate separate random walks

When assuming separate time effects across strata it might nevertheless be plausible to assume some correlation.

$\Rightarrow$ Use of correlated random walks

Illustration for $R = 3$:

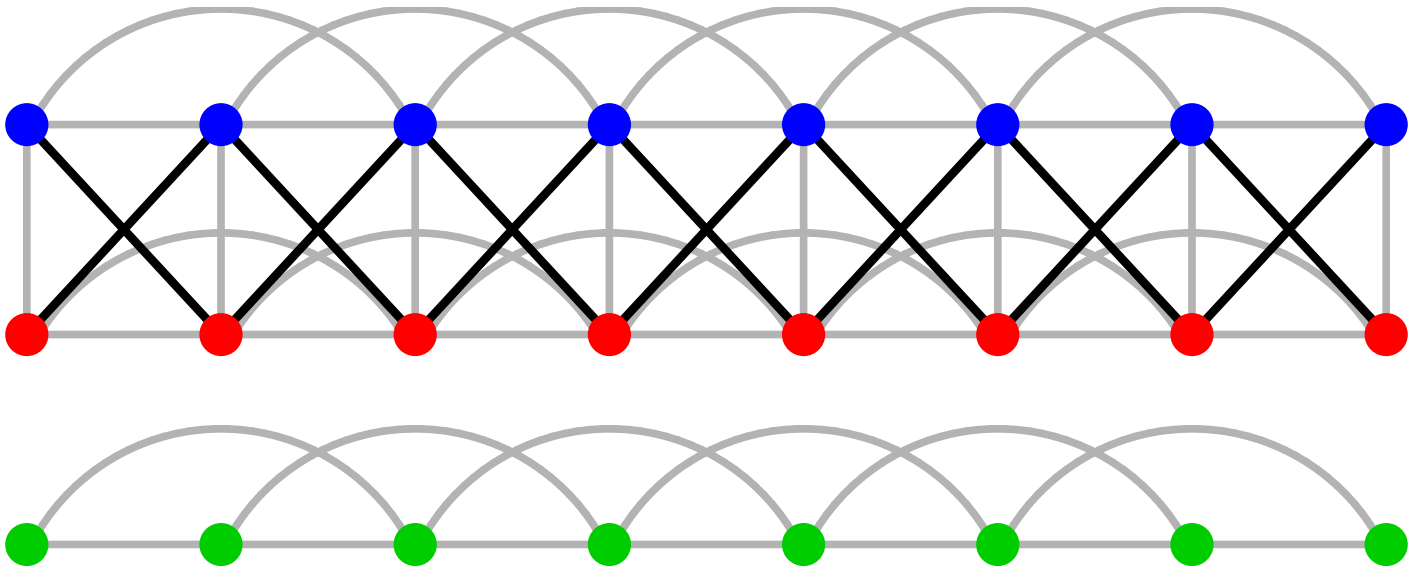


Correlate separate random walks

When assuming separate time effects across strata it might nevertheless be plausible to assume some correlation.

$\Rightarrow$ Use of correlated random walks

Illustration for $R = 3$:

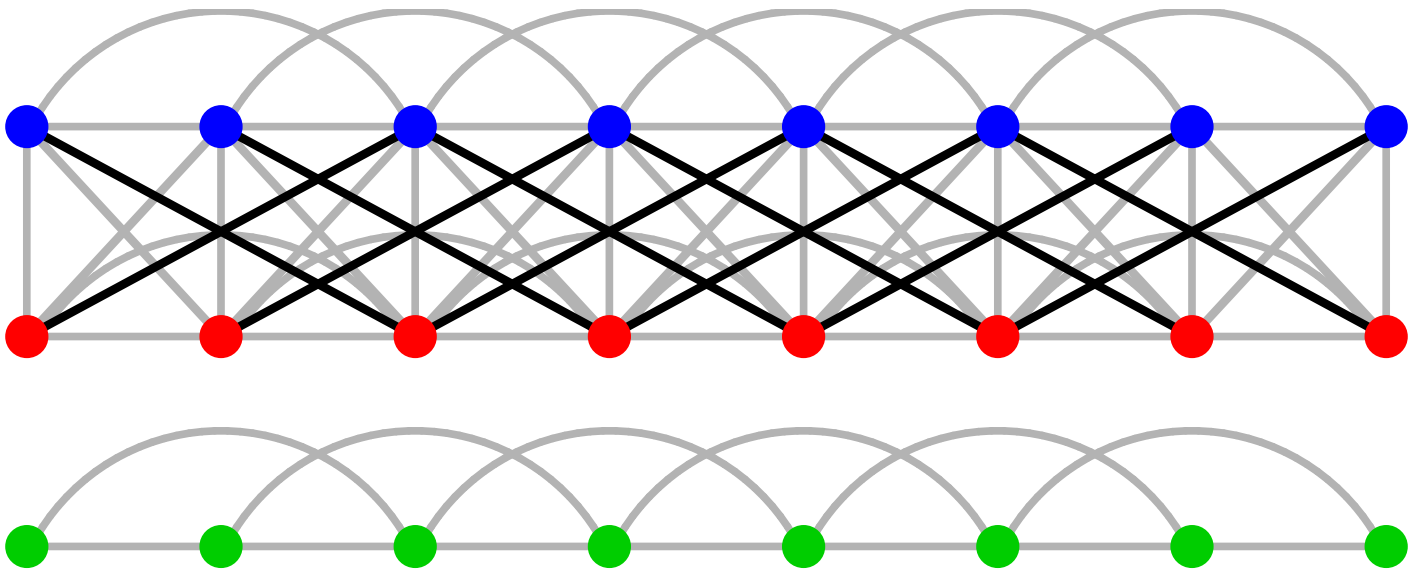


Correlate separate random walks

When assuming separate time effects across strata it might nevertheless be plausible to assume some correlation.

$\Rightarrow$ Use of correlated random walks

Illustration for $R = 3$:

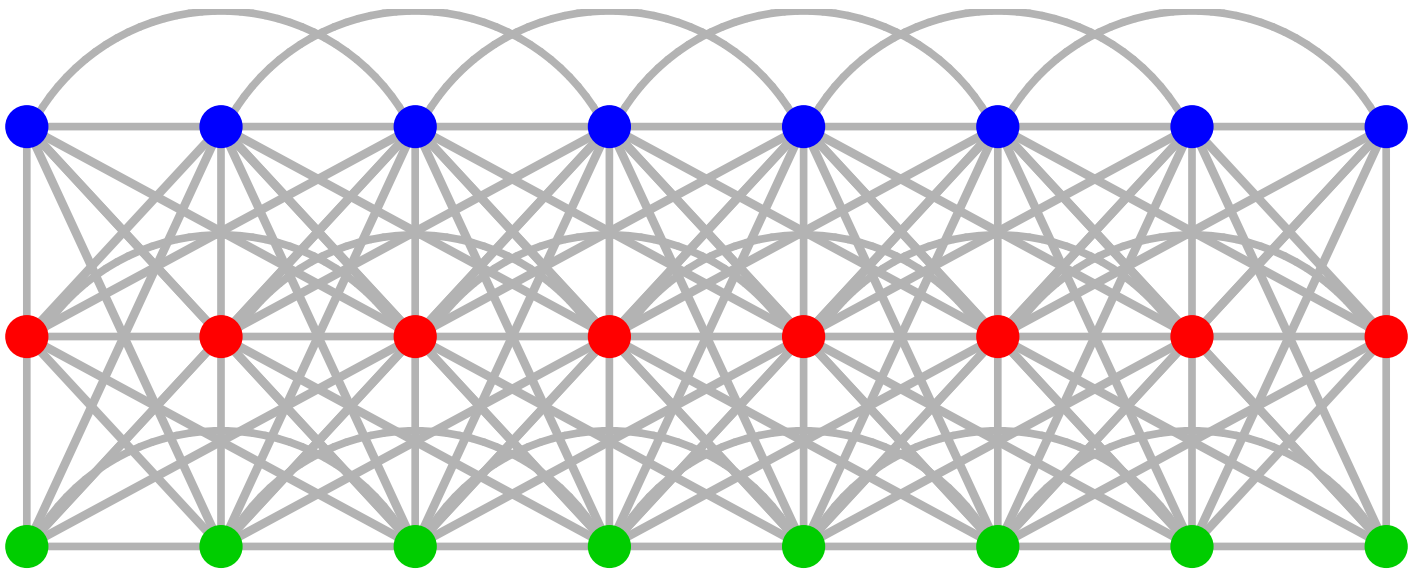


Correlate separate random walks

When assuming separate time effects across strata it might nevertheless be plausible to assume some correlation.

