The analysis of heterogeneous time trends in multivariate age-period-cohort models

Leonhard Held & Andrea Riebler

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

5 March 2009

Financial support by the Swiss National Science Foundation (SNF) is gratefully acknowledged.

Introduction

Area-level incidence or mortality counts are typically stratified by gender, age and period.

Our goal lies in detecting heterogeneous time trends across areas as they could be related with possible causes of the disease.

Method:

Age-period-cohort (APC) model

to describe incidence or mortality rates using three time scales

- **A**ge: age at diagnosis/death
- **P**eriod: date of diagnosis/death
- **C**ohort: date of birth

Universität Zürich

Univariate age-period-cohort model

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

$$y_{ij} \sim \text{Poisson}(n_{ij} \exp(\xi_{ij})) \qquad \xi_{ij} = \mu + \alpha_i + \beta_j + \gamma_k + z_{ij}$$

with age, period and cohort effect α_i, β_j, and γ_k and additional random effect $z_{ij} \sim \mathcal{N}(0, \delta^{-1})$ to adjust for overdispersion.

To assure identifiability of the intercept μ, we set

$$\sum_{i=1}^{I} \alpha_i = \sum_{j=1}^{J} \beta_j = \sum_{k=1}^{K} \gamma_k = 0.$$

Note: Because of the linear relationship $k = I - i + j$, the age, period and cohort effects are still not identifiable.

Non-identifiability

- For any value of c, the linear transformations

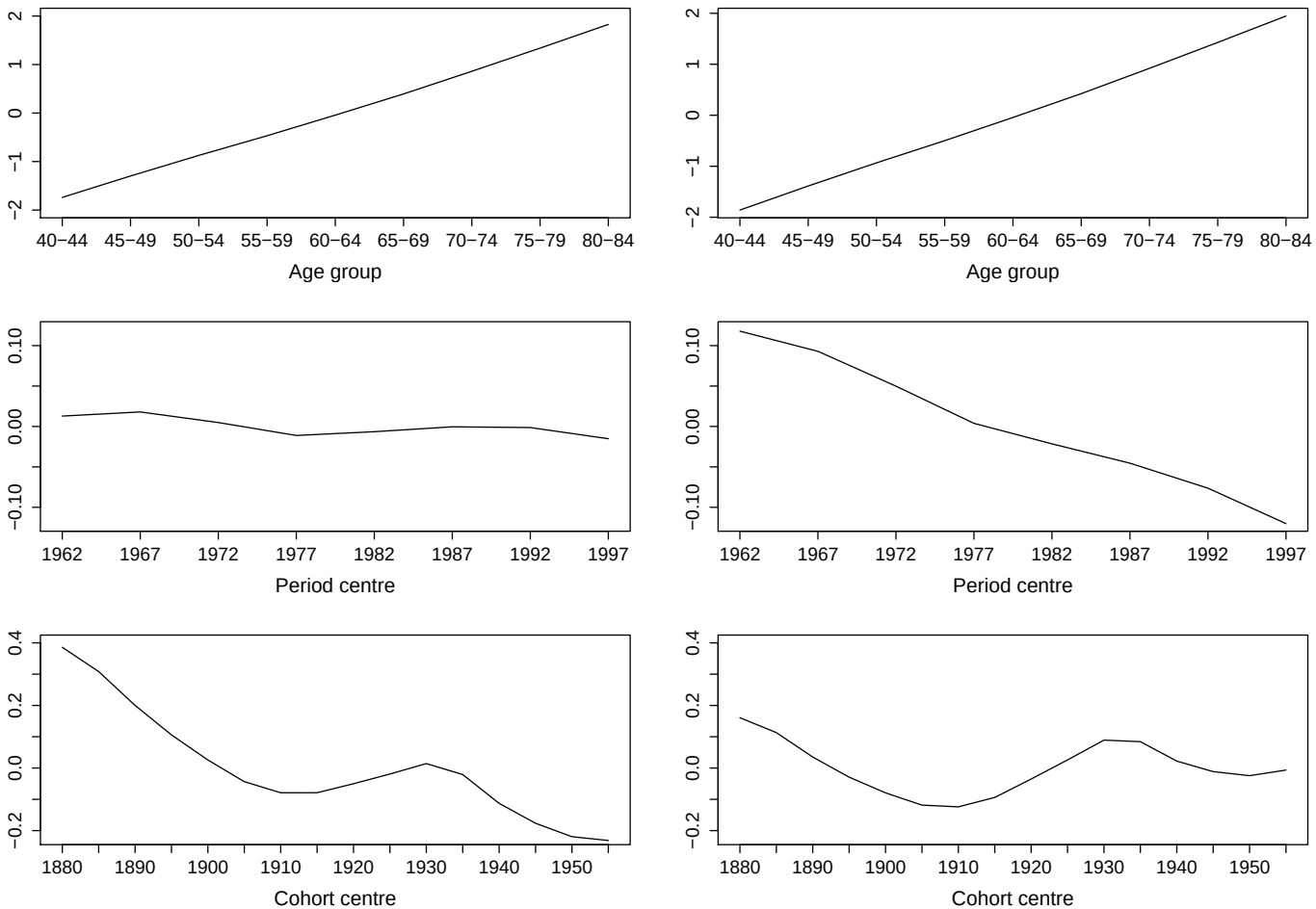
$$\begin{aligned}
\alpha_i &\rightarrow \alpha_i + c\left(i - \frac{I+1}{2}\right) \\
\beta_j &\rightarrow \beta_j - c\left(j - \frac{J+1}{2}\right) \\
\gamma_k &\rightarrow \gamma_k + c\left(k - \frac{K+1}{2}\right)
\end{aligned}$$

