Disease Mapping

HOMEWORK 3 PROJECT

Spatial distribution of cervix cancer incidence in women in the Antwerp district in Belgium: age-standardised impact of urbanisation and identification of high risk areas

By
Wim van der Elst
Nico Alemngu
Soutrik Banerjee

Master of Science in Biostatistics
2009 – 2010

CONTENTS

LIST OF TABLES

LIST OF FIGURES

1 ABSTRACT

The aim of the present study was to evaluate the geographical distribution and the impact of urbanisation on cervix cancer incidence in $N = 70$ municipalities in the Antwerp district.

Age-standardised expected cancer rates were computed according to the demographic distribution of the entire Flanders region. Exploratory analyses were conducted and spatial models were fitted to the data (*i.e.*, with Poisson-gamma, Poisson-lognormal and Conditional Auto-Regressive (CAR) models).

This study provides evidence that there is a possibility of increased risk of cervix cancer incidence mainly in the western parts of the Antwerp district when adjusted for the urbanisation index. In addition, no major spatial correlation was detected. More urban areas were found to carry a negative impact on the cervix cancer incidence overall over less urban areas.

2 INTRODUCTION

2.1 OVERVIEW OF CERVIX CANCER IN WOMEN

Cervix cancer is an important public health problem. It represents the second most common cancer in woman worldwide (Bosch *et al.*, 1995). Age-adjusted incidence rates vary from about 10 per 100,000 per year (in industrialised nations) to about 40 per 100,000 per year (in developing countries) (Bosch *et al.*, 1995). Five-year survival rates vary from 20% to 95%, depending on the stage of the disease at diagnosis (Rohan & Shah, 2004).

The central aetiological factors of cervix cancer are genital Human Papilloma Virus (HPV), of which at least 20 different types exist, and Herpes Simplex Virus, type 2 (HSV-2) (Bosch *et al.*, 1995; Clifford *et al.*, 2003; Walboomers *et al.*, 1999). Risk factors for cervical cancer include a large number of sexual partners, a low socio-economic status, long-term use of oral contraceptives, older age, and smoking (Bosch *et al.*, 1992; Smith *et al.*, 2003). Recently, anti-HPV vaccines have been developed and approved for use (*e.g.*, *Gardasil*® and *Cervarix*®). At present, large vaccination programmes are being conducted in Belgium and other countries (mainly targeted at pre-adolescent girls).

The aims of the present study were threefold. First, we explored the geographical distribution of cervix cancer in the Antwerp district in Belgium. Second, we evaluated whether urbanisation was associated with an increased risk of cervix cancer (because at least some of the risk factors for cervix cancer were also associated with the urbanisation level of a municipality, *e.g.*, socio-economical status). However, as the distribution of age could vary according to the degree of urbanisation within a municipality, which in itself is a confounder for cervix cancer, we standardised the expected rates by considering 5-year age groups, taking into account the entire Flanders region of Belgium. And third, we tried to identify high risk municipalities within the Antwerp district.

2.2 DATA DESCRIPTION

Data on cervix cancer incidence were collected in $N = 308$ municipalities in the Flanders region in Belgium. For each municipality, data on disease counts and number of people at risk were available as a function of age groups (in 5-year age groups). Indices which reflected the functional and morphological urbanisation, as well as combined information on these two types of urbanisation, named "agglo" for each municipality were available (coded as 1 = agglomeration, 2 = suburb, 3 = residential area, and 4 = not an urban area).

In the present study, we were mainly interested in estimating the risk ratios (RR) of cervix cancer incidence in the Antwerp district in Belgium. However to obtain more stable rates for cervix cancer incidence, pooled data from the entire Flanders region was used to obtain the expected age-standardised cervix cancer incidence in the Antwerp district. A map of the municipalities of the Antwerp region is provided in Figure 1.

Figure 1. Map of the n = 70 municipalities in the Antwerp district.

2.3 STUDY OBJECTIVES

The main objectives of the present study were: (*i*) to explore the geographical distribution of cervix cancer in the Antwerp district, (*ii*) to evaluate whether urbanisation was associated with an increased risk of cervix cancer, and if so, to quantify the magnitude of this effect, and (*iii*) to identify clusters of areas in which cervix cancer had a high incidence.

3 METHODS

3.1 EXPLORATORY DATA ANALYSIS

Smoothing

A map with crude counts (or crude rates) of cervix cancer is not the best tool for drawing definite conclusions, because regions with a larger population are also likely to have a larger number of cases. Rates account for population differences within municipalities, but a map of rates may still obscure the spatial pattern (*e.g.*, rates based on small populations may be elevated artificially). Three different spatial smoothing techniques were used to avoid this problem.

Locally weighted averages. In this method, smoothed values are obtained by averaging values of neighbouring regions. We used disk smoothing procedures (with radius 0.10 degree), with binary weights and with weights based on population sizes.

Nonparametric smoothing. In this method, the locally smoothed rate is computed as:

$\hat{r}_i = \frac{\sum_{i=1}^{N} w_{ij} Y_j}{\sum_{i=1}^{N} w_{ij} n_j}$, with $w_{ij} = kern\left(\frac{s_i - s_j}{b}\right)$

s_i and s_j are the spatial locations of the centroids of regions *i* and *j*, Y_i is the number of diseases in the area and n_i is the number of women in the area. We used a small value of b in the present paper to keep smoothing minimal, *i.e.*, $b = 0.01$.

Empirical Bayes (EB) smoothing (global and local). In EB smoothing, the shrinkage-factor $C_i = v_{\theta i} / ((v_{\theta i} + m_{\theta i})/n_i)$ defines how much the crude rates "shrink" toward the prior mean (with $v_{\theta i}$ = the variance of the prior distribution, $m_{\theta i}$ = the mean of the prior distribution, and n_i = the number of women in the area). Global EB smoothes towards the global mean (particularly, if the population size is small). Local EB smoothes to the neighbourhood mean (particularly, if the population size is small).