$\Rightarrow$ Use of correlated random walks

Illustration for $R = 3$:



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 β_1, β_2, β_3 can be correlated using the stacked vector $\tilde{\beta} = (\beta_1^\top, \beta_2^\top, \beta_3^\top)^\top$:

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

- Multivariate RW2 with correlated increments.
- Correlated overdispersion can also be incorporated:

$$\boldsymbol{z}_{ij} = (z_{ij1}, z_{ij2}, z_{ij3})^\top \sim \mathcal{N}(\mathbf{0}, \tau_{\boldsymbol{z}}^{-1}\mathbf{C}_{\boldsymbol{z}}).$$

- All correlations are treated as unknown.

Prior on correlation parameters

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

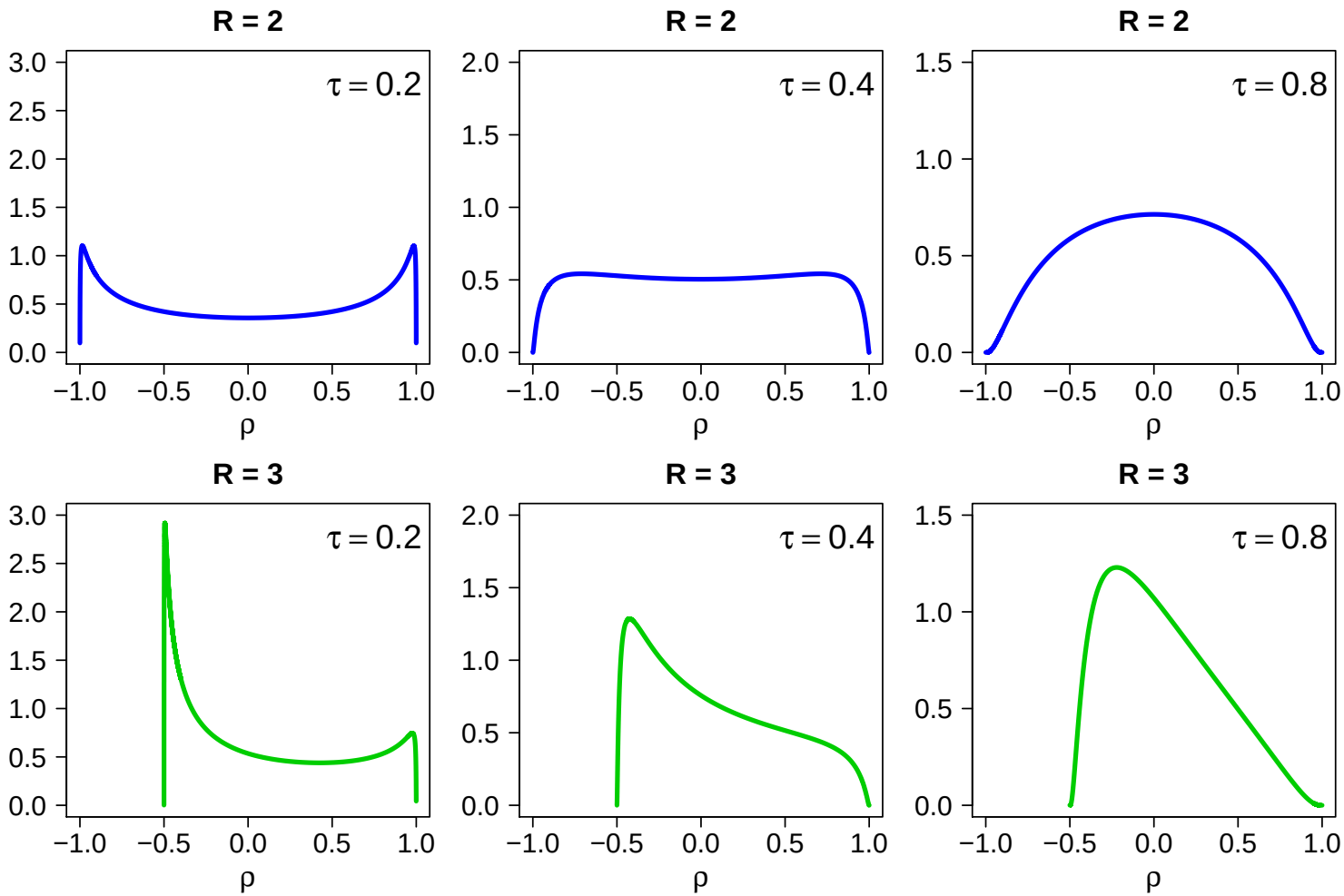
$$\rho = \frac{\exp(\rho^\star) - 1}{\exp(\rho^\star) + R - 1} \qquad \rho^\star = \log\left(\frac{1 + \rho \cdot (R-1)}{1-\rho}\right),$$

