

مدل‌های نقشه‌بندی بیماری‌ها

یدالله محرابی، بهزاد مهکی

گروه آمار زیستی و اپیدمیولوژی، دانشگاه علوم پزشکی شهید بهشتی

نقشه‌بندی بیماری شامل مجموعه‌ای از روش‌های آماری است که به تهیه نقشه‌های دقیق براساس برآوردهایی از میزان‌های بروز، شیوع و مرگ و میر منجر می‌شوند تا کاربران بتوانند با اطمینان توزیع جغرافیایی بیماری موردنظر را برآورد کنند. نمایش و تحلیل نقشه‌های بیماری‌ها اکنون به‌عنوان یک ابزار اساسی در تحلیل‌های بهداشت عمومی در نظر گرفته می‌شود. در سالیان اخیر موضوع نقشه‌بندی بیماری‌ها به شکل قابل ملاحظه‌ای گسترش یافته است. این گسترش منجر به استفاده بیشتر از ابزارهای آمار فضایی در تحلیل داده‌ها گردیده است. نقشه‌ها روشی کارا و منحصر به فرد برای پاسخگویی به پرسشهایی درباره مکان و میزان بیماری ایفا می‌کنند. اهداف اصلی نقشه‌بندی بیماری عبارتند از: تشریح تغییرات فضایی در شیوع بیماری به منