UHASSELT
KNOWLEDGE IN ACTION

Maastricht University

Faculty of Sciences
School for Information Technology

Master of Statistics

Masterthesis

The bivariate spatial modelling of breast and ovary cancer in Limburg

Leyla Kodalci

Thesis presented in fulfillment of the requirements for the degree of Master of Statistics, specialization Epidemiology & Public Health Methodology

SUPERVISOR :

dr. Thomas NEYENS

2017
2018

Transnational University Limburg is a unique collaboration of two universities in two countries: the University of Hasselt and Maastricht University.

UHASSELT
KNOWLEDGE IN ACTION

www.uhasselt.be
Universiteit Hasselt
Campus Hasselt:
Martelarenlaan 42 | 3500 Hasselt
Campus Diepenbeek:
Agoralaan Gebouw D | 3590 Diepenbeek

Maastricht University

Faculty of Sciences
School for Information Technology
Master of Statistics

Masterthesis

The bivariate spatial modelling of breast and ovary cancer in Limburg

Leyla Kodalci

Thesis presented in fulfillment of the requirements for the degree of Master of Statistics, specialization Epidemiology & Public Health Methodology

SUPERVISOR :

dr. Thomas NEYENS

Abstract

The aim of this study is to investigate the spatial distribution of breast and ovary cancer in Limburg and the correlation between these spatial distributions.
The data used for this study is the incidence data of breast and ovary cancer between the years 1996 and 2005 for the female population in the 44 municipalities of Limburg.
To describe the spatial distribution of the cancers, a number of methods have been suggested to model relative risks within municipalities. Typically, Bayesian hierarchical models are used, which account for spatially correlated and uncorrelated heterogeneity present in the data by including random effects. In the univariate setting, the conditional autoregressive (CAR) convolution model and alternative combined model account for the spatially correlated and uncorrelated heterogeneity. Extending the univariate convolution and combined model with bivariate distributions for the random effects, makes it possible to model the relative risks of two diseases and investigate the correlation between the spatial distributions.
In the univariate analysis for breast cancer, the CAR convolution model performed the best based on DIC and showed a pattern of increased risk in southern Limburg. For ovary cancer, the CAR model performed the best based on DIC and also shows a pattern of increased risk in southern Limburg. In the bivariate analysis, the bivariate combined model with disease specific CAR distribution for the spatial random effects performed the best based on DIC and MSPE. The relative risks show a significant correlation of 0.599 and this correlation is a result of the spatially uncorrelated heterogeneity.

Keywords
Bivariate Modeling; Disease Mapping; Overdispersion; Multivariate Condition Autoregressive Model; Bivariate combined Model

Contents

1 Introduction

Breast cancer is the most common cancer in the female population. Important risk factors of breast cancer are : age, family history of cancer, reproductive factors, age at menarche and menopause. McPherson et al. (2000) suggested that environmental factors are of greater importance in the risk of breast cancer than genetic factors.[16]

Ovary cancer is less common in the female population than breast cancer. An important risk factor for ovary cancer is a family history of breast or ovary cancer. Other important risk factors are: age at menarche and menopause, increasing age.[12]

The genes BRCA1 and BRCA2 are associated with risk of breast and ovary cancer, a mutation in one of these genes increases the risk strongly for both cancers. [11]
Both types of cancer show common risk factors, which suggests that they are related but they are not equally present in the population suggesting that they don't behave the same.

Disease mapping investigates the risk of a disease in the regions of an area, it can also detect if the disease has a spatial trend in the area.
In the early-stages of disease mapping, methods for spatial modeling were mainly developed in the univariate setting. But in the recent years more investigation is done in methods to model multiple diseases. [14]

In this study, an area is subdivided into fixed spatial units, in which cancer counts are observed. This type of data show more variability than would be expected from the underlying distributions, which results from different possible causes. This extra variability is also known as overdispersion.[20] Uncorrelated heterogeneity is specified as the extra variability caused by unknown factors and correlated heterogeneity is specified as the extra variability caused by known spatial structure of the data. Hierarchical Bayesian models are used to account for overdispersion.[14]

For the univariate setting, the Poisson-gamma and Poisson-lognormal models account for overdispersion by including a random effects term. The Conditional Autoregressive (CAR) convolution model has been very popular to account for correlated heterogeneity.[6] Alternatively, Molenberghs et al. (2007) developed the combined model by combining the CAR convolution model with the Poisson-gamma model.
For the bivariate setting, methods based on the extension of the convolution model were proposed by including a multivariate CAR distribution and a multivariate normal distribution to account for correlated and uncorrelated heterogeneity respectively.[4,6] Alternatively, Neyens et al. (2016) proposed an extension of the combined model by including a multivariate CAR distribution and a bivariate gamma distribution to account for correlated and uncorrelated heterogeneity respectively.

In section 2, the data used in this study is described. Section 3 reviews the methods used for the spatial analysis, and section 4 the results of spatial the analysis. In section 5, the results are discussed and in section 6 the conclusion is provided.

The aim of this study is to investigate the spatial distribution of breast and ovary cancer in Limburg and the correlation between these spatial distributions. The occurrence of a spatial trend of breast and ovary cancer in Limburg will be investigated in the univariate analysis. The bivariate analysis, which is of interest in this study, will investigate the spatial distribution of breast and ovary cancer and the correlation between these distributions.

2 Data

The dataset used in this study is part of the Limburgs Cancer Registry (LIKAR). LIKAR registers all cancers present in the province of Limburg (Belgium) between the years 1996 and 2005, with additional information about each cancer type per region, age and gender (http://likas.edm.uhasselt.be/).
The province of Limburg is situated in the northeast of Belgium. Limburg consists of 44 municipalities with a wide range of population densities.

The data used for this study is the incidence data of breast and ovary cancer between the years 1996 and 2005 for the female population, which is the number of new cancer cases occurring in cancer-free females. The data was distributed in 18 age-groups ranging from age 0 to age 85+.

3 Methods

3.1 Exploratory Analysis

An indirect age standardisation was conducted on the data to adjust observed rates to reflect rates that would be observed if the population standard's age-specific rates (r_g^S) applied to the study population. These adjusted rates were used to estimate indirectly the number of cases expected in each age group in the observed study population (E_{gi}). The number of expected cases were summed over the different age groups for each municipality, which is represented by the following equation[25]:

$$E_i = \sum_g E_{gi} = \sum_g r_g^S n_{gi} = \sum_g \frac{y_g^S}{n_g^S} n_{gi}, \tag{1}$$

with g representing the number of age-groups, i specifies the municipality, n_{gi} specifies the number of people at risk in age-group g for municipality i, n_g^S specifies the number of people at risk in age-group g for the standard population with y_g^S specifying the number of observed cancer counts in the standard population.
The standardized incidence ratio (SIR) defined by the following equation was calculated by taking the ratio of the observed counts for each municipality and the expected counts for each municipality. If the SIR has a value larger than 1 it indicates that there are more cases observed than what was expected, if the value is smaller than 1 it indicates that there are less cases observed than what was expected.[14,25]

$$SIR_i = \frac{y_i}{E_i} \tag{2}$$

The Error factor method was used to calculate the 95% confidence interval(CI) of SIR, these CI were used to investigate the significance of the increased or decreased risk[5]:

$$\left[SIR_i / exp\left(-1.96 * \sqrt{\frac{1}{Y_i}} \right) ; SIR_i / exp\left(1.96 * \sqrt{\frac{1}{Y_i}} \right) \right]. \tag{3}$$

A test for spatial autocorrelation was used to measure the tendency of observations from nearby municipalities to be more or less alike than observations from municipalities farther apart. This was done because most statistics are based on the assumption that the values of observations are independent.[7] If there is any systematic pattern in the spatial distribution, it's spatially autocorrelated: positive spatial autocorrelation is a result of nearby observations being more alike than distant ones and negative spatial autocorrelation is a result of nearby observations being more different from nearby areas than area's further away. A random spatial pattern results in no spatial autocorrelation.[26]

Moran (1950) proposed a spatial autocorrelation indicator, which is described by the following equation:

$$I = \frac{n \sum_i \sum_j W_{ij}(Y_i - \bar{Y})(Y_j - \bar{Y})}{(\sum_i \sum_j W_{ij}) \sum_i (Y_i - \bar{Y})^2}, \tag{4}$$

where W_{ij} denotes the elements of the row-standardized weights matrix $\mathbf{W}$. The weights matrix was based on sharing common boundaries and thus defined which municipalities are adjacent. This statistic compares the value of a variable at any municipality with the value of all other municipalities. [19]
When spatial independence is assumed, Moran's spatial autocorrelation indicator is as asymptotically normally distributed with mean and variance[26]:

$$E[I] = -\frac{1}{n-1}, \tag{5}$$

$$Var(I) = \frac{n^2 \sum_{ij} W_{ij}^2 + 3(\sum_{ij} W_{ij})^2 - n \sum_i (\sum_j W_{ij})^2}{(n^2-1)(\sum_{ij} W_{ij})^2}. \tag{6}$$

Moran's I indicates the presence of positive autocorrelation when Moran's I is larger than the expected value under independence, greater values of a Morans I indicate stronger positive clustering pattern. Moran's I indicates the presence of negative autocorrelation when Moran's I is smaller than the expected value under independence, smaller values of a Morans I indicate a regular pattern.[26]

Geary (1954) proposed an alternative indicator for spatial autocorrelation, which is defined by the following equation:

$$C = \frac{(n-1) \sum_i \sum_j W_{ij}(Y_i - Y_j)^2}{2(\sum_i \sum_j W_{ij}) \sum_i (Y - \bar{Y})^2}. \tag{7}$$

Under the spatial independence assumption, the expected value of Geary's C statistic is equal to 1. Geary's C statistic indicates a perfect positive spatial autocorrelation when the statistic is zero and a perfect negative spatial autocorrelation when the statistic has a value 2. [8]

This statistic differs from Moran's I statistic: Moran's I is a cross product based on the deviations from the overall mean and Geary's C cross product is based on the deviation of the actual values of each observation location with one another. [26]

To test the significance of Moran's I and Geary's C statistics, randomization and Normal approximation was used. Randomization is done by producing a randomization distribution where the data values are reassigned among the fixed location. If the statistic lies in the tails of this distribution, the assumption of independence is rejected. The normal approximation compares the observed Z-score to the standard normal distribution.[22]

3.2 Univariate Analysis

The models used in the univariate analysis will be based on the Poisson model. This model was chosen because the data contained a relatively low count of cancer cases and a relatively large population size in each small area (municipality).
A basic Poisson model assumes that the counts y_i are independently Poisson distributed for each municipality in Limburg ($i = 1...44$)[14]:

$$y_i \sim Poisson(e_i\theta_i), \tag{8}$$

with e_i defined as the expected number of cancer cases for each municipality and θ_i represents the relative risk for each municipality. The SIR is the maximum likelihood estimator for the relative risk, with standard deviation $\frac{\sqrt{y_i}}{e_i}$.[14]

The drawback of using a traditional Poisson model is that it often doesn't capture all of the variability that can be present in count data caused by unknown factors, this is called overdispersion. Another drawback is the assumption of independence between the municipalities, which is not valid with diseases that are environmental related.[20]
These drawbacks are solved by extending the Poisson model to include a prior distribution for the relative risk or to include an extra random effect to the model.[20] These models are described in the following sections.

3.2.1 Poisson-Gamma model

The Poisson-gamma model is an extension of the basic Poisson model that takes overdispersion into account, but it doesn't take spatial correlation into account and including covariates is difficult.[20] This was done by including a gamma prior distribution for the relative risk θ_i for each municipality

$$\theta_i \sim Gamma(a, b), \tag{9}$$

with scale parameter a, rate parameter b,prior mean $\frac{a}{b}$ and prior variance $\frac{a}{b^2}$.[20]

The Poisson and gamma distribution are conjugate distributions, this makes it mathematical convenient to analytically derive the posterior distribution.[18] When the scale and rate parameter are fixed and known, the relative risk has the following posterior distribution for each municipality[14]:

$$\theta_i \sim Gamma(a + y_i, b + e_i), \tag{10}$$

with a posterior mean and variance equal to $\frac{a+y_i}{b+e_i}$ and $\frac{a+y_i}{(b+e_i)^2}$ respectively. As a consequence the relative risks are no longer estimated based on the data alone but the assumed prior has also an impact.[14]

In this setting, the scale and rate parameter are unknown and prior distributions for these parameters were specified to express the ignorance or prior knowledge. These prior distributions are more specifically called hyperprior distributions. The hyperprior distibutions were chosen to be on the positive real line with parameters chosen such that a large variance is underlying these parameters ($\sim Exp(0.01)$).[1]

3.2.2 Poisson-Lognormal model

As in the previous model, the Poisson-lognormal model is an extension of the basic Poisson model to take overdispersion into account.[14] This is done by incorporating a normal uncorrelated random effect in the linear predictor of the relative risk for each municipality

$$log(\theta_i) = \alpha + \nu_i, \tag{11}$$

$$\nu_i \sim N(0, \sigma_\nu^2), \tag{12}$$

with α defining the overall mean risk and ν_i is defined as an uncorrelated random effect with a normal prior distribution (5).[1]
The random effects (ν_i) take the overdispersion into account and represent the residual relative risk in each municipality after adjusting for the overall mean. The variance of the random effects (σ_ν^2) reflects the amount of overdispersion in the data.[1]

The Poisson-lognormal model adjusts for covariates much easier in contrast with the Poisson-gamma model but is mathematical less convenient, both models do not account for possible spatial correlation.[20]

The fixed effect (α) was assigned a vague prior and the variance parameter (σ_ν^2) was assigned a vague hyperprior.

3.2.3 Conditional Autoregressive model

In contrast with the previous models, the intrinsic conditional autoregressive (CAR) model extends the basic Poisson model to take spatial correlation into account (correlated heterogeneity). The linear predictor of the relative risk for each municipality includes the overall mean risk (α), which was defined by an uninformative prior and the spatial random effects term (υ_i) for each municipality[14]

$$log(\theta_i) = \alpha + \upsilon_i. \tag{13}$$

The spatial random effects term was assigned a spatially structured prior distribution to allow for spatial correlation[3]:

$$\upsilon_i|\upsilon_{j\neq i} \sim N(\frac{\sum_{j\neq i} c_{ij}\upsilon_j}{\sum_{j\neq i} c_{ij}}, \frac{\sigma_\upsilon^2}{\sum_{j\neq i} c_{ij}}), \tag{14}$$

with c_{ij} representing weights that define the influence of region u_j on the prior mean of u_i. The weights were based on the Queen's case proximity matrix, which is based on sharing common boundaries and thus defines which regions j are adjacent to region i. These weights can take the value 1 if regions i and j are adjacent and 0 otherwise, which leads to the following reformulation of the CAR prior distribution:

$$\upsilon_i|\upsilon_{j\neq i} \sim N(\bar{\upsilon_i}, \sigma_i^2), \tag{15}$$

with the conditional mean $\bar{\upsilon_i}$ equal to $\frac{\sum_{j\in\delta_i}\upsilon_j}{n_{\delta_i}}$ where δ_i is the of neighbors of region i and n_{δ_i} is the number of neighbors n this set. The conditional variance σ_i^2 equals $\frac{\sigma_\upsilon^2}{n_{\delta_i}}$ and was defined with an uninformative hyperprior.[3]

3.2.4 Convolution model

The convolution model is an extension of the basic Poisson model, the Poisson-lognormal model and the CAR model. This model takes both the correlated heterogeneity and uncorrelated heterogeneity random effects into account, and thus taking spatial correlation and overdispersion into account.[14]
The relative risk for each municipality is defined by the following equation

$$log(\theta_i) = \alpha + \upsilon_i + \nu_i, \tag{16}$$

with α defined as the overall mean risk, υ_i is the correlated heterogeneity (spatial random effect) and ν_i is the uncorrelated heterogeneity (non-spatial random effect).[14]
The spatial random effects term is assumed to follow an intrinsic CAR prior distribution as defined in (15), the non-spatial random effects term is assumed to follow a lognormal prior distribution with zero mean and variance σ_v^2 as in (11).[15]
By looking at the ratio between the random effect terms it's possible to investigate if the residual relative risk is due to correlated or uncorrelated heterogeneity.

3.2.5 Combined model

Molenberghs et al. (2007) proposed the combined model as an alternative to the convolution model, in terms of the way they model the uncorrelated heterogeneity. The combined model uses a gamma distribution in stead of a lognormal distribution. Therefore, this model can also be seen as an extension of the Poisson-gamma model.[17]

The relative risk for each municipality is defined by

$$\theta_i = g_i exp(\alpha + \upsilon_i), \tag{17}$$

with α defined as the overall mean risk, defined by an uninformative prior distribution. The spatial random effects term (υ_i) was assumed to follow a CAR prior distribution as defined in (15). The non-spatial random effects term (g_i) was defined by a gamma prior distribution as defined in (8).[20]

Adjusting to a gamma distribution in this model gains the advantage of the strong conjugacy of the Poisson and gamma distribution, which makes it easy to analytically define the posterior mean of y_i conditional on the correlated heterogeneity random effect as $\frac{a+y_i}{b+\kappa_i e_i}$ with κ_i equal to $\alpha + x_i\beta + \upsilon_i$. In contrast with the Poisson-gamma model, the smoothing is spatially structured.[18]

3.3 Bivariate Analysis

To investigate the spatial distribution of both cancer types as well as the correlation between the spatial distributions, an extension of the univariate setting to the bivariate setting was required.

There are three possible approaches to define the relationship between the random effects; models with common random effects, models with shared random effects and models with correlated random effects.[14] The first approach uses the same random effects for both cancer types,

$$\begin{aligned} \theta_{1i} &= exp(\alpha_1 + W_{1i}), \\ \theta_{2i} &= exp(\alpha_2 + W_{1i}), \end{aligned} \tag{18}$$

with W_{1i} being the common random effects component between the models.[14] This approach cannot be used when the cancer types are not equally common.
Alternatively, the shared random effects approach uses the same random effects for both cancer types but adjusts this with a scaling component[13]:

$$\begin{aligned} \theta_{1i} &= exp(\alpha_1 + \delta W_{1i}), \\ \theta_{2i} &= exp(\alpha_2 + W_{1i}/\delta), \\ log(\delta) &\sim N(0, 0.17), \end{aligned} \tag{19}$$

with W_{1i} being the shared random effects component between the models. The scaling component is represented by δ, which is defined by a prior distribution.[13]
The random effects for both approaches can be spatial random effects or non-spatial random effects and both models may also have cancer-specific terms.
These models have common or shared random effect terms, which are too restrictive in this study because breast and ovary cancer act alike but not the same.
The third approach uses correlated random effects by extending the univariate convolution and combined model with bivariate distributions for the random effects.[21]

The multivariate CAR distribution was used for the spatial random effects and for the non-spatial random effects a multivariate normal distribution or alternative a bivariate gamma distribution were used. By using these bivariate distributions for the random effects it's possible to investigate the correlation between the spatial distributions of breast and ovary cancer.

The empirically based Pearson correlations were calculated using $r_{x1,x_2} = \frac{cov(x_1,x_2)}{sd(x_1)sd(x_2)}$, where x_1 and x_2 were the relative risks, the spatial random effects and the non-spatial random effects.

3.3.1 Bivariate Convolution model

The bivariate convolution model is an extension of the previously described univariate convolution model. The relative risks for each municipality for the bivariate convolution model are described by the following model[14]:

$$\theta_{li} = exp(\alpha_l + \upsilon_{li} + \nu_{li}), \tag{20}$$

with α_l defining the overall cancer-specific risk. The non-spatial random effects term (ν_{li}) can be defined univariately as in (11) or by a multivariate normal distribution $\sim MVN(0, \Sigma_\nu)$ (Table 1).[2] The covariance matrix was defined by a vague Wishart hyperprior distribution.
The spatial random effects term (υ_{li}) can be defined univariately with a intrinsic CAR prior, or it can be defined by a multivariate CAR prior (Table 1). The multivariate CAR prior distribution is defined by[6]:

$$\upsilon_{\mathbf{l}} \sim \mathbf{N}(\mathbf{0}, \mathbf{\Sigma}_\upsilon), \tag{21}$$

where $\upsilon_{\mathbf{l}}$ contains the cancer-specific spatial random effects for all municipalities. The covariance matrix ($\mathbf{\Sigma}_\upsilon$) is valid as long as it is symmetric and positive definite, which was defined by a vague Wishart hyperprior distribution.[10]
The marginal distribution of the spatial random effects of breast cancer ($\upsilon_{\mathbf{1}}$) and the conditional distribution the spatial random effects of ovary cancer ($\upsilon_{\mathbf{2}}$) is described by[24]:

$$\begin{aligned} \upsilon_{\mathbf{1}} &\sim N(0, (\mathbf{D} - \gamma_{\mathbf{1}}\mathbf{C})^{-\mathbf{1}}\sigma_{\mathbf{1}}^{\mathbf{2}})), \\ \upsilon_{\mathbf{2}}|\upsilon_{\mathbf{1}} &\sim N(\mathbf{A}\upsilon_{\mathbf{1}}, (\mathbf{D} - \gamma_{\mathbf{2}}\mathbf{C})^{-\mathbf{1}}\sigma_{\mathbf{2}}^{\mathbf{2}})), \end{aligned} \tag{22}$$

with $\mathbf{D}$ representing a matrix with the number of neighbors of the municipalities on the diagonal, $\mathbf{C}$ represents the adjacency matrix with values that take the value 1 if the municipalities are adjacent and zero otherwise. The smoothing parameter and variance for the marginal distribution of υ_1 is represented by γ_1 and σ_1^2 respectively. The smoothing parameter and variance for the conditional distribution of $\upsilon_2|\upsilon_1$ is represented by γ_2 and σ_2^2 respectively.[24] The matrix $\mathbf{A}$ determines the relationship between the spatial random effects of breast and ovary cancer,

$$\mathbf{A} = \xi_{\mathbf{0}}\mathbf{I} + \xi_{\mathbf{1}}\mathbf{C}, \tag{23}$$

with $\mathbf{I}$ representing an identity matrix (n x n), ξ_0 and ξ_1 are defined as the bridging parameters. The bridging parameter ξ_0 defines the relationship between υ_{2i} and υ_{1i} if the municipalities are the same. The bridging parameter ξ_1 determines the relationship between υ_{2j} and υ_{1i} if municipality i and j are adjacent.[10]
The models fitted based on the bivariate convolution model are described in table 1.