still fulfill $\sum \alpha_i = 0$, $\sum \beta_i = 0$ and $\sum \gamma_i = 0$, but leave ξ_{ij} unchanged for all possible combinations of i and j.

- Hence, only nonlinear trends of the age, period or cohort effects are interpretable, but linear trends are not.

Universität Zürich

Illustration of non-identifiability



Unequally spaced intervals

- The APC model is not directly applicable when age group and period do not have the same interval lengths, typically the age group length is a multiple M of the period length.
- Cohort indices $k = M \cdot (I - i) + j$ as proposed in Heuer (1997) can be then used, e.g. for $M = 5$:

Age group									
≥ 85	1	2	3	4	5	6	7	...	20
80-84	6	7	8	9	10	11	12	...	25
75-79	11	12	13	14	15	16	17	...	30
70-74	16	17	18	19	20	21	22	...	35
:	:	:	:	:	:	:	:		:
35-39	51	52	53	54	55	56	57	...	70
	1981	1982	1983	1984	1985	1986	1987	...	2000
	Period (years)								

Multivariate age-period-cohort model

Data are available for strata $r = 1, \ldots, R$, e.g. gender, regions, etc.

Idea: Linking different strata, e.g. through common age effects, period or cohort effects can be separate (Riebler and Held, 2009):

- ApC: Separate period effects:

$$\xi_{ijr} = \mu_r + \alpha_i + \beta_{jr} + \gamma_k + z_{ijr}$$

- APc: Separate cohort effects:

$$\xi_{ijr} = \mu_r + \alpha_i + \beta_j + \gamma_{kr} + z_{ijr}$$

- Apc: Separate period effects and separate cohort effects:

$$\xi_{ijr} = \mu_r + \alpha_i + \beta_{jr} + \gamma_{kr} + z_{ijr}$$

Relative risks

- Differences

$$\begin{aligned} \Delta_j &= \beta_{j,r_1} - \beta_{j,r_2} && \text{in the ApC model} \\ \Delta_k &= \gamma_{k,r_1} - \gamma_{k,r_2} && \text{in the APc model} \\ \Delta_{jk} &= \Delta_j + \Delta_k && \text{in the Apc model} \end{aligned}$$

  for two arbitrary strata r_1 and r_2 are identifiable.

- Let $\Delta_\mu = \mu_{r_1} - \mu_{r_2}$. If available, the adjusted differences

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

  can be interpreted as (average) log relative risk.

Inference

Up to now only maximum likelihood (ML) estimation has been used (Hansell *and others*, 2003; Jacobsen *and others*, 2004) for the multivariate APC model.
Disadvantages:

- Additional constraints are necessary for identifiability.
- The model must provide exact fits to the cells which correspond to the single observations on cohorts 1 and K (Besag *and others*, 1995).
- For unequally spaced intervals of age and period, ML estimates become very unstable resulting in an artificial saw-tooth pattern (Holford, 2006).

Here we use a Bayesian approach based on smoothing priors.

Bayesian age-period-cohort model

Non-parametric smoothing priors are used for the main effects with gamma hyperpriors for the associated smoothing parameters.

Second-order random walk (RW2)

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

RW2 penalises deviations from a linear trend $\alpha_i = 2\alpha_{i-1} - \alpha_{i-2}$.

Note:

- Non-identifiability of the latent parameters remains, but does not require further constraints.

Model choice

- Marshall and Spiegelhalter (2003) proposed cross-validatory predictive checks for Bayesian hierarchical models.
- Approach can be used to provide sound model choice criteria based on proper scoring rules (Gneiting and Raftery, 2007).
- Here we use the ranked probability score (RPS) and the David-Sebastiani score (DSS). Smaller score values are better.

I: Mortality among Females in Denmark and Norway

- Kesteloot (2001) speculated that the increase in mortality in Denmark could be linked to a role model effect of Queen Margrethe II of Denmark who has been a known cigarette smoker.
- If there is a link we would expect a pronounced period effect around her ascension to the throne in 1972.

- 8 Periods: 1960-1964, 1965-1969, . . ., 1995-1999
- 17 Age groups: 0-4, 5-9, . . . , 75-79, 80+
- 2 Regions: Denmark and Norway

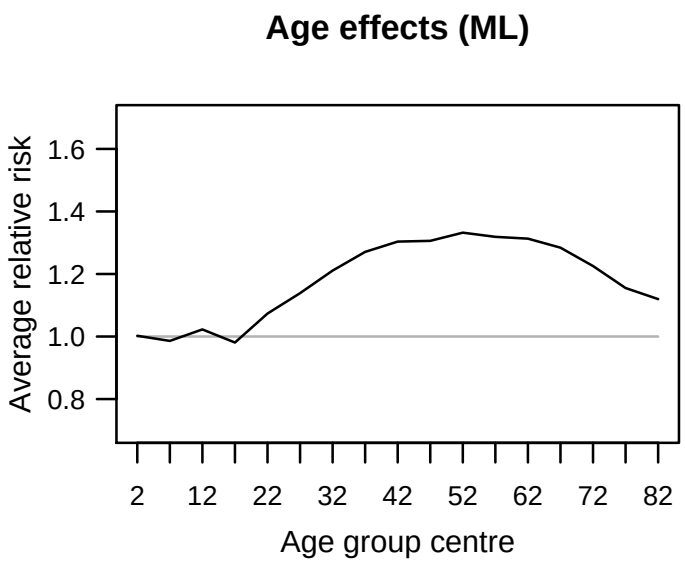
Model choice

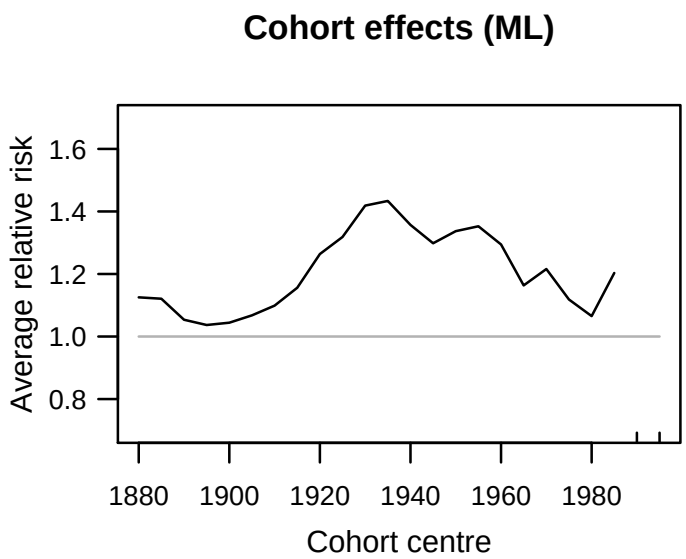
	APC	aPC	ApC	APc
$\overline{\text{RPS}}$	231.73	134.85	226.14	122.00
$\overline{\text{DSS}}$	11.09	10.34	11.08	10.29

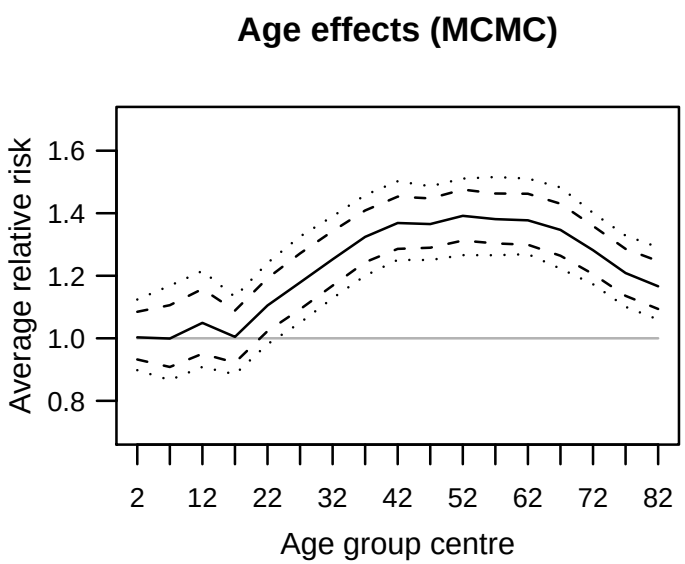
	apC	aPc	Apc	apc
$\overline{\text{RPS}}$	120.14	86.81	134.50	89.08
$\overline{\text{DSS}}$	10.27	9.91	10.30	9.93

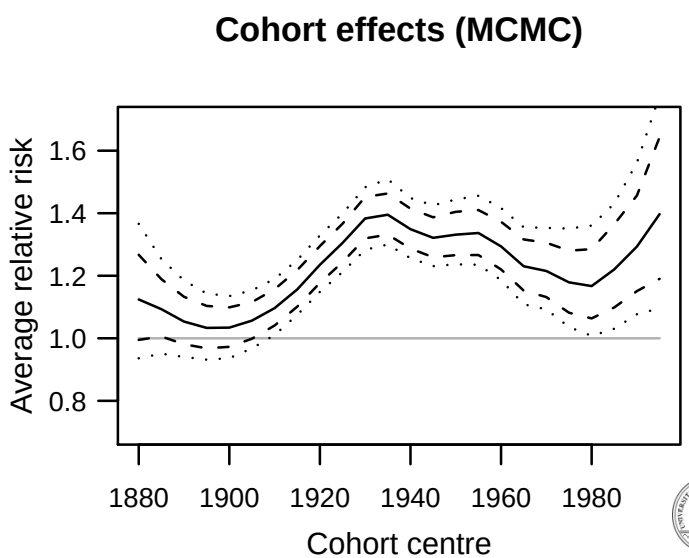
Universität Zürich

Estimates of relative risk (aPc model)









Universität Zürich

II: COPD mortality among Males in England and Wales

Chronic obstructive pulmonary disease (COPD) is a serious lung disease making it hard to breathe.

Factors implicated in COPD development:

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

Data (Hansell *and others*, 2003):

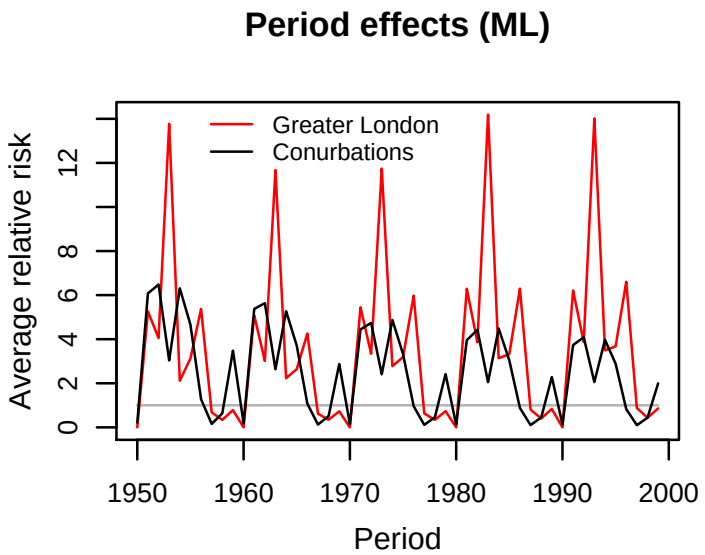
- 50 Periods: 1950 - 1999 on a yearly basis
- 7 Age groups: 15-24, 25-34, ..., 75+
- 3 Regions: conurbations excluding Greater London, Greater London and non-conurbations

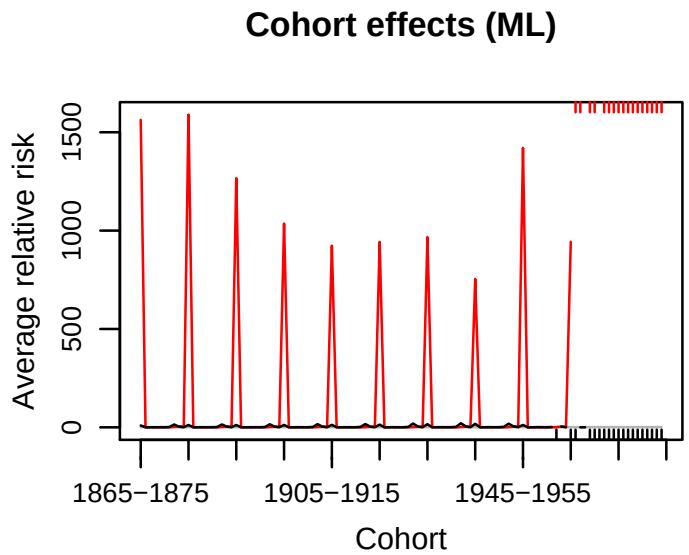
Model choice

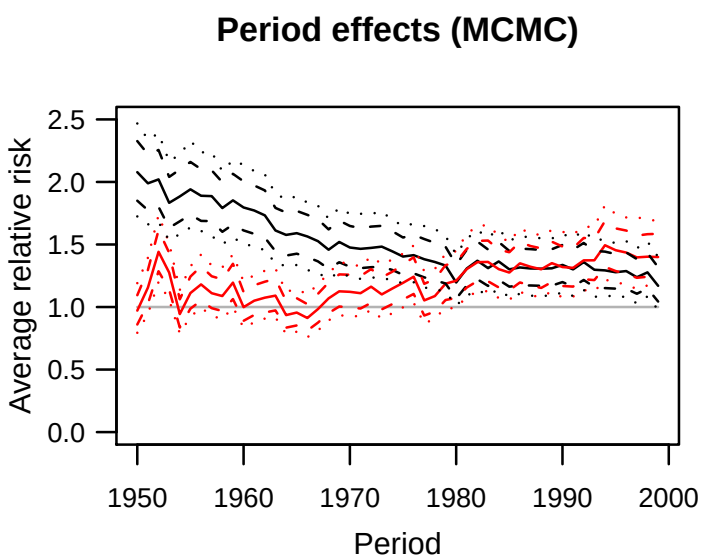
	APC	aPC	ApC	APc
$\overline{\text{RPS}}$	51.52	39.37	50.42	41.26
$\overline{\text{DSS}}$	7.19	6.82	7.14	7.01

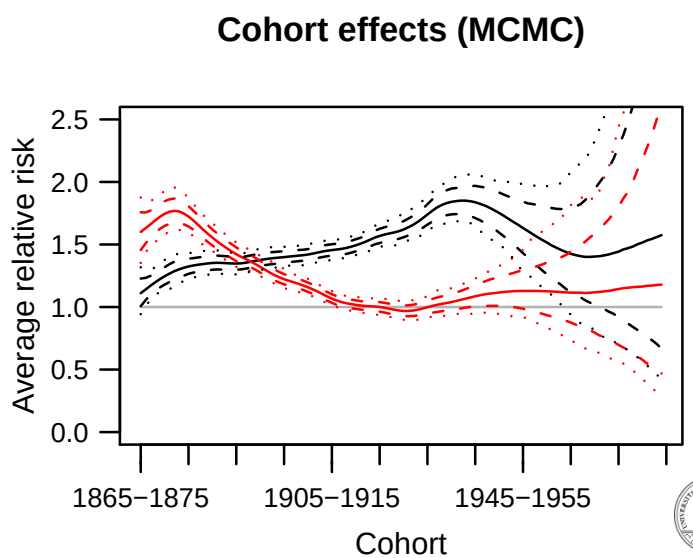
	apC	aPc	Apc	apc
$\overline{\text{RPS}}$	30.74	27.45	21.23	20.13
$\overline{\text{DSS}}$	6.56	6.56	6.30	6.18

Estimates of relative risk (Apc model)










Universität Zürich

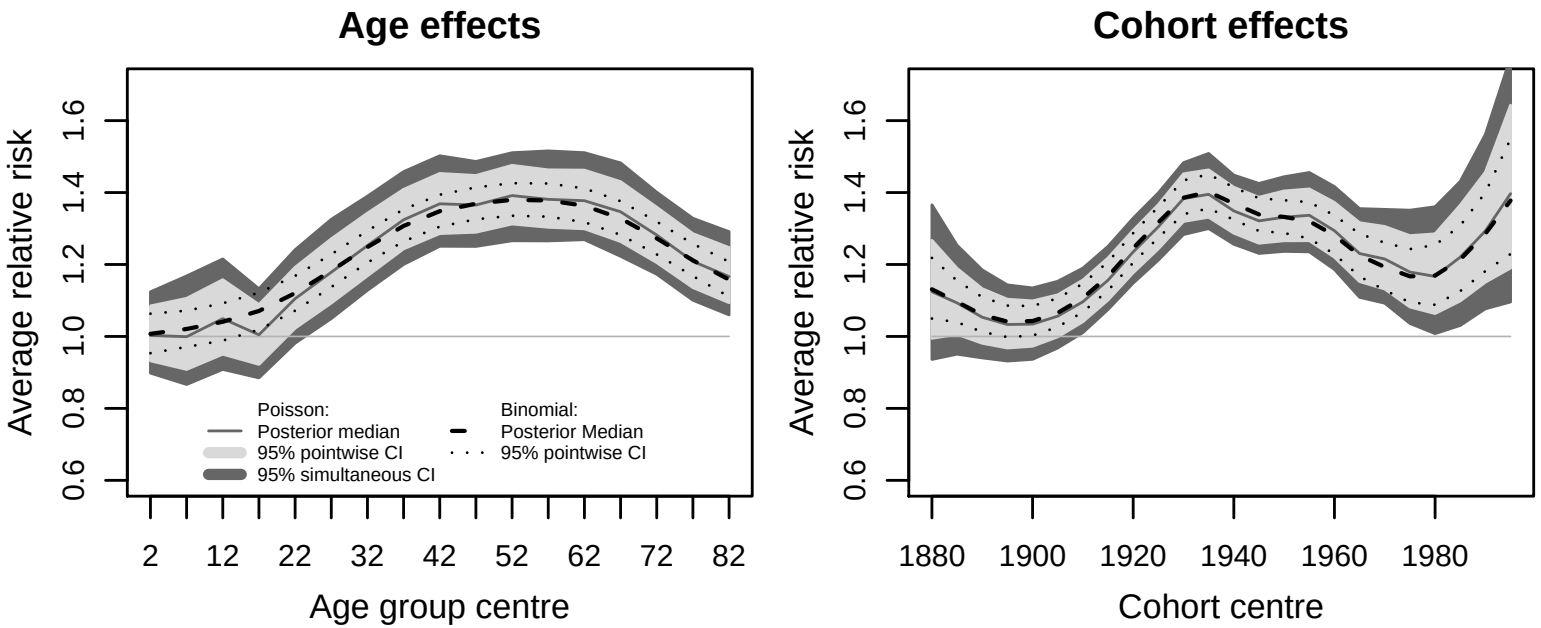
A conditional approach

- For simplicity, consider the case of $R = 2$ strata and let $y_{ij} = y_{ij1} + y_{ij2}$.
- It is easy to see that the Apc model for y_{ijr} implies that $y_{ij1}|y_{ij}$ is binomial with success probability

$$\pi_{ij1} = \text{logit}^{-1}\left(\log\left(\frac{n_{ij1}}{n_{ij2}}\right) + \Delta_\mu + \Delta_j + \Delta_k + \Delta_z\right)$$

- Note that through conditioning, the original age, period and cohort parameters are replaced by the differences Δ_j and Δ_k.
- Age effects are no longer present in this formulation.
- All parameters are identifiable and can be estimated with ML or Bayesian techniques (MCMC or INLA).

I: Mortality among Females in Denmark and Norway (aPc model)



- Similar point estimates
- Slightly narrower credible intervals

Universität Zürich

II: COPD Mortality in England and Wales
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

