British Journal of Medical and Health Sciences
URL: http://www.bjmhs.baar.org.uk
ISSN: 2326 – 4276 Vol. 1, No. 7, pp 37-53, **May 2013**

MAPPING DISTRIBUTION OF DISEASE PAIRS IN SPACE AND TIME

Elek Dinya, Tamás Tóth, Gergely Tóth, Sándor Kabos*, Gabriella Merth, György Surján

* **Corresponding author** to whom correspondence should be addressed: Sándor Kabos, ELTE TATK,
Address: H-1117 Pázmány Péter sétány 1/A. Budapest, Hungary.
E-mail: kabos@tatk.elte.hu
Tel.: +3613722500 ext. 6828
Fax: +3613722912

Author's affiliation:
Elek Dinya[1], Tamás Tóth[1],
(1) SOTE Semmelweis University, Department of Health Informatics, H-1091 Üllői út 25. Budapest, Hungary

Gabriella Merth[3], György Surján[3]
(3) GYEMSZI National Institute for Quality- and Organizational Development in Healthcare and Medicines, H-1054 Hold utca 1. Budapest, Hungary.

Gergely Tóth[2], Sándor Kabos[2*],
(2) ELTE Eötvös Loránd University, Faculty of Social Sciences, Department of Statistics, H-1117 Pázmány Péter sétány 1/A. Budapest, Hungary.

E-mail to authors:
dinya.elek@public.semmelweis-univ.hu; toth.tamas@public.semmelweis-univ.hu; toth.gergo@gmail.com; merth.gabriella@gyemszi.hu; surjan.gyorgy@gyemszi.hu

ABSTRACT

Problem Statement: 350-420 new cases of gastric and duodenal ulcer per year per 100,000 people have been recorded in Hungary in the years 2004-10. The aim is to give a detailed characterization of the joint distribution of these two types of peptic ulcer in space and time. The empirical distribution seems to be far from the uniform and it shows similar spatial patterns to that of some socio-economical factors. Common procedures of mapping standardized incidence ratios (SIR) appeared to be inapplicable due to low case numbers in some cells when analyzing detailed data of yearly incidence classified by age, gender and place of patient's residence.
Approach: Our approach is twofold: we estimate parameters of a multilevel Poisson-Binomial regression model, and we use interactive mapping tools for generating hypotheses and for representing the estimated parameters.
Results: The family of Bayesian multilevel regression models proved to be suitable to test hypotheses formulating interaction effects between time and spatial factors. An OpenBUGS source code of the final model is also given. We found that the socio-economically deprived micro-regions of North-East Hungary are less favoured in terms of public health; they have high SIR values of both gastric (K25) and duodenal ulcer (K26), moreover the ratio K26/K25 is also extremely high. Spatial differences in reaction to changes in 2007-8 are also characterized.
Conclusions: The composed use of disease mapping and statistical modeling is demonstrated as an efficient tool of data mining for exploring unexpected space-time effects of epidemiological processes.

Keywords: disease mapping, peptic ulcer, socio-economical factors in public health, Poisson regression.

Conflict of interest

The authors declare that they do not have any conflict of interests or financial interests pertaining to this paper.

ISSN: 2326 – 4276 Vol. 1, No. 7, pp 37-53, **May 2013**

Introduction

Peptic ulcers have two common forms: gastric ulcer (ICD-10 code: K25) and duodenal ulcer (ICD-10 code: K26). The occurrence of the diseases show high geographic and temporal variations: in Western populations duodenal ulcer is more common while in Asia, especially in Japan, gastric ulcers dominate [1]. ICD (International Classification of Diseases) is the most widely used international disease classification system, maintained by the World Health Organisation. Hungary is using ICD-10, the current, tenth version of ICD both for mortality and morbidity statistics.

In both diseases, ulcer is consequence of the imbalance between protective and aggressive factors of the mucous membrane, mainly caused by inflammation. Earlier, peptic ulcers were considered psychosomatic diseases mainly caused by stress [2]. In 1983, two researchers identified a bacterium, Helicobacter pylori (H. pylori) in patients with chronic gastritis [3]. Since then, a relationship has been discovered between H. pylori and several gastrointestinal diseases including peptic ulcers and some forms of tumors [4]. However, the cause-effect link is not clear: over 80 percent of infected people never develop an ulcer, and at least in10 per cent of ulcer cases the presence of H. Pylori cannot be proven [5]. Consequently, peptic ulcers supposed to be of multifactorial origin influenced by bacterial infection, psychological factors, as well as behavioural factors and some drugs [6, 7].

Duodenal ulcer is more related to H. pylori infection, and is caused mainly by an increase in acid and pepsin load, and gastric metaplasia in the duodenal cap. Gastric ulcer, especially in Western countries is rather associated with NSAID ingestion, but H. pylori might be present in these patients as well. Gastritis is predominating in gastric ulcers [8]. H. pylori and NSAID use are independent risk factors of peptic ulcers and they have synergistic effects. These two factors together are responsible for approximately 90% of peptic ulcer cases. However some studies report the growing proportion of ulcers not caused by these two factors, especially in the US [9]. This might be a consequence of the decreasing prevalence of H. pylori infection, but might also be caused by undetected NSAID use and/or inaccurate diagnosis of the infection, thus the results are uncertain. In areas with high prevalence of H. pylori such as Asia, this type of ulcers is rare [10].

Several studies inspected the prevalence of H. Pylori infection and the associated diseases. The results show a high variety among countries [11]. The studies inspected by the review were conducted in various age ranges, but in general the prevalence of the bacterium is high in less developed countries. In these countries, the infection rate can be higher than 70 per cent, and the infection appears already in early childhood. In developed countries, the infection is rare among children, but the prevalence increases with age. The overall incidence of H. pylori infection as well as the diagnosis and procedures associated with end-stage peptic ulcers show a decreasing trend in the US [12].

