# Accounting for residential history in disease mapping

**Christel Faes**

I-Biostat, Hasselt University

Interuniversity Institute for Biostatistics
and statistical Bioinformatics

26 November 2018

In collaboration with: Petrof, O., Neyens, T., Nuyts, V., Nackaerts, K., Nemery, B.

# Introduction

- Interest in: geographical variation of disease risk
- Disease mapping
- Bayesian hierarchical modelling
- Aggregate number of cases per area for certain time period
- Area: residential location at time of diagnosis
- Valid if disease has long latency period?

# Case study

- Mesothelioma is a rare and aggressive type of cancer
- Exposure to asbestos
- Latency period of 20 to 40 years
- Long history of asbestos use in Belgium
- 2,076 (male) mesothelioma patients who died between 2004 and 2015

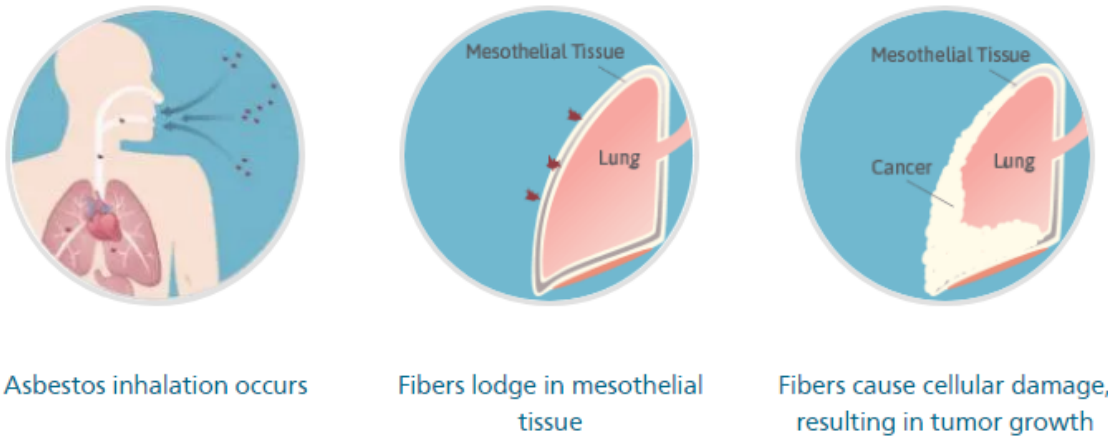



Figure 1: *The process of mesothelioma cancer exposure and development in the case of pleural mesothelioma.*

# Standard disease mapping model

- $Y_i$: observed number of cases in residential area $i$
- $e_i$: expected number of cases (according to standard population)
- Convolution model, BYM model (Besag, 1995)

$$Y_i \sim Poisson(e_i\theta_i),$$

$$log(\theta_i) = \alpha + v_i + u_i,$$

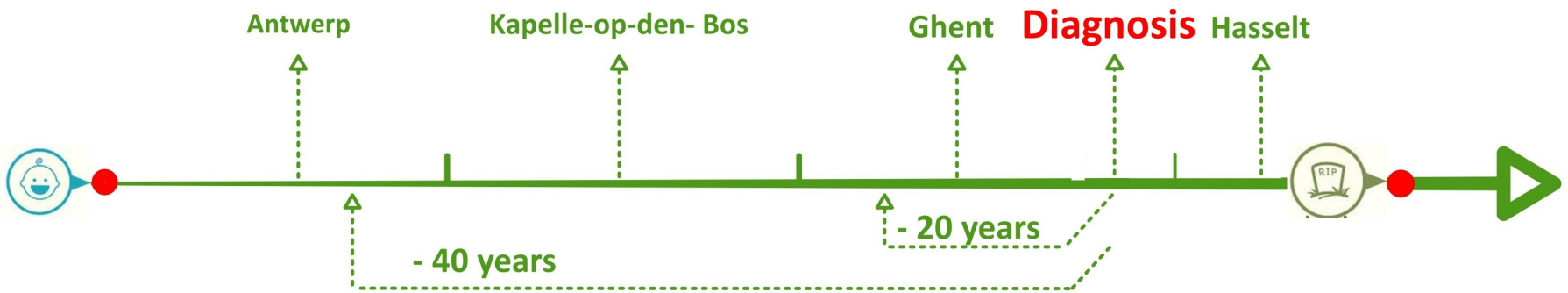
where

- $\alpha$ is an overall level of the relative risk
- $v_i$ is an uncorrelated heterogeneity factor $v_i \sim N(0, \sigma_v^2)$
- $u_i$ is a spatially correlated heterogeneity factor $u_i|u_k \sim N(\bar{u}_i, \sigma_i^2)$

