UNIVERSITÄT BERN

Discrete and continuous domain models for disease mapping and applications on childhood cancers

Garyfallos Konstantinoudis

Institute of Social and Preventive Medicine,
University of Bern

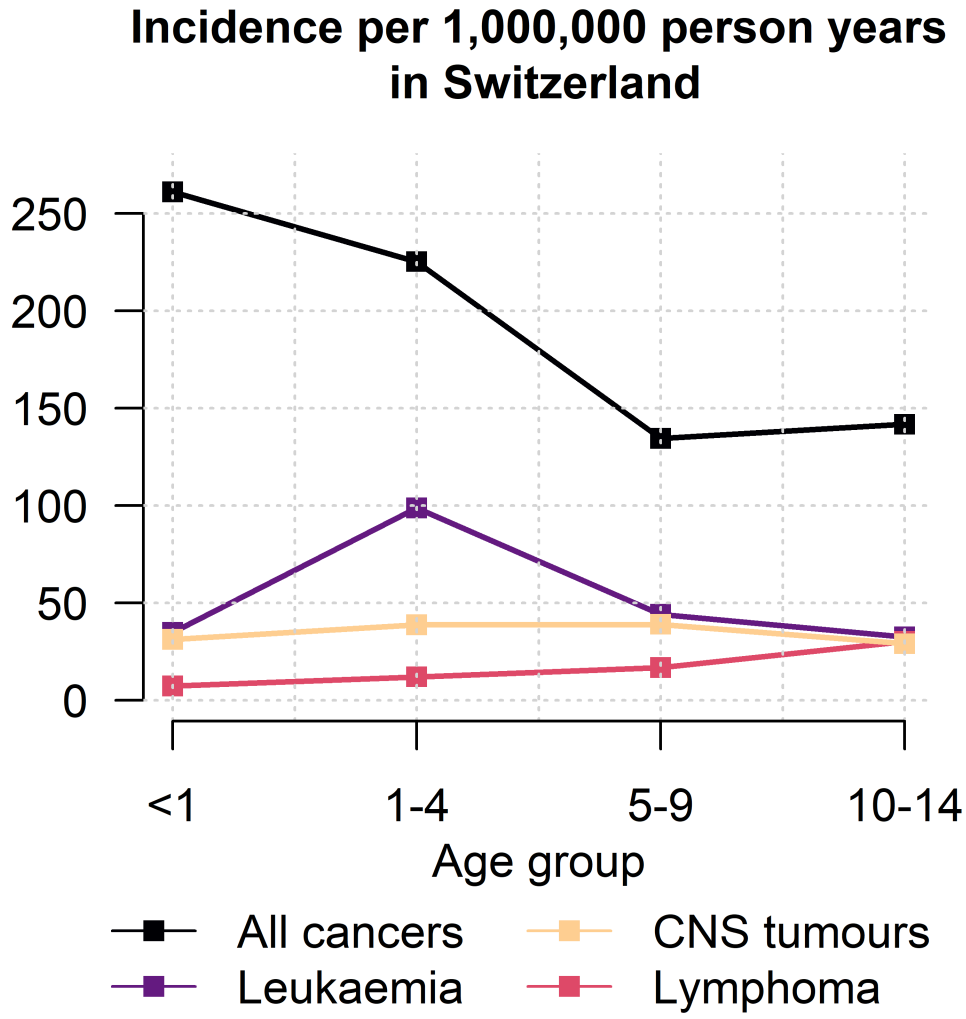
August 19, 2019

Outline

Background

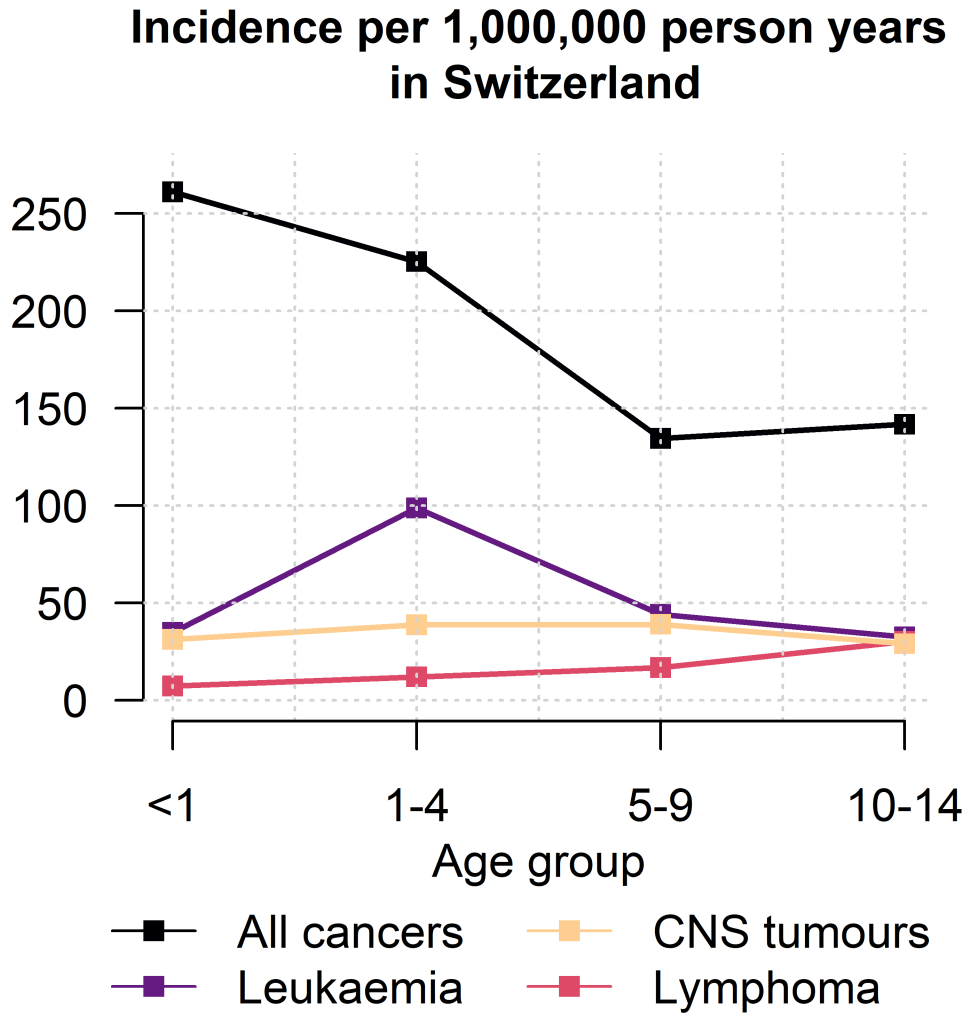
Childhood Cancers

- 200 children (0-14) diagnosed with cancer every years in Switzerland
- Ionizing radiation in high doses cause childhood cancer
- Hypothesised environmental risk factors: air-pollution, pesticide exposure, infections etc.