The impact of socio-economic inequalities on health can be modeled in many ways [13]. The first step is to define “welfare” as a comprehensive index for measuring socio-economic factors. Several indices of welfare are proposed by regional scientist [14, 15], but there is no generally accepted one, because the given nature of such an indicator could be considered to only a limited number of components. Despite the different paradigmatic approaches, it can be clearly stated that the components of the social welfare are mainly stable in correlations to

ISSN: 2326 – 4276 Vol. 1, No. 7, pp 37-53, **May 2013**

each other, as well as the fact that these elements are fairly constant in time. The central and Western regions of Hungary are historically more improved and prosperous contrast to the Eastern and North-Eastern regions.

Disease mapping techniques are widely used in public health research for exploring regional differences. Standardization is an elementary tool for removing effects which result from the different age-by-gender distribution of the population in different spatial units [16]. Mapping spatial units by its standardized incidence ratio (SIR) value and modeling SIR values in Bayesian framework became a popular tool of spatial epidemiology. This technique descends from the classical paper [17] and is applied by a huge amount of case studies including our previous work [18].

Gastric and duodenal ulcer are non-contagious diseases, therefore it is straightforward to model the incidence numbers as conditionally independent realizations of a two-dimensional Poisson random variables. The idea to model the joint distribution of several diseases by Bayesian decomposition of the joint likelihood is popularized by [19]. The full scale of marginal and conditional decompositions of the joint likelihood is given by [20]. The dynamic Poisson-Binomial model proposed here is a special case of the previously mentioned ones and is relatively simple because we deal with joint distribution of only two diseases. An application of this model in disease mapping and statistical testing will be discussed below.

Data and Methods

Medical data

Hospital admissions (more precisely, departmental cases) in 2004-10 classified by admission diagnosis (using ICD-10 codes) were analyzed. The Hungarian national data repository (hosted by GYEMSZI) stores records of all hospital cases. Variables are taken into account: admission date and diagnosis, age, gender and ZIP code of patient's place of residence. Patient identifier is replaced by a pseudo code that does not identify the real person but enables to match the records belonging to the same patient. According to current legislation in Hungary, such data are not considered as personal as far as the user of the data is not in possession of any tools that enables re-identify the subject of the data.

Records coming from healthcare providers may consist a certain amount of missing or invalid fields. We confine ourselves to those records that have valid ICD codes, so the effects of diagnostic uncertainty are neglected. The cases with invalid age, gender and ZIP code of patient's residence (and the non-residents of Hungary) are also excluded. The total losses due to the incorrect coding are about 15-40 percents. The risks given below may have underestimated to this extent, but this effect is uniform, so does not distort the main points of this paper. No sign of spatial or temporal accumulation of coding errors have been detected by our data screening system.

Owing to the fact that only the first occurrence of a patient identifier has been taken into account by basic method of calculating incidence counts, we cannot calculate valid yearly incidence data for 2004. About 10 % of cases are common in two consecutive years.

ISSN: 2326 – 4276 Vol. 1, No. 7, pp 37-53, **May 2013**

The standardized incidence ratio (SIR) is the ratio of observed incidence number and the expected one. The latter is calculated using age-by-gender distribution of the current spatial unit and the Hungarian countrywide age- and gender-specific incidence rates as shown in Table 1. This method is referred to as indirect standardization [16].

Demographical and socio-economical data

The welfare indicator is a composite index and originally based on the technique of paper of Quadrado et al. [14], but in the absence of some administrative data, the applied version of the index is slightly different from it. In this manner the indicator with the use of several different variables on micro-regional level conveys information about infrastructural, educational, labor market, unemployment, social assistance, income and taxation data for each year of the study period.

An interactive tool for mapping and modeling: Rapporter

Rapporter [21] is a scalable and extensible web application helping users to create, edit and publish comprehensive, reliable statistical reports on PC or any mobile device, using an intuitive user interface. The software is based on the power of the R programming language besides other open-source technologies, and is intended to be used in any modern web-browser, with the maths and heavy computations carried out on the server side. The main service of the application is freely available for non-profit academic purposes at rapporter.net.

As Rapporter also provides an extensible application-programming interface (API), which seamlessly integrates statistical methods into any web page, a specific web application was created for the scientific community to run further analysis on the dataset used in this paper. This Rapporter application calls OpenBUGS to perform the Bayesian model fitting. Readers are encouraged to test this application from this link: http://web.tatk.elte.hu/~eregr/kabos/Lset_Welfare.html

The multilevel Poisson-Binomial model

Bayesian multilevel model applied here takes the observed case numbers as conditionally independent realizations of a Poisson distribution. The dependencies of Poisson parameters on age, gender and spatial units are specified in BUGS language [22]. A source code in OpenBUGS is given in Appendix 3.

The idea behind our Poisson-Binomial model is that if Y_1 and Y_2 are two independent Poisson then $Y_1 + Y_2$ is also a Poisson and the conditional distribution of $Y_2 \mid Y_1 + Y_2$ is Binomial. This fact is widely used in the log-linear analysis of cross-tables for parametrization of a two dimensional distribution by its marginals and odds ratio [23]. The general theory given in [20] has much more parametrization possibilities, but the simple special case we are using here has a straightforward epidemiological interpretation: $Y_1 + Y_2$ means the risk of being hospitalized by peptic ulcer (either by K25 or K26) while $Y_2 \mid Y_1 + Y_2$ means the risk of having K26 provided having either K25 or K26. Our basic model is specified formally in Appendix 1, and a dynamic version is given in Appendix 2.

British Journal of Medical and Health Sciences
URL: http://www.bjmhs.baar.org.uk
ISSN: 2326 – 4276 Vol. 1, No. 7, pp 37-53, **May 2013**

Results

Regional inequalities of K25 and K26 incidence rates

The North-Eastern part of Hungary consists of 25-30 micro-regions with extremely low socio-economical indicators forming a contiguous zone. As a first approximation, this zone can be identified by counties Szabolcs-Szatmár-Bereg and Borsod-Abaúj-Zemplén and will be referred to as NE part of Hungary. Total population of NE part is about 1.2 million people (12% of Hungary).