Lawson, Banerjee, Haining, Ugarte (2016) *Handbook of spatial epidemiology*, CRC Press.
Besag and Kooperberg (1995) On conditional and intrinsic autoregression, *Biometrika*, 82(4).

# Challenge

- Long latency period of disease
- Mobility of patients



# Mobility of patients

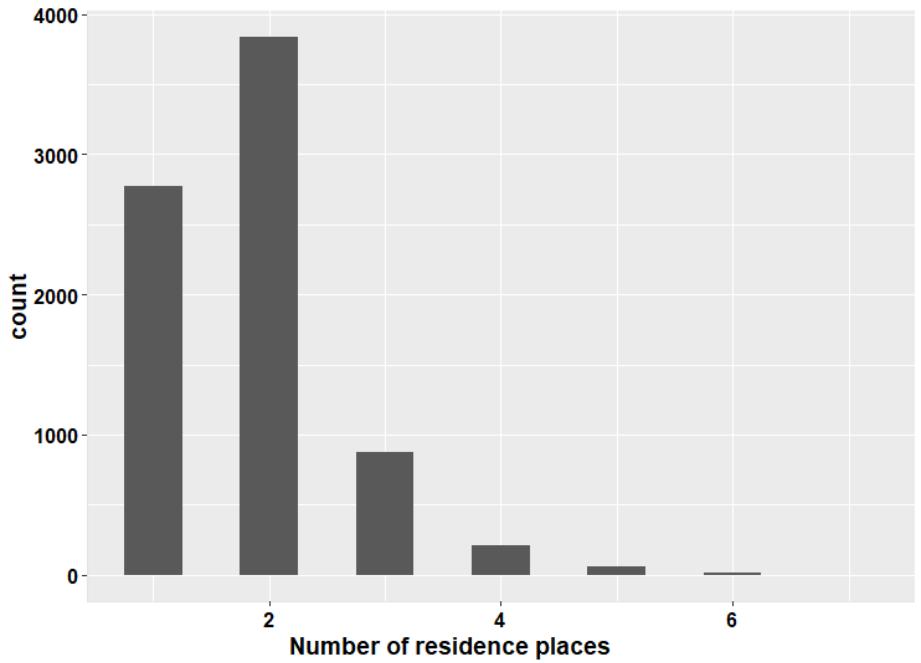



Figure 2: *The total number of residence places for males.*

# Case Study

- Case - control study
- Information on residential history of patients
- Control disease: Pancreatic cancer
  - no evidence of an environmental link
  - has the same population at risk as mesothelioma
  - 5,689 pancreatic cancer cases
  - patients who died between 2004 and 2015