3.3.2 Bivariate Combined model

As in the univariate setting, the bivariate combined model is an alternative to the bivariate convolution model. Neyens et al. (2016) proposed to model the relative risks for each municipality using the bivariate combined model,

$$\theta_{il} = g_{il} exp(\alpha_l + \upsilon_{il}), \tag{24}$$

with α_l describing the overall cancer-specific risk. The spatial random effects term (υ_{il}) can be defined univariately with an intrinsic CAR prior, or it can be defined by a multivariate CAR prior (Table 1).
The non-spatial random effects term (g_{il}) can be defined univariately or by a bivariate gamma distribution (Table 1), with the bivariate specification defined as follows:

$$g_{il} = \frac{1}{k_0 + k_1}(k_0\gamma_{i0} + k_l\gamma_{il}), \tag{25}$$

where k_0 and k_l are real positive variables, γ_{i0} and γ_{il} are assumed gamma distributed random variables $\sim \Gamma(1,1)$. This bivariate specification produces correlated non-spatial random effects.[21]
The models fitted based on the bivariate combined model are described in table 1.

Table 1: Overview fitted models with models 1,2,3 based on the bivariate convolution model and models 4,5,6 based on the bivariate combined model.

Model	RR	Spatial random effects	Non-spatial random effects
1		$\upsilon_{li} \sim CAR(\tau_{\upsilon_l})^*$	$\nu_{li} \sim MVN(0, \Omega)$
2	$\theta_{li} = exp(\alpha_l + \upsilon_{li} + \nu_{li})$	$\upsilon_{li} \sim MCAR(1, \Omega)$	$\nu_{li} \sim N(0, \tau_{\nu_l})^*$
3		$\upsilon_{li} \sim MCAR(1, \Omega)$	$\nu_{li} \sim MVN(0, \Omega)$
4		$\upsilon_{li} \sim CAR(\tau_{\upsilon_l})^*$	$g_{il} = \frac{1}{k_0+k_1}(k_0\gamma_{i0} + k_l\gamma_{il})$
5	$\theta_{il} = g_{il} exp(\alpha_l + \upsilon_{il})$	$\upsilon_{il} \sim MCAR(1, \Omega)$	$g_{il} \sim Gamma(a, b)^*$
6		$\upsilon_{il} \sim MCAR(1, \Omega)$	$g_{il} = \frac{1}{k_0+k_1}(k_0\gamma_{i0} + k_l\gamma_{il})$

* cancer specific random effects
τ represents the precision
Ω represents the precision matrix

3.4 Model Comparison

To compare the different models separately in the univariate and bivariate setting the goodness-of-fit for each model was tested.
Spiegelhalter et al. (2002) proposed the use of a deviance information criteria as a measure for goodness-of-fit for Bayesian hierarchical models, which is defined in the following equation,

$$\begin{aligned} DIC &= E_{\theta|y}(D) + pD, \\ pD &= E_{\theta|y}(D) - D(E_{\theta|y}(\theta)). \end{aligned} \tag{26}$$

The DIC is calculated based on the posterior mean deviance ($E_{\theta|y}(D)$) and the estimated effective number of parameters (pD). The estimated effective number of parameters is defined as the difference between the posterior mean deviance and the deviance of posterior expected parameter estimates ($D(E_{\theta|y})(\theta)$). The estimated effective number of parameters penalizes the deviance for complexity.[23]
If there are models with very similar DIC values, less complex models having lower effective number of parameters were chosen as 'the best' model.

The mean squared predictive error (MSPE) was used as a second goodness-of-fit measure and is defined by the following equation. The mean squared predictive error (MSPE) was used to asses the information loss in the data by investigating the predictive ability of the model, which is the ability to produce replicated data similar to the data that was observed.[9]

$$MSPE = \sum_i \sum_j \frac{(y_i - y_{ij}^{pred})^2}{m * G}, \tag{27}$$

with y_{ij}^{pred} defining the predicted response at MCMC iteration j, the number of observations m and the sampler sample size G.

In contrast with DIC who penalizes more complex models, MSPE will prefer more complex models because they tend to predict more precise. However, the prediction will become less precise if the complexity is too high due to more variability in the replicated data and a less complex model will perform better.[4]

3.5 Model fitting on WinBUGS

A burn-in of 50 000 iterations was used to achieve convergence and an additional 50 000 iterations were conducted for inference.
All models were visually investigated for MCMC convergence issues by looking at the history plots in WinBUGS. There were no convergence issues for the selected models.

4 Results

4.1 Exploratory Analysis

Table 2 gives the summary statistics for the observed counts in breast and ovary cancer, it shows that breast cancer is more present in the population than ovary cancer. The distribution of the observed counts in Limburg are shown in figure 1, it also indicates that breast cancer is more present in the population than ovary cancer but shows no clear pattern in the distribution.

Table 2: Summary Statistics observed counts for breast and ovary cancer

	Breast	Ovary
Mean	130.430	9.409
Standard devation	118.810	8.379
Minimum	1	0
Maximum	653	41

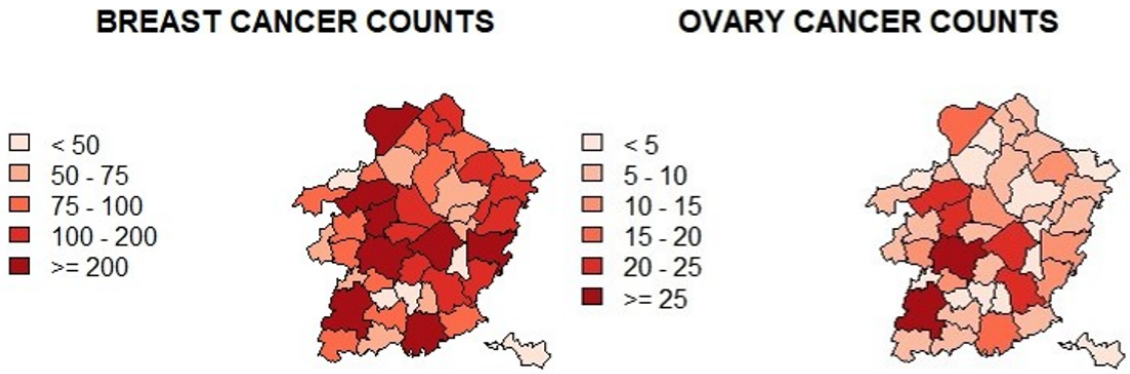

Figure 1: *Left:* Observed breast cancer counts in Limburg. *Right:* Observed ovary cancer counts in Limburg.

After conducting the indirect standardisation the SIR values were calculated, confidence intervals were calculated using the error factor method. In the breast cancer setting the following municipalities show an significant decreased risk: Bilzen, Ham, Maasmechelen, Riemst and Tessenderlo. Gingelom, Hasselt and Sint-Truiden show a significant increased risk of breast cancer. In the ovary cancer setting there were no municipalities with a significant increased or decreased risk. The SIR estimates with 95% credible interval can be found in the appendix table 11 for breast cancer and table 12 for ovary cancer. Figure 2 shows the spatial distribution of SIR in Limburg, there seems to be a pattern in the south of Limburg for both cancers.

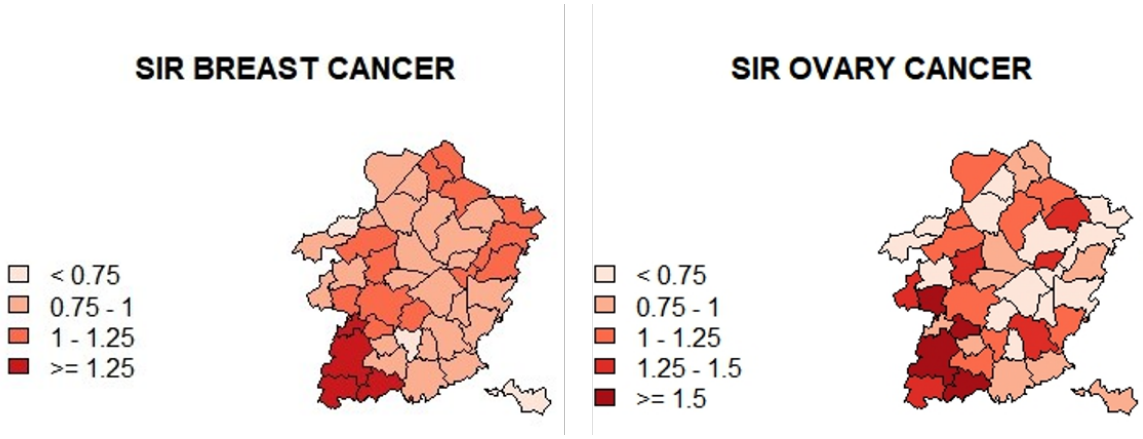

Figure 2: *Left:* SIR values for breast cancer in Limburg. *Right:* SIR values for ovary cancer in Limburg.

Table 3 summarizes the spatial autocorrelation estimators, p-values of Moran's I and Geary's C under the normality and randomization assumption based on a Queen's case proximity matrix (sharing boundaries). The expected value under independence is for the Moran's I statistic is equal to -0.023, and for Geary's C statistic, this is equal to 1.