Table 1. Age-standardised observed rates, expected rates, Standardised Incidence Rates (SIR) (95% LL, 95% UL) per community in the Antwerp district of Belgium showing the cervix cancer incidence data.

Municipality	Observed	Expected	SIR	95% LL	95% UL	Municipality	Observed	Expected	SIR	95% LL	95% UL
Aartselaar	2	1,639242	1,22	0,3051	4,8785	Lier	2	4,436634	0,4508	0,112739	1,802511
Antwerpen	112	67,88539	**1,65**	**1,3709**	**1,9855**	Mechelen	11	10,34318	1,0635	0,588962	1,920394
Boechout	0	1,428631	0			Nijlen	1	2,367783	0,4223	0,05949	2,998301
Boom	4	2,202352	1,816	0,6817	4,8393	Putte	1	1,909689	0,5236	0,07376	3,71753
Borsbeek	1	1,183442	0,845	0,119	5,9989	Puurs	1	1,908699	0,5239	0,073798	3,719459
Brasscaat	5	4,39764	1,137	0,4732	2,7317	Sint-Amands	0	0,967931	0		
Brecht	6	2,326415	**2,579**	**1,1587**	**5,7408**	Sint-Katelijne-Waver	1	2,403984	0,416	0,058594	2,95315
Edegem	5	2,940501	1,7	0,7077	4,0853	Willebroek	4	3,025025	1,3223	0,496275	3,523219
Essen	1	1,715606	0,583	0,0821	4,1381	Arendonk	2	1,281193	1,561	0,390404	6,241902
Hemiksem	4	1,23387	**3,242**	**1,2167**	**8,6377**	Baarle-Hertog	0	0,251625	0		
Hove	0	0,99218	0			Balen	0	2,157589	0		
Kalmthout	0	1,974544	0			Beerse	0	1,450582	0		
Kapellen	3	3,31662	0,905	0,2917	2,8046	Dessel	0	0,886994	0		
Kontich	1	2,198536	0,455	0,0641	3,2291	Geel	2	3,904935	0,5122	0,12809	2,047942
Lint	1	0,718707	1,391	0,196	9,8779	Grobbendonk	0	1,108488	0		
Mortsel	2	3,483703	0,574	0,1436	2,2956	Herentals	5	2,969118	1,684	0,700917	4,045931
Niel	1	1,163778	0,859	0,121	6,1002	Herenthout	0	0,904311	0		
Ranst	3	1,854956	1,617	0,5216	5,0146	Herselt	1	1,553368	0,6438	0,090679	4,570281
Rumst	2	1,787475	1,119	0,2798	4,474	Hoogstraten	0	1,667319	0		
Schelle	1	0,904987	1,105	0,1556	7,8447	Hulshout	1	0,890803	1,1226	0,158125	7,969586
Schilde	1	2,3762	0,421	0,0593	2,9877	Kasterlee	0	1,610961	0		
Schoten	5	4,182418	1,195	0,4976	2,8722	Lille	0	1,383069	0		
Stabroek	5	1,679637	**2,977**	**1,239**	**7,152**	Meerhout	1	1,063083	0,9407	0,1325	6,678057
Wijnegem	0	1,261311	0			Merksplas	1	0,665519	1,5026	0,211652	10,66735

Municipality	Observed	Expected	SIR	95% LL	95% UL	Municipality	Observed	Expected	SIR	95% LL	95% UL
Wommelgem	3	1,38784	2,162	0,6972	6,7024	Mol	2	3,620961	0,5523	0,138135	2,208552
Wuustwezel	1	1,59803	0,626	0,0881	4,4426	Olen	0	1,062296	0		
Zandhoven	0	1,23868	0			Oud-Turnhout	1	1,265351	0,7903	0,11132	5,610561
Zoersel	2	2,248693	0,889	0,2224	3,5563	Ravels	1	1,289194	0,7757	0,109261	5,506796
Zwijndrecht	2	2,279316	0,877	0,2194	3,5085	Retie	2	0,925419	2,1612	0,540493	8,64158
Malle	1	1,432476	0,698	0,0983	4,956	Rijkevorsel	0	1,008175	0		
Berlaar	3	1,305983	2,297	0,7409	7,1225	Turnhout	1	4,797326	0,2084	0,029362	1,479851
Bonheiden	0	1,888993	0			Vorselaar	2	0,783764	2,5518	0,63818	10,20343
Bornem	0	2,378889	0			Vosselaar	1	0,962224	1,0393	0,146388	7,378036
Duffel	2	2,134294	0,937	0,2344	3,7469	Westerlo	1	2,271375	0,4403	0,062015	3,125564
Heist-op-den-Berg	2	4,570852	0,438	0,1094	1,7496	Laakdal	0	1,590077	0		

UL = upper confidence limit; LL = lower confidence limit.

3.2 SPATIAL MODELS

The *Poisson-Gamma model* assumes that the observed number of cervix cancer cases per municipality follow a Poisson distribution with parameter $E_i\theta_i$ (likelihood), and that the risk ratios (θ_i) follow a gamma distribution (prior). Due to the conjugacy of the Poisson and the Gamma distributions, the risk ratios (θ_i) have a closed form, *i.e.*, gamma distribution (posterior). The posterior mean of the risk ratio is given by:

$$[\theta_i / Y_i] = C_i SMR_i + (1 - C_i)\frac{a}{b}.$$

It is thus the weighted average of the data-based SMR for area *i* and the risk ratio in the overall map standardised for the age-groups (the prior mean θ_i). The Poisson-Gamma model has two main disadvantages: (i) spatial autocorrelation is *not* taken into account, and (ii) adding covariates to the model is *not* straightforward.

The *Poisson-lognormal model* accounts for additional heterogeneity (extra-Poisson) as in the previous model, however it is convenient to incorporate covariates easily into the model. The Poisson-lognormal model assumes that the observed number of cervix cancer cases per municipality follow a Poisson distribution (likelihood) with parameter $E_i\theta_i$. The (log-) risk ratios ($\log(\theta_i)$) are assumed to follow a normal distribution (prior), *i.e.*,

$$\text{Log}(\theta_i) = \alpha + X_i\beta + v_i,$$

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

with v_i = the overall heterogeneity random-effects (capturing extra-Poisson (or unexplained) variance) and x_i = the explanatory covariates (*e.g.*, the urban area covariate in this study).

The *Conditional Auto-Regressive (CAR)* model again assumes that the observed counts follow a Poisson distribution with parameter $E_i\theta_i$. The (log-) risk ratios ($\log(\theta_i)$) are modelled by the following equation:

$$\text{Log}(\theta_i) = \alpha + \beta X_i + u_i + v_i,$$

with α = the overall level of risk ratios, v_i = the uncorrelated heterogeneity (unexplained variance), and u_i is the correlated heterogeneity (a spatial correlation structure which reflects the spatial dependency). The risk in area *i* depends on its neighbouring areas, as is expressed by the following equation:

$$[u_i / u_j, i \neq j, \tau_u^2] \sim N(\bar{u}_i, \tau_{u_i}^2).$$

Thus u_i is smoothed towards the mean risk in neighbouring areas, with the variance inversely proportional to the number of neighbours. Area-specific random-effects are thus decomposed into two heterogeneity components, referring to the effects that vary in a spatially structured manner and the effects that vary in an spatially unstructured manner, respectively. Thus, the CAR model does not only takes covariate effects into account, but also allows to evaluate how much of the residual disease risk is attributable to spatially structured variation and how much is attributable to unstructured variation. When the σ^2_u / σ^2_v ratio is large, it can be concluded that much spatially structured variance is present (*i.e.*, differences in risk ratios are mainly attributable to spatial effects).

Following a suitable burn-in period, model convergence was assessed with the Brooke-Gelman-Rubin (BGR) plots (with 2 chains). Goodness-of-fit was assessed as the

model with the smallest Deviance Information Criterion (DIC). The Bayesian residuals were computed ($r_i = y_i - e_i\theta_i^{pred}$); one such example is shown in the APPENDIX section, which shows that the region 2 (Anterwerp municipality within the Antwerp district) as an 'outlier'.

The spatial scan statistic of Kulldorf was used to detect clusters of area's in which cervix cancer had a high incidence. In this procedure, a circular window is gradually scanned across the map. The base of the window is centred across each of the possible centroids (of the different municipalities). For each centroid, the radius of the windows varies continuously in size from zero to an upper limit (*i.e.*, an interval ≤ half of the area of the entire study area). The window is then moved into each possible location is visited. The circular window is thus flexible in both location and size. For every window, the number of cases inside and outside the window is noted, together with the expected number of cases. Based on these values, the likelihood ratio is calculated for each window, to determine whether the number of cervix cancer cases is larger than expected. The circle with the maximum likelihood is defined as the most likely cluster (the least likely to occur by chance). The maximum observed value for the test statistic is obtained through Monte Carlo (MC) hypothesis testing (999 MC replications were used), where the null hypothesis is that no clustering occurs. CAR smoothed risk ratios were determined.

Moran's *I* is an indicator of spatial autocorrelation. It compares the values of the outcome variable (*i.e.*, cervix cancer counts) at any one location with the values at all other locations. Moran's *I* is computed as:

$$I = \frac{n \sum_i \sum_j W_{ij}(Z_i - \hat{Z})(Z_j - \hat{Z})}{(\sum_i \sum_j W_{ij})(Z_i - \hat{Z})^2}.$$

It has a range between -1 (indicating perfect dispersion) and 1 (indicating perfect correlation).

All analyses were preformed with SAS 9.1.3 (SAS *Inc.*, NC Cary, USA), R 11.0 (CRAN), WinBUGS 1.4.3 (MRC), OpenBUGS 3.1.0., GeoBUGS (MRC).

4 RESULTS

4.1 EXPLORATORY DATA ANALYSIS

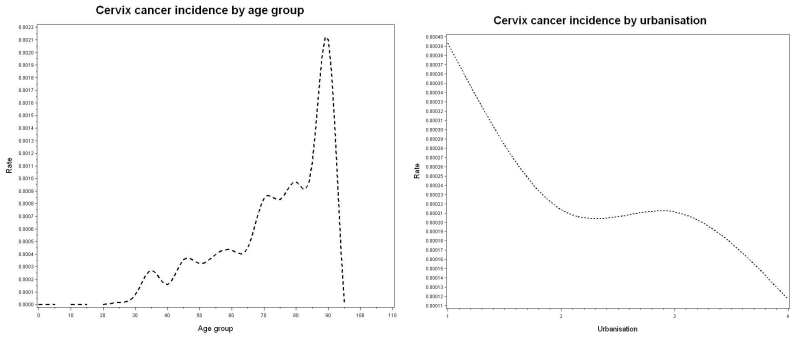


Figure 2. Cervix cancer incidence risk ratio by age group (left) and by urbanisation (right).
Note: lower value indicates more urban area in the figure (right).

Smoothing
Figure 3 presents the SIRs of cervix cancer (left) and the number of female inhabitants (right) per municipality in the Antwerp region. The locally weighted averages are provided in Figure 4, using binary weights (left) and weights based on population sizes (right). In both figures, the disc radius = 0.10 degree (see Figure 5, neighbourhood map). The results of the nonparametric smoothing are presented in Figure 6. The results of the global (left) and local (right) EB smoothing are presented in Figure 7, with disc radius = 0.05 degree.