K25 in Hungary	male	female
18-34	12.33	7.29
35-64	113.81	67.37
65-xxx	335.48	245.11

K26 in Hungary	male	female
18-34	14.91	6.57
35-64	123.63	71.79
65-xxx	298.87	186.68

Table 1. Age- and gender-specific incidence rates of gastric (K25) and duodenal ulcer (K26) in Hungary, yearly averages in 2005-10.

K25 in NE part	male	female
18-34	13.69	9.11
35-64	134.33	84.25
65-xxx	365.36	260.87

K25 out of NE	male	female
18-34	12.12	7.02
35-64	110.87	64.99
65-xxx	331.72	242.99

Table 2.a. Age- and gender-specific incidence rates of gastric ulcer (K25) in North-East part of Hungary compared with the other parts, yearly averages in 2005-10.

K26 in NE part	male	female
18-34	26.04	11.86
35-64	201.38	141.92
65-xxx	405.08	268.14

K26 out of NE	male	female
18-34	13.18	5.77
35-64	112.49	61.93
65-xxx	285.50	175.70

Table 2.b. Age- and gender-specific incidence rates of duodenal ulcer (K26) in North-East part of Hungary compared with the other parts, yearly averages in 2005-10.

An obvious observation from these tables is that for each age-by-gender category risks are higher in the North-Eastern part than outside. Differences between the two types of peptic ulcer can be achieved by taking a closer look. We need some mapping and statistical modeling tools to get more detailed results.

British Journal of Medical and Health Sciences
URL: http://www.bjmhs.baar.org.uk
ISSN: 2326 – 4276
Vol. 1, No. 7, pp 37-53, **May 2013**

Mapping socio-economical factors and SIR values

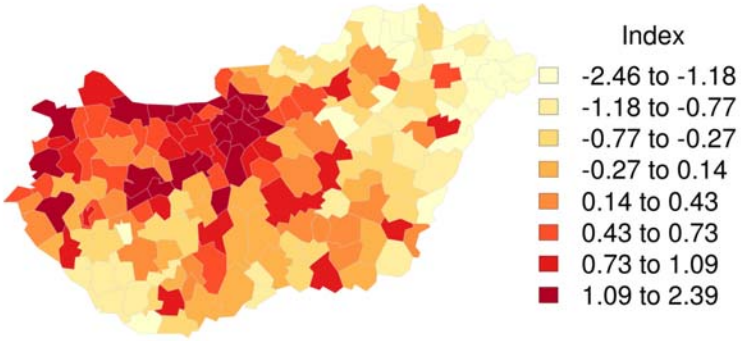

Figure 1. A welfare index of micro-regions of Hungary, 2005-10, composed by G. Tóth

The central and western regions of the country are historically more developed and prosperous in contrast to the eastern and northeastern regions. However these differences had been existed for decades, some slow changes could be observed inside the regions [24].

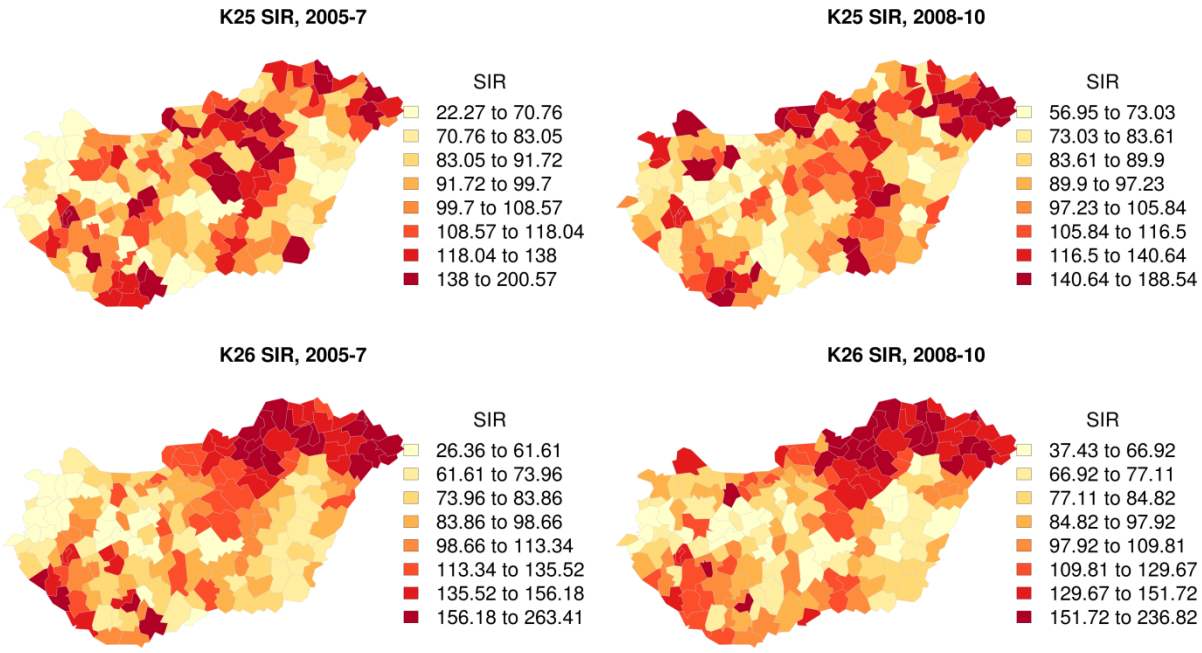

Figure 2. Gastric ulcer (K25) and duodenal ulcer (K26) SIR values by micro-regions, Hungary, in 2005-7 and 2008-10.

Both types of peptic ulcer are related to the socio-economical factors, but K26 has closer links with welfare than K25. All the comparisons have the same meaning: the lower the welfare, the higher the SIR level.