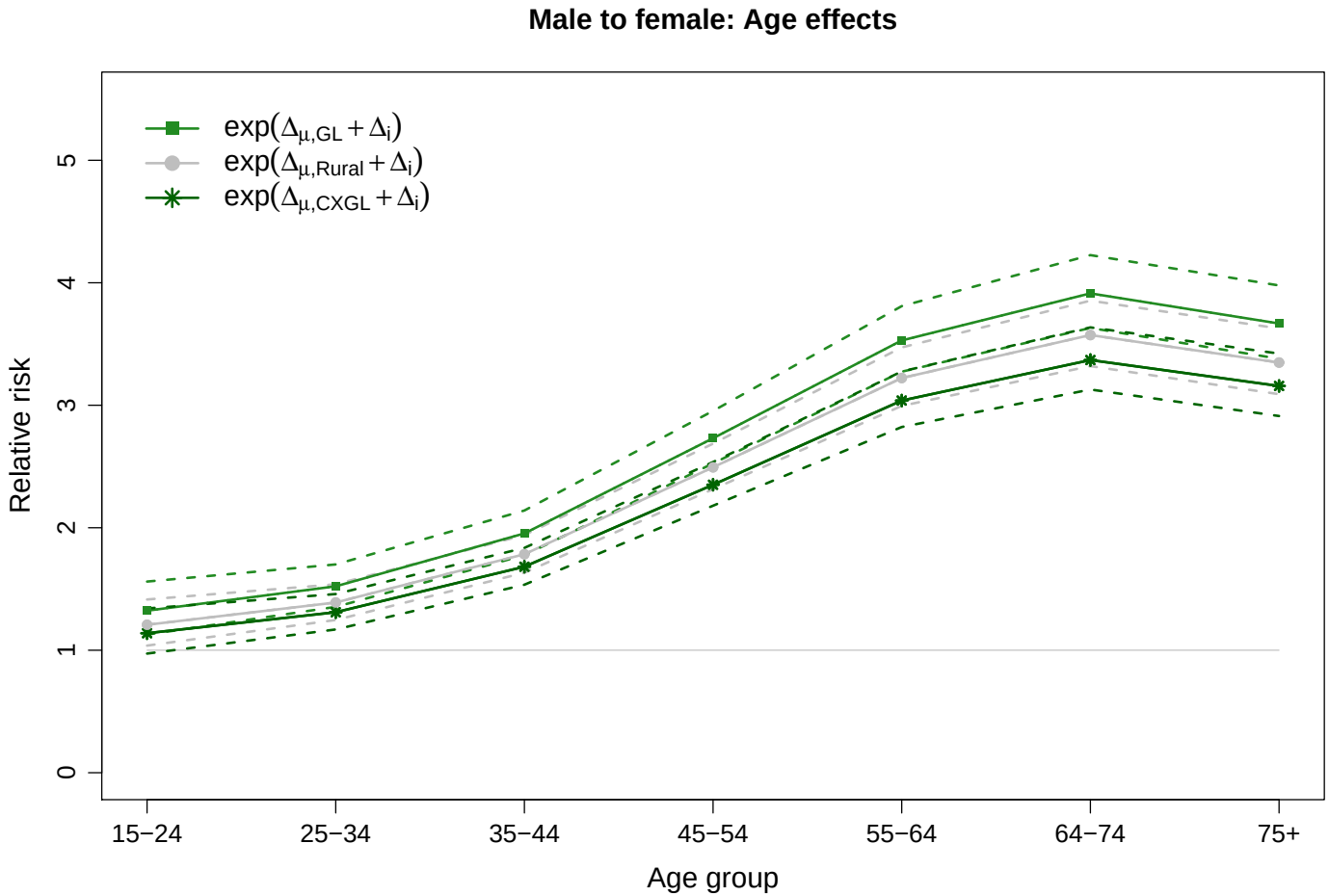
$$\xi_{ijrg} = \mu_{rg} + \alpha_{ig} + \beta_{jr} + \gamma_{kr} + z_{ijrg}$$

- A conditional formulation can be deduced for $y_{ijr,♂} | y_{ijr,♂} + y_{ijr,♀}$ with success probability

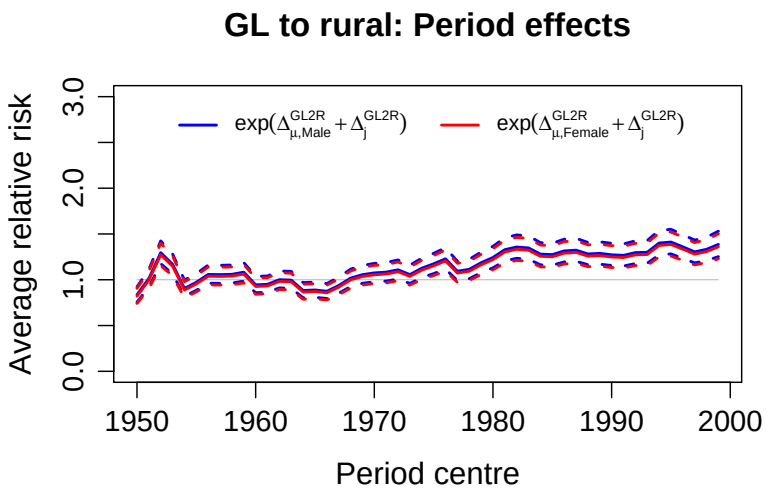
$$\pi_{ijr,♂} = \text{logit}^{-1}\left(\log \frac{n_{ijr,♂}}{n_{ijr,♀}} + \Delta_{\mu,r} + \Delta_i + \Delta_z\right)$$

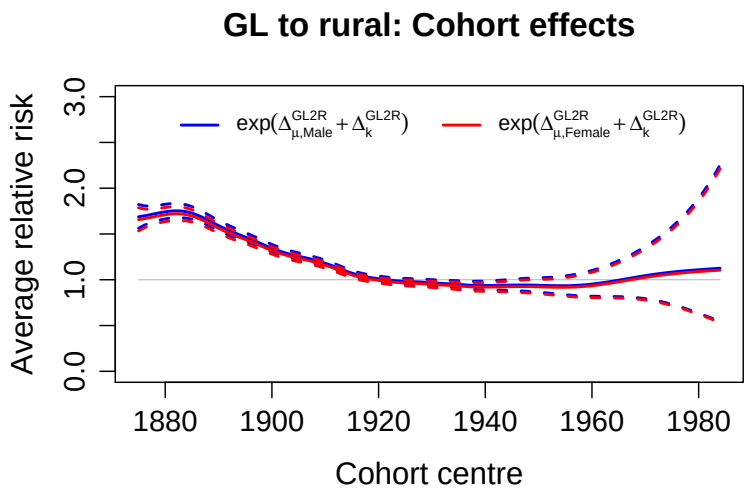
and similarly for $y_{ij,1,♂} | y_{ij,1,♂} + y_{ij,2,♂}, \ldots$