In the breast cancer setting, the Moran's I statistic is bigger than the expected value, which indicates that there seems to be a positive spatial autocorrelation. The p-values show that there seems to be a deviation from the independence assumption. The Geary's C statistic is smaller than the expected value under independence which indicates that there seems to be a negative spatial autocorrelation, this is supported by the significant p-values.
In the ovary cancer setting, the Moran's I statistic is bigger than the expected value which indicates that there seems to be a positive spatial autocorrelation. The p-values are borderline not significant which tells that there seems to be no deviation of the dependence assumption. The Geary's C statistic is smaller than the expected value under independence which indicates that there seems to be a negative spatial autocorrelation, this is supported by the significant p-values.
Both statistics give different results but give an indication of possible positive spatial autocorrelation.

Table 3: Spatial autocorrelation statistics for breast and ovary cancer in Limburg.

	Breast			Ovary		
	Statistic	P-value Rand.	P-value Norm.	Statistic	P-value Rand.	P-value Norm.
Moran's I	0.368	2.639e-05	0.001	0.134	0.053	0.066
Geary's C	0.026	1.628e-05	0.001	0.037	0.02022	0.025

4.2 Univariate Analysis

4.2.1 Breast cancer

Table 4 summarizes the goodness-of-fit statistics for all the univariate models for breast cancer in Limburg. Goodness-of-fit based on the DIC criterion shows that the convolution model and the CAR model performs the best of all models. The convolution model was chosen as 'best' model because it has a lower value for the MSPE than the CAR model.

Table 4: Model Fits in the univariate setting for breast cancer.

Model	Model Fit		
	DIC	pD	MSPE
Poisson-gamma	353.768	25.469	266.100
Poisson-lognormal	353.793	26.801	262.000
CAR	351.218	22.506	269.700
Convolution	**350.517**	23.154	267.600
Combined	358.797	37.804	**254.500**

The parameter estimates obtained from the convolution model are summarized in table 5, parameter estimate α represents the global mean ($e^{-0.019}$). The global mean suggests that the distribution of the relative risks lies around a value of 1. The ratio of the variance of the spatial random effects (σ_u^2) and the non-spatial random effects (σ_v^2) is large which reflects that the variability of the relative risks is more attributed to the spatially structured heterogeneity. The variance of the spatial random effects(σ_u^2) is small, this indicates stronger spatial dependence between neighboring municipalities.

Table 5: Posterior parameter estimates for the convolution model

Parameter	Estimate	sd	MC error	Lower limit	Upper limit
α	-0.019	0.018	1.326E-4	-0.056	0.015
Global Mean	0.982	0.018	1.29E-4	0.946	1.016
σ_u^2	0.026	0.015	3.196E-4	0.001	0.061
σ_v^2	0.004	0.004	1.096E-4	2.1E-4	0.015
σ_u^2/σ_v^2	31.120	52.380	1.038	0.099	179.300

Figure 3 shows the relative risk distribution of breast cancer in Limburg (left) and the exceedence probability of the relative risks being greater than 1 (right). The maps seem to show a pattern in the south of Limburg with an increased risk and probability of the relative risks. Gingelom, Hasselt and Sint-Truiden show a significant increased relative risk. The following municipalities show a significant decreased relative risk: Bilzen, Ham, Maasmechelen and Tessenderlo. The relative risk estimates with 95% credible interval can be found in the appendix (Table 13).

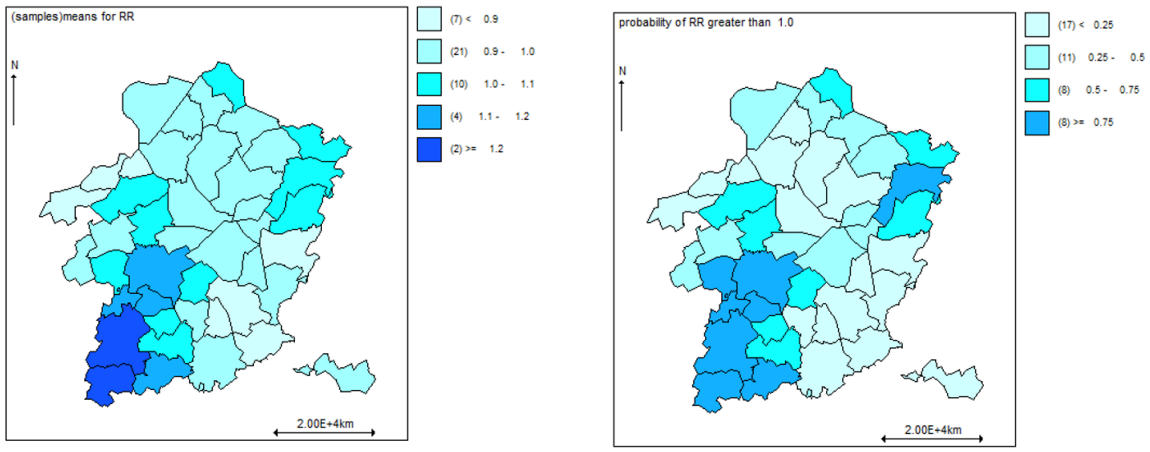

Figure 3: *Left:* Relative risk estimates yielded by the Convolution model for breast cancer in Limburg. *Right:* The posterior expected exceedence probability based on the convolution model for breast cancer in Limburg.

4.2.2 Ovary cancer

Table 6 summarizes the goodness-of-fit statistics for all the univariate models for ovary cancer in Limburg.Goodness-of-fit based on the DIC criterion shows that the CAR model performs the best of all models. The Combined model has the lowest MSPE value but performs the worst based on DIC.

Table 6: Model Fits in the univariate setting for ovary cancer.

Model	Model Fit		
	DIC	pD	MSPE
Poisson-gamma	210.682	6.885	19.530
Poisson-lognormal	211.468	9.527	18.830
CAR	**208.555**	8.405	18.490
Convolution	210.471	5.931	20.190
Combined	219.984	22.797	**17.020**

The parameter estimates obtained from the CAR model are summarized in table 7, parameter estimate α represents the global mean ($e^{-0.011}$). The global mean suggests that the distribution of the relative risks lies around a value of 1. The variance of the spatial random effects(σ_u^2) is small, this indicates stronger spatial dependence between neighboring municipalities.

Table 7: Posterior parameter estimates for the CAR model

Parameter	Estimate	sd	MC error	Lower limit	Upper limit
α	-0.011	0.052	2.10E-04	-0.113	0.089
Global Mean	0.991	0.051	2.07E-04	0.893	1.093
σ_u^2	0.085	0.067	0.001	0.010	0.258

Figure 4 shows the relative risk distribution of ovary cancer in Limburg (left) and the exceedence probability of the relative risks being greater than 1 (right). The maps seem to show a pattern in the south of Limburg with an increased risk and probability of the relative risks of ovary cancer in Limburg. There are no municipalities with a significant increased or decreased risk of ovary cancer in Limburg. The relative risk estimates with 95% credible interval can be found in the appendix (Table 14).

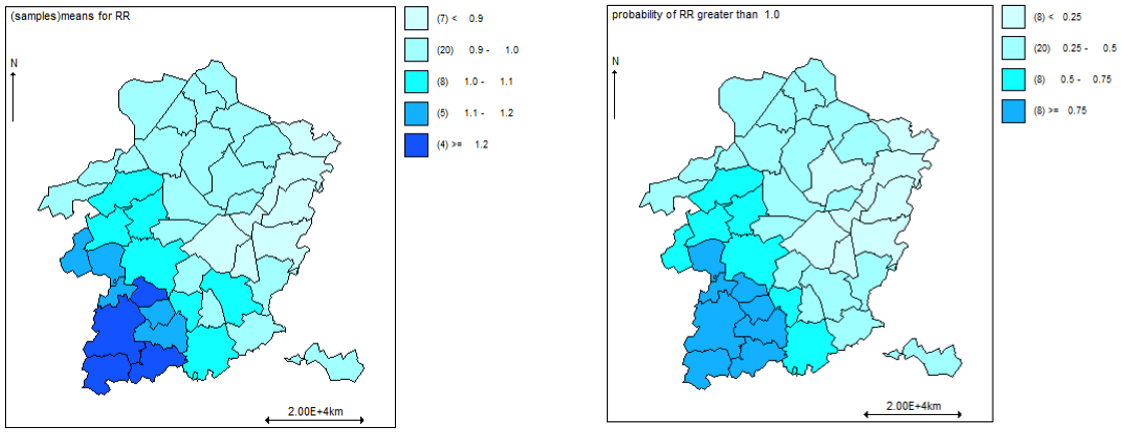

Figure 4: *Left:* Relative risk estimates based on the CAR model for ovary cancer in Limburg. *Right:* The posterior expected exceedence probability based on the car model for ovary cancer in Limburg.

4.3 Bivariate Analysis

In the bivariate analysis 6 different models were fitted which are summarized in table 8. The first three models are based on the bivariate convolution model and the next three models are based on the bivariate combined model with different combinations of the random effects.
The model with the smallest DIC value is the bivariate gamma model with spatial random effects based on the multivariate CAR distribution (model 6). The bivariate gamma model with spatial random effects based on the univariate CAR distribution (model 4) has a similar DIC value but also has a predictive ability which is much better than model 6 and is thus chosen as 'best' model.

Table 8: Overview of the fitted models in the bivariate setting and model fits.

Model	Family	Random Effects		Model Fit		
		Spatial	Non-Spatial	DIC	pD	MSPE
1	Bivariate Convolution	UCAR[a]	bivariate normal	584.766	65.152	276.300
2	Bivariate Convolution	MCAR	univariate normal[a]	584.634	57.726	289.300
3	Bivariate Convolution	MCAR	bivariate normal	595.124	72.158	289.400
4	Bivariate Combined	UCAR[a]	bivariate gamma	**576.408**	**58.656**	**274.200**
5	Bivariate Combined	MCAR	univariate gamma[a]	589.138	67.835	286.500
6	Bivariate Combined	MCAR	bivariate gamma	**575.723**	**56.632**	284.800

[a] disease-specific univariate random effects

The following table describes the parameter estimates of the bivariate gamma model with spatial random effects based on the univariate CAR distribution. The global means of both cancers are described by α_l and results in values 0.971 ($exp(-0.029)$) and 0.977 ($exp(-0.023)$) for the global mean of breast and ovary cancer respectively. The amount of non-spatial variability that is common for both cancers in specified by k_0 and the non-spatial variability specific for the cancer is specified by k_1 for breast cancer and k_2 for ovary cancer. The variance of the non-spatial random effects for breast (var_1) and ovary (var_2) cancer are equal to 0.599 and 0.676 respectively. The variance of the univariate spatial random effects based on the CAR distribution (σ_{ul}) has small values for both cancers, this indicates that there is not much spatial variability and there could be some spatial dependency.

Table 9: Parameter estimates, standard deviations, MC error and 95% CI.