British Journal of Medical and Health Sciences
URL: http://www.bjmhs.baar.org.uk
ISSN: 2326 – 4276 Vol. 1, No. 7, pp 37-53, **May 2013**

Fitting the multilevel Poisson-Binomial model and mapping the estimated parameters

A Poisson distribution with parameter μ_{ij} (indices refer to the year j and spatial unit i) is used to model the number of new cases diagnosed with peptic ulcer. The following equation specifies the "mu-part" of the model $log(\mu_{ij}) = log(E_i) + a.mu_j + b.mu \cdot X.MU_i + \varepsilon.mu_{ij}$ where E_i is the expected incidence number in spatial unit i as calculated by the indirect standardization, $a.mu_j$ is a correction term for year j, and $X.MU_i$ is the welfare index characterized in the previous subsection with regression coefficient $b.mu$.

The "p-part" of the model is $logit(p_{ij}) = a.p_j + b.p \cdot X.P_i + d.p_i \cdot X.D_j + \varepsilon.p_{ij}$ where the conditional probability of having duodenal ulcer provided having any type of peptic ulcer is denoted by p_{ij}. This part has two explanatory variables: $X.P_i$ is the indicator variable of welfare at level -1 (equals 1 for micro-regions with welfare index less than -1, and equals 0 elsewhere) with regression coefficient $b.p$ while $X.D_j$ is the indicator of the change point in time with regression coefficient $d.p_i$. Some more modeling issues will be discussed in the next section.

The multilevel Poisson-Binomial model outlined below stated as Model 7 in the Table 3. Other models are made by omitting some terms as it is identified in the columns of model formulation. For example, Model 1 consists of $log(E_i)+a.mu_j$ in its mu-part, while the p-part consists of all the explanatory variables, but the error term.

British Journal of Medical and Health Sciences
URL: http://www.bjmhs.baar.org.uk
ISSN: 2326 – 4276
Vol. 1, No. 7, pp 37-53, **May 2013**

model no.	model formulation		criteria for model selection			
	mu-part	p-part	Dbar	Dhat	pD	DIC
Model 1.	E+a.mu	a.p+X.P+D.P	26684.1	26565.9	118.2	26800
Model 2.	E+a.mu +X.MU	a.p+X.P+D.P	25858.6	25738.6	120	25980
Model 3.	E+a.mu	a.p+X.P+D.P+err.p	24619.4	23902.7	716.7	25340
Model 4.	E+a.mu +X.MU	a.p+X.P+D.P+err.p	23795.8	23077.9	718	24510
Model 5.	E+a.mu +X.MU+err.mu	a.p+D.P	17611.7	16477.9	1134	18750
Model 6.	E+a.mu +X.MU+err.mu	a.p+X.P+D.P	17324.7	16191.6	1133	18460
Model 7.	E+a.mu +X.MU+err.mu	a.p+X.P+D.P+err.p	15257.5	13532.3	1725	16980

Table 3. The posterior mean of the deviance (Dbar), the point estimate of deviance (Dhat), the complexity of model (pD) and the Deviance Information Criterion (DIC) of several models as estimated by OpenBUGS

Table 3 shows that the lack-of-fit of data to the model (Dhat) decreases as the complexity of model (pD) increases. ***Deviance Information Criterion (DIC) is based on trade-off between these criteria, so it can be used to pick the optimal model [25]. Model 7 was chosen as optimal according to this criterion. The OpenBUGS results for other parameters of this model are given in the following table.***

	mean	sd	2.5%	25%	50%	75%	97.5%	Rhat
a.mu[1]	0.1725	0.0248	0.1242	0.1556	0.1733	0.1894	0.2192	1.0055
a.mu[2]	0.0039	0.0244	-0.0425	-0.0130	0.0041	0.0211	0.0504	1.0236
a.mu[3]	-0.0992	0.0249	-0.1484	-0.1159	-0.0986	-0.0820	-0.0532	1.0046
a.mu[4]	-0.0963	0.0249	-0.1452	-0.1131	-0.0966	-0.0794	-0.0450	1.0149
a.mu[5]	-0.1921	0.0245	-0.2412	-0.2081	-0.1917	-0.1752	-0.1450	1.0034
a.mu[6]	-0.2876	0.0258	-0.3371	-0.3049	-0.2882	-0.2699	-0.2379	1.0016
b.mu	-0.1291	0.0103	-0.1501	-0.1360	-0.1287	-0.1218	-0.1097	1.0204
a.p[1]	0.0299	0.0300	-0.0295	0.0099	0.0298	0.0494	0.0886	1.0030
a.p[2]	0.0022	0.0302	-0.0576	-0.0186	0.0030	0.0227	0.0596	1.0053
a.p[3]	-0.1226	0.0323	-0.1865	-0.1437	-0.1225	-0.1015	-0.0585	1.0016
a.p[4]	-0.1474	0.0313	-0.2084	-0.1692	-0.1470	-0.1266	-0.0857	1.0010
a.p[5]	-0.1977	0.0316	-0.2586	-0.2183	-0.1977	-0.1760	-0.1330	1.0009
a.p[6]	-0.1824	0.0333	-0.2476	-0.2053	-0.1822	-0.1599	-0.1190	1.0083
b.p	0.2474	0.0311	0.1876	0.2257	0.2479	0.2690	0.3065	1.0025
deviance	15257.3	70.0	15120.0	15210.0	15260.0	15310.0	15420.0	1.0020

Table 4. Posterior estimates for parameters of Model 7.

Both *b.mu* and *b.p* are significant (i.e. zero is out of their 95% level confidence interval), this is the most important information given by Table 4. This result can be interpreted as a statistical evidence for relationship between the welfare and SIR values.

One can notice some differences between spatial configuration of hot-spots of K26 in 2005-7 and in 2008-10 when analysing maps on Figure 2. This effect is described by the term $d.p_i \cdot X.D_j$ in the p-part of model. The estimates cannot be given as a table because this model has as many *d.p* parameters as the number of spatial units.

