Spatial and temporal variation in the Heinz Nixdorf Recall study and their effects on the risk of depression at the district level

Dany Djeudeu[1,2], Susanne Moebus[2] , Karl-Heinz Jöckel[3] , Katja Ickstadt[1]

[1]Faculty of Statistics, TU Dortmund

[2] Centre for Urban Epidemiology (CUE), Institute for Medical Informatics, Biometry and Epidemiology (IMIBE), University Hospital Essen, University Duisburg-Essen

[3] Institute for Medical Informatics, Biometry and Epidemiology (IMIBE), University Hospital Essen, University Duisburg-Essen

08. Dezember 2017

technische universität dortmund

CENTRE FOR URBAN EPIDEMIOLOGY

Universitätsklinikum Essen

Background

- Increasing studies investigating urban effects on health

- Spatial variation(dependencies) often not considered

- Change of spatial variation over time often not considered in longitudinal studies

$\downarrow$

Errors in the covariate effects

Aims

- Estimate the risk of a selected outcome at district level adjusting for district level covariates
- Estimate the change of spatial structure in health outcome
- Investigate the effects of spatial and temporal variation on covariate effects

↓

Example:
Analysis of the effect of urban greenness on depression at district level

Orban et al., *Urban Residential Greenness and Depressive Symptoms: Results from the Heinz Nixdorf Recall Study*,2017, Journal of Transport and Health, 5, 3 – 63

Data set

Data of the Heinz Nixdorf Recall Study

- Population-based cohort study of $4,814$ randomly selected men and women
- $45 - 75$ years at baseline ($2000 - 2003$)
- From Essen, Mülheim and Bochum in the metropolitan Ruhr area, Germany

Modeling approach

Variables, all on district level

- Outcome : depression, aggregated
- Exposure: greenness
- Covariates: unemployment in districts, Body Mass Index, multi-morbidity, education level, changed addresses

Modeling approach

Statistical methods

- Traditional Poisson model + incorporating covariate effects
- Moran's I statistic to test for the remaining spatial clustering in the residuals
- Besag-Newell method to detect clusters
- Smoothing the previous risk: weighted, Besag-york Molie Model smoothing r for each follow-up
- **spatio-temporal autocorrelation via random effects**

Definition-greenness

- Normalized difference vegetation index (NDVI)
- From satellite data
- Values between -1 and 1, here only $0 - 1$
- District level neighborhood greenness
- Measurements: 2003, 2006, 2009

Definition depression

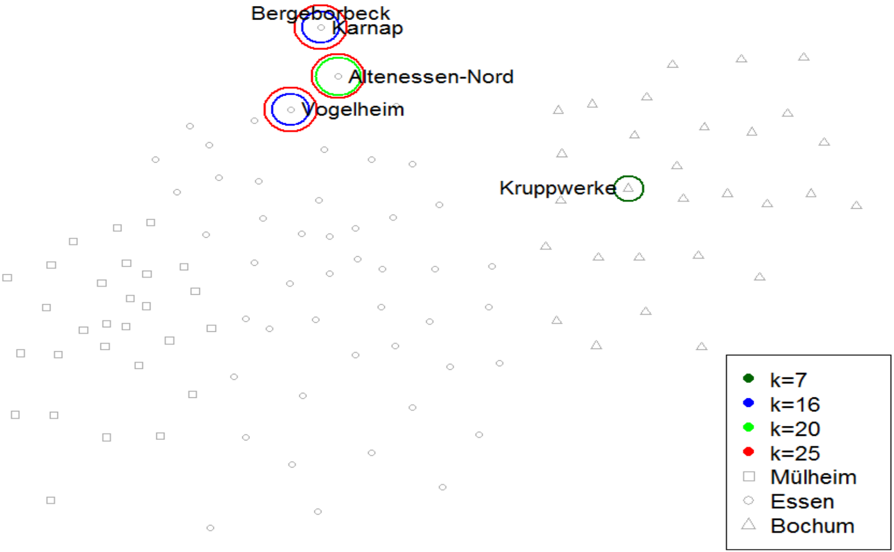
- Assessed using a 15-item short-form questionnaire of the CES-D
- Scores 0 – 45
- Cut point: $>= 17$
- Cases of depression aggregated at the district level
- 9 measurement time points (between 2000 – 2013)