Parameter	Estimate	sd	MC error	Lower limit	Upper limit
α_1	-0.029	0.126	0.004	-0.263	0.227
α_2	-0.023	0.145	0.004	-0.295	0.275
k_0	1.860	1.170	0.035	0.345	4.809
k_1	0.678	0.468	0.014	0.118	1.894
k_2	0.394	0.352	0.008	0.020	1.312
var_1	0.599	0.521	0.018	0.157	2.026
var_2	0.676	0.603	0.020	0.178	2.276
σ^2_{u1}	0.004	0.008	0.001	0.001	0.024
σ^2_{u2}	0.005	0.012	0.001	0.001	0.035

The correlation estimates are summarized in table 10. The correlation between the relative risks of both cancers is significant with a value of 0.599. The correlation estimate between the spatial random effects has a small value of 0.007 and is not significant. The correlation between non-spatial random effects is significant with a value of 0.606.

Table 10: Empirically-based correlation estimates between random effects and relative risks with standard deviations, MC errors and 95% CI.

EB Correlation	Estimate	sd	MC error	Lower limit	Upper limit
Relative Risks	0.599	0.146	0.002	0.283	0.856
Spatial random effects	0.007	0.240	0.002	-0.458	0.469
Non-spatial random effects	0.606	0.149	0.002	0.286	0.867

The high correlation in the non-spatial random effects is visible in figure 5. Both maps show the correlation pattern in southern Limburg. The maps of the non-spatial variability terms ($\gamma_{0,1,2}$) are shown in the appendix (Table 9), the common UH term (γ_0) again shows the correlation of the relative risks in the southern of Limburg. The maps of the CH for breast and ovary cancer can be found in the appendix (Table 8), these maps show no common pattern which is in line with the low value of the correlation of the spatial random effects.

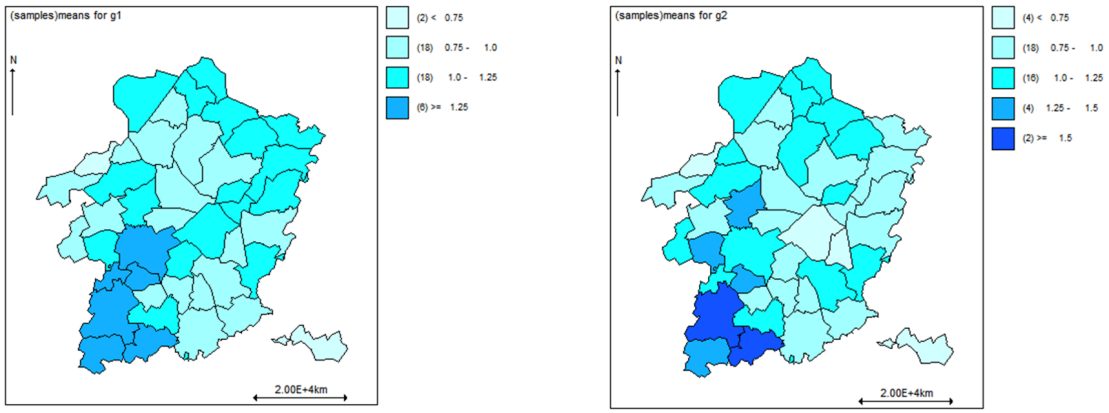

Figure 5: *Left:* Map of UH (g_{i1}) for breast cancer in Limburg. *Right:*Map of UH (g_{i2}) for ovary cancer in Limburg.

The relative risk of breast cancer is shown on the left map of figure 6 and the probability of risk exceedence is shown on the right map. The map of the relative risks show a pattern in southern Limburg and the map of the exceedence probability indicates the significance of this pattern. The relative risks of breast cancer show a significant decreased risk of breast cancer for the following municipalities: Bilzen, Ham, Hoeselt, Maasmechelen, Riemst, Tessenderlo and Tongeren. The relative risk of breast cancer show significant increased risk of breast cancer for the following municipalities: Gingelom, Hasselt, Heers and Sint-Truiden. The relative risk estimates with 95% credible intervals can be found in the appendix (Table 15).

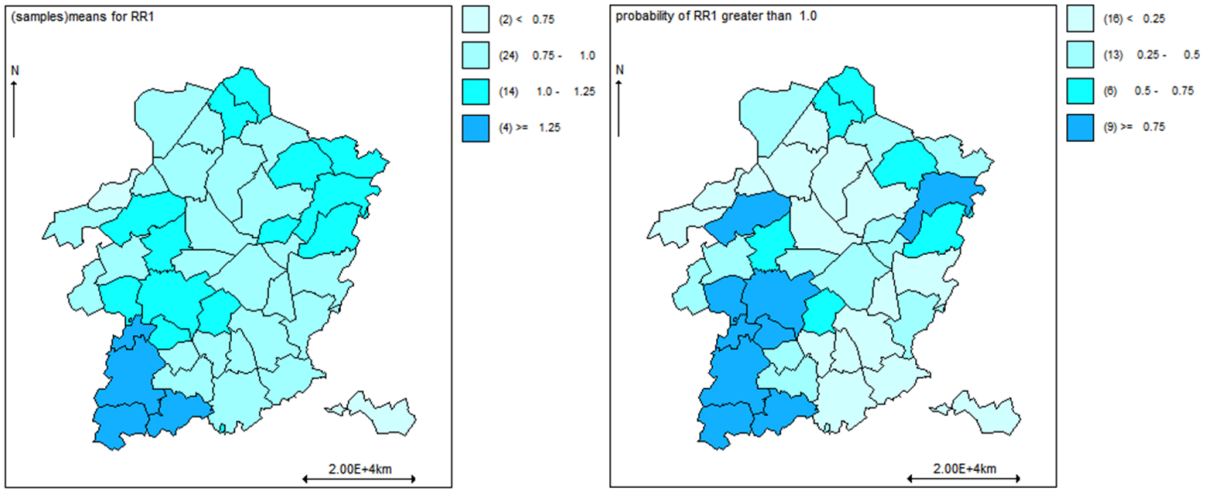

Figure 6: *Left:*Relative risk map for breast cancer in Limburg. *Right:* Probability of exceedence with threshold 1 for breast cancer in Limburg.

The relative risk of ovary cancer is shown on the left map of figure 7 and the probability of risk exceedence is shown on the right map. The map of the relative risks show again a pattern in southern Limburg and the map of the exceedence probability indicates the significance of this pattern. Sint-Truiden shows a significant increased risk of ovary cancer, there were no municipalities with a decreased risk of ovary cancer. The relative risk estimates with 95% credible intervals can be found in the appendix (Table 16).

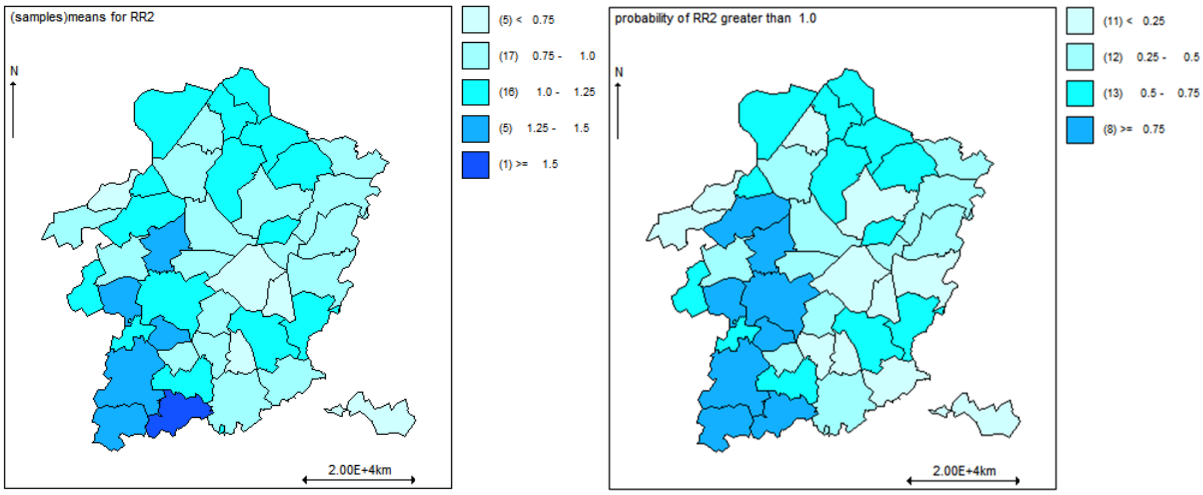

Figure 7: *Left:*Relative risk map for ovary cancer in Limburg. *Right:* Probability of exceedence with threshold 1 for ovary cancer in Limburg.

5 Discussion

When investigating the spatial autocorrelation with Moran's I and Geary's C statistics, it resulted in spatial trends for both cancer types. These trends are more apparent in breast cancer than ovary cancer. By modeling the relative risk this spatial correlation is accounted for.

Model comparison based on DIC for breast cancer in Limburg shows that models including a spatial term in the model performs better than models without spatial term (Poisson-gamma, Poisson-lognormal) in contrast with MSPE where these models without spatial term performs better. But extending the model to the combined model makes the model not perform well based on DIC but has the best predictive ability.
Model comparison based on DIC for ovary cancer in Limburg shows that only considering the spatial random effects performs the best, but extending the model to the combined model makes the model perform the worst based on DIC but has the best predictive ability.
Model comparison based on DIC for the bivariate distribution of breast and ovary cancer in Limburg shows that models based on bivariate distributions for the non-spatial random effects perform better than if these models were based on a univariate distribution. The bivariate gamma distribution for the non-spatial random effects performs better based on DIC in comparison with the bivariate normal distribution. The combination of non-spatial random effects based on the bivariate normal distribution with spatial random effects based on the univariate CAR distribution performs better based on DIC than when the spatial random effects are based on the multivariate CAR distribution, this also holds for the MSPE. The combination of non-spatial random effects based on the bivariate gamma distribution with spatial random effects based on the univariate CAR distribution performs better based on MSPE than when the spatial random effects are based on the multivariate CAR distribution.

The empirically-based correlation between the relative risks is significant which indicates that the bivariate model is needed to understand the association between the cancers which cannot be done with univariate models. Investigation of the empirically-based correlations between the random effects help to understand the association between the cancers. The empirically-based correlation of the spatial random effects is not significant and the correlation of the non-spatial random effects is significant. The maps in figure 5 and 8 confirm these results and show that the non-spatial random effects are the source of the correlation in the relative risks, which leads to the conclusion there is no environmental cause of the correlation between the cancers.

Based on the bivariate results, the municipalities with an increased risk in breast cancer are: Gingelom, Hasselt, Heers and Sint-Truiden. The municipalities with a decreased risk in breast cancer are: Bilzen, Ham, Hoeselt, Maasmechelen, Riemst, Tessenderlo and Tongeren. Sint-Truiden has an increased risk in ovary cancer are, there were no municipalities with a decreased risk in ovary cancer.