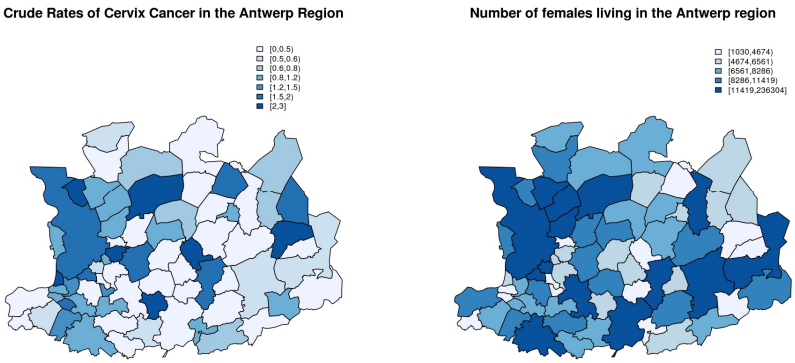


Figure 3. SIRs of cervix cancer (left) and number of females (study population at risk) (right) in the Antwerp region.

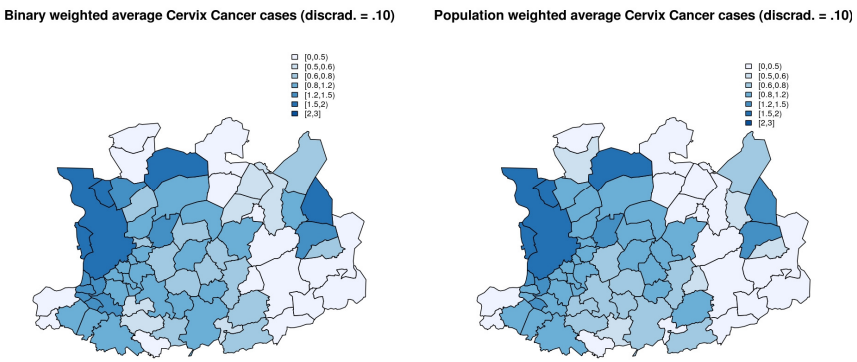


Figure 4. Locally weighted averages using binary weights (left) and population size based (right).

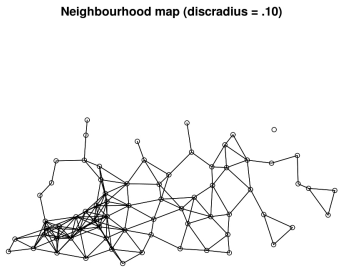


Figure 5. Neighbourhood map with a disc radius = 0.10 degree.

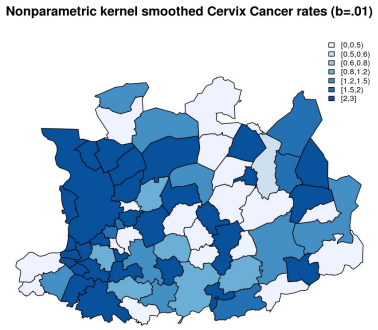


Figure 6. Nonparametric kernel smoothed Cervix Cancer rates (bandwidth = 0.01)

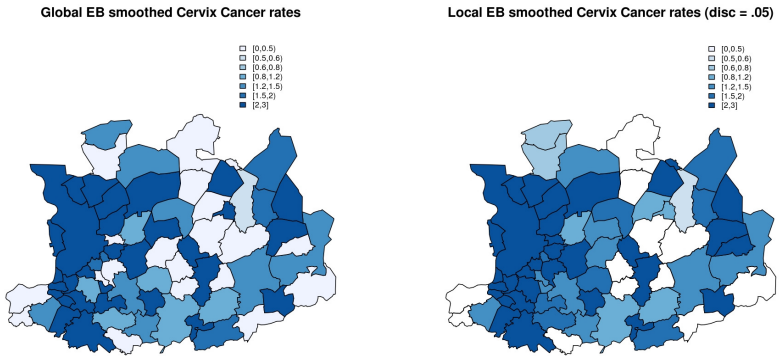


Figure 7. Global (left) and local (right) Empirical Bayes smoothed Cervix Cancer rates.

Based on the results of the exploratory analyses, it can be observed that the incidence of cervix cancer is especially high in the west of the Antwerp district.

4.2 SPATIAL MODELS

Poisson-gamma model

The mean (95% CrI) risk ratio for cervix cancer equalled 0.88 (0.71-1.06). The variance of the gamma distribution (posterior) was .08.

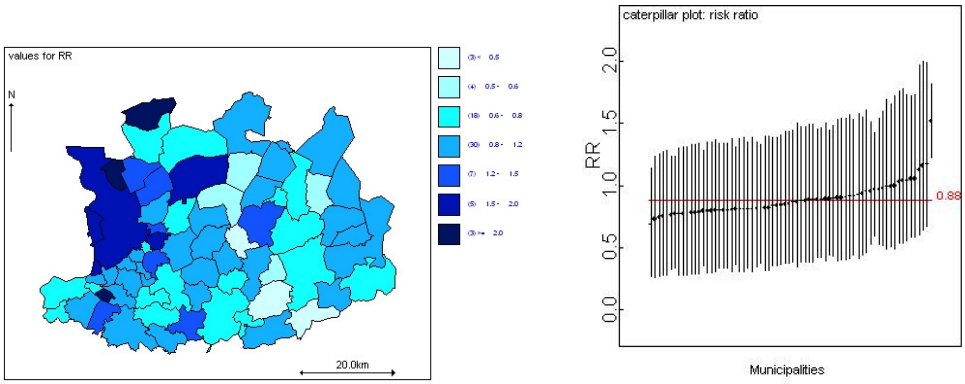


Figure 8. Map (left) and ranked caterpillar plot (right) of the risk ratio of cervix cancer as based on the Poisson-gamma model.