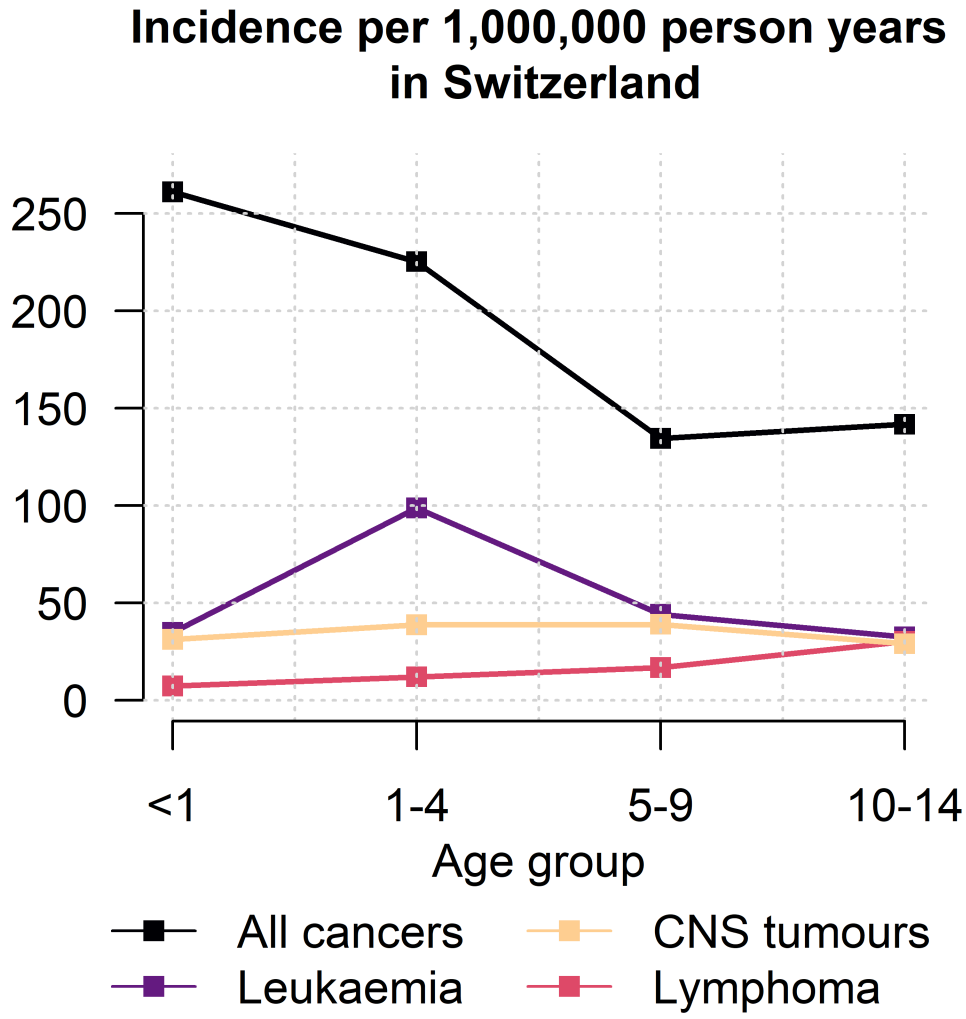
Childhood Cancers

- 200 children (0-14) diagnosed with cancer every years in Switzerland
- Ionizing radiation in high doses cause childhood cancer
- Hypothesised environmental risk factors: air-pollution, pesticide exposure, infections etc.



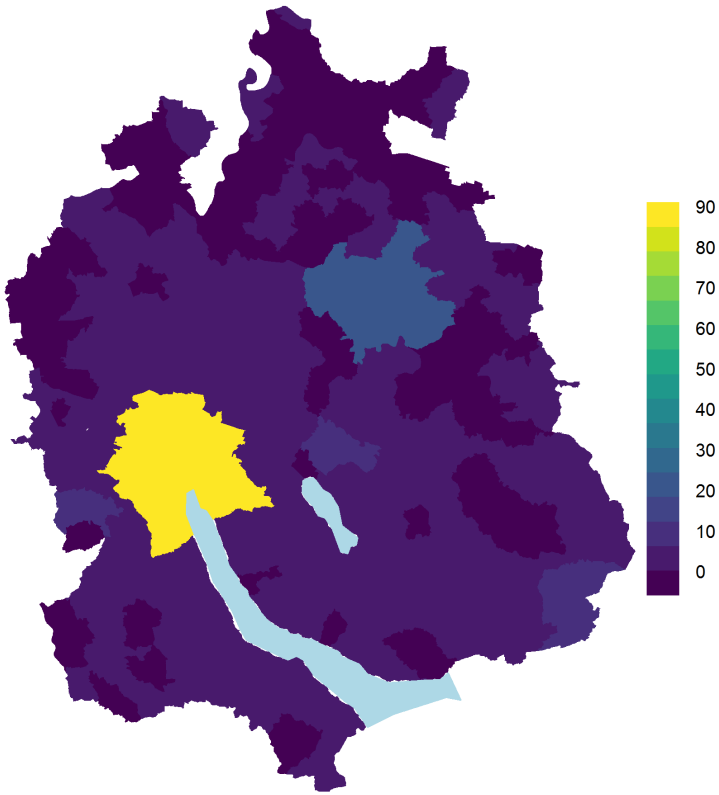
Childhood Cancers

- 200 children (0-14) diagnosed with cancer every years in Switzerland
- Ionizing radiation in high doses cause childhood cancer
- Hypothesised environmental risk factors: air-pollution, pesticide exposure, infections etc.



Data availability

A. Areal data

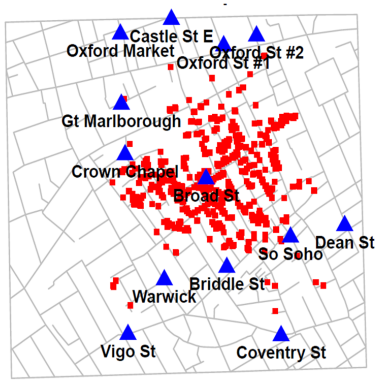


B. Point data

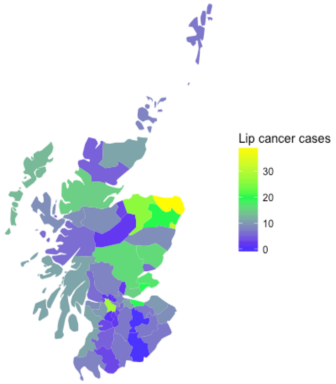
Available tools

Disease Mapping

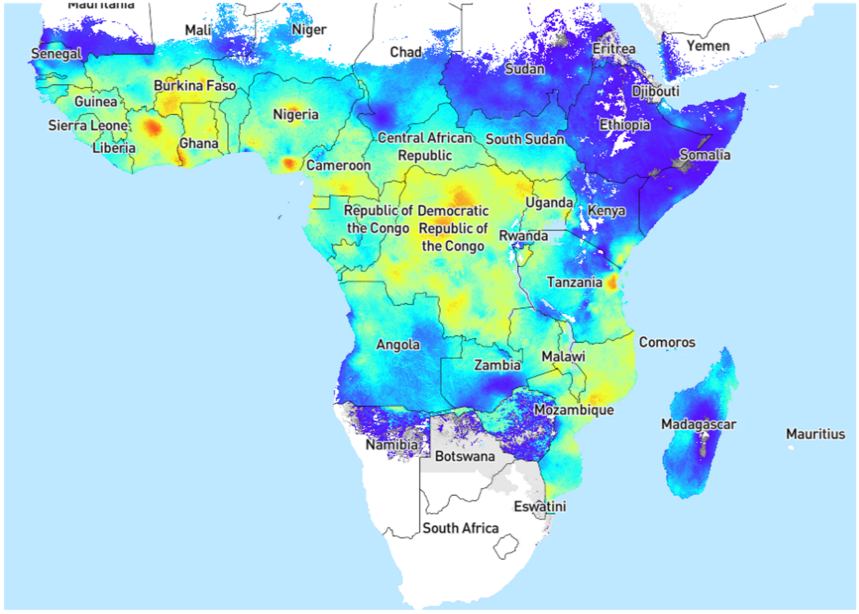
Snow's Cholera London



Scottish lip cancer

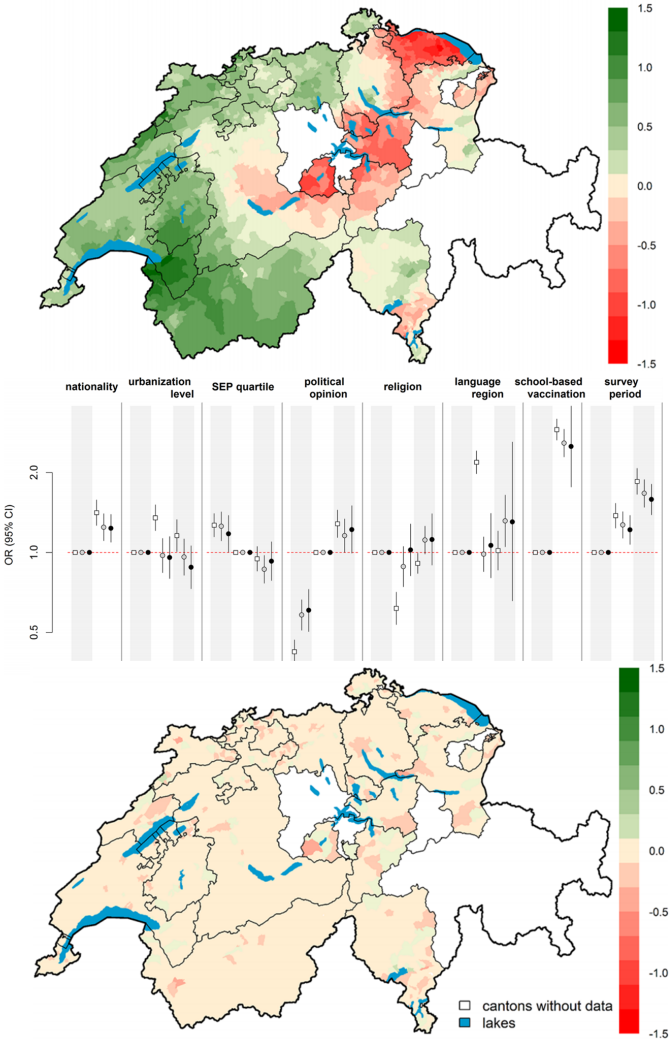


Malaria in Africa in 2007



Spatial Regression

HPV vaccination uptake in Switzerland



Previous studies

- Areal data: BYM
  - Besag *Ann Inst Statist Math* 1991
  - Acute Leukaemia in France (Faure *Eur J Cancer Prev* 2009), Childhood leukaemia and Type 1 Diabetes in Yorkshire (Manda *Eur J Epidemiol* 2009)
- Point data: Log Gaussian Cox Processes (LGCPs)
  - Møller et al. Scand J Stat 1998
  - Cancer mapping: Lung cancer in Spain (Diggle *Stat Sci* 2013), Colon and rectum in Minnesota (Liang *Ann Appl Stat* 2008)
  - none for childhood cancers
- Compared these methods:
  - Lung and stomach cancer (Li *J R Stat Soc C-Appl* 2012)
  - Syphilis (Li *Methods in Medical Research* 2012)
  - Cancer mortality (Kang *PLOS one* 2013)

Previous studies

- Areal data: BYM
  - Besag *Ann Inst Statist Math* 1991
  - Acute Leukaemia in France (Faure *Eur J Cancer Prev* 2009), Childhood leukaemia and Type 1 Diabetes in Yorkshire (Manda *Eur J Epidemiol* 2009)
- Point data: Log Gaussian Cox Processes (LGCPs)
  - Møller et al. Scand J Stat 1998
  - Cancer mapping: Lung cancer in Spain (Diggle *Stat Sci* 2013), Colon and rectum in Minnesota (Liang *Ann Appl Stat* 2008)
  - none for childhood cancers
- Compared these methods:
  - Lung and stomach cancer (Li *J R Stat Soc C-Appl* 2012)
  - Syphilis (Li *Methods in Medical Research* 2012)
  - Cancer mortality (Kang *PLOS one* 2013)

Previous studies

- Areal data: BYM
  - Besag *Ann Inst Statist Math* 1991
  - Acute Leukaemia in France (Faure *Eur J Cancer Prev* 2009), Childhood leukaemia and Type 1 Diabetes in Yorkshire (Manda *Eur J Epidemiol* 2009)
- Point data: Log Gaussian Cox Processes (LGCPs)
  - Møller et al. Scand J Stat 1998
  - Cancer mapping: Lung cancer in Spain (Diggle *Stat Sci* 2013), Colon and rectum in Minnesota (Liang *Ann Appl Stat* 2008)
  - none for childhood cancers
- Compared these methods:
  - Lung and stomach cancer (Li *J R Stat Soc C-Appl* 2012)
  - Syphilis (Li *Methods in Medical Research* 2012)
  - Cancer mortality (Kang *PLOS one* 2013)

BYM model

Let $\mathcal{W}$ an observation window, $A_1, ..., A_N$ a partition of $\mathcal{W}$, Y_i be the disease counts P_i the population and λ_i the risk in A_i:

$$Y_i|\lambda_i, P_i \sim \text{Poisson}(\lambda_i P_i)$$

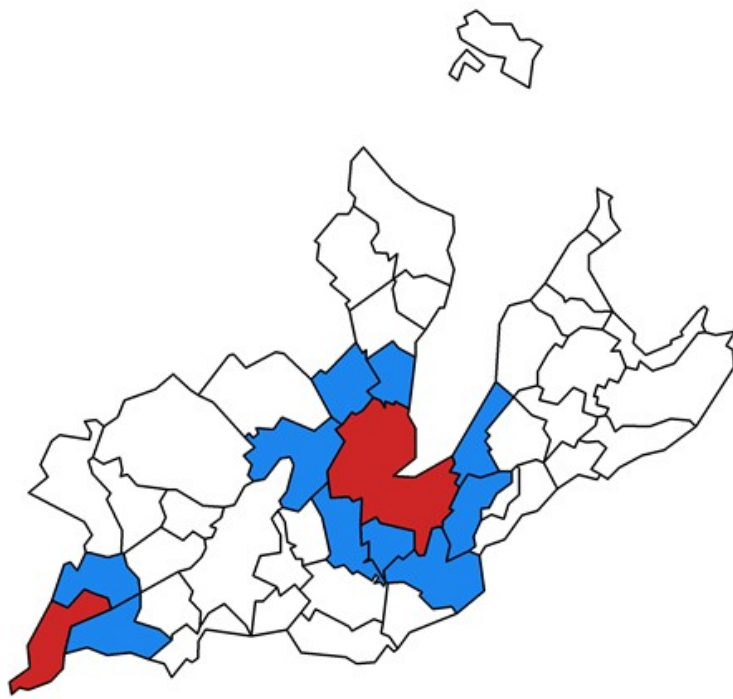
$$\log(\lambda_i) = \beta_0 + z_i^T \beta + u_i + v_i$$

$$u_i|\mathbf{u}_{-i} \sim \mathcal{N}\Big(\frac{\sum_{j=1}^{N} w_{ij} u_j}{\sum_{j=1}^{N} w_{ij}}, \frac{1}{\tau_1 \sum_{j=1}^{N} w_{ij}}\Big)$$

$$v_i \sim \mathcal{N}(0, \tau_2^{-1})$$

$$\beta_0, \beta \sim \mathcal{N}(0, \sigma^2)$$

$$\tau_1, \tau_2 \sim \text{PCpriors}$$



Log-Gaussian Cox Processes

Let $\mathcal{W}$ an observation window and Ξ a point process with intensity $\lambda(s)$ on the s location:

$$\Xi|\lambda(s) \sim \text{Poisson}(\int_{\mathcal{W}} \lambda(s)ds)$$

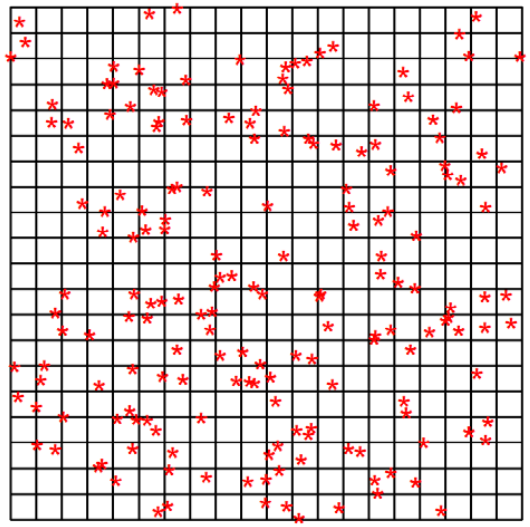
$$\log(\lambda(s)) = log(\lambda_0(s)) + \beta_0 + z(s)^T\beta + u(s)$$