Result: Spatial autocorrelation over time

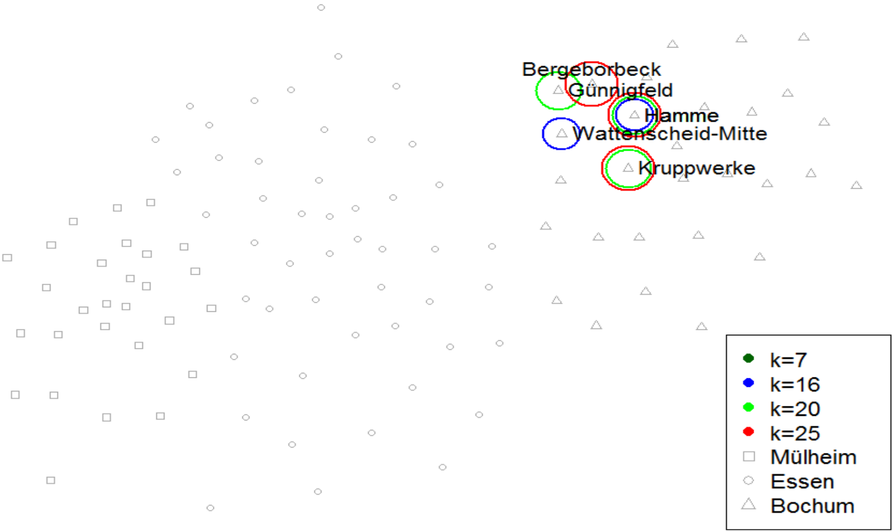
- Closer neighbour districts tend to have similar observations compared to districts farther away.
- Between -1 and 1
- Positive values indicate spatial autocorrelation

	Moran's I	p-value
$Jahr_0$	0.17	0.0024
$Jahr_5$	0.05	0.16
⋮	⋮	⋮
$Jahr_{10}$	-0.08	0.9
⋮	⋮	⋮

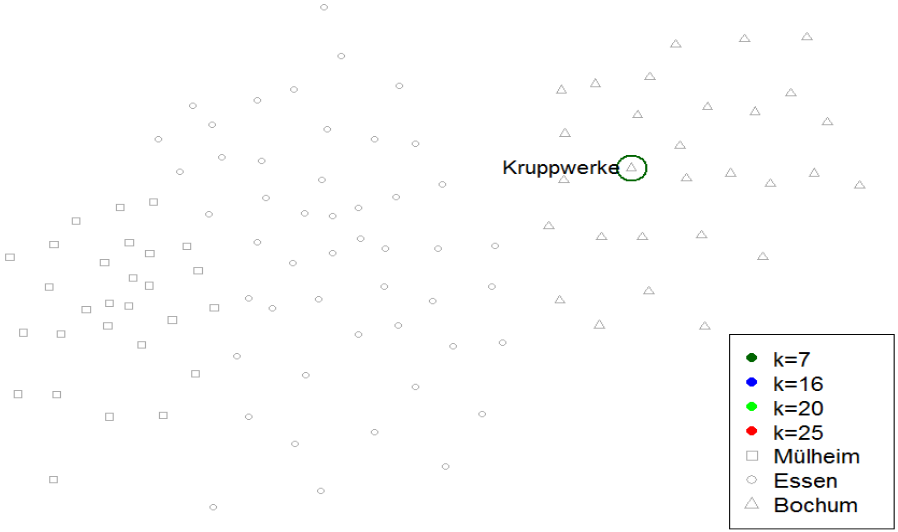
Local clusters Besag-Newell, first follow up



Local clusters second follow up



Local clusters third follow up



SIR

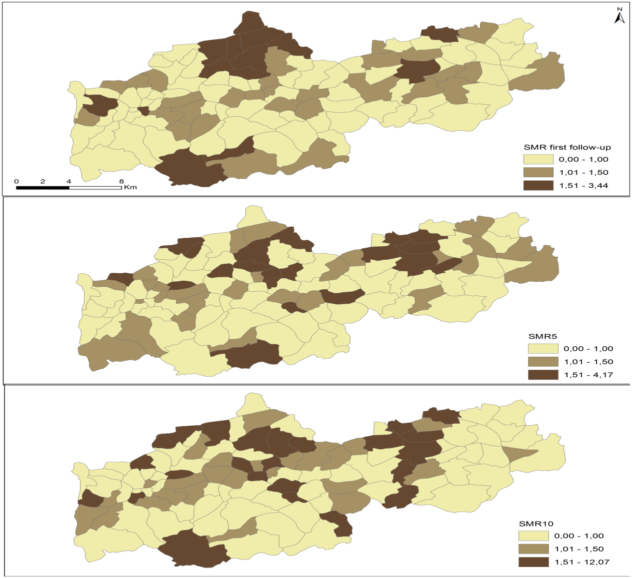


Figure: The standardized incidence rate

Smoothed (weighted) SIR

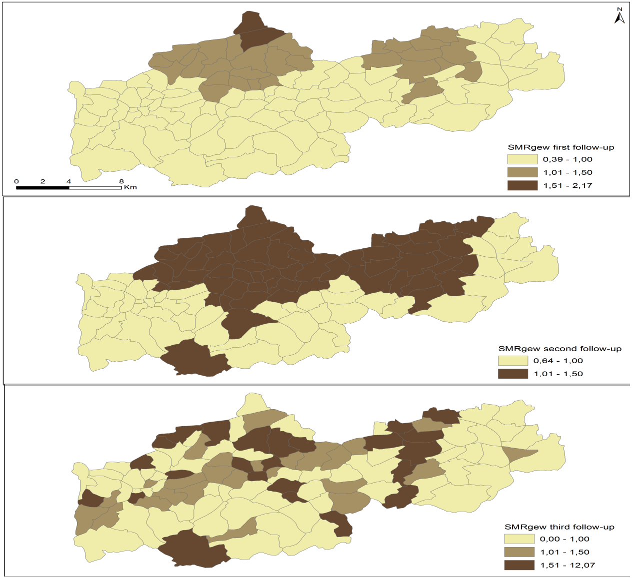
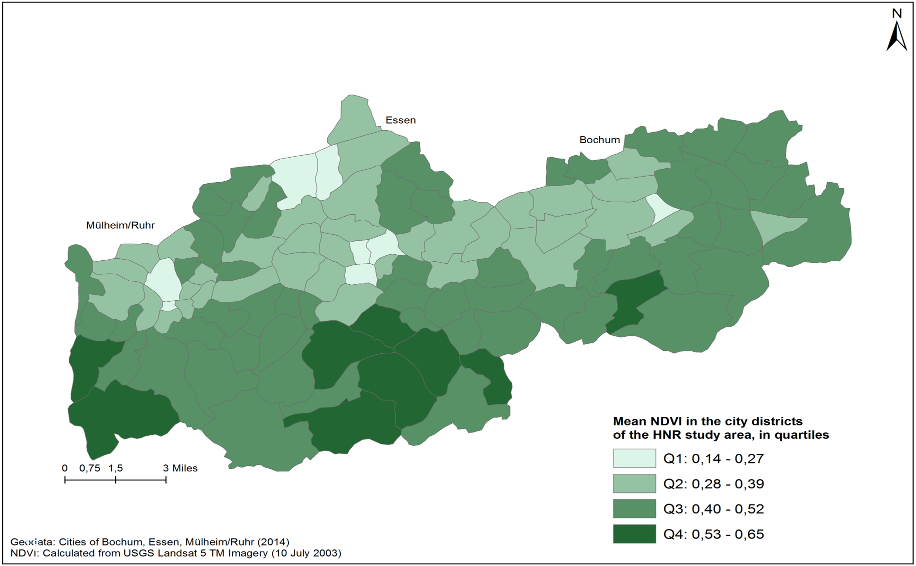


Figure: The standardized incidence rate, weighted smoothing

Spatial distribution of NDVI 2006



Traditional Poisson model:Besag York Molie

$$\begin{cases} Y_k & \sim Poisson(\theta_k E_k) \\ ln(\theta_k) & = X_k^T \beta + U_k + V_k \end{cases}$$

- $V_i \sim N(0, \tau_v^2)$: clustering in each spatial unit
- U_i connection between adjacent units: using CAR (**CARspatial**)

$$[U_i | U_i, j \neq i, \tau_u^2 \sim N(\bar{u}_i, \tau_i^2),$$

- $\bar{u}_i = \frac{\sum_{j=1}^{K} w_{ij} u_j}{\sum_{j=1}^{K} wij}$ and $\tau_i^2 = \frac{\tau_u^2}{\sum_{j=1}^{K} wij}$
- $W = (w_{ij})_{i,j=1\ldots K}$ = adjacent matrix
- $X_k^T \beta = V_k = U_k = 0 \Rightarrow$ Traditional Poisson model

Smoothed risk : Besag-York-Mollie (BYM) model

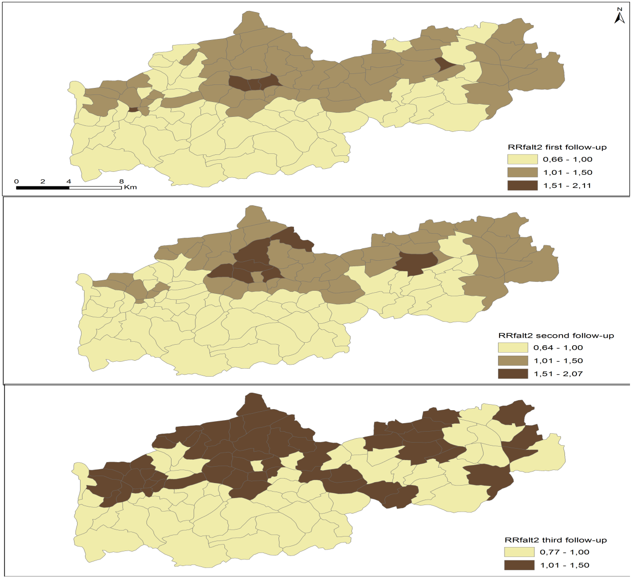


Figure: Risk estimate Besag-York-Mollie (BYM)