Poisson-lognormal model without covariates

The mean (95% CrI) risk ratio for cervix cancer equalled 0.88 (0.77-0.99). The random-effects parameters v_i represent the unexplained (log-) risk ratio in area *i* that is not accounted for by α. The random variance, σ_v^2 equalled 0.20 reflecting the amount of extra-Poisson variation.

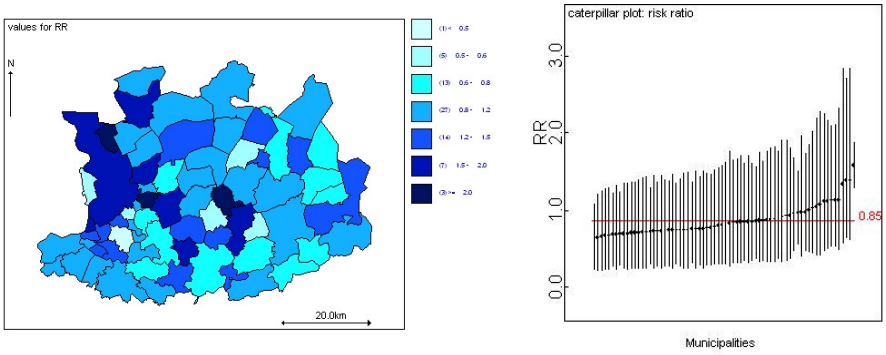


Figure 9. Map (left) and ranked caterpillar plot (right) of the risk ratio of cervix cancer as based on the Poisson-lognormal model without covariates.

Poisson-lognormal model with covariates

The mean (95% CrI) risk ratio for cervix cancer equalled 1.30 (1.03-1.55). The random-effects parameters v_i represent the unexplained (log) risk ratio in area *i* that is not accounted for by α. The random variance, σ_v^2 equalled 0.06 (reduced from that in the previous model by the addition of urbanisation in this model as covariate) reflecting the amount of extra-Poisson variation.

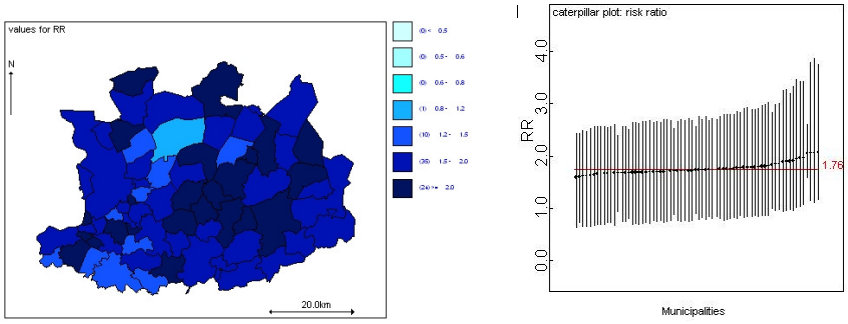


Figure 10. Map (left) and ranked caterpillar plot (right) of the risk ratio of cervix cancer as based on the Poisson-lognormal model with covariates.

Conditional Autoregressive (CAR) model without covariates

The mean (95% CrI) risk ratio for cervix cancer equalled 0.87 (0.78-0.97). The random-effects parameters were decomposed into unstructured overdispersion (v_i) and an structured spatial (u_i) heterogeneity components. The $1/\tau_u$ and $1/\tau_v$ (*i.e.*, σ_u^2 and σ_v^2) equalled 0.10 and 0.31, respectively. The ratio $\tau^2_u/\ \tau^2_v$ was about 3, which indicates that the variability of the risk ratio is attributed less to the unstructured heterogeneity than to the structured spatial heterogeneity.

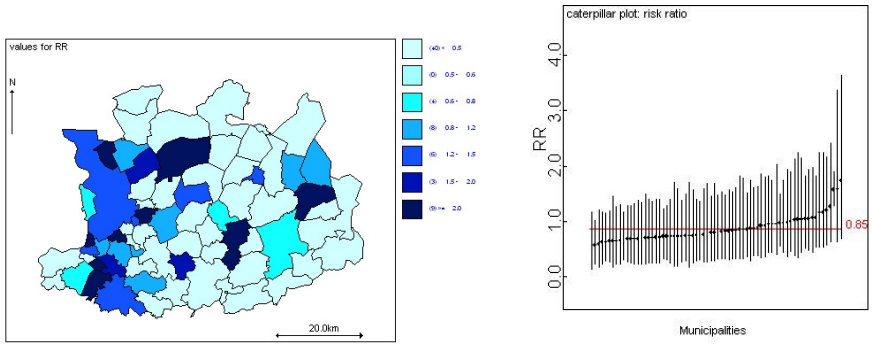


Figure 11. Map (left) and ranked caterpillar plot (right) of the risk ratio of cervix cancer as based on the Conditional Auto-Regressive model without covariates.

Conditional Autoregressive (CAR) model with covariates

The mean (95% CrI) risk ratio for cervix cancer equalled 1.29 (1.00-1.55). The random-effect parameters were decomposed into unstructured overdispersion (v_i) and an structured spatial (u_i) heterogeneity components. The $1/\tau_u$ and $1/\tau_v$ (*i.e.*, σ_u^2 and σ_v^2) equalled 0.05 and 0.07, respectively. The ratio $\tau^2_u/\ \tau^2_v$ was quite close to 1, which indicates that the variability of the risk ratio is attributed equally to the unstructured

heterogeneity as to the structured spatial heterogeneity. Again, we can observe that both these random variances were reduced (especially, u_i) due to the urbanisation as covariate in this model.

