

A joint spatial analysis of gastrointestinal infectious diseases

Leonhard Held

Biostatistics Unit
Institute of Social and Preventive Medicine
University of Zurich

Financial support by the German Research Foundation (DFG), SFB 386

Remagen/Bonn, October 6, 2006

Outline

- 1 Introduction
- 2 Adjustments for Local Outbreaks
- 3 Separate Spatial Analysis
- 4 Joint Spatial Analysis

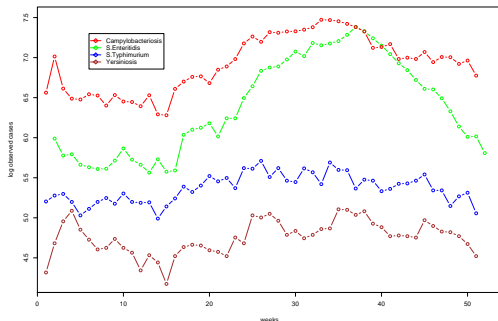
Data

Surveillance data on (mainly endemic) gastrointestinal infections in Germany in 2004 on district level:

- 1 Campylobacteriosis (55676 cases)
- 2 Salmonellosis
 - a) *Salmonella enteritidis* (36615 cases)
 - b) *Salmonella typhimurium* (11298 cases)
- 3 Yersiniosis (6175 cases)
 - For each disease, the number of cases y_{it} is available in district $i = 1, \dots, 438$ and week $t = 1, \dots, 52$
 - Population counts n_i are also available

Time trends

Log number of notified cases in 2004 in Germany for the different diseases



Epidemiological background

- Ingestion of raw/undercooked **poultry** is considered as the dominant risk factor for infection with *Campylobacteriosis* and *S. enteritidis*
- Ingestion of raw/undercooked contaminated **pork** is considered as the dominant risk factor for infection with *S. typhimurium* and *Yersiniosis*

Underreporting

- Huge problem!
- In principle, local health authorities are required to report all cases to the Robert Koch Institute in Berlin
- However, willingness of GP's to investigate and report cases is very variable, in particular for gastrointestinal diseases
- Also, many patients may not go to a GP at all

⇒ spatial pattern will be dominated by heterogeneity in underreporting rates

Goal of analysis

- Investigate for a **common spatial pattern** (to be interpreted as heterogeneity in underreporting), that does depend on the particular disease only through a **different scaling factor**
- Investigate if **residual spatial variation** can be decomposed into components that reflect spatial variation of contaminated poultry and pork product consumption, respectively

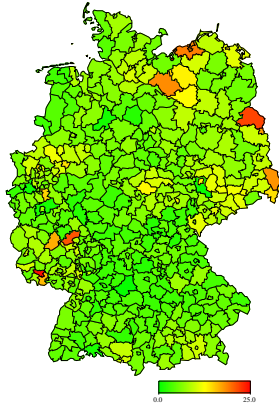
⇒ **joint disease mapping** approach

- Model will allow for overdispersion, however, **local outbreaks** are not accounted for; may further dilute the spatial pattern

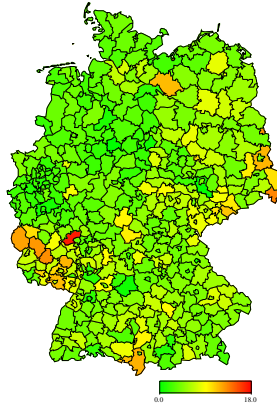
⇒ initial **adjustments** based on time series model for y_{it} (Held et al. 2005)

Incidence Rates (per 10,000 inhabitants)

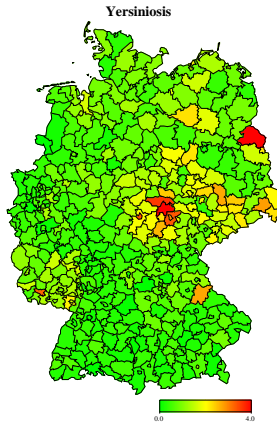
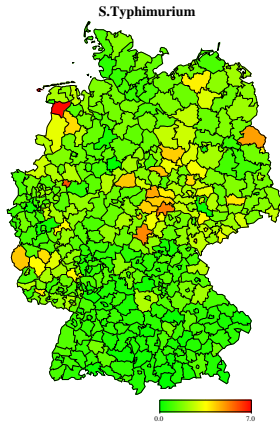
Campylobacteriosis



S.Enteritidis



Incidence Rates (per 10,000 inhabitants)