Fisher, 1958, page 219

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 $\mathbf{C}$, is fulfilled.
- In addition, $P(\rho > 0) = 0.5$, independent of R.

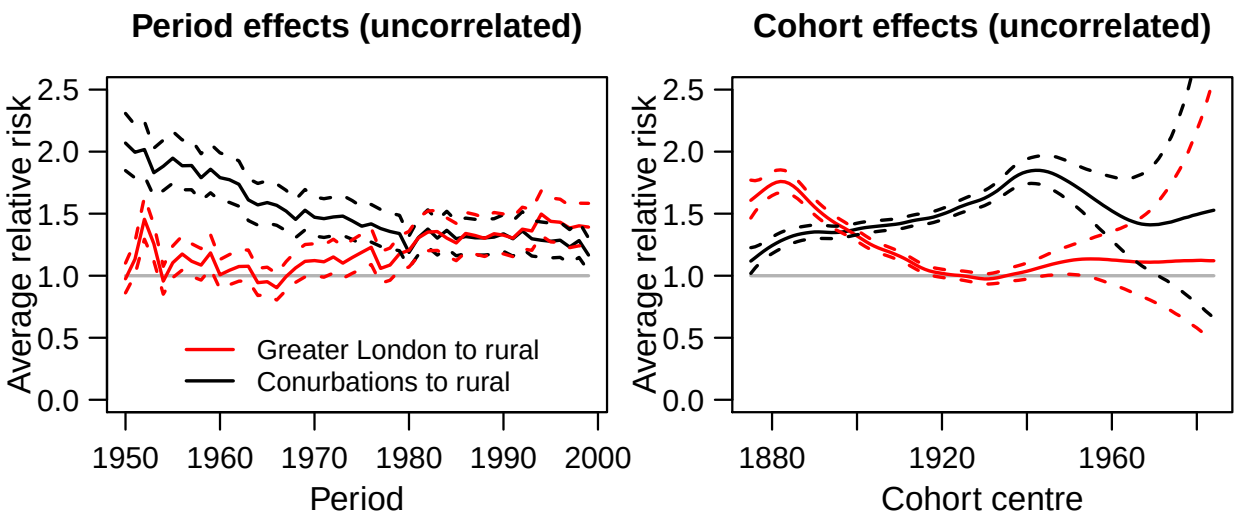
Illustration of prior on correlation parameters