Comparing these results with the results of the univariate analysis of breast and ovary cancer, it shows that the results of the univariate analysis doesn't include all the municipalities defined by the bivariate analysis. This shows that there is a correlation in risk between breast and ovary cancer and the need of the bivariate analysis to discover this correlation.

A drawback in this study is the possible effect of the choice of the prior and hyperprior distributions on the model. To see if there is an effect a sensitivity analysis could be executed with different prior distributions.
Another drawback is that the bivariate combined model only functions in the case of two diseases not more which is not a problem for the multivariate convolution model.

6 Conclusion

In effort to investigate the spatial distribution of breast and ovary cancer in Limburg and the correlation between these spatial distributions, different bivariate models were fitted. The bivariate combined model with spatial random effects defined by the univariate CAR distribution and non-spatial effects defined by the bivariate gamma distribution performed the best based on DIC and MSPE. The relative risks show a significant correlation of 0.599 and this correlation is a result of the spatially uncorrelated heterogeneity. The spatial effects were not correlated which tells us that there is no environmental cause of the correlation.

References

[1] Aguero-Valverde, J. (2013). Full Bayes Poisson gamma, Poisson lognormal, and zero inflated random effects models: Comparing the precision of crash frequency estimates. *Accident Analysis & Prevention*, **50**, pp.289-297.

[2] Barua, S., El-Basyouny, K. and Islam, M. (2014). A Full Bayesian multivariate count data model of collision severity with spatial correlation. *Analytic Methods in Accident Research*, **3-4**, pp.28-43.

[3] Besag J. and Kooperberg, C. (1995) On conditional and intrinsic autoregressions. *Biometrika*, **82**, 733746.

[4] Carroll, C. and D. S. Johnson. (2008) The importance of being spa- tial (and reserved): assessing Northern Spotted Owl habitat rela- tionships with hierarchical Bayesian models. *Conservation Biology*

[5] Clayton, D. and Hills, M. (2013). *Statistical models in Epidemiology*. Oxford: Oxford University Press.

[6] Gelfand, A. and Vounatsou, P. (2003). Proper multivariate conditional autoregressive models for spatial data analysis. *Biostatistics*, **4**, pp.1125. **22**(4), pp.1026-1036.

[7] F. Dormann, C., M. McPherson, J., B. Arajo, M., Bivand, R., Bolliger, J., Carl, G., G. Davies, R., Hirzel, A., Jetz, W., Daniel Kissling, W., Khn, I., Ohlemller, R., R. Peres-Neto, P., Reineking, B., Schrder, B., M. Schurr, F. and Wilson, R. (2007). Methods to account for spatial autocorrelation in the analysis of species distributional data: a review. *Ecography*, **30**(5), pp.609-628.

[8] Geary, R. (1954). The Contiguity Ratio and Statistical Mapping. *The Incorporated Statistician*, **5**(3), pp.115-146.

[9] Gelman, A., Carline, B., Stern, S., Rubin D., B. (2004) *Bayesian Data analysis*, 2nd ed,Boca Rotan: Chapman & Hall\CrC Press.

[10] Greco, F. and Trivisano, C. (2009). A multivariate CAR model for improving the estimation of relative risks. *Statistics in Medicine*, **28**(12), pp.1707-1724.

[11] Hall, P. (2012). Current knowledge and tomorrows challenges of breast, ovarian and prostate cancer genetics. *Journal Of Internal Medicine*, **271**(4), 318-320.

[12] Hennessy, B., Coleman, R., and Markman, M. (2009). Ovarian cancer. *The Lancet*, **374**(9698), pp.1371-1382.

[13] Knorr-Held, L. and Best, N. G. (2001) A shared component model for detecting joint and selective clustering of two diseases. *Journal of the Royal Statistical Society*,**164**, 7385.

[14] Lawson, A. B. (2013) Bayesian Disease Mapping: Hierarchical Modeling in Spatial Epidemiology,2nd ed, Boca Rotan: Chapman & Hall.

[15] Lawson, A. B., Biggeri, A. B., Boehning, D., Lesaffre, E., Viel, J.-F., Clark, A., Schlattmann, P., and Divino, F. (2000) Disease mapping models: an empirical evaluation. *Statistics in Medicine*, **19**, pp.22172241.

[16] McPherson, K., Steel, C. M., and Dixon, J. M. (2000). Breast cancerepidemiology, risk factors, and genetics. *BMJ: British Medical Journal*, **321**(7261), pp.624628.

[17] Molenberghs, G., Verbeke, G., and Demtrio, G. (2007) An extended random-effects approach to modeling repeated, overdispersed count data. *Lifetime Data Analysis*, **13**, pp.513531.

[18] Molenberghs, G., Verbeke, G., Demtrio, C. and Vieira, A. (2010). A Family of Generalized Linear Models for Repeated Measures with Normal and Conjugate Random Effects. *Statistical Science*, **25**(3), pp.325-347.

[19] Moran, P. (1950). Notes on Continuous Stochastic Phenomena. *Biometrika*, **37**(1/2), pp.17-23.

[20] Neyens, T., Faes, C. and Molenberghs, G. (2012). A generalized Poisson-gamma model for spatially overdispersed data. *Spatial and Spatio-temporal Epidemiology*, **3**(3), pp.185-194.

[21] Neyens, T., Lawson, A., Kirby, R. and Faes, C. (2016). The bivariate combined model for spatial data analysis. *Statistics in Medicine*, **35**(18), pp.3189-3202.

[22] Xu, Y., Kennedy E. (2015). An Introduction to Spatial Analysis in Social Science Research. *Tutorials in Quantitative Methods for Psychology*, **11**(1), pp.22-31.

[23] Spiegelhalter, D., Best, N., Carlin, B. and van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **64**(4), pp.583-639.

[24] Torabi, M. (2014). Spatial generalized linear mixed models with multivariate CAR models for areal data. *Spatial Statistics*, **10**, pp.12-26.

[25] Tripepi, G., Jager, K., Dekker, F. and Zoccali, C. (2010). Stratification for Confounding Part 2: Direct and Indirect Standardization. *Nephron Clinical Practice*, **116**(4), pp.c322-c325.

[26] Waller, L. and Gotway, C. (2004). *Applied spatial statistics for public health data*. Hoboken, N.J.: John Wiley & Sons.

Appendix

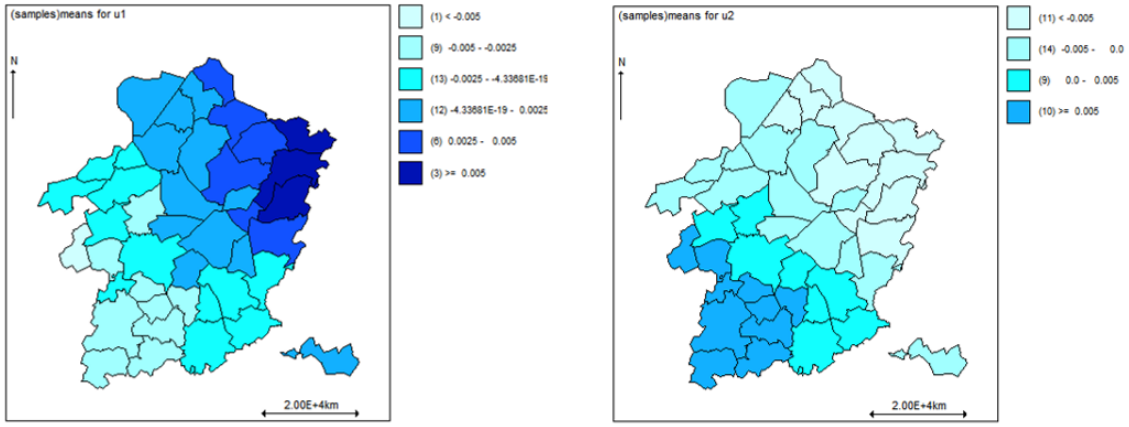

Figure 8: *Left:* Map of CH (υ_{i1}) for breast cancer in Limburg. *Right:*Map of CH (υ_{i2}) for ovary cancer in Limburg.

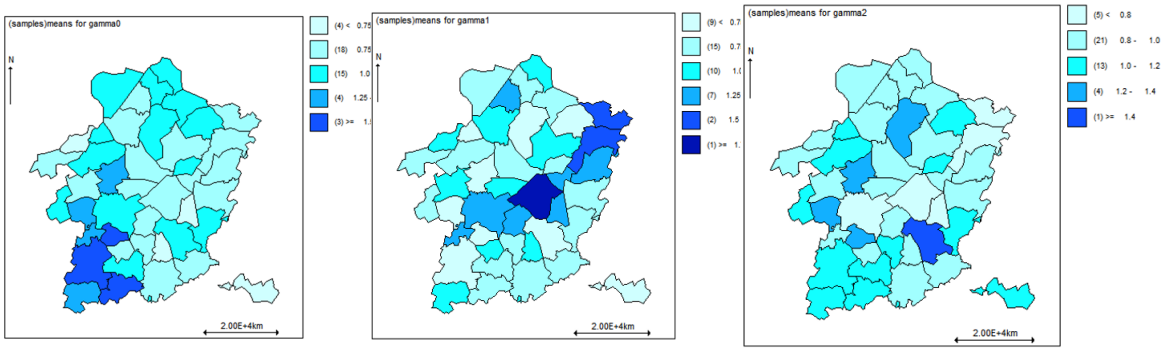

Figure 9: *Left:* Map of common UH term (γ_0). *Middle:* Map of breast cancer specific UH term(γ_1). *Right:* Map of ovary cancer specific UH term (γ_2).

Code Bivariate combined model

```
 ## UCAR BIGAM ##
model
{for (i in 1:N)
{
Y1[i] ~dpois(mu1[i])
log(mu1[i]) <- log(E1[i]) + alpha1 + u[1,i]+ log(g1[i])

Y2[i] ~dpois(mu2[i])
log(mu2[i]) <- log(E2[i]) + alpha2 + u[2,i]+ log(g2[i])

RR1[i]<-exp(alpha1+u[1,i]+log(g1[i]))
RR2[i]<-exp(alpha2+u[2,i]+log(g2[i]))
```

```
## Goodness of fit
Opred1[i] ~ dpois(mu1[i])
pres1[i] <- Y1[i] - Opred1[i]
SPE1[i] <- pow(pres1[i],2)
Opred2[i] ~ dpois(mu2[i])
pres2[i] <- Y2[i] - Opred2[i]
SPE2[i] <- pow(pres2[i],2)

#Bivariate Gamma specification
g1[i] <-var1*(k0*gamma0[i] + k1*gamma1[i])
g2[i] <-var2*(k0*gamma0[i] + k2*gamma2[i])
U1[i]<-u[1,i]
U2[i]<-u[2,i]
gamma0[i]~dgamma(1,1)
gamma1[i]~dgamma(1,1)
gamma2[i]~dgamma(1,1)}

# CAR for the prior distrbution of the random effects
u[1,1:N] ~ car.normal(adj[ ], weights1[ ], num[ ], tau.u1)
for (k in 1:sumNumNeigh)
{weights1[k]<-1}
u[2,1:N] ~ car.normal(adj[ ], weights2[ ], num[ ], tau.u2)
for (k in 1:sumNumNeigh)
{weights2[k]<-1}

# CAR - Prior Specification
tau.u1 ~ dgamma(0.5, 0.0005) # prior on precision
tau.u2 ~ dgamma(0.5, 0.0005)
sigma.u1 <- 1/tau.u1
sigma.u2<-1/tau.u2

# Empirically based Correlations
##RR- correlation
mu.1<-mean(RR1[])
mu.2<-mean(RR2[])
sd1<-sd(RR1[])
sd2<-sd(RR2[])
mu12<-inprod(RR1[],RR2[])/N
CRR12<-(mu12-mu.1*mu.2)/(sd1*sd2)
##Spatial random effects corr
Spmu.1<-mean(u[1,])
Spmu.2<-mean(u[2,])
SPsd1<-sd(u[1,])
SPsd2<-sd(u[2,])
SPmu12<-inprod(u[1,],u[2,])/N
SPC12<-(SPmu12-Spmu.1*Spmu.2)/(SPsd1*SPsd2)
##Non-spatial Random effects
nsmu.1<-mean(g1[])
```

```
nsmu.2<-mean(g2[])
nssd1<-sd(g1[])
nssd2<-sd(g2[])
NSPmu12<-inprod(g1[],g2[])/N
NSPC12<-(NSPmu12-nsmu.1*nsmu.2)/(nssd1*nssd2)

# other priors
alpha1 ~ dflat()
alpha2 ~ dflat()
var1<-1/(k0+k1)
var2<-1/(k0+k2)
k0~dexp(1)
k1~dexp(1)
k2~dexp(1)

MSPE1 <- mean(SPE1[])
MSPE2 <- mean(SPE2[])
MSPE<-MSPE1+MSPE2}
```

Table 11: The observed counts, the expected counts and SIR with 95% CI based on the error factor method for breast cancer in Limburg.

Municipality	**Observed**	**Expected**	**SIR**	**Lower CI**	**Upper CI**
ALKEN	92	77.480	1.187	0.968	1.457
AS	51	50.955	1.001	0.761	1.317
BERINGEN	286	272.691	1.049	0.934	1.178
BILZEN	174	210.891	0.825	0.711	0.957
BOCHOLT	83	82.965	1.000	0.807	1.241
BORGLOON	76	78.659	0.966	0.772	1.210
BREE	102	103.127	0.989	0.815	1.201
DIEPENBEEK	127	121.719	1.043	0.877	1.242
DILSEN-STOKKEM	134	129.269	1.037	0.875	1.228
GENK	427	434.294	0.983	0.894	1.081
GINGELOM	78	58.744	1.328	1.064	1.658
HALEN	58	63.634	0.911	0.705	1.179
HAM	46	67.041	0.686	0.514	0.916
HAMONT-ACHEL	106	99.545	1.065	0.880	1.288
HASSELT	653	537.144	1.216	1.126	1.313
HECHTEL-EKSEL	69	77.070	0.895	0.707	1.134
HEERS	68	51.054	1.332	1.050	1.689
HERK-DE-STAD	96	84.426	1.137	0.931	1.389
HERSTAPPE	1	0.699	1.430	0.201	10.149
HEUSDEN-ZOLDER	209	205.192	1.019	0.889	1.166
HOESELT	51	65.931	0.774	0.588	1.018
HOUTHALEN-HELCHTEREN	175	192.333	0.910	0.785	1.055
KINROOI	82	80.008	1.025	0.825	1.273
KORTESSEM	42	56.051	0.749	0.554	1.014
LANAKEN	172	178.801	0.962	0.828	1.117
LEOPOLDSBURG	90	101.809	0.884	0.719	1.087
LOMMEL	215	221.050	0.973	0.851	1.112
LUMMEN	94	97.219	0.967	0.790	1.184
MAASEIK	184	167.433	1.099	0.951	1.270
MAASMECHELEN	200	248.382	0.805	0.701	0.925
MEEUWEN-GRUITRODE	72	81.947	0.879	0.697	1.107
NEERPELT	113	112.195	1.007	0.838	1.211
NIEUWERKERKEN	63	48.656	1.295	1.011	1.657
OPGLABBEEK	61	61.729	0.988	0.769	1.270
OVERPELT	85	92.652	0.917	0.742	1.135
PEER	87	100.362	0.867	0.703	1.070
RIEMST	90	116.756	0.771	0.627	0.948
SINT-TRUIDEN	384	297.840	1.289	1.167	1.425
TESSENDERLO	96	116.337	0.825	0.676	1.008
TONGEREN	217	232.355	0.934	0.818	1.067
VOEREN	14	30.315	0.462	0.274	0.780
WELLEN	45	47.939	0.939	0.701	1.257
ZONHOVEN	130	136.397	0.953	0.803	1.132
ZUTENDAAL	41	47.905	0.856	0.630	1.162

Table 12: The observed counts, the expected counts and SIR with 95% CI based on the error factor method for ovary cancer in Limburg.

Municipality	**Observed**	**Expected**	**SIR**	**Lower CI**	**Upper CI**
ALKEN	11	5.552	1.981	1.097	3.578
AS	2	3.657	0.547	0.137	2.187
BERINGEN	23	19.738	1.165	0.774	1.754
BILZEN	20	15.205	1.315	0.849	2.039
BOCHOLT	6	5.939	1.010	0.454	2.249
BORGLOON	7	5.624	1.245	0.593	2.611
BREE	10	7.518	1.330	0.716	2.472
DIEPENBEEK	6	8.767	0.684	0.307	1.523
DILSEN-STOKKEM	7	9.226	0.759	0.362	1.591
GENK	20	31.833	0.628	0.405	0.974
GINGELOM	6	4.267	1.406	0.632	3.130
HALEN	6	4.605	1.303	0.585	2.900
HAM	2	4.875	0.410	0.103	1.640
HAMONT-ACHEL	7	7.188	0.974	0.464	2.043
HASSELT	41	38.997	1.051	0.774	1.428
HECHTEL-EKSEL	4	5.534	0.723	0.271	1.926
HEERS	7	3.688	1.898	0.905	3.982
HERK-DE-STAD	10	6.015	1.662	0.894	3.090
HERSTAPPE	0	0.046	0.000	0.000	NaN
HEUSDEN-ZOLDER	22	14.919	1.475	0.971	2.239
HOESELT	3	4.731	0.634	0.205	1.966
HOUTHALEN-HELCHTEREN	12	13.842	0.867	0.492	1.526
KINROOI	3	5.619	0.534	0.172	1.656
KORTESSEM	4	3.945	1.014	0.381	2.702
LANAKEN	14	12.788	1.095	0.648	1.849
LEOPOLDSBURG	9	7.468	1.205	0.627	2.316
LOMMEL	16	15.900	1.006	0.616	1.643
LUMMEN	5	6.951	0.719	0.299	1.728
MAASEIK	8	12.110	0.661	0.330	1.321
MAASMECHELEN	13	17.933	0.725	0.421	1.248
MEEUWEN-GRUITRODE	4	5.865	0.682	0.256	1.817
NEERPELT	8	8.035	0.996	0.498	1.991
NIEUWERKERKEN	3	3.468	0.865	0.279	2.682
OPGLABBEEK	6	4.455	1.347	0.605	2.998
OVERPELT	4	6.748	0.593	0.222	1.579
PEER	9	7.229	1.245	0.648	2.393
RIEMST	7	8.426	0.831	0.396	1.743
SINT-TRUIDEN	34	21.416	1.588	1.134	2.222
TESSENDERLO	6	8.407	0.714	0.321	1.589
TONGEREN	15	16.766	0.895	0.539	1.484
VOEREN	2	2.156	0.928	0.232	3.709
WELLEN	3	3.430	0.875	0.282	2.712
ZONHOVEN	8	9.720	0.823	0.412	1.646
ZUTENDAAL	1	3.397	0.294	0.041	2.090

Table 13: The estimated relative risks and standard deviations with 95% CI based on the convolution model for breast cancer in Limburg.

Municipality	**Mean**	**sd**	**Lower CI**	**Upper CI**
ALKEN	1.128	0.084	0.971	1.299
AS	0.969	0.077	0.825	1.128
BERINGEN	1.004	0.052	0.906	1.112
BILZEN	0.871	0.051	0.772	0.974
BOCHOLT	0.985	0.071	0.851	1.132
BORGLOON	1.018	0.077	0.871	1.174
BREE	0.987	0.072	0.853	1.134
DIEPENBEEK	1.004	0.067	0.880	1.141
DILSEN-STOKKEM	1.003	0.071	0.871	1.148
GENK	0.975	0.041	0.897	1.058
GINGELOM	1.243	0.117	1.029	1.488
HALEN	0.975	0.091	0.804	1.161
HAM	0.823	0.077	0.676	0.976
HAMONT-ACHEL	1.030	0.083	0.877	1.201
HASSELT	1.169	0.045	1.084	1.260
HECHTEL-EKSEL	0.926	0.069	0.797	1.065
HEERS	1.176	0.100	0.994	1.388
HERK-DE-STAD	1.093	0.083	0.939	1.263
HERSTAPPE	0.978	0.174	0.674	1.365
HEUSDEN-ZOLDER	1.011	0.057	0.902	1.126
HOESELT	0.882	0.072	0.743	1.027
HOUTHALEN-HELCHTEREN	0.935	0.053	0.832	1.041
KINROOI	1.008	0.081	0.857	1.176
KORTESSEM	0.942	0.076	0.789	1.088
LANAKEN	0.936	0.060	0.824	1.058
LEOPOLDSBURG	0.899	0.070	0.767	1.043
LOMMEL	0.965	0.057	0.856	1.081
LUMMEN	0.999	0.071	0.864	1.142
MAASEIK	1.044	0.063	0.927	1.172
MAASMECHELEN	0.856	0.050	0.760	0.956
MEEUWEN-GRUITRODE	0.937	0.068	0.806	1.074
NEERPELT	0.986	0.068	0.858	1.126
NIEUWERKERKEN	1.176	0.102	0.989	1.390
OPGLABBEEK	0.976	0.076	0.834	1.133
OVERPELT	0.936	0.071	0.802	1.080
PEER	0.918	0.064	0.796	1.048
RIEMST	0.847	0.064	0.723	0.975
SINT-TRUIDEN	1.238	0.059	1.127	1.358
TESSENDERLO	0.858	0.070	0.725	0.999
TONGEREN	0.946	0.052	0.847	1.050
VOEREN	0.937	0.068	0.767	1.037
WELLEN	1.033	0.087	0.865	1.207
ZONHOVEN	0.974	0.065	0.851	1.105
ZUTENDAAL	0.910	0.076	0.767	1.065

Table 14: The estimated relative risks and standard deviations with 95% CI based on the CAR model for ovary cancer in Limburg.