Data from the Belgian Cancer Registry

# Case Study

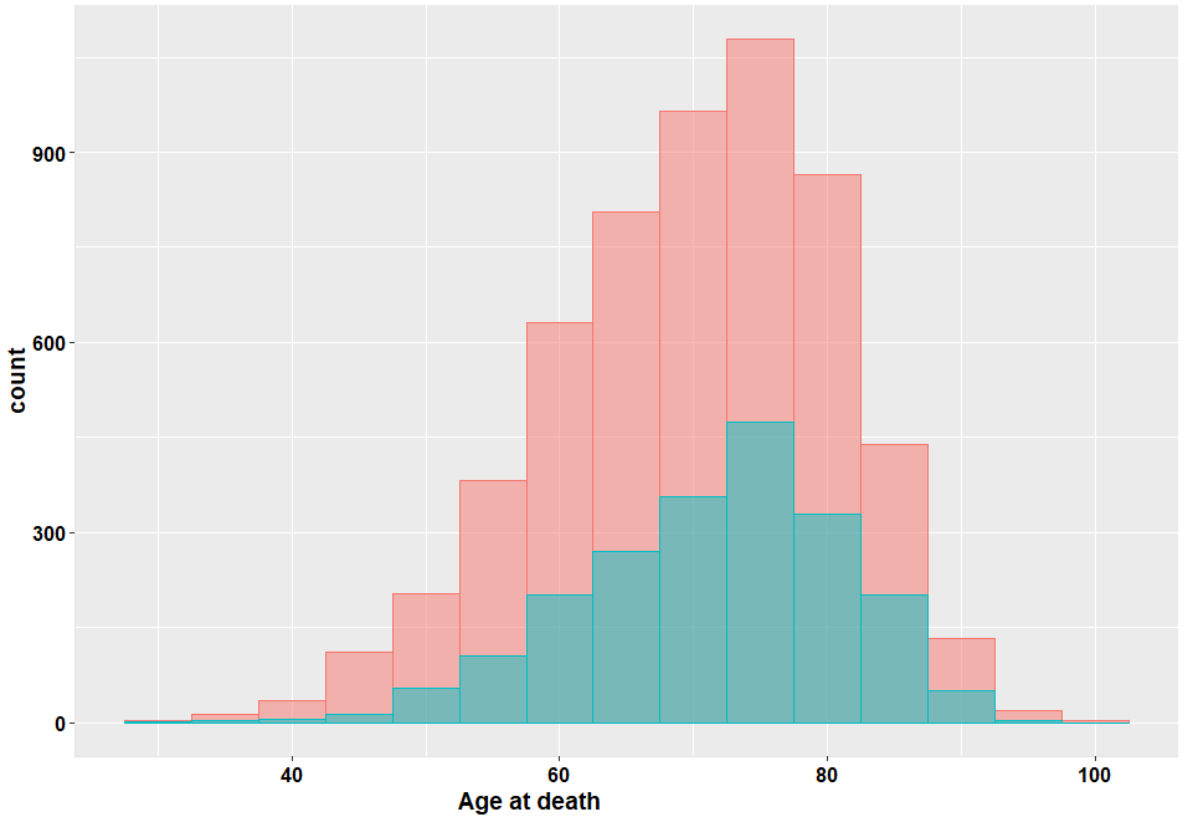



Figure 3: *Age at death for mesothelioma (blue) and pancreatic cancer cases (pink).*

# Model Specification: one-location model

- $Y_i = \begin{cases} 1 & \text{if the patient is a case} \\ 0 & \text{if the patient is a control.} \end{cases}$
- $x$ is the location of the patient
- We assume the model

$$Y_i \sim Bernoulli(\pi_g(x)),$$

with the probability that an event at location $x$ is a case is given by

$$\pi_g(x) = \frac{\rho_g f(x)}{1 + \rho_g f(x)}$$

  - $\rho_g$ reflects the overall prevalence of the disease of interest relative to the control disease,
  - $f(x)$ describes the elevation in risk as a function of the residential location.
- This is similar to assuming an inhomogeneous Poisson process with intensity function $\lambda_1(x) = \rho_g \lambda_0(x) f(x)$

Diggle, Morris, Elliott, Shaddick (1997) Regression modelling of disease risk in relation to point sources, *J R Stat Soc A*, 160(3).

# Model Specification: one-location model

Latent process model:

$$logit(\pi_g(x)) = \log(\rho_g) + \log(f(x))$$
$$= \left(\alpha + \sum_{k=1}^{4} \gamma_k * Age_k\right) + (v_k + u_k),$$

$$v_k \sim N(0, \sigma_v^2)$$

$$[u_k | u'_k, k' \neq k, \sigma_u^2] \sim N(\bar{u}_k, \sigma_k^2),$$

$$\bar{u}_k = \frac{1}{n_k} \sum_{k' \sim k} u_{k'}$$

where $n_k$ is the number of neighboring municipalities.