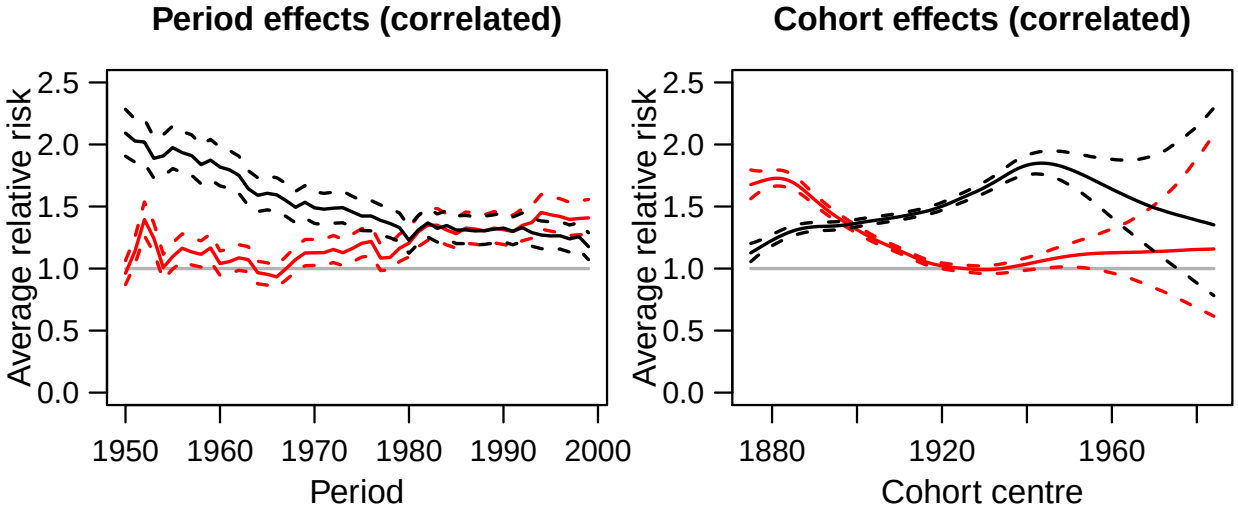
INLA call in R-INLA package

```
> library(INLA)
> ## define the grouping index
> g <- rep(c(1,2,3), each=I*J)
> ## Apc model with correlated time effects & overdispersion.
> my.hyper.rho <- list(rho=list(param=c(0, 0.2)))
> model_Apc <- y ~ -1 + mu1 + mu2 + mu3 +
   f(age.group, model="rw2", hyper=my.hyper) +
   f(period, model="rw2", hyper=my.hyper, constr=TRUE, rankdef=2,
     control.group=list(hyper=my.hyper.rho, model="exchangeable"),
     group=g) +
   f(cohort, model="rw2", ...) +
   f(index, model="iid", hyper=my.hyper,
     control.group=list(hyper=my.hyper.rho, model="exchangeable"),
     group=g)

> results <- inla(model_Apc, family="poisson", E=n, data=data,
   control.compute=list(hyperpar =TRUE))
```

Relative risks accounting for correlation



Wide credible intervals for younger birth cohorts.



Adjusting for correlation improves precision of the relative risk estimates!

Imputation/extrapolation of rates

Imputation of missing data for one stratum by taking advantage of corresponding observations in other strata.

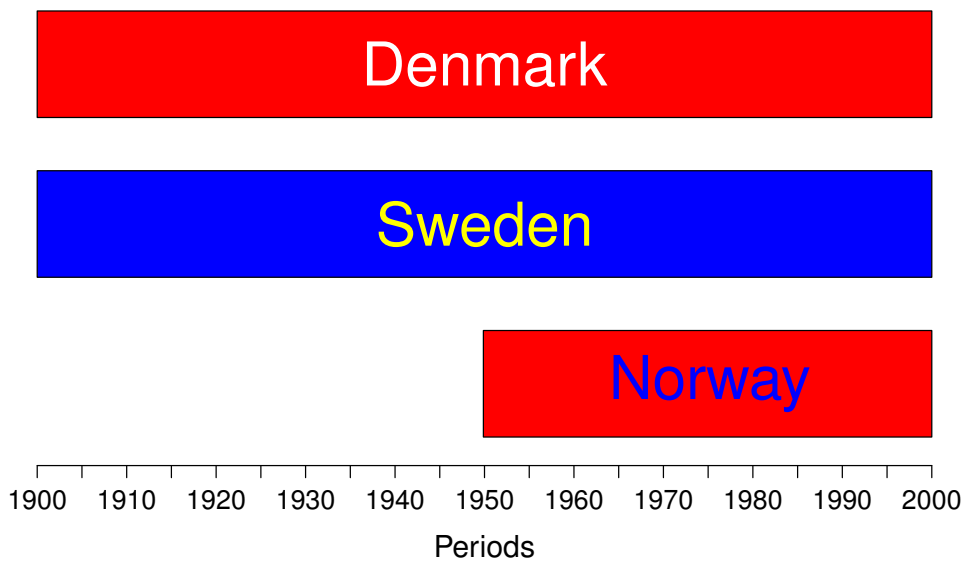
- In particular interesting for short term projections, historic data or to adjust for varying collection periods.
- We are able to consider the most flexible apc model.