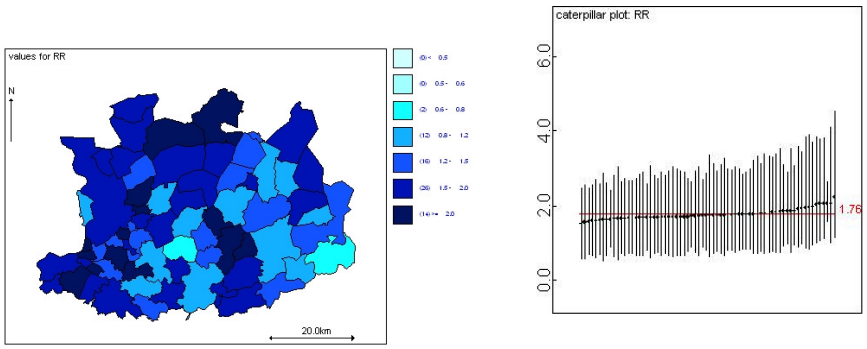
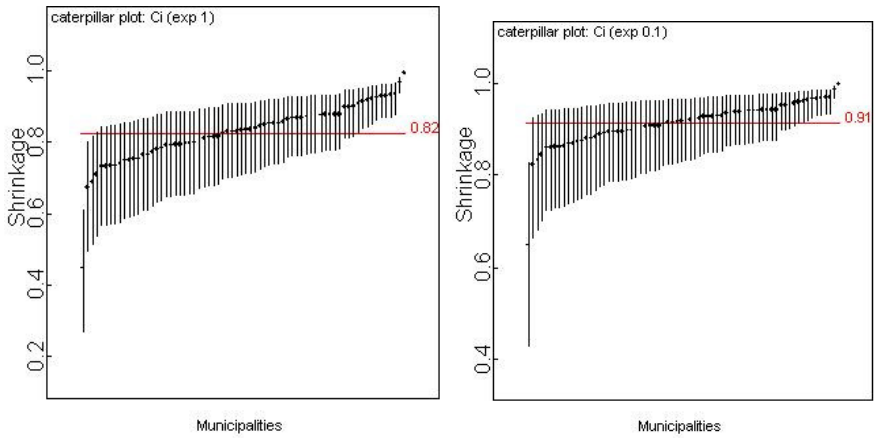


Figure 12. Map (left) and ranked caterpillar plot (right) of the risk ratio of cervix cancer as based on the Conditional Auto-Regressive model with covariates.

Poisson-gamma model – a sensitivity analysis with two different priors
We restricted our sensitivity analysis to only to the Poisson-gamma model only and observed the impact of using 2 different priors. Notably, we changed the priors on the alpha and beta parameters of the gamma distribution from exp(0.01) previously, in each case, to exp(0.1) and exp(1), respectively.

- For exp(0.1) as prior to alpha and beta, we observed the mean shrinkage (C_i) by 9%, with least impact on the RR at region 2 (Antwerp municipality), which had the highest population size. For exp(1) as prior to alpha and beta, we observed the mean shrinkage (C_i) by 18%, with again least impact on the RR at region 2 (Antwerp municipality).
- The model fits improved marginally (DIC from 224.7 to 222.7 and 222.3, respectively).
- The overall mean (*e.g.*, decreased from 0.88 (SD, 0.09) to 0.86 (SD, 0.09) and 0.86 (SD, 0.11), respectively).
- The overall variance increased from 0.08 (SD, 0.06) to 0.14 (SD, 0.07) and 0.34 (SD, 0.14), respectively.



Note: the right most tiny looking observation in both images corresponding to the region 2 (Antwerp municipality) shows the least shrinkage.

Figure 13. Shrinkages observed by exp(1) (left) and exp(0.1) (right) on the alpha, beta parameters of the gamma distribution in the Poisson-gamma model.

Table 2. Parameters estimates on age-standardised data.

Parameters	Poisson (mean, 95% CI)	Poisson-gamma (mean, 95% CrI)	Poisson-lognormal (mean, 95% CrI)	Poisson-lognormal with urbanisation (mean, 95% CrI)	Conditional autoregressive (mean, 95% CrI)	Conditional autoregressive with urbanisation (mean, 95% CrI)
DIC[1]		224.70	224.80	221.50	210.50	216.10
Alpha		15.23 (3.20 – 54.00)	-0.13 (-0.26 – -0.01)	0.26 (0.03 – 0.44)	-0.13 (-0.25 – -0.03)	0.25 (0.00 – 0.46)
Beta		16.84 (3.76 – 56.39)				
Random-effects variance			0.20 (0.04 – 0.50)	0.06 (0.00 – 0.27)	0.10 (0.00 – 0.06)	0.05 (0.00 – 0.26)
Spatial Random-effects variance					0.31 (0.00 – 1.17)	0.07 (0.00 – 0.49)
Overall mean	1.09 (0.95 – 1.24)	0.88 (0.71 – 1.06)	**0.88** **(0.77 – 0.99)[4]**	**1.30** **(1.03 – 1.55)[4]**	**0.87** **(0.78 – 0.97)[4]**	**1.29** **(1.00 – 1.55)[4]**
Variance (gamma distribution)		0.08 (0.02 – 0.23)				
RR [2][2]	**1.65** **(1.37 – 1.99)**	**1.51** **(1.22 – 1.83)**	**1.58** **(1.29 – 1.90)**	**2.06** **(1.58 – 2.63)**	**1.57** **(1.27 – 1.90)**	**2.07** **(1.58 – 2.68)**
Urbanisation[3]				-0.29 (-0.44 – -0.13)		-0.29 (-0.45 – -0.11)
Exp(urbanisation)				**0.75** **(0.64 – 0.88)**		**0.75** **(0.64 – 0.90)**

Note: only risk ratio for region 2 (Antwerp municipality) is presented in this table. Non-informative prior was used in all the models where applicable. All estimates are age-standardised by groups of 5-year interval with respect to the entire Flanders region. Only the estimates in the first column are computed in a frequentist way assuming Poisson distribution of the observed rates in the different communities. All other estimates are computed by Monte Carlo integration in BUGS.
CI = Confidence Interval; CrI = Equal-tail Credible Interval.
1 Deviance Information Criterion.
2 Risk Ratio [region no. 2 (Antwerp municipality)].
3 Coded as 1 = agglomeration, 2 = suburb, 3 = residential area, 4 = not an urban area.
4 exp(alpha).