British Journal of Medical and Health Sciences
URL: http://www.bjmhs.baar.org.uk
ISSN: 2326 – 4276 Vol. 1, No. 7, pp 37-53, **May 2013**

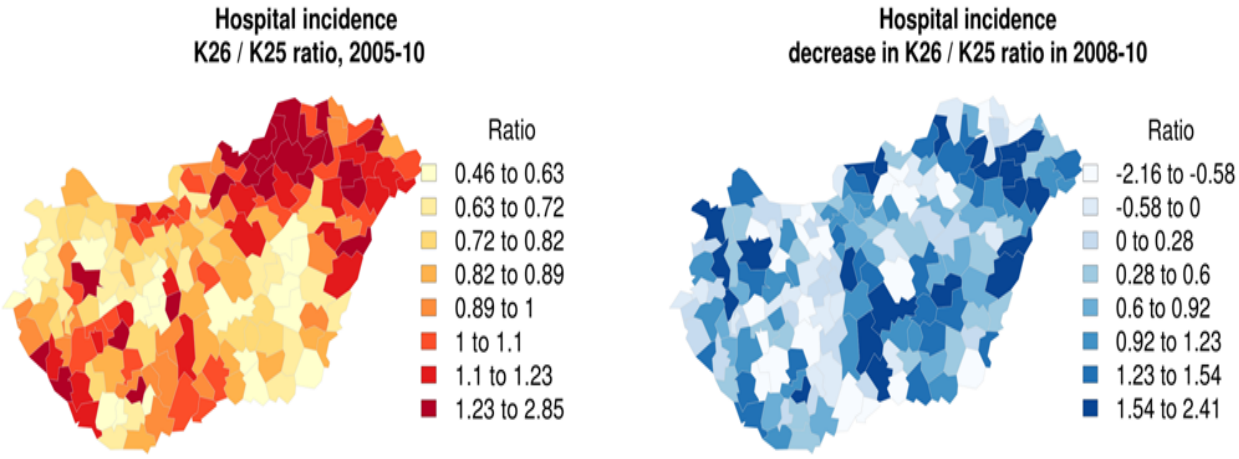

Figure 3. K26 SIR / K25 SIR ratio in 2005-10 (left) and *t*-values of parameter estimation *-d.p* as given by OpenBUGS

Note that the ratio visualized on the left hand side is not equal to the ratio of incidences of K26 and K25 because of the different age-by-gender specific rates of these two diseases. The *t*-value represented by the right hand side map is simply defined as the ratio of MEAN and SD values as seen on the Table 3. The distribution is approximately standard normal because of the high degrees-of-freedom so the values above 1.65 are significant (at level 0.05 of significance using one sided confidence limits).

Hungary joined the EU in 2004 which made it possible to involve high capital and EU funds in the country, thus launched regional changes for a few years, but the global economic crisis in 2008 has highly influenced the Hungarian economic performance. The fallback of the underperforming regions was accelerated and therefore the regional differences became more obvious, thus we investigated the differences between the before and after crisis periods, albeit there are generally slow changes in territorial level in Hungary. The results displayed above show mixed trends: the micro-regions of North-Eastern part changed in different ways in 2008. One can notice that micro-regions decreasing in K26/K25 (the dark blue ones in the right side of Figure 3) have a specific spatial pattern. This corresponds to the hypothesis that regional developments may strongly be influenced by the effects of the new section of motorway M3 opened in 2007, as similar results have been reported by K. Gkritza [26]. This correspondence cannot be confirmed here (due to the model over-dispersion discussed below) but it may be an initial point of further investigations based on more detailed regional data.

Discussion

Decision making support for public health policy

Current trends in "evidence based health policy" require sound statistical analyses of morbidity data. Distribution of incidence data both in time and space are crucial for health care capacity planning and optimal use of scarce resources. Most of published studies however concentrate on single diseases and limited number

ISSN: 2326 – 4276 Vol. 1, No. 7, pp 37-53, **May 2013**

of predicting factors. Experts in health policy and quality of health care delivery are continuously seeking for relevant indicators. Presentation of their data in geographical information systems is very common but often fails to provide proving of statistical significance of the demonstrated special differences and inequities.

"All maps of parameter estimates are misleading"

There are some common mistakes in interpreting disease maps as a paper of Gelman and Price put it sharply [27].The main objection against maps representing raw rates comes from the different subsample sizes. The different spatial units with the same color category may lead to different levels of statistical significance. This is because such maps cannot be used in statistical reasoning. We must admit that this remark applies to the SIR values on the maps of Figure 2 and to the ratios of SIR values on the left map of Figure 3.

There are also problems with maps based on statistical significance, due to their tendency to overemphasize the role of sample size. Next figure helps us to compare these methods by displaying the same content in two different ways.