Model equation + spatio-temporal autocorrelation via random effects

$$\begin{cases} Y_{kt} & \sim Poisson(\theta_{kt} E_{kt}) \\ ln(\theta_{kt}) & = X_{kt}^T \beta + \boxed{\psi}_{kt} \\ \beta & \sim N(\mu_\beta, \Sigma_\beta) \quad \text{(prior for } \beta) \end{cases}$$

- $\mathbf{Y} = (\mathbf{Y}_1, \ldots, \mathbf{Y}_N)_{K \times N}$, $\mathbf{Y}_t = (Y_{1t}, \ldots, Y_{Kt})$ = the $K \times 1$ column vector of observations for all K spatial units for time period t

- θ_{kt} = risk (of depression) at time t in spatial unit k

- $\beta = (\beta_1, \ldots \beta_p)$: covariate regression parameters

- $\Psi = (\Psi_1, \ldots, \Psi_N)$, $\Psi_t = (\psi_{1t}, \ldots, \psi_{Kt})$

- ψ_{kt}: random component for areal unit k and time period t

CAR-Model for random effect ψ

$$\begin{cases} \psi_{kt} & = \phi_{kt} \\ \phi_t|\phi_{t-1} & \sim N(\rho_T\phi_{t-1}, \tau^2 Q(W,\rho_S)^{-1}), t = 2,\ldots,N \\ \phi_1 & \sim N(0, \tau^2 Q(W,\rho_S)^{-1}) \\ \tau^2 & \sim Inverse - Gamma(a,b) \\ \rho_S, \rho_T & \sim Uniform(0,1) \\ Q(W,\rho_S) & = \rho_S[diag(W.1 - W)] + (1-\rho_S)I \end{cases}$$

- ρ_S, ρ_T: spatial and temporal autoregressive parameter resp.
- $W = (w_{kj}) =$ neighborhood matrix, $w_{kj} =$ spatial closeness between the two areas
- $Q(W,\rho_S)$: precision matrix

Lee et al., *CARBayesST: Spatio-Temporal Generalised Linear Mixed Models for Areal Unit Data*,2017, https://CRAN.R-project.org/package=CARBayesST

Results Spatio-temporal model

model I	model I'	model II	model II'	model III	model III'
Greenness alone	model I with $\rho_S = 0$	model I + other covariates	model II with $\rho_S = 0$	ModelII + unemployment status	Model III Without spatial effect

	(ρ_T)		(ρ_S)		τ^2		NDVI	
	est.	95% credible interval	est.	95% credible interval	est.	95% credible interval	est.	95% credible interval
Model I	0.98	$(0.90, 0.99)$	0.05	$(0.004, 0.33)$	0.02	$(0.01, 0.05)$	0.91	$(0.86, 0.96)$
Model I'	0.96	$(0.87, 0.99)$	0	$(0.00, 0.00)$	0.02	$(0.01, 0, 03)$	0.91	$(0.86, 0.97)$
Model II	0.98	$(0.90, 0.99)$	0.05	$(0.004, 0.16)$	0.02	$(0.01, 0.03)$	0.91	$(0.86, 0.98)$
Model II'	0.97	$(0.91, 0.99)$	0	$(0.00, 0.00)$	0.02	$(0.01, 0.03)$	0.91	$(0.85, 0.98)$
Model III	0.98	$(0.90, 0.99)$	0.08	$(0.006, 0.28)$	0.02	$(0.01, 0.03)$	0.96	$(0.90, 1.01)$
Model III'	0.98	$(0.88, 0.99)$	0	$(0.00, 0.00)$	0.02	$(0.01, 0.03)$	0.97	$(0.91, 1.04)$

Conclusion and outlook

- Strong temporal trend
- Weak spatial trend, suggestive of neglecting it
- Greenness and depression negatively associated district unit

$\longrightarrow$ Random effects should be taken into account in observational studies when analysing health outcomes and environmental (risk)factors

- Data limitation and missing values (dropout)
- Spatial unit of analysis: appropriateness of aggregation
- Next step: Analysing both individual and district-level covariates for the risk estimate at individual level