Kulldorff's Scan statistic

With a maximum window size of 50%, 999 Monte Carlo simulations (on age-standardised data), the most likely 9 municipalities (high-risk cluster) identified in Figure 14 with overall risk ratio significantly >1: Zwijndrecht, Antwerpen, Hemiksem, Schelle, Aartselaar, Niel, Stabroek, Edegem and Boom. It had an overall risk ratio = 2.33; *p*-value < 0.001 compared to the other regions. This corroborates well with the Choynowki's probability map below (Figure 15). Both figures were created using age-standardised data.

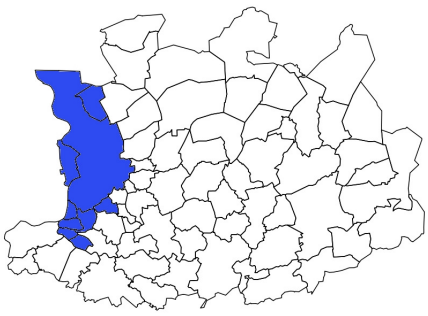


Figure 14. Most likely cluster (with Kulldorff's scan statistic) in blue with an overall risk ratio = 2.33.

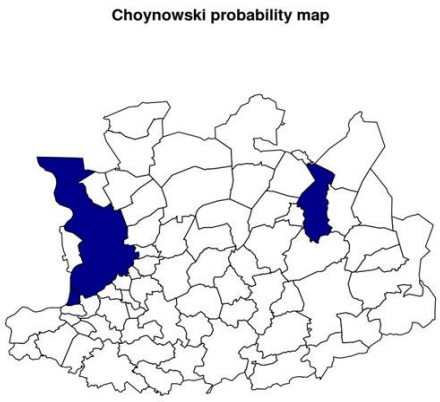


Figure 15. Choynowski probability map showing areas with high risk (in blue).

Finally, Moran's *I* statistic was not significant, including the simulated Moran's *I* statistic, indicating lack of spatial autocorrelation on the age-standardised data. This is corroborated in the Moran's plot (see APPENDIX section).

5 DISCUSSION

The present study points towards the possibility of an increased risk of cervix cancer incidence in women mainly in the western parts of the Antwerp district in Belgium when adjusted for the degree of urbanisation and standardised for age. Exploratory analysis of the data revealed that the overall incidence of cervix cancer increased with age and degree of urbanisation. Since the distribution of these two potential confounders could vary from one municipality to another, we adopted the following approach: the expected rates for each municipality were obtained by standardising for age (in 5-year age groups) taking into account the age group distribution of the entire Flanders region in Belgium to provide more stability in computing the expected disease rates. Next, crude (based on Poisson approximation) risk ratio per municipality were calculated. This gave an idea of overall crude disease rates distribution in the Antwerp district. Following, we computed by Markov Chain Monte Carlo simulation using Bayesian approach, the risk ratios in each municipality, with Poisson-gamma, Poisson-lognormal and Conditional Auto-Regressive (CAR) models. These models did not account for urbanisation. All these models showed that overall risk ratios were less than unity (Antwerp municipality was the exception). In the last step, we added urbanisation as a covariate in order to observe any confounding and also to measure the impact of this variable in the models (*viz.*, Poisson-lognormal, and CAR). Interestingly, we observed that some other, especially western, municipalities (*i.e.*, Boom, Brecht, Hemiksem, Stabroek, and Herentals) in addition to Antwerp had significantly increased risk (risk ratios significantly greater than 1). Furthermore, the degree of urbanisation became significantly increased with increased risk.

Comparing the goodness of fit between the different models (*i.e.*, using Deviance Information Criterion (DIC)), we observed that the Poisson-gamma and Poisson-lognormal models have almost similar model fit statistics. Addition of urbanisation did only marginally improve the model fit in the Poisson-lognormal model; a part of the overall random heterogeneity was "absorbed" by urbanisation (random variance decreased from 0.20 to 06). The advantage of this model over the Poisson-gamma is that the former allows easy incorporation of covariates, however the disadvantage is that the overall heterogeneity cannot be partitioned into structured and unstructured. For this reason, we fitted the CAR models in the next step. There was some additional improvement in DIC by fitting CAR models, which splits the overall random heterogeneity into overdispersion (unstructured) and spatially structured heterogeneity by taking into account the adjacency structure. Further, in the CAR (without urbanisation) showed a larger proportion of spatial heterogeneity compared to overdispersion (0.31 vs. 0.10). Interestingly, the latent spatial heterogeneity seemed to attenuate by a substantial amount when urbanisation was entered in this model; that is both the structured and unstructured heterogeneity were then comparable and lower than the CAR model without urbanisation (0.07 vs. 0.05).

In the current study urbanisation variable was used as a continuous variable, *i.e.*, assuming a linear impact on the outcome. In future analysis, this assumption could be relaxed in letting this variable as categorical. Again for visual comparison of model fits, a map on residuals could be plotted to identify any trend left out in the fitted models.

The main strengths of this study was a quality-controlled by independent double programming ensured that the derivation of the analysis datasets and results are accurate. The caveats for this analysis were certain known risk factors, like smoking, no. of partners in the past were not entered in the models. So potential confounding by these factors cannot be ruled out.