Municipality	**Mean**	**sd**	**Lower CI**	**Upper CI**
ALKEN	1.216	0.198	0.920	1.690
AS	0.870	0.130	0.608	1.119
BERINGEN	1.009	0.122	0.790	1.274
BILZEN	1.006	0.122	0.784	1.272
BOCHOLT	0.942	0.138	0.688	1.239
BORGLOON	1.153	0.177	0.864	1.562
BREE	0.958	0.155	0.685	1.305
DIEPENBEEK	0.962	0.133	0.698	1.232
DILSEN-STOKKEM	0.858	0.149	0.569	1.154
GENK	0.881	0.102	0.671	1.070
GINGELOM	1.284	0.282	0.858	1.961
HALEN	1.144	0.244	0.754	1.725
HAM	0.915	0.173	0.584	1.274
HAMONT-ACHEL	0.959	0.189	0.627	1.378
HASSELT	1.055	0.102	0.870	1.272
HECHTEL-EKSEL	0.950	0.132	0.701	1.229
HEERS	1.241	0.229	0.898	1.785
HERK-DE-STAD	1.165	0.198	0.858	1.634
HERSTAPPE	1.067	0.350	0.531	1.893
HEUSDEN-ZOLDER	1.096	0.150	0.854	1.448
HOESELT	0.984	0.144	0.704	1.285
HOUTHALEN-HELCHTEREN	0.950	0.108	0.740	1.169
KINROOI	0.883	0.160	0.577	1.211
KORTESSEM	1.068	0.140	0.818	1.377
LANAKEN	0.966	0.139	0.710	1.265
LEOPOLDSBURG	0.999	0.171	0.701	1.384
LOMMEL	0.960	0.148	0.691	1.280
LUMMEN	1.048	0.148	0.779	1.376
MAASEIK	0.872	0.123	0.627	1.112
MAASMECHELEN	0.864	0.122	0.620	1.096
MEEUWEN-GRUITRODE	0.918	0.127	0.673	1.180
NEERPELT	0.952	0.142	0.691	1.260
NIEUWERKERKEN	1.153	0.190	0.846	1.593
OPGLABBEEK	0.929	0.133	0.677	1.208
OVERPELT	0.918	0.148	0.635	1.226
PEER	0.964	0.132	0.726	1.252
RIEMST	0.976	0.147	0.698	1.285
SINT-TRUIDEN	1.280	0.184	0.983	1.696
TESSENDERLO	0.911	0.178	0.574	1.284
TONGEREN	1.030	0.125	0.795	1.295
VOEREN	0.991	0.051	0.893	1.093
WELLEN	1.134	0.171	0.849	1.530
ZONHOVEN	0.968	0.136	0.708	1.254
ZUTENDAAL	0.889	0.136	0.616	1.151

Table 15: The estimated relative risks and standard deviations with 95% CI based on the bivariate gamma model with univariate CAR spatial random effects for breast cancer in Limburg.

Municipality	**Mean**	**sd**	**Lower CI**	**Upper CI**
ALKEN	1.218	0.121	0.992	1.467
AS	0.979	0.134	0.734	1.258
BERINGEN	1.050	0.061	0.935	1.174
BILZEN	0.842	0.063	0.723	0.970
BOCHOLT	1.000	0.106	0.803	1.219
BORGLOON	0.977	0.108	0.776	1.200
BREE	01.003	0.096	0.825	1.199
DIEPENBEEK	1.029	0.090	0.860	1.213
DILSEN-STOKKEM	1.025	0.087	0.862	1.202
GENK	0.977	0.048	0.886	1.072
GINGELOM	1.321	0.145	1.050	1.619
HALEN	0.930	0.117	0.716	1.172
HAM	0.681	0.097	0.505	0.884
HAMONT-ACHEL	1.059	0.100	0.871	1.263
HASSELT	1.210	0.048	1.119	1.306
HECHTEL-EKSEL	0.889	0.104	0.697	1.104
HEERS	1.348	0.156	1.059	1.672
HERK-DE-STAD	1.155	0.114	0.943	1.388
HERSTAPPE	1.070	0.677	0.199	2.773
HEUSDEN-ZOLDER	1.033	0.070	0.901	1.173
HOESELT	0.771	0.104	0.581	0.990
HOUTHALEN-HELCHTEREN	0.909	0.067	0.781	1.045
KINROOI	1.004	0.108	0.802	1.227
KORTESSEM	0.766	0.113	0.561	1.002
LANAKEN	0.966	0.072	0.830	1.112
LEOPOLDSBURG	0.899	0.091	0.728	1.086
LOMMEL	0.973	0.065	0.849	1.104
LUMMEN	0.956	0.096	0.777	1.155
MAASEIK	1.083	0.079	0.934	1.244
MAASMECHELEN	0.803	0.056	0.698	0.916
MEEUWEN-GRUITRODE	0.872	0.100	0.686	1.079
NEERPELT	1.005	0.092	0.834	1.193
NIEUWERKERKEN	1.261	0.155	0.977	1.583
OPGLABBEEK	1.005	0.123	0.778	1.261
OVERPELT	0.905	0.096	0.726	1.105
PEER	0.884	0.091	0.713	1.071
RIEMST	0.775	0.079	0.627	0.937
SINT-TRUIDEN	1.295	0.065	1.171	1.424
TESSENDERLO	0.822	0.082	0.670	0.989
TONGEREN	0.932	0.062	0.814	1.057
VOEREN	0.509	0.123	0.296	0.778
WELLEN	0.935	0.134	0.693	1.217
ZONHOVEN	0.948	0.082	0.795	1.114
ZUTENDAAL	0.833	0.127	0.604	1.096

Table 16: The estimated relative risks and standard deviations with 95% CI based on the bivariate gamma model with univariate CAR spatial random effects for ovary cancer in Limburg.

Municipality	**Mean**	**sd**	**Lower CI**	**Upper CI**
ALKEN	1.460	0.296	0.954	2.132
AS	0.842	0.302	0.269	1.443
BERINGEN	1.141	0.181	0.801	1.514
BILZEN	1.087	0.205	0.751	1.559
BOCHOLT	1.015	0.253	0.533	1.534
BORGLOON	1.086	0.258	0.613	1.645
BREE	1.133	0.241	0.701	1.660
DIEPENBEEK	0.876	0.250	0.404	1.369
DILSEN-STOKKEM	0.909	0.238	0.454	1.379
GENK	0.710	0.151	0.440	1.023
GINGELOM	1.378	0.329	0.746	2.056
HALEN	1.059	0.271	0.580	1.661
HAM	0.597	0.229	0.190	1.074
HAMONT-ACHEL	1.039	0.252	0.557	1.551
HASSELT	1.115	0.158	0.814	1.431
HECHTEL-EKSEL	0.843	0.248	0.370	1.342
HEERS	1.506	0.335	0.892	2.227
HERK-DE-STAD	1.339	0.276	0.843	1.945
HERSTAPPE	1.064	0.779	0.113	3.036
HEUSDEN-ZOLDER	1.262	0.212	0.899	1.742
HOESELT	0.737	0.242	0.292	1.238
HOUTHALEN-HELCHTEREN	0.908	0.189	0.545	1.291
KINROOI	0.807	0.281	0.282	1.357
KORTESSEM	0.850	0.253	0.406	1.411
LANAKEN	1.052	0.201	0.680	1.476
LEOPOLDSBURG	1.026	0.232	0.614	1.541
LOMMEL	1.011	0.187	0.658	1.397
LUMMEN	0.874	0.250	0.394	1.366
MAASEIK	0.847	0.236	0.410	1.312
MAASMECHELEN	0.777	0.164	0.465	1.109
MEEUWEN-GRUITRODE	0.811	0.241	0.352	1.295
NEERPELT	1.016	0.233	0.569	1.492
NIEUWERKERKEN	1.162	0.355	0.462	1.856
OPGLABBEEK	1.105	0.274	0.611	1.699
OVERPELT	0.780	0.246	0.317	1.270
PEER	1.023	0.233	0.619	1.552
RIEMST	0.818	0.205	0.439	1.255
SINT-TRUIDEN	1.489	0.201	1.119	1.914
TESSENDERLO	0.793	0.214	0.390	1.229
TONGEREN	0.938	0.182	0.590	1.304
VOEREN	0.610	0.256	0.216	1.225
WELLEN	0.932	0.289	0.383	1.530
ZONHOVEN	0.908	0.220	0.483	1.350
ZUTENDAAL	0.655	0.285	0.150	1.229

Auteursrechtelijke overeenkomst

Ik/wij verlenen het wereldwijde auteursrecht voor de ingediende eindverhandeling:
The bivariate spatial modelling of breast and ovary cancer in Limburg

Richting: **Master of Statistics-Epidemiology & Public Health Methodology**
Jaar: **2018**

in alle mogelijke mediaformaten, - bestaande en in de toekomst te ontwikkelen - , aan de Universiteit Hasselt.

Niet tegenstaand deze toekenning van het auteursrecht aan de Universiteit Hasselt behoud ik als auteur het recht om de eindverhandeling, - in zijn geheel of gedeeltelijk -, vrij te reproduceren, (her)publiceren of distribueren zonder de toelating te moeten verkrijgen van de Universiteit Hasselt.

Ik bevestig dat de eindverhandeling mijn origineel werk is, en dat ik het recht heb om de rechten te verlenen die in deze overeenkomst worden beschreven. Ik verklaar tevens dat de eindverhandeling, naar mijn weten, het auteursrecht van anderen niet overtreedt.

Ik verklaar tevens dat ik voor het materiaal in de eindverhandeling dat beschermd wordt door het auteursrecht, de nodige toelatingen heb verkregen zodat ik deze ook aan de Universiteit Hasselt kan overdragen en dat dit duidelijk in de tekst en inhoud van de eindverhandeling werd genotificeerd.

Universiteit Hasselt zal mij als auteur(s) van de eindverhandeling identificeren en zal geen wijzigingen aanbrengen aan de eindverhandeling, uitgezonderd deze toegelaten door deze overeenkomst.

Voor akkoord,

Kodalci, Leyla

Datum: **15/06/2018**