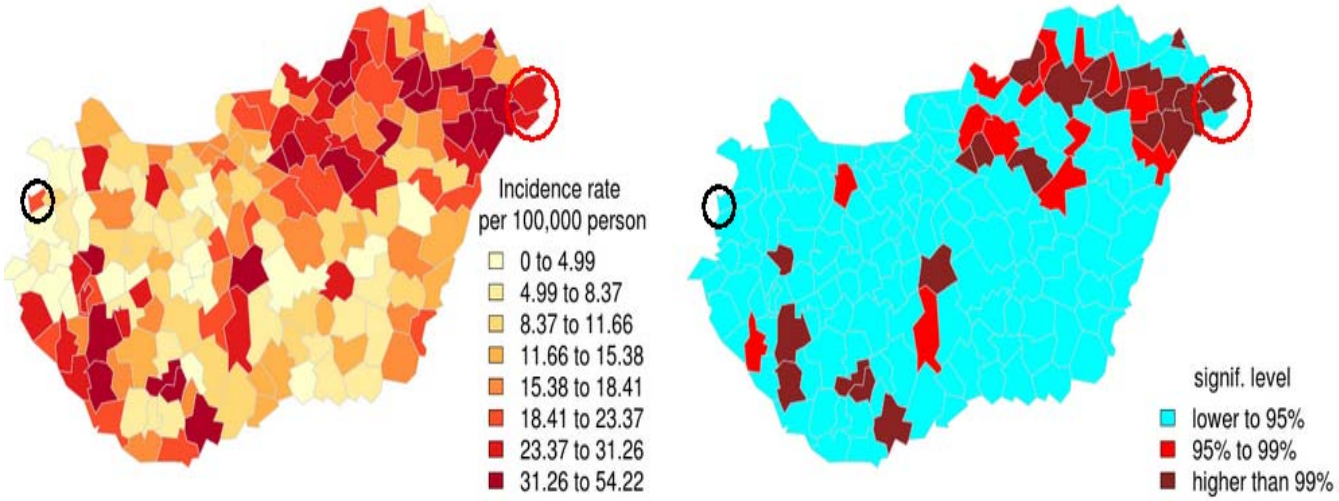

Figure 4. Maps of duodenal ulcer (K26) incidences among men aged 18-34 in 2005-10; displayed by SIR values (on the left) and by rejection probabilities of the null hypothesis that the observed numbers of incidences are independent realizations of a homogeneous Poisson process (on the right).

Let us take a look at the two micro-regions circled by red. They were in the same category on the left map while one of them (the bigger one) became dark red on the right map and the smaller one became light blue. One can easily see (from the detailed population data not included here) that this is the case of overemphasizing we have mentioned below. The micro-region circled by black is a counterexample, while the left map may be misleading. Mapping method of displaying posterior z-scores (like the map on the right side of Figure 3) commits the same errors and there are some additional problems by neglecting the rules of simultaneous decision making [27]. Readers are welcome to visit Rapporter application http://web.tatk.elte.hu/~eregr/kabos/Compare_Maps.html for testing the possibilities to compare many other maps.

ISSN: 2326 – 4276 Vol. 1, No. 7, pp 37-53, **May 2013**

Professor Gelman and others have published a number of papers explaining many faces of this general problem in displaying spatial distributions [28-30]. Nevertheless, there are no alternatives to the use of maps of parameter estimates in public health research. According to the considerations summarized above these maps are used as explanatory tools for formulating hypotheses to be tested by standard statistical methods.

The Poisson-Binomial model in more detail

Poisson parameter of the sum of K25 and K26 incidences is denoted by μ and is specified by the "mu-part" of models. One of the differences between two models formulated in Appendix 1 and 2 is the different use of standardization. The first model has no explicit reference to the age and gender, but in the standardization, which is performed by years according to the age-by-gender distribution. The second model specifies the mu-part as $log(\mu_{ijk\ell}) = log(E_{ik\ell}) + a.mu_j + b.mu_{k\ell} \cdot X.MU_i + \varepsilon.mu_{ijk\ell}$ for i -th spatial unit, j -th year, k -th age-group of the patient and ℓ -th gender of the patient. The first term on the right hand side is an offset (that is a variable with fixed regression coefficient 1) and it does not depend on year (j) but the second term (the intercept of regression) may vary from year to year. The indicator of spatial differences $X.MU$ has cofficient $b.mu$ which may depend on age and gender while $\varepsilon.mu$ is an independent and identically distributed (i.i.d.) error term.

Given the sum of K25 + K26 the conditional distribution of K26 is binomial with parameter p which is specified by the "p-part" of models. Let us take a closer look at the p-part of the dynamic model: $logit(p_{ijk\ell}) = a.p_j + b.p_{k\ell} \cdot X.P_i + d.p_{ik} \cdot X.D_j + \varepsilon.p_{ijk\ell}$ where coefficients $a.p$ and $b.p$ are similar (but not the same) to that of the mu-part.

Explanatory variable $X.D_j$ is designed to detect a possible change point in 2008 regarding the dynamics of incidences of K26. Analyzing maps on Figure 2 we noticed that spatial patterns of K26 incidence and K26/K25 SIR ratio has been changed in 2008. This type of hypotheses can be tested by our model because coefficients $d.p$ may vary from micro-region to micro-region so an indicator area of this effect can be identified by the level of significance of these coefficients. Of course this is not a statistically correct method of hypothesis testing as explained below (a multiple test is needed) but this method helps us to narrow the area to be investigated involving other variables. Another possible aim for a further analysis is to test the dependency of $d.p$ on age-group. Some age-specific maps (not included here) suggest that the age-group 35-64 is more exposed to the change in 2008 than others.

Although Model 7 was chosen as the best of models according to its DIC value in Table 3, we must not interpret it as a final model. Individual diagnostics of parameters of this model are acceptable (e.g. *Rhat* values in Table 4 are less than 1.1) but there is an overall lack of fit. *Dhat* should follow an approximate chi-squared distribution with n-p_D degrees of freedom if we assume the model, but n = no. of diseases * no. of age groups * no. of gender groups * no. of years * no. of spatial units = 2*3*2*6*174 = 12528 and n-p_D = 12528 - 1725.0 = 10803 is less than *Dhat* = 13532.3 showing that the model is over-dispersed [25] . One can hardly get rid of over-dispersion by involving other explanatory variables into the model because it increases the model complexity (p_D), but using more flexible error structures may lead to models fit better [31-33].

ISSN: 2326 – 4276 Vol. 1, No. 7, pp 37-53, **May 2013**

Conclusion

Maps representing raw incidence rates or SIR values are essential in public health research, but one must be careful in their interpretation. The models treated here may be useful to describe changes in space and time. The proper use of maps may be an efficient tool of visual data mining while representing estimated model parameters and suggest hypotheses for further research.