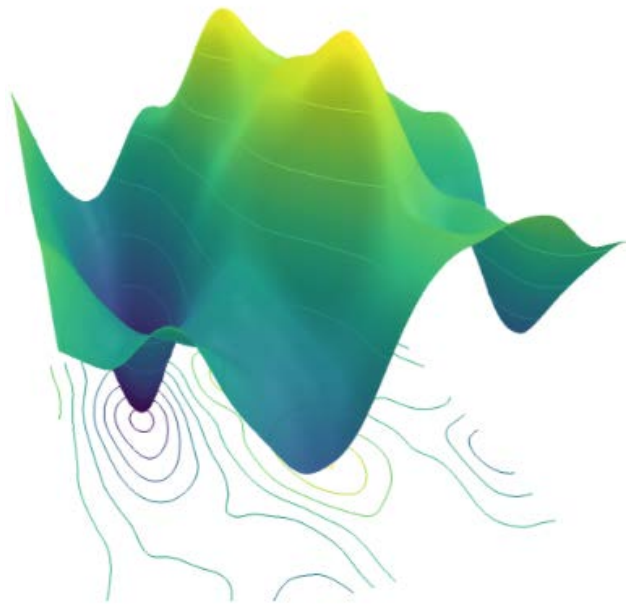
$$u(s) \sim \text{GF}(0, \mathbf{\Sigma}(h, \tau, \phi))$$

$$\kappa(h) = \tau^2\rho_\nu(h/\phi), \rho_\nu(\cdot) \text{ Matérn}$$

$$\beta_0, \beta \sim \mathcal{N}(0, \sigma^2)$$

$$\tau, \phi \sim \text{PCpriors}$$





Simulation study

Simulation Study

- Canton of Zurich
- $N = 205,242$ (15%) children
- Leukaemia incidence 1985-2015 ($n = 334$)

Radius	RR	times n	decay
1km	2	1	step function
5km	5	5	smooth function
10km	-	10	-

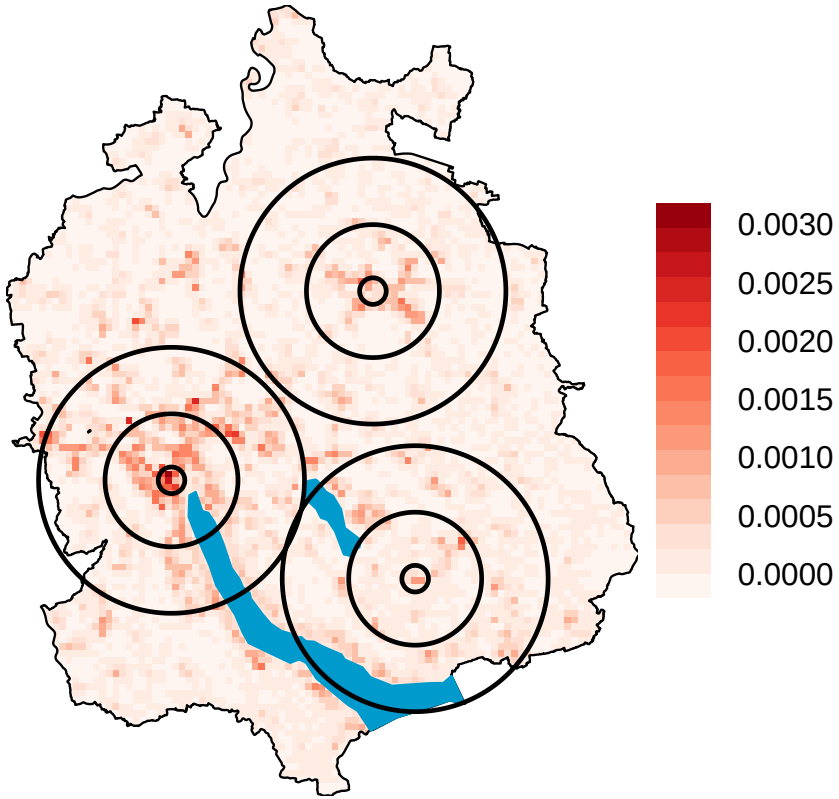


Figure: Population density and circles

Simulation Study; Metrics

Quantify the risk in space

- Root mean integrated square error (RMISE)

$$\text{RMISE} = \left(\mathbb{E} \int_{\mathcal{W}} b(s)(log(\hat{\lambda}(s)) - log(\lambda(s)))^2 ds \right)^{1/2} =$$

$$\left(\mathbb{E} \sum_{g=1}^{G} b_g |D_g| (log(\hat{\lambda}_g) - log(\lambda_g))^2 \right)^{1/2}$$

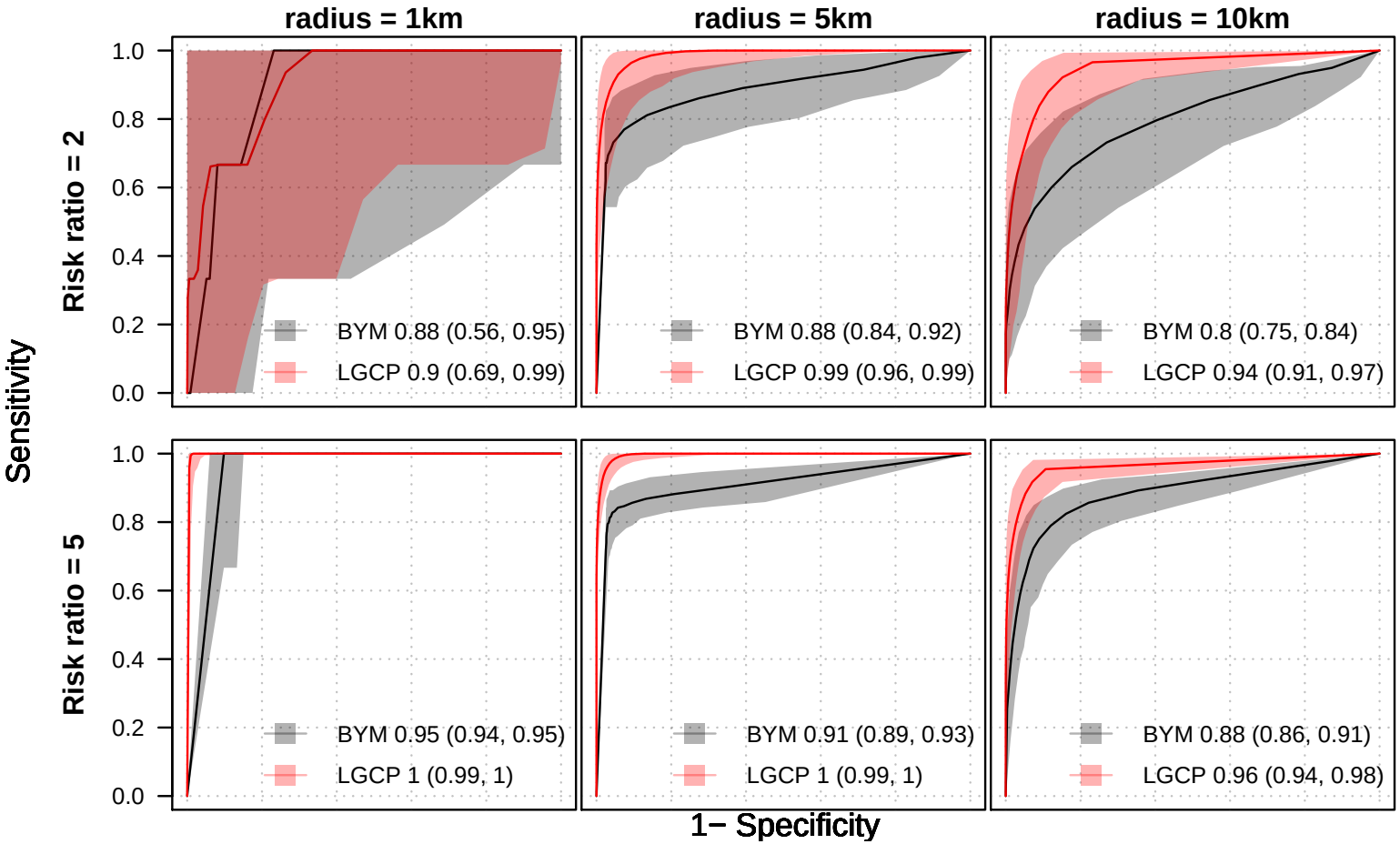
Identify high-risk areas

- Sensitivity, Specificity and area under the curve (AUC)