Mortality extrapolation

Riebler et al., 2012, AOAS

Data: Overall mortality counts among females stratified by

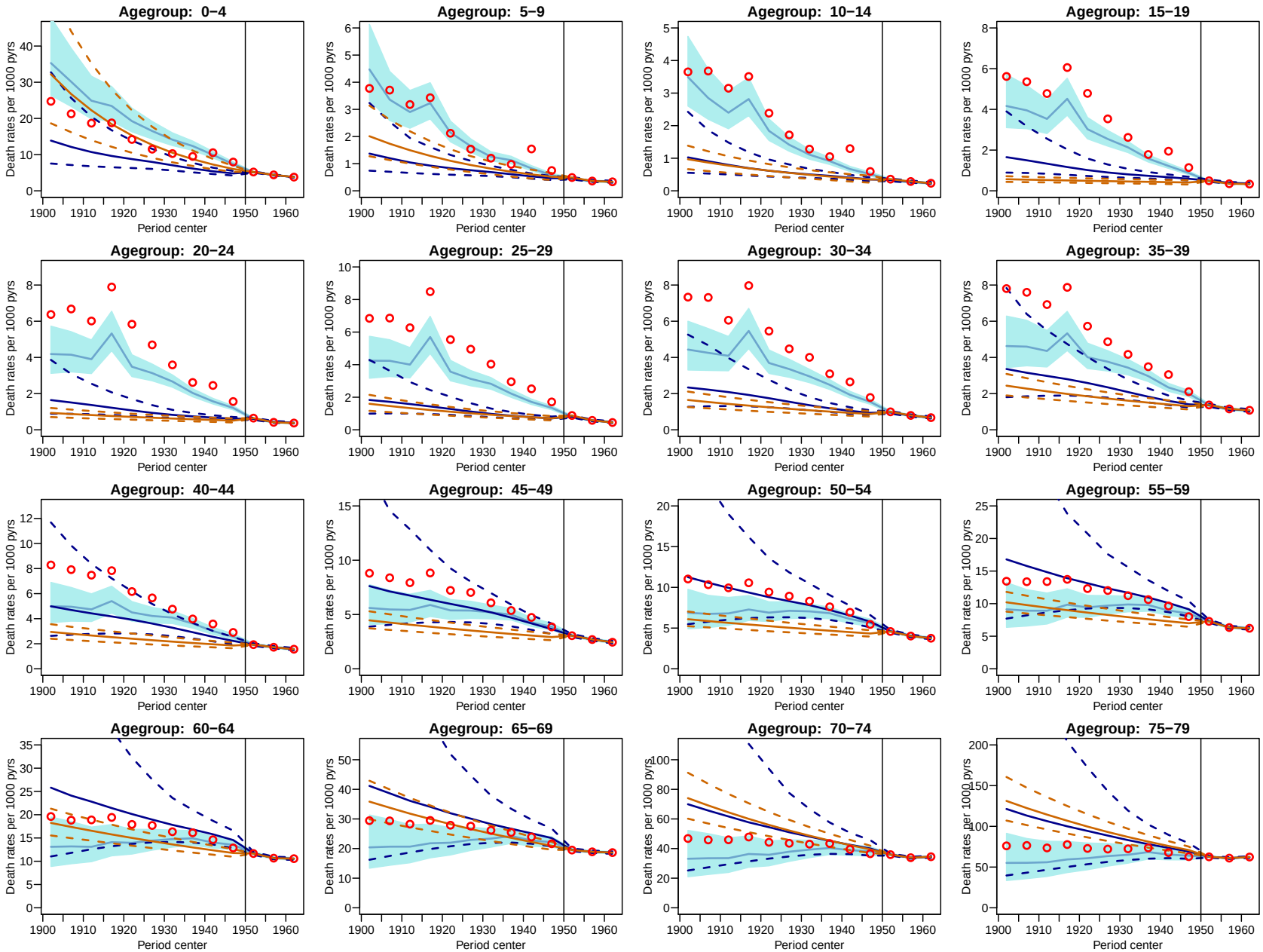
- $R = 3$ regions: Denmark, Sweden and Norway.
- $I = 17$ age groups: $0-4, 5-9, \ldots, 75-79, 80-84$.
- $J = 20$ periods from $1900-1999$ for Denmark and Sweden (only 10 periods for Norway)



Mortality extrapolation II

- Projection for Norwegian women for 1900–1949 by borrowing strength of full mortality rates of Sweden and Denmark.
- Comparison to a univariate APC model and an extended Lee-Carter demographic forecasting approach.
- The correlated model performs better in terms of mean squared error, coverage probabilities and proper scoring rules.