Appendix 1. Formal specification of Poisson-Binomial model

$$Z_{ij} \mid \mu_{ij} \sim Poisson\left(\mu_{ij}\right)$$

$$log(\mu_{ij}) = log(E_{ij}) + a.mu + b.mu \cdot X.MU_i + \varepsilon.mu_{ij}$$

where $\varepsilon.mu_{ij} \sim Normal\ (0\ ,\ \sigma_{mu})$ (i.i.d.)

$$Y_{ij} \mid Z_{ij}\ ,\ p_{ij} \sim Binomial\ (Z_{ij}\ ,\ p_{ij})$$

$$logit(p_{ij}) = a.p + b.p \cdot X.P_i + \varepsilon.p_{ij}$$

where $\varepsilon.p_{ij} \sim Normal\ (0\ ,\ \sigma_p)$ (i.i.d.) for $i=1..I$, $j=1..J$

cell indices: i-th region, j-th year,

Z_{ij} are observed incidences of gastric and duodenal ulcer (K25 + K26),

Y_{ij} are observed incidences of duodenal ulcer (K26),

E_{ij} are age-gender standardized expected incidences of (K25 + K26),

$a.mu$, $b.mu$, σ_{mu} are first level unknown parameters,

$a.p$, $b.p$, σ_p are second level unknown parameters,

$X.MU$ and $X.P$ are explanatory variables characterizing spatial differences in welfare.

British Journal of Medical and Health Sciences
URL: http://www.bjmhs.baar.org.uk
ISSN: 2326 – 4276 Vol. 1, No. 7, pp 37-53, **May 2013**

Appendix 2. Formal specification of dynamic Poisson-Binomial model

$$Z_{ijk\ell} \mid \mu_{ijk\ell} \sim Poisson\left(\mu_{ijk\ell}\right)$$

$$log(\mu_{ijk\ell}) = log(E_{ik\ell}) + a.mu_j + b.mu_{k\ell} \cdot X.MU_i + \varepsilon.mu_{ijk\ell}$$

where $\varepsilon.mu_{ijk\ell} \sim Normal\ (\ 0\ ,\ \sigma_{mu}\ )$ (i.i.d.)

$$Y_{ijk\ell} \mid Z_{ijk\ell}\ ,\ p_{ijk\ell} \sim Binomial\ (\ Z_{ijk\ell}\ ,\ p_{ijk\ell}\ )$$

$$logit(p_{ijk\ell}) = a.p_j + b.p_{k\ell} \cdot X.P_i + d.p_{ik} \cdot X.D_j + \varepsilon.p_{ijk\ell}$$

where $\varepsilon.p_{ijk\ell} \sim Normal\ (\ 0\ ,\ \sigma_p\ )$ (i.i.d.) for i=1..I , j=1..J , k=1..K , ℓ=1..L

cell indices: i-th region, j-th year, k-th age group, ℓ-th gender,

$Z_{ijk\ell}$ are observed incidences of gastric and duodenal ulcer (K25 + K26),

$Y_{ijk\ell}$ are observed incidences of duodenal ulcer (K26),

$E_{ik\ell}$ are age-gender specific expected incidences of (K25 + K26) for region i

$a.mu$, $b.mu$, σ_{mu} are first level unknown parameters,

$a.p$, $b.p$, $d.p$, σ_p are second level unknown parameters,

$X.MU$ and $X.P$ are explanatory variables characterizing spatial differences in welfare,

$X.D$ is an explanatory variable characterizing change point in time.

ISSN: 2326 – 4276 Vol. 1, No. 7, pp 37-53, **May 2013**

Appendix 3. OpenBUGS source code of dynamic Poisson-Binomial model

```
model (){
#### Likelihood MU part
         for (l in 1:Ngender) {
         for (k in 1:Nage) {
         for (j in 1:Nyears) {
         for (i in 1:Nareas) {
                     Z[l,k,j,i] ~ dpois(mu[l,k,j,i])
                     ZZ[l,k,j,i] <- max(Z[l,k,j,i] , .1)
                     log(mu[l,k,j,i]) <- log(E[l,k,j,i]) +
                    a.mu[j] + b.mu*  (X.MU[i] - X.MU.MEAN) + err.mu[l,k,j,i]
         }
         }
         }
         }

#### MU part error term
         for (l in 1:Ngender) {
         for (k in 1:Nage) {
         for (j in 1:Nyears) {
         for (i in 1:Nareas) {
                     err.mu[l,k,j,i]  ~ dnorm(0, tau.err.mu)
         }
         }
         }
         }


#### Likelihood P part
         for (l in 1:Ngender) {
         for (k in 1:Nage) {
         for (j in 1:Nyears) {
         for (i in 1:Nareas) {
               Y[l,k,j,i] ~ dbin( p[l,k,j,i] , ZZ[l,k,j,i])
               logit(p[l,k,j,i]) <- a.p + b.p * (X.P[i] - X.P.MEAN) +
              d[k,i]*(X.D[j] - X.D.MEAN) + err.p[l,k,j,i]
         }
         }
         }
         }

#### P part error term
         for (l in 1:Ngender) {
         for (k in 1:Nage) {
         for (j in 1:Nyears) {
         for (i in 1:Nareas) {
               err.p[l,k,j,i] ~ dnorm(0,tau.err.p)
         }
         }
         }
         }
#### dynamics
         for (k in 1:Nage) {
         for (i in 1:Nareas) {
                     d[k,i] ~ dnorm(0,tau.d)
         }
         }
```

ISSN: 2326 – 4276 Vol. 1, No. 7, pp 37-53, **May 2013**

```
#### inits and stoch constraints
        for (j in 1:Nyears) {
              a.mu[j] ~ dflat()
        }
        X.D.MEAN <- mean(X.D[])
        X.P.MEAN <- mean(X.P[])
        X.MU.MEAN <- mean(X.MU[])
        b.mu ~ dflat()
        b.p ~ dflat()
        a.p ~ dflat()
        tau.err.mu ~ dgamma(.005,.005)
        tau.err.p ~ dgamma(.005,.005)
        tau.d ~ dgamma(.005,.005)

}
```