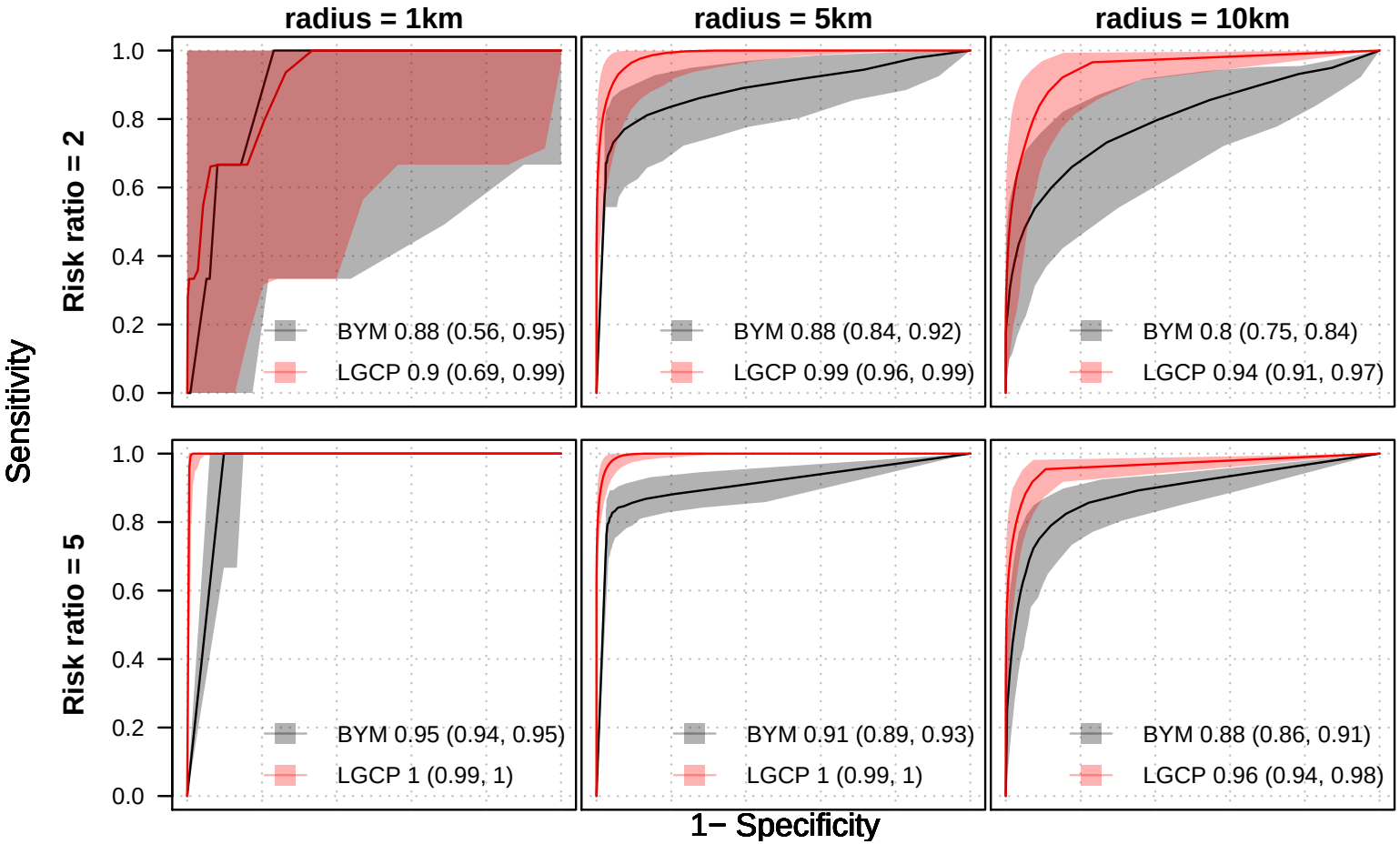
Results in a nutshell; RMISE, $5n$

	BYM	LGCP
Step function		
Radius = 1km		
RR = 2	4.47 (3.17, 6.81)	6.62 (4.24, 9.88)
RR = 5	10.4 (8.77, 12.5)	14.8 (13.1, 17.1)
Radius = 5km		
RR = 2	11.6 (10.6, 13.1)	12.2 (10.8, 14.7)
RR = 5	22.8 (21.4, 24.5)	21.5 (19.6, 24.6)
Radius = 10km		
RR = 2	14.9 (14.3, 15.8)	12.1 (11, 14.4)
RR = 5	28.4 (27.3, 29.8)	22.3 (20.8, 24.6)
Smooth function		
Radius = 1km		
RR = 2	4.48 (3.1, 6.88)	6.51 (4.27, 9.9)
RR = 5	10.8 (8.82, 12.5)	14.8 (13, 16.8)
Radius = 5km		
RR = 2	10.4 (9.32, 12)	11 (9.33, 14.3)
RR = 5	19.2 (18, 20.6)	16.8 (14.8, 19.9)
Radius = 10km		
RR = 2	12.3 (11.5, 13.4)	10.1 (8.57, 12.7)
RR = 5	21.8 (21, 22.8)	13.9 (12.1, 17)

Results in a nutshell; ROC-curves, Step-function, $5n$



Results in a nutshell; ROC-curves, Step-function, $5n$



Clear Message

There are important gains to be made from the use of LGCPs

Application on childhood cancers

Aims

1. Create smooth maps of childhood cancer risk in Switzerland (disease mapping)
2. Examine the sources of the observed spatial variation (spatial regression)
3. Assess the residual spatial variation

Methods; Population

Cases

- Children aged 16 $<$ years at diagnosis
- Registered in SCCR
- Diagnosed in Switzerland during 1985-2015

Outcomes

- All cancers, Leukaemia, Lymphoma, CNS tumours
- Address at birth and diagnosis

Population at risk

- Entire Swiss population from censuses (1990, 2000, 2010 onwards)
- Calculate expected number of cases per grid cell to adjust for age, population density and year of diagnosis

Methods; Population

Cases

- Children aged 16 $<$ years at diagnosis
- Registered in SCCR
- Diagnosed in Switzerland during 1985-2015

Outcomes

- All cancers, Leukaemia, Lymphoma, CNS tumours
- Address at birth and diagnosis

Population at risk

- Entire Swiss population from censuses (1990, 2000, 2010 onwards)
- Calculate expected number of cases per grid cell to adjust for age, population density and year of diagnosis

Methods; Population

Cases

- Children aged 16 $<$ years at diagnosis
- Registered in SCCR
- Diagnosed in Switzerland during 1985-2015