6 REFERENCES

Bosch, F. X., Munoz, N., de Sanjosé, S., Izarzugaza, I., Gili, M., Viladiu, P., et al. (1992). Risk factors for cervical cancer in Colombia and Spain. *International Journal of Cancer, 52,* 750-758.

Bosch, F. X., Manos, M. M., Munoz, N., Sherman, M., Jansen, A. M., Peto, J., et al. (1995). Prevalence of Human Papillomavirus in Cervical Cancer: a Worldwide perspective. *Journal of the National Cancer Institute, 87,* 796-802.

Clifford, G. M., Smith, J. S., Plummer, M., Munoz, N., & Franceschi, S. (2003). Human papillomavirus types in invasive cervical cancer worldwide: a meta-analysis. *British Journal of Cancer, 88,* 63-73.

Rohan, T. E., & Shah, K. V. (2004). *Cervical cancer: from etiology to prevention.* Norwell, MA: Kluwer Academic Publishers.

Smith, J. S., Green, J., de Gonzalez, A. B., Appleby, P., Peto, J., Plummer, M., Franceschi, S., & Beral, V. (2003). Cervical cancer and use of hormonal contraceptives: a systematic review. *The Lancet, 363,* 1159-1167.

Walboomers, J. M., Jacobs, M. V., Manos, M. M., Bosch, F. X., Kummer, J. A., Shah, K. V., Snijders, P. J. F., et al. (1999). Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. *The Journal of Pathology, 189,* 12-19.

7 APPENDIX

BUGS codes

Poisson-gamma model

```
model
        {
                for (i in 1 : N)

                        { # Poisson likelihood for observed counts
                                O[i] ~ dpois(mu[i])
                                mu[i] <- E[i]*theta[i]
                        # Risk Ratios
                                theta[i] ~ dgamma(alpha, beta)
                                RR[i] <- mu[i]/E[i]
                        }

                        # Vague prior distributions
                                alpha ~ dexp(0.01)
                                beta ~ dexp(0.01)

                        # Additional estimates
                        mean <- alpha/beta
                        variance <- alpha/pow(beta, 2)
        }
```

Poisson-logormal model with agglo

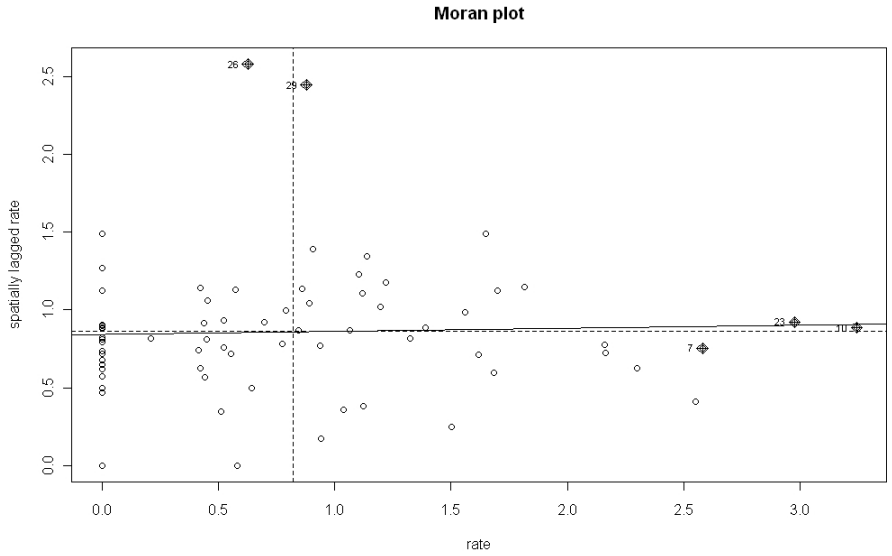
```
model
        {
                for (i in 1 : N)

                { # Poisson likelihood for observed counts
                        O[i] ~ dpois(mu[i])
                        log(mu[i]) <- log(E[i]) + alpha + alpha2*agglo[i] + H[i]
                # Heterogeneity random-effects
                        H[i] ~ dnorm(alpha, v.inv) # hierarchical centring
                # Risk Ratios
                        RR[i] <- exp(alpha + H[i])
                }

                # Vague prior distribution for intercept, agglo
                        alpha ~ dnorm(0.0, 1.0E-6)
                        alpha2 ~ dnorm(0.0, 1.0E-6)

                # Hyperprior distibutions on inverse variance parameter
                        v.inv ~ dgamma(0.0001, 0.0001)
                        v <- 1/v.inv
        }
```

Conditional-autoregressive model with agglo

```
model
        {
                for (i in 1 : N)

                { # Poisson likelihood for observed counts
```

```
        O[i] ~ dpois(mu[i])
        log(mu[i]) <- log(E[i]) + alpha + alpha2*agglo[i] +u[i] + v[i]
    # Heterogeneity random-effects
        v[i] ~ dnorm(alpha, tau.v) # hierarchical centring
    # Risk Ratios
        RR[i] <- exp(alpha + u[i] + v[i])
    }

    # CAR prior distribution for spatial random-effects
        u[1 : N] ~ car.normal(adj[], weights[], num[], tau.u)
        for(k in 1 : sumNumNeigh)
        {
            weights[k] <- 1
        }

    # Other priors
        alpha ~ dnorm(0.0, 1.0E-6)
        alpha2 ~ dnorm(0.0, 1.0E-6)
        tau.u ~ dgamma(0.0001, 0.0001)
        tau.v ~ dgamma(0.0001, 0.0001)
        sigma2.v <- 1 / tau.v
        sigma2.u <- 1 / tau.u
}
```

Moran's plot



Example of an index plot of Bayesian residuals in the Poisson-lognormal model without agglo showing the 'outlying' region 2.