Observed and predicted mortality rates



Summary

- Applying APC models you need to be aware of identifiability problems.
- (Multivariate) APC models are well suited to estimate and project time trends in registry data.
- A Bayesian APC analysis can easily be done using INLA. No MCMC necessary.

Acknowledgements

- **Leonhard Held** & the Biostatistics group, UZH.
- **Håvard Rue** & the Department of Mathematical Sciences, NTNU.

- Financial support by the “Swiss National Science Foundation (SNSF)” and the “Research Council of Norway” is gratefully acknowledged.

Thank you for your attention!

Backup Slides

INLA: non-Gaussian observations

Main goal: Compute the posterior marginals

$$\pi(x_i|\boldsymbol{y}) = \int_{\boldsymbol{\theta}} \underbrace{\int_{\boldsymbol{x}_{-i}} \pi(\boldsymbol{x},\boldsymbol{\theta}|\boldsymbol{y}) d\boldsymbol{x}_{-i}}_{\pi(x_i,\boldsymbol{\theta}|\boldsymbol{y})=\pi(x_i|\boldsymbol{\theta},\boldsymbol{y})\pi(\boldsymbol{\theta}|\boldsymbol{y})} d\boldsymbol{\theta},$$

$$\pi(\theta_j|\boldsymbol{y}) = \int_{\boldsymbol{\theta}_{-j}} \underbrace{\int_{\boldsymbol{x}} \pi(\boldsymbol{x},\boldsymbol{\theta}|\boldsymbol{y}) d\boldsymbol{x}}_{\pi(\boldsymbol{\theta}|\boldsymbol{y})} d\boldsymbol{\theta}_{-j}.$$

From $\pi(\boldsymbol{x},\boldsymbol{\theta},\boldsymbol{y}) = \pi(\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{y}) \times \pi(\boldsymbol{\theta}|\boldsymbol{y}) \times \pi(\boldsymbol{y})$ it follows that:

$$\pi(\boldsymbol{\theta}|\boldsymbol{y}) \propto \frac{\pi(\boldsymbol{x},\boldsymbol{\theta},\boldsymbol{y})}{\pi(\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{y})} \approx \frac{\pi(\boldsymbol{x},\boldsymbol{\theta},\boldsymbol{y})}{\tilde{\pi}_G(\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{y})} \propto \tilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y}) \quad \text{(Laplace approximation)}$$

with $\tilde{\pi}_G(\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{y})$ Gaussian approximation to $\pi(\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{y})$.

INLA

Step I Build a Laplace approximation to

$$\pi(\boldsymbol{\theta}|\boldsymbol{y}) \propto \frac{\pi(\boldsymbol{x},\boldsymbol{\theta},\boldsymbol{y})}{\pi(\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{y})} \approx \frac{\pi(\boldsymbol{x},\boldsymbol{\theta},\boldsymbol{y})}{\tilde{\pi}_G(\boldsymbol{x}|\boldsymbol{\theta},\boldsymbol{y})} \propto \tilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y})$$

and "explore it numerically" to obtain good support points $\boldsymbol{\theta}_k$.

Step II Approximate $\pi(x_i|\boldsymbol{y},\boldsymbol{\theta}_k)$ for each $\boldsymbol{\theta}_k$.

Step III For each i, sum out $\boldsymbol{\theta}_k$

$$\tilde{\pi}(x_i|\boldsymbol{y}) = \sum_k \tilde{\pi}(x_i|\boldsymbol{\theta}_k,\boldsymbol{y}) \times \tilde{\pi}(\boldsymbol{\theta}_k|\boldsymbol{y}) \times \Delta_k.$$

Step IV Approximate $\pi(\theta_k|\boldsymbol{y})$ using numerical integration.

Posterior correlation in Apc model