Outcomes

- All cancers, Leukaemia, Lymphoma, CNS tumours
- Address at birth and diagnosis

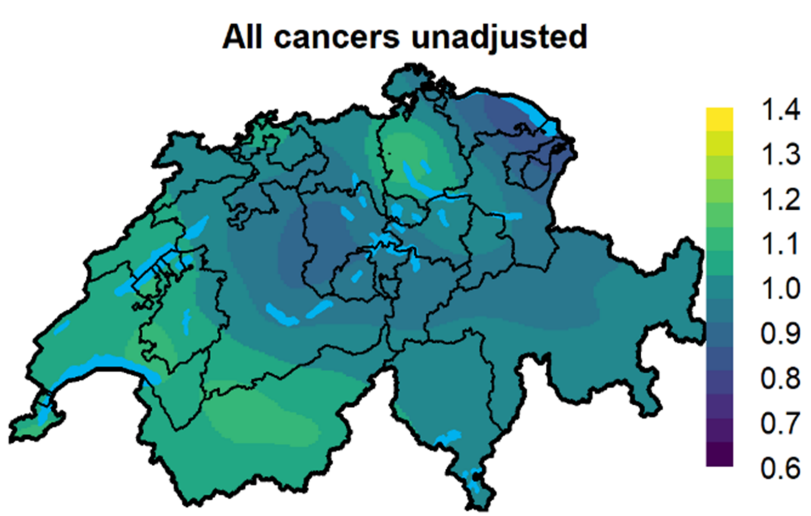
Population at risk

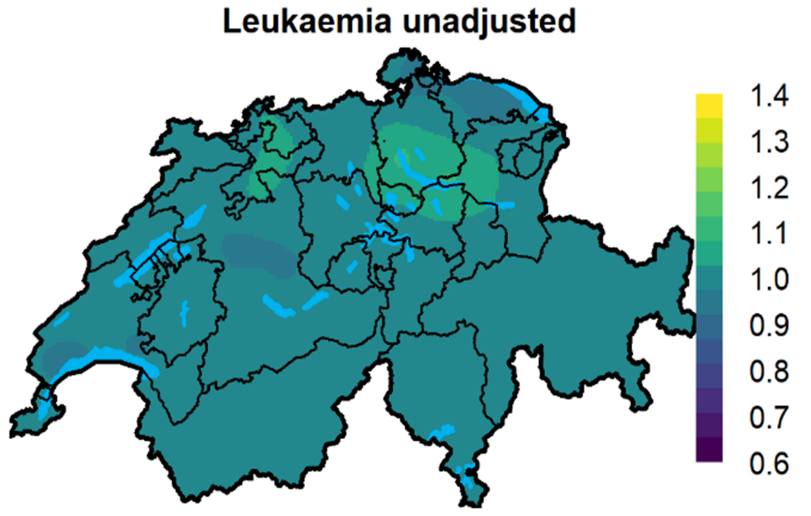
- Entire Swiss population from censuses (1990, 2000, 2010 onwards)
- Calculate expected number of cases per grid cell to adjust for age, population density and year of diagnosis

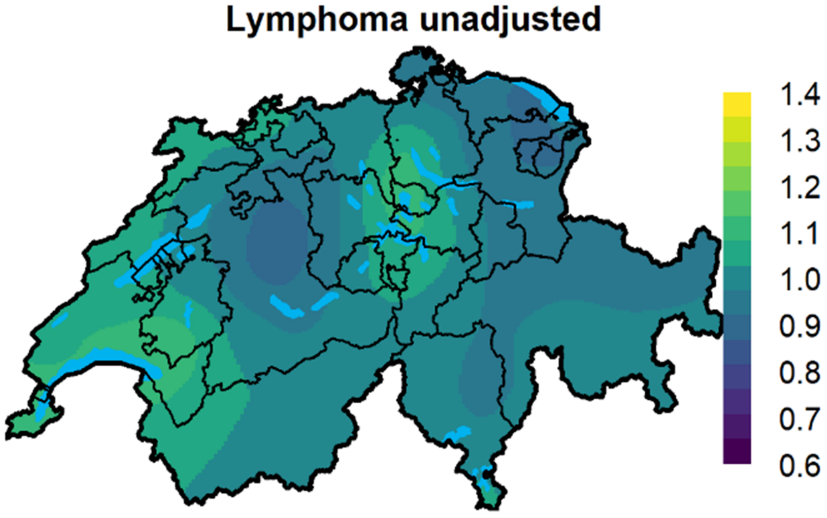
Methods; Covariates

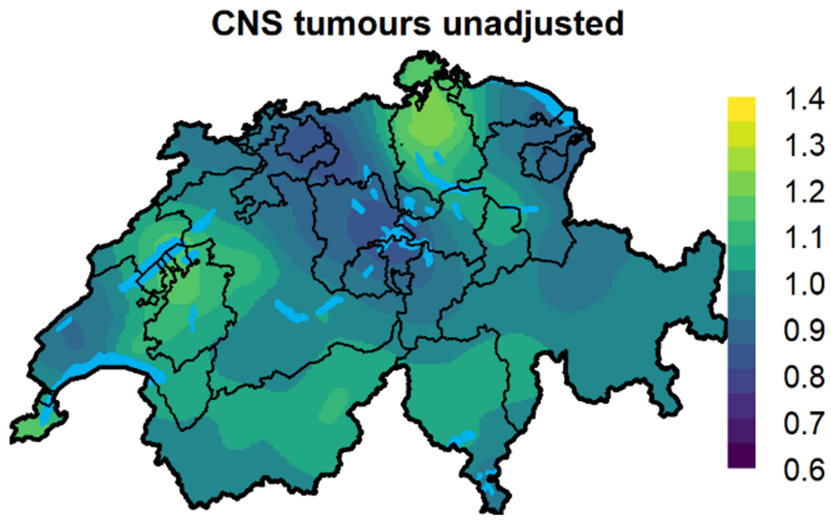
Variable	Unit	Type	Spatial unit	Year
NO_2	$\mu g/m^3$	Continuous	200x200m^2	1990, 2000 2010
Total radiation	nSv/h	Continuous	2x2km^2	1960-95
Swiss-SEP	index	Continuous	Building level	2000
Cantonal registry	Years [y]	Continuous	canton	2015
Language region	-	Ge, Fr, It	Municipality	2012
Urbanisation level	-	rural, semi, urban	Municipality	2012