Branching process with immigration

The observed number of counts Y_t is decomposed into two (unknown) components, V_t and W_t :

$$Y_t = V_t + W_t, \quad t = 1, \dots, n$$

$$V_t \sim \text{Po}(\nu)$$

$$W_t | Y_{t-1} \sim \text{Po}(\lambda Y_{t-1})$$

- V_t : endemic component
- W_t : autoregressive epidemic component ($\lambda > 0$)
- Y_t is stationary with mean $\nu/(1 - \lambda)$ if $\lambda < 1$
- \rightsquigarrow Interpretation of λ as epidemic proportion of total incidence

Generalized branching process with immigration

Now ν is time-dependent:

$$Y_t = V_t + W_t, \quad t = 1, \dots, n$$
$$V_t \sim \text{Po}(\nu_t)$$
$$W_t | Y_{t-1} \sim \text{Po}(\lambda Y_{t-1})$$

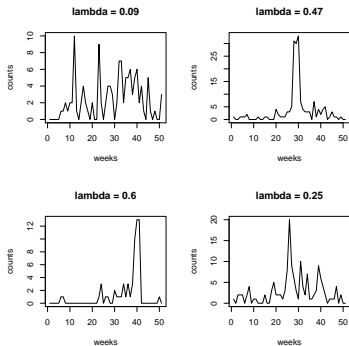
- $\log \nu_t$ is modelled parametrically as in log-linear Poisson regression; will include terms for **seasonality**
- Interpretation of λ still valid

Adjusting for local outbreaks

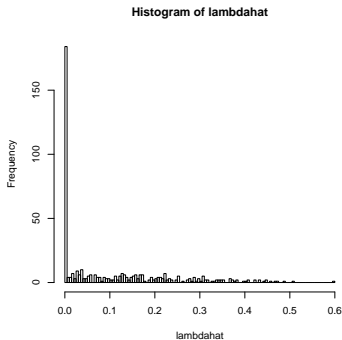
- Fit time series model separately in each district by Maximum Likelihood (ML)
 - ↪ ML-estimates of epidemic proportion λ_i
- Consider only (suitably rounded) endemic proportion $(1 - \hat{\lambda}_i)$ of total incidence (the **endemic incidence**)
- Data reduction between 3% (Yersiniosis) and 12% (Salmonella Enteritidis).

Example: Salmonella Enteritidis

Time series in four selected districts

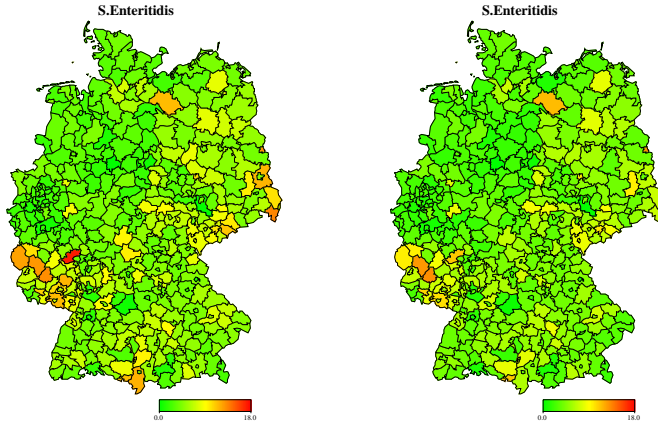


Estimated epidemic proportion



Example: Salmonella Enteritidis

Total and estimated endemic incidence rate



Separate Spatial Model

$$y_i \sim \text{Bin}(n_i, \text{logit}^{-1}(\eta_i))$$

with the observed yearly number of counts y_i in area i .

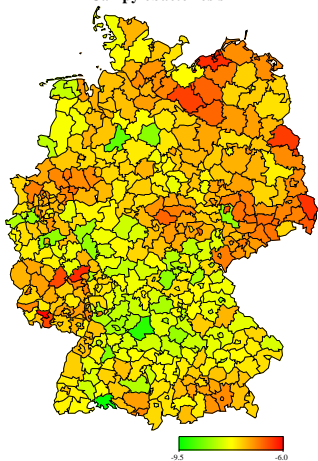
The log odds η_i is assumed to be normal with mean

$$E(\eta_i) = a + u_i \quad \text{where}$$

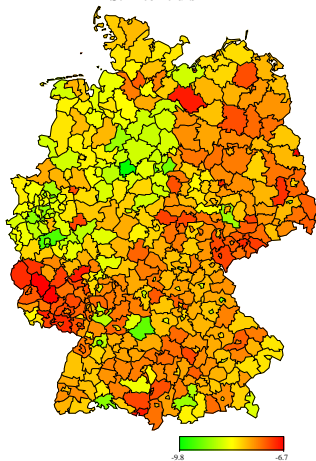
- \mathbf{u} is a (centered) **Gaussian Markov random field**
 - $\text{Var}(\eta_i) = \sigma^2$ allows for **overdispersion**
 - Model dates back to Besag, York and Mollié (1991)
- ↪ Spatial smoothing of incidence rates