- Special cases follow
- Uninformative hyperpriors assumed - MCMC

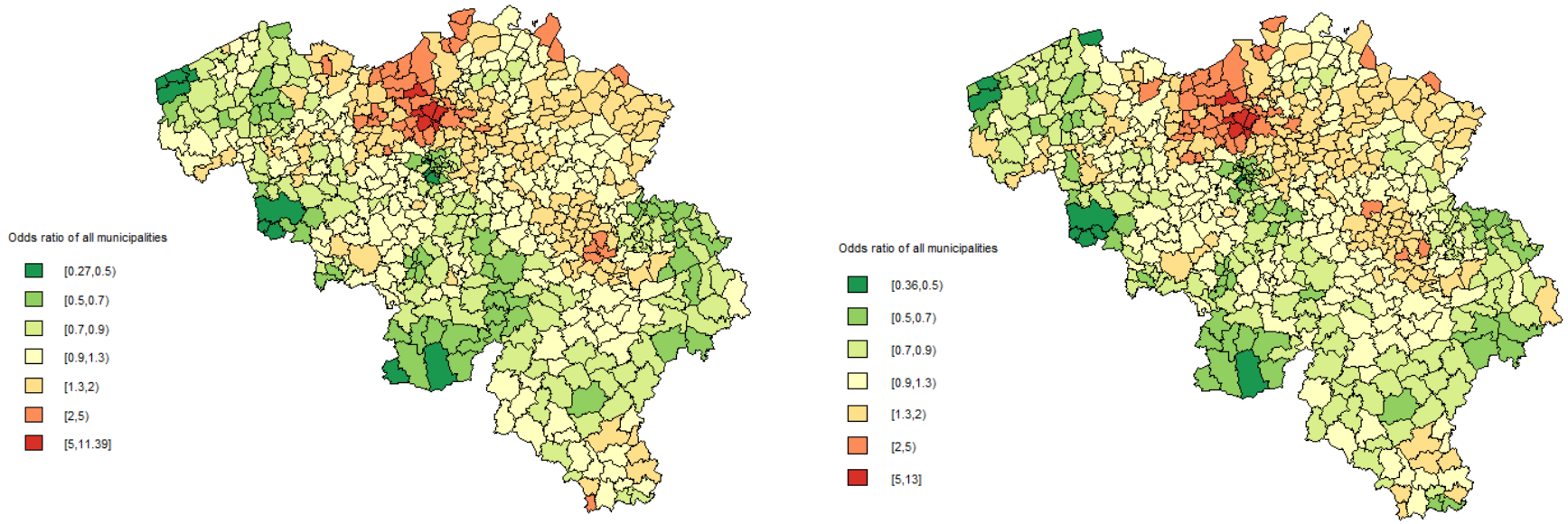



Figure 4: *Maps of the odds ratio for the two level convolution model based on the last residential location (left) and 20 years before diagnosis (right).*

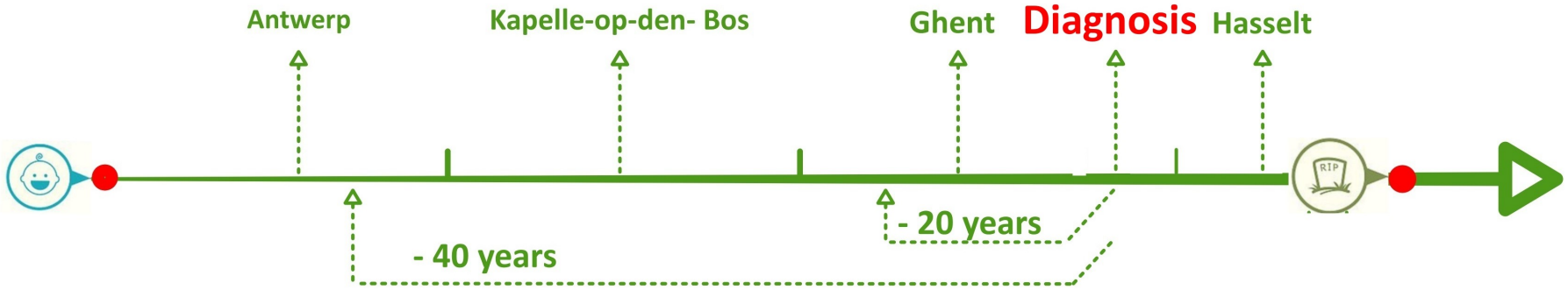
# Model Specification: multiple-membership model

- Previous model: pure hierarchical structure
- Asbest exposure took place 20 to 40 years before diagnosis
- More complex data structure
  - Many patients lived in multiple municipalities
  - Time spent in municipalities might be different
  - Residential history is patient-specific
- Use of multiple-membership model

Goldstein (2011) *Multilevel Statistical Models*, John Wiley & Sons

# Model Specification: multiple-membership model

Definition weight $w_{k,i}$ according to the proportion of time a patient $i$ lived in area $k$ between 20 up to 40 years prior to diagnosis:




- Every patient $i$ has its own weights
- $\sum_k w_{k,i} = \sum_{k \in \mathcal{H}_{(i)}} w_{k,i} = 1$
- Many $w_{k,i} = 0$

# Model Specification: multiple-membership model

We propose the use of the following multiple membership model:

$$Y_i|\mathcal{H}_i \sim Bernoulli\left(\pi_i(\mathcal{H}_i)\right),$$

$$\frac{\pi_i(\mathcal{H}_i)}{1-\pi_i(\mathcal{H}_i)} = \rho_g \prod_{x\in\mathcal{H}_i} f(x)^{w_{x,i}}$$

Depending on the form assumed for $f(x)$, this gives rise to different models.

# Model Specification: multiple-membership model

| $f(x)$ | Equation | Model |
|---|---|---|
| $\exp(v_k)$ | $logit\left(\pi_i(\mathcal{H}_i)\right) = \log(\rho_g) + \sum_{k\in\mathcal{H}_i} w_{k,i} v_k$ | Unstructured multiple membership (MMM) |
| $\exp(u_k)$ | $logit\left(\pi_i(\mathcal{H}_i)\right) = \log(\rho_g) + \sum_{k\in\mathcal{H}_i} w_{k,i} u_k$ | CAR multiple membership (CAR MMM) |
| $\exp(v_k + u_k)$ | $logit\left(\pi_i(\mathcal{H}_i)\right) = \log(\rho_g) + \sum_{k\in\mathcal{H}_i} w_{k,i}(u_k + v_k)$ | Convolution multiple membership type I (Convolution MMM type I) |
| $\exp(v_k)\exp(u_k)$ | $logit\left(\pi_i(\mathcal{H}_i)\right) = \log(\rho_g) + \sum_{k\in\mathcal{H}_i} w_{k,i} u_k + v_k$ | Convolution multiple membership type II (Convolution MMM type II) |

- Uninformative (vague) priors
- Make use of sparseness of weight-matrix in computations
- Model comparison using DIC
- MCMC via OpenBugs