Results; Maps LGCPs

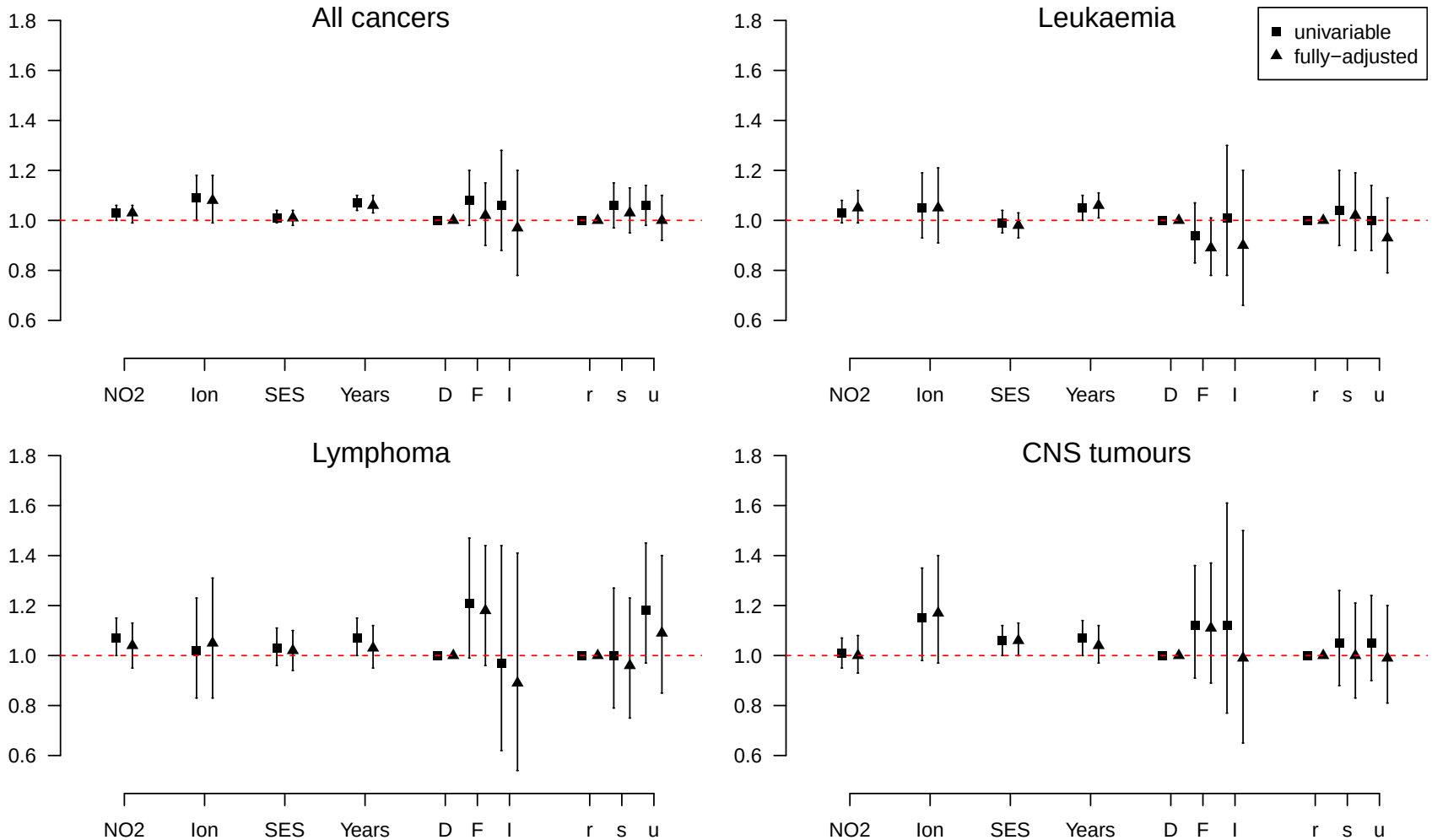




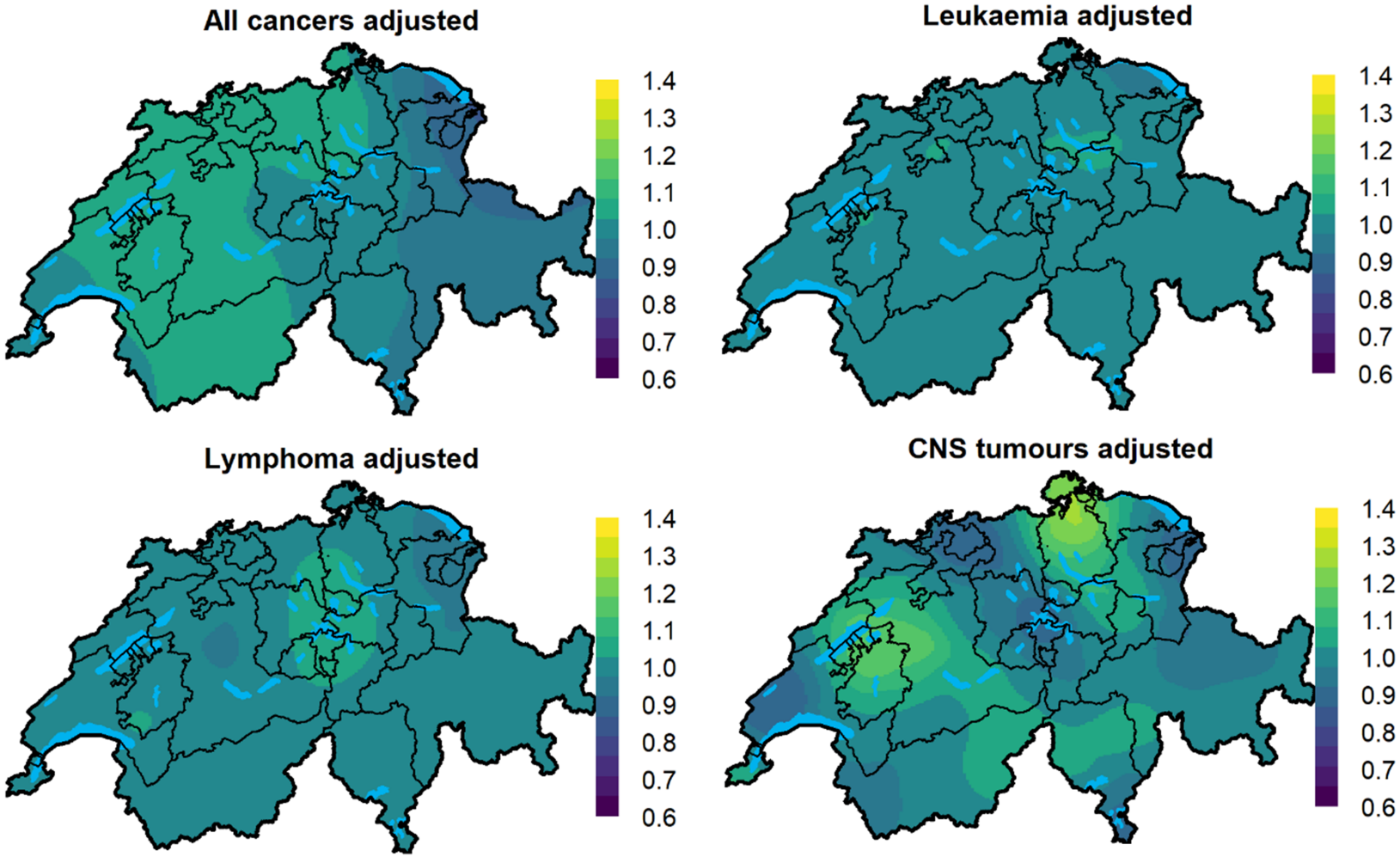




Results; Association at diagnosis



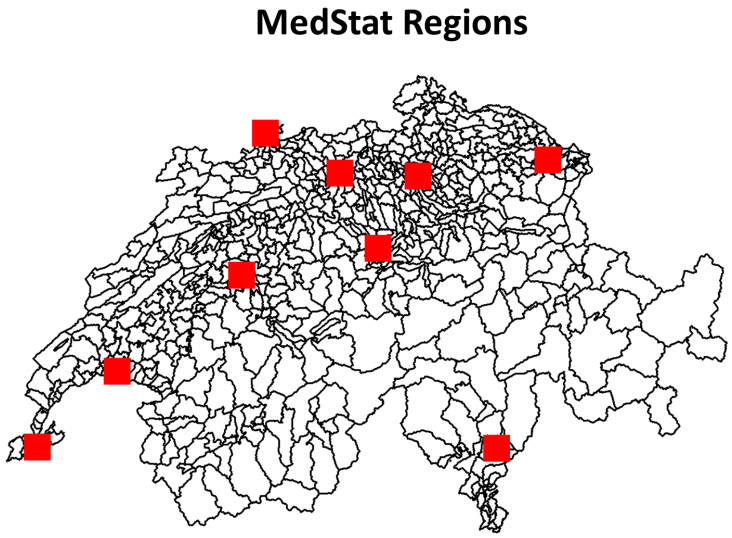
Results; Residual variation at diagnosis

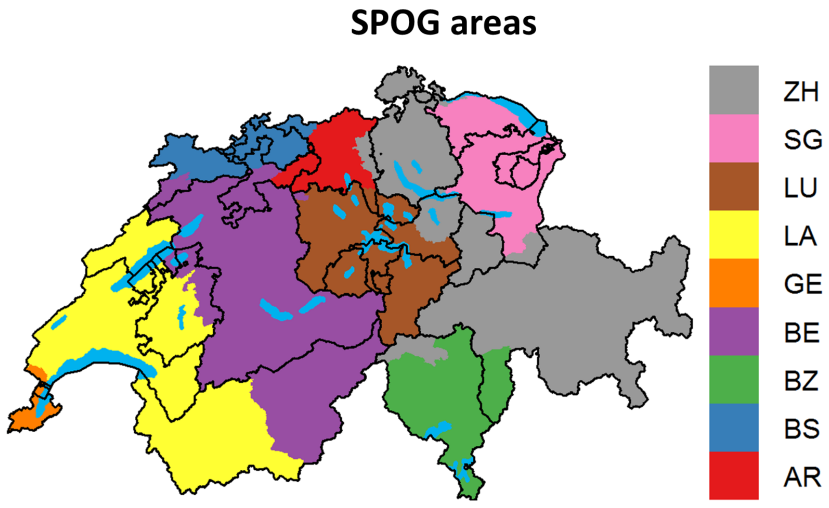


Post-hoc analyses: CNS tumours

Differences in case ascertainment:

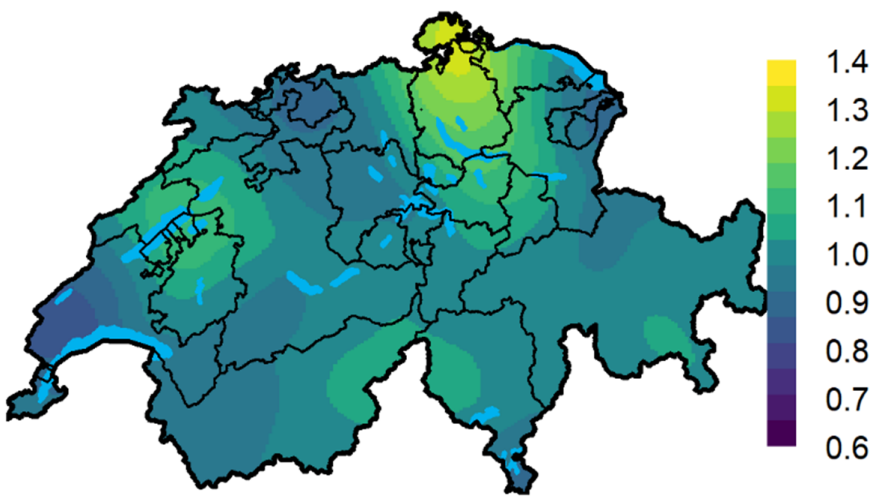
- Restrict to 1995-2015 ($>$ 95% completeness)
- SPOG clinics areas
  - Medstat regions
  - Proportion of cases reported in j-th SPOG clinic
  - Flag area based on max probability
  - Denoise based on the neighboring areas



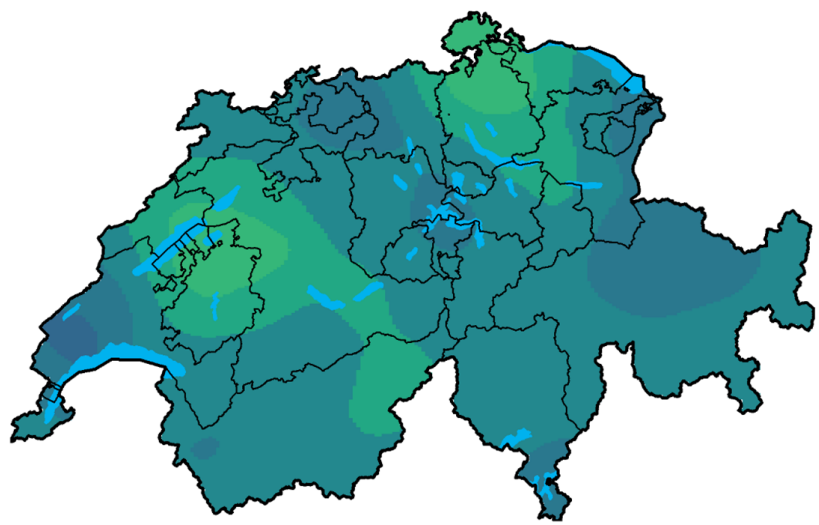


Post-hoc analyses: CNS tumours

1995-2015



SPOG catchment areas



Take home message

- Use LGCPs when precise data is available
- Childhood cancers and in particular CNS tumours vary in space

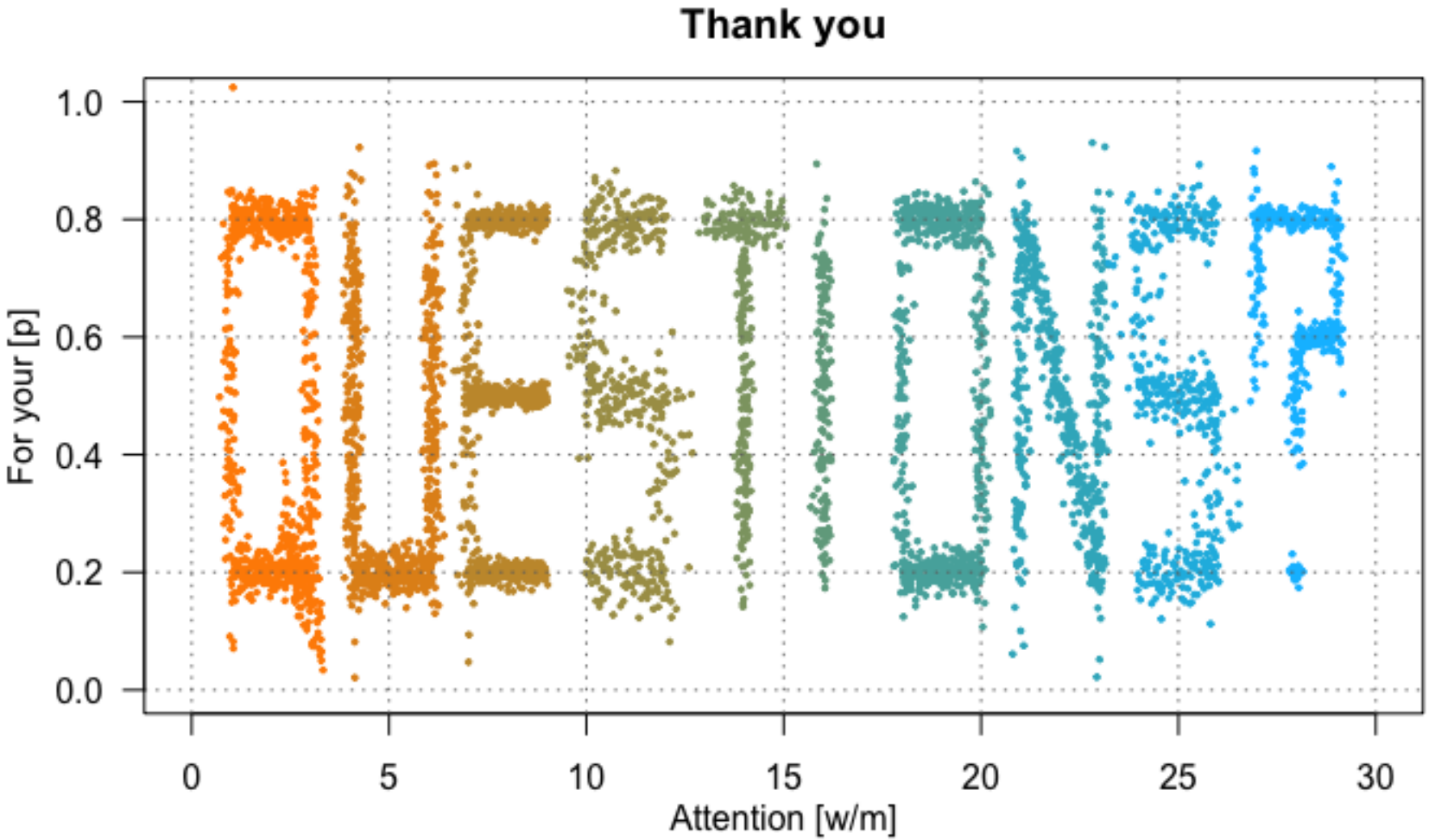
References

Konstantinoudis G, Schuhmacher D, Ammann R, Diesch T, Kuehni C, Spycher D. **Bayesian spatial modelling of childhood cancer incidence in Switzerland using exact point data: A nationwide study during 1985-2015**. medRxiv 19001545. https://doi.org/10.1101/19001545. 2019

Konstantinoudis G, Schuhmacher D, Rue H, Spycher B. **Discrete versus continuous domain models for disease mapping**. arXiv preprint arXiv:1808.04765. 2018

Email : `garyfallos.konstantinoudis@ispm.unibe.ch`

Twitter : `@konstantinoudis`


Thank you
For your [p]
1.0
0.8
0.6
0.4
0.2
0.0
0
5
10
15
20
25
30
Attention [w/m]