Results from separate spatial analyses

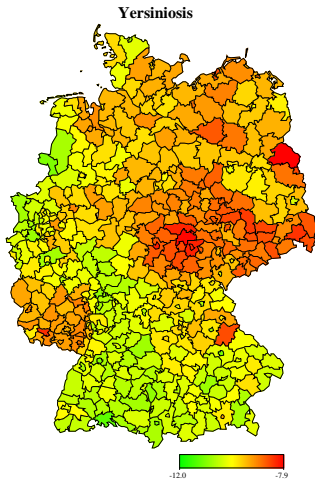
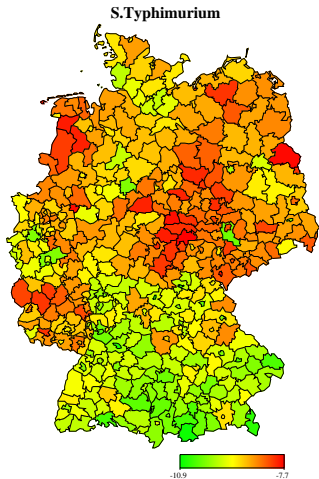
Campylobacteriosis



S.Enteritidis



Results from separate spatial analyses



Joint Spatial Model

$$y_{di} \sim \text{Bin}(n_i, \text{logit}^{-1}(\eta_{di}))$$

with the observed yearly number of counts y_{di} in area i for disease j . The log odds η_{di} is assumed to be normal with mean

$$E(\eta_{di}) = a_d + \sum_k \delta_{dk} u_{ki} \quad \text{where}$$

- \mathbf{u}_k are (centered) **Gaussian Markov random fields** and
- $\delta_{dk} > 0$ are (unknown) **scaling factors** for pre-selected diseases $d \in N_k$ to determine the relative contribution of the spatial field \mathbf{u}_k to each relevant disease
- **Identifiability constraints** are placed on δ_{dk} , $d \in N_k$
- $\text{Var}(\eta_{di}) = \sigma_d^2$ allows for disease-specific **overdispersion**

Model and inference

Specific model includes

- one shared component u_1 for all diseases (“underreporting”)
- one component u_2 for Campylobacter and S.Enteritidis (“poultry”)
- one component u_3 for S.Typhimurium and Yersiniosis (“pork”)

$$E(\eta_{Ci}) = a_C + \delta_{C1} u_{1i} + \delta_{C2} u_{2i}$$

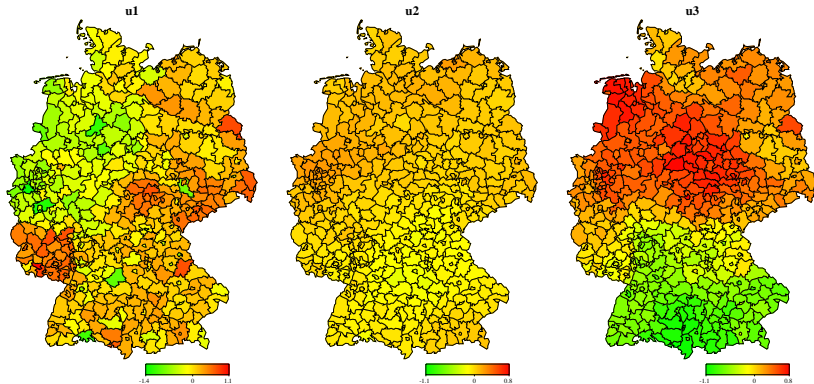
$$E(\eta_{Ei}) = a_E + \delta_{E1} u_{1i} + \delta_{E2} u_{2i}$$

$$E(\eta_{Ti}) = a_T + \delta_{T1} u_{1i} + \delta_{T3} u_{3i}$$

$$E(\eta_{Yi}) = a_Y + \delta_{Y1} u_{1i} + \delta_{Y3} u_{3i}$$

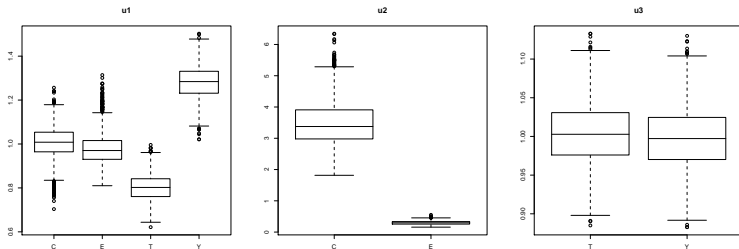
Bayesian inference by block-sampling MCMC algorithm (Rue and Held, 2005)