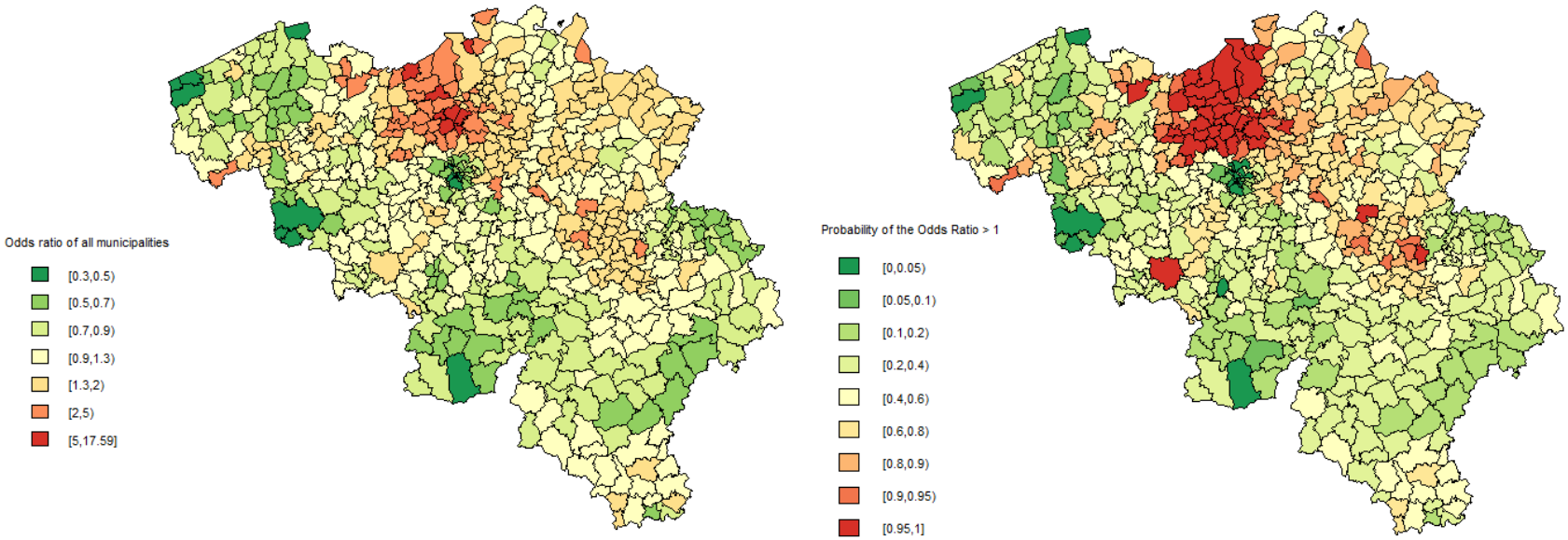



Figure 5: *Maps of the odds ratio for the Convolution multiple membership model type II (left). Maps of the exceedance probabilities for the Convolution multiple membership model type II (right).*

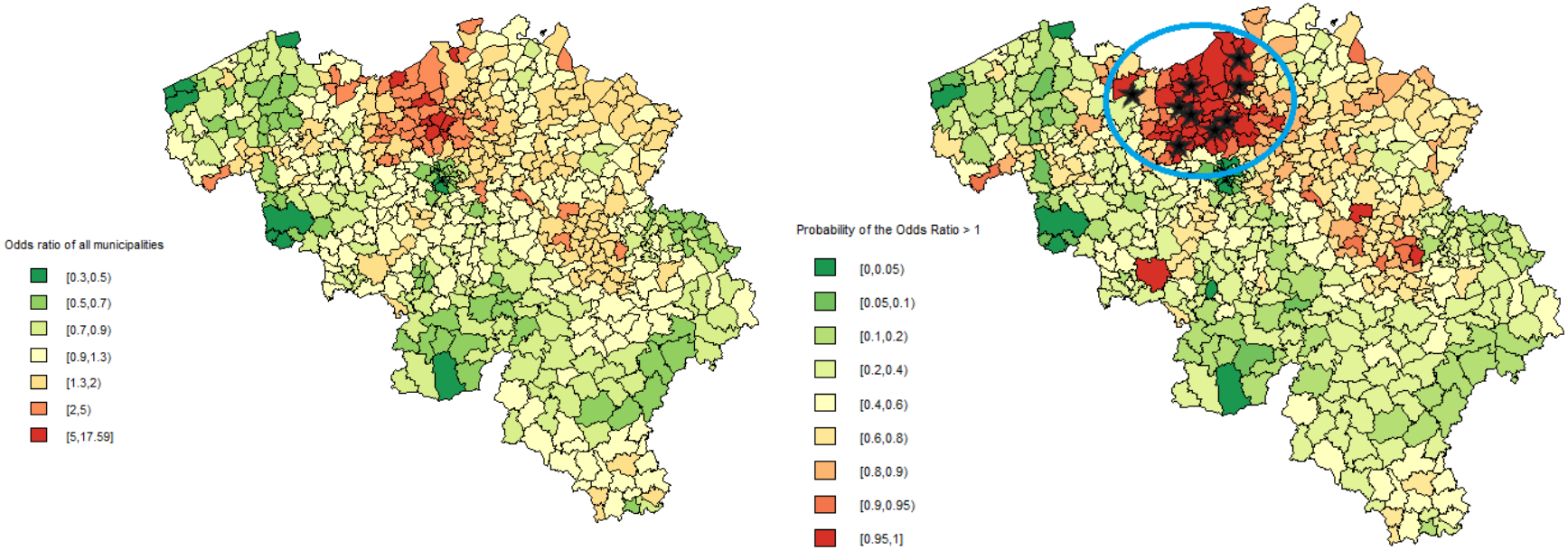



Figure 6: *Maps of the odds ratio for the Convolution multiple membership model type II (left). Maps of the exceedance probabilities for the Convolution multiple membership model type II (right).*