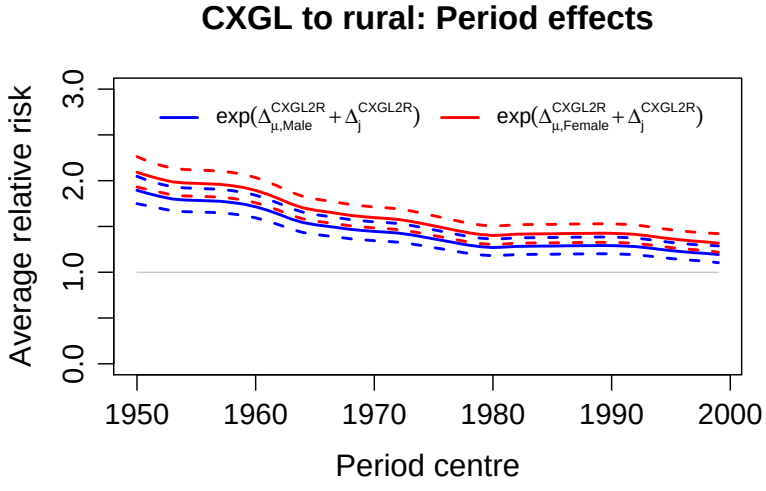
Relative risk estimates for gender

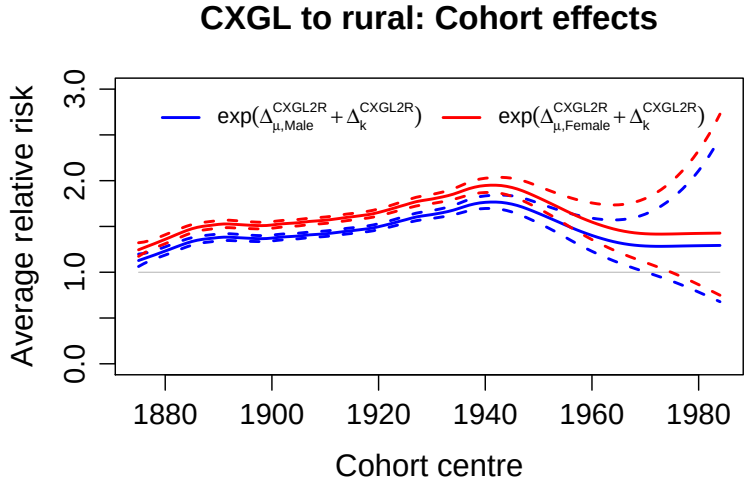


Relative risk estimates for period and cohort










Universität Zürich

Summary and outlook

Summary

- A multivariate analysis of APC models makes it possible to identify relative time trends.
- Only the Bayesian approach provides plausible estimates for unequally spaced data.
- A conditional formulation simplifies the analysis.

Work in progress

- Model choice for conditional formulation?
- Integrated model selection to find the most appropriate multivariate APC model for a large number of regions.

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

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

Hansell, A., Knorr-Held, L., Best, N., Schmid, V. and Aylin, P. (2003). COPD mortality trends 1950-1999 in England & Wales - Did the 1956 Clean Air Act make a detectable difference?, *Epidemiology* **14**(5): S55.

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

Holford, T. R. (2006). Approaches to fitting age-period-cohort models with unequal intervals, *Statistics in Medicine* **25**: 977–993.

Jacobsen, R., von Euler, M., Osler, M., Lynge, E. and Keiding, N. (2004). Women's death in Scandinavia - what makes Denmark different?, *European Journal of Epidemiology* **19**: 117–121.

Kesteloot, H. (2001). Queen Margrethe II and mortality in Danish women, *Lancet* **357**: 871–872.

Marshall, E. C. and Spiegelhalter, D. J. (2003). Approximate cross-validatory predictive checks in disease-mapping methods, *Statistics in Medicine* **22**(4): 1649–1660.

Riebler, A. and Held, L. (2009). The analysis of heterogeneous time trends in multivariate age-period-cohort models, *Technical report*. In revision for *Biostatistics*.