Estimated spatial fields in the joint model

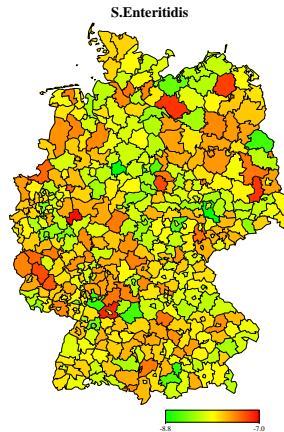
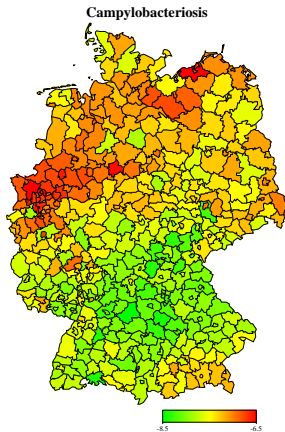


Estimated scaling factors

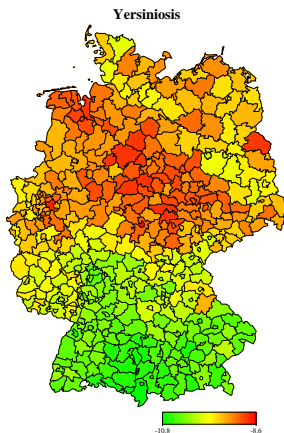
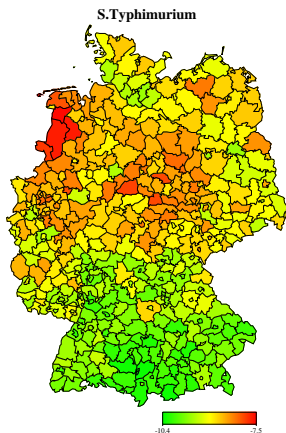
Boxplots of posterior distribution of δ_{dk}



Disease incidence (log odds scale) adjusted for underreporting



Disease incidence (log odds scale) adjusted for underreporting



Model fit

Comparison of model fit (\bar{D}), model complexity (p_D), and DIC values between separate and joint modelling.

| Disease | Model | \bar{D} | p_D | DIC |
|--------------------|----------|-----------|--------|--------|
| Campylobacteriosis | Separate | 460.2 | 402.4 | 862.6 |
| | Joint | 461.8 | 388.4 | 850.3 |
| S. Enteritidis | Separate | 474.1 | 391.7 | 865.8 |
| | Joint | 470.1 | 382.0 | 852.2 |
| S. Typhimurium | Separate | 469.3 | 328.3 | 797.6 |
| | Joint | 466.7 | 306.2 | 772.9 |
| Yersiniosis | Separate | 462.2 | 280.2 | 742.5 |
| | Joint | 464.7 | 237.7 | 702.3 |
| Total | Separate | 1865.8 | 1402.6 | 3268.4 |
| | Joint | 1863.4 | 1314.4 | 3177.7 |

Discussion

- **Joint spatial analysis**
 - allows to estimate a joint spatial pattern and to adjust for spatial variation in underreporting
 - can detect (epidemiologically plausible) spatial variation in the residual rates
 - improves considerably in terms of model fit
- **Parsimonious model** ($\#$ spatial fields $<$ $\#$ diseases) ensures identifiability
 - identifiability of “contaminated pork consumption pattern” would not be possible for, say, a joint model for Salmonella Typhimurium and Yersiniosis alone

Discussion

Two-stage approach to investigate spatial variation of infectious diseases

- Ideally: link time series model with spatial model to a **joint spatio-temporal model** for infectious disease data
- Goal: Parallel recent statistical methodology for the spatio-temporal analysis of chronic diseases, for example **spatial ecological regression** for infectious disease surveillance data

References

- Besag, J. E., York, J. and Mollié, A. (1991).
Bayesian image restoration with two applications in spatial statistics (with discussion).
Annals of the Institute of Statistical Mathematics, **43**, 1–59.
- Held, L., Graziano, G., Frank, C. and Rue, H. (2006)
Joint spatial analysis of gastrointestinal infectious diseases.
Statistical Methods in Medical Research, **15**, 465-480.
- Held, L., Höhle, M. and Hofmann, M. (2005).
A statistical framework for the analysis of multivariate infectious disease surveillance data.
Statistical Modelling, **5**, 187-199.
- Held, L., Natario, I., Fenton, S., Rue, H. and Becker, N. (2005)
Towards joint disease mapping.
Statistical Methods in Medical Research, **14**, 61-82.
- Rue, H. and Held, L. (2005).
Gaussian Markov Random Fields. Theory and Applications.
CRC/Chapman and Hall.