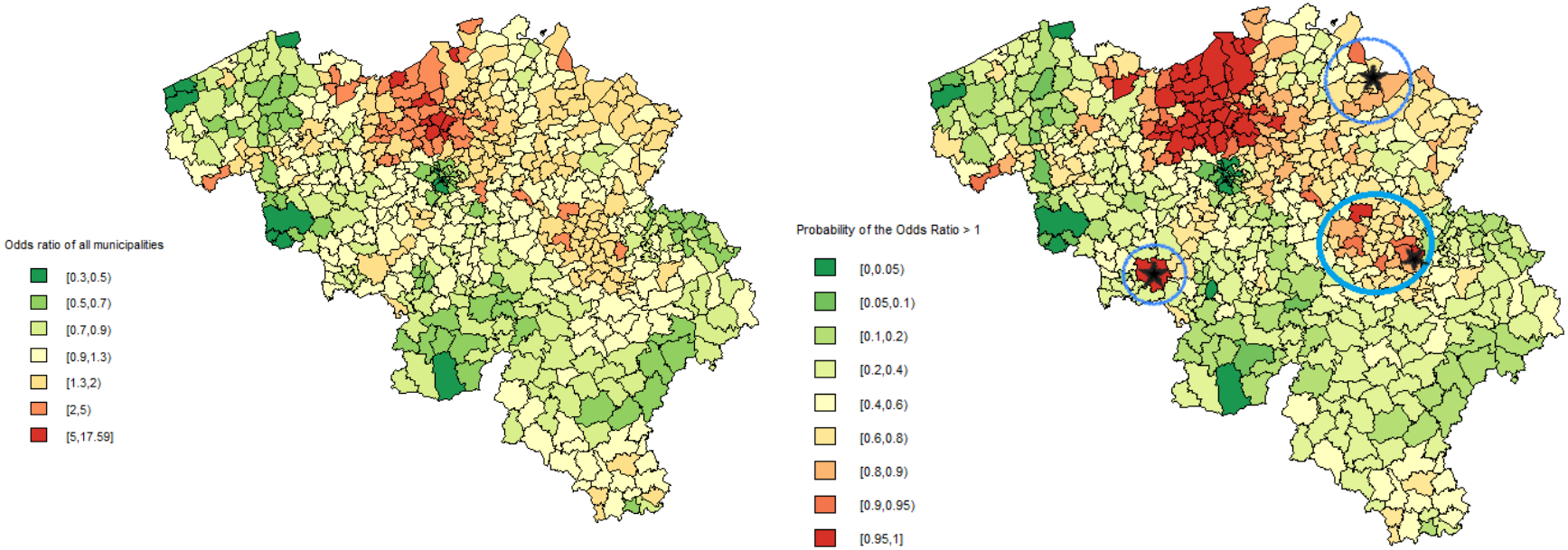



Figure 7: *Maps of the odds ratio for the Convolution multiple membership model type II (left). Maps of the exceedance probabilities for the Convolution multiple membership model type II (right).*

# Conclusion

- Time of residential location has impact on the risk status of areas
- We propose the use of an interval of 20 to 40 years in which the patients lived prior to diagnosis of the disease.
- Multiple membership models are preferred over the classical multilevel approach
- Pancreatic cancer used as control disease (as historical residential locations not routinely available)
- Lower DIC values are found for the models incorporating a multiple membership structure.

# Conclusion

- A cluster of municipalities in the Northern Central part of Belgium presents a highly elevated risk of mesothelioma, as well as municipalities in the Central Eastern part of the country.
- Municipalities with a lot of certainty of a decreased risk of mesothelioma are mainly located in the Southern and Western part of Belgium.
- Assumptions:
  - Exposure in 20 years interval in different areas
  - Ordering of residential locations not important
  - Exposure is constant in time
  - Residential location is proxy for workplace

**THANK YOU!**