ISSN: 2326 – 4276 Vol. 1, No. 7, pp 37-53, **May 2013**

References

1. Sonnenberg, A. (1985). Geographic and temporal variations in the occurrence of peptic ulcer disease. *Scandinavian Journal of Gastroenterology*, *20*(S110), 11-24.
2. Mirsky, I. A. (1958). Physiologic, psychologic, and social determinants in the etiology of duodenal ulcer. *Digestive Diseases and Sciences*, *3*(4), 285-314.
3. Marshall, B., & Warren, J. R. (1984). Unidentified curved bacilli in the stomach of patients with gastritis and peptic ulceration. *The Lancet*, *323*(8390), 1311-1315.
4. Jennifer, M. N., Richard, M. P. Jr. Helicobacter pylori (2013). An Overview in JeanMarie Houghton (ed.), *Helicobacter Species: Methods and Protocols, Methods in Molecular Biology*, vol. 921
5. Peterson WL, Graham DY. ***Helicobacter pylori***. In: Feldman M, Scharschmidt B, Sleisenger MH, eds. ***Gastrointestinal and Liver Disease: Pathophysiology, Diagnosis, Management***. 6th ed. Philadelphia, Pa: WB Saunders; 1997:604-619.
6. Levenstein, S., Ackerman, S., Kiecolt-Glaser, J. K., & Dubois, A. (1999). Stress and peptic ulcer disease. *JAMA: the journal of the American Medical Association*, *281*(1), 10-11.
7. Nakai Y, Fukunaga M. Peptic Ulcer. *Japan Medical Association Journal* 2003;46(2):61-5.
8. Yuan Y, Padol IT, Hunt RH (2006): Peptic ulcer disease today. *Nature Clinical Practice Gastroenterology & Hepatology* 3: 80-89.
9. Quan C and Talley NJ (2002) Management of peptic ulcer disease not related to Helicobacter pylori or NSAIDs. *Am J Gastroenterol* 97: 2950–2961.
10. Nishikawa, K., Sugiyama, T., Kato, M., Ishizuka, J., Komatsu, Y., Kagaya, H. & Asaka, M. (2000). Non-Helicobacter pylori and non-NSAID peptic ulcer disease in the Japanese population. *European journal of gastroenterology & hepatology*, 12(6), 635-640.
11. Muhammad, J. S., Zaidi, S. F., & Sugiyama, T. (2012). Epidemiological Ins and Outs of Helicobacter pylori: a review. JPMA. *The Journal of the Pakistan Medical Association*, 62(9), 955-959.
12. Bashinskaya B, Nahed BV, Redjal N, Kahle KT, Walcott BP: Trends in Peptic Ulcer Disease and the Identification of Helicobacter Pylori as a Causative Organism: Population-based Estimates from the US Nationwide Inpatient Sample, *J Glob Infect Dis.* 2011 Oct-Dec; 3(4): 366–370.
13. Hunt, K. and Batty, G. D. (2009). Gender and socio-economic inequalities in mortality and health behaviours: an overview. in: H. Graham (editor). *Understanding health inequalities.* Open University Press.
14. Quadrado, L., Heijman, W., & Folmer, H. (2001). Multidimensional analysis of regional inequality: The case of Hungary. *Social Indicators Research*, 56(1), 21-42.
15. Hungarian Central Statistical Office, KSH. (2010). Regional distribution of gross domestic product (GDP), http://www.ksh.hu/docs/eng/xftp/idoszaki/gdpter/egdpter10.pdf
16. Waller, L. A., Gotway, C. A. (2004). *Applied spatial statistics for public health data* (Vol. 368). Wiley-Interscience.
17. Clayton, D., Kaldor, J. (1987). Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics*, 671-681.
18. Tóth, G., Kabos, S., Surján, G. (2012). Joint Modeling of Disease Pairs. *Applied Medical Informatics*, *30*(1), 29-33.
19. Knorr-Held L, Best NG. A shared component model for detecting joint and selective clustering of two diseases. *J R Statist Soc A* 2001;164(1):73-85.

ISSN: 2326 – 4276 Vol. 1, No. 7, pp 37-53, **May 2013**

20. Tassone, E. C., Miranda, M. L., Gelfand, A. E. (2010). Disaggregated spatial modelling for areal unit categorical data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *59*(1), 175-190.
21. Rapporter open source package on CRAN (2011) http://cran.r-project.org/web/packages/rapport/index.html
22. Lunn DJ, Thomas A, Best N, Spiegelhalter D. WinBUGS - a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing* 2000;10:325-37.
23. Agresti A. *Categorical Data Analysis.* New York: John Wiley; 1990.
24. Gál, R. (2010). Pensions, Health and Long-term Care, asisp: Annual National Report, Hungary 2010. European Commission
25. Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4), 583-639.
26. Gkritza, K., Sinha, K. C., Labi, S., & Mannering, F. L. (2008). Influence of highway construction projects on economic development: an empirical assessment. The Annals of Regional Science, 42(3), 545-563.
27. Gelman, A., Price, P. N. (1999). All maps of parameter estimates are misleading. *Statistics in Medicine*, *18*(23), 3221-3234.
28. Gelman, A. (2004). Exploratory data analysis for complex models. *Journal of Computational and Graphical Statistics*, *13*(4).
29. Liu, J., Louis, T. A., Pan, W., Ma, J. Z., Collins, A. J. (2003). Methods for estimating and interpreting provider-specific standardized mortality ratios. *Health Services and Outcomes Research Methodology*, *4*(3), 135-149.
30. Meliker, J. R., Sloan, C. D. (2011). Spatio-temporal epidemiology: principles and opportunities. *Spatial and Spatio-temporal Epidemiology*, *2*(1), 1-9.
31. Ver Hoef, J. M., & Boveng, P. L. (2007). Quasi-Poisson vs. negative binomial regression: how should we model overdispersed count data?. *Ecology*, *88*(11), 2766-2772.
32. Bandyopadhyay, D., Reich, B. J., & Slate, E. H. (2011). A spatial beta-binomial model for clustered count data on dental caries. *Statistical methods in medical research*, *20*(2), 85-102.
33. Best, N., Richardson, S., & Thomson, A. (2005). A comparison of Bayesian spatial models for disease mapping. *Statistical Methods in Medical Research*,*14*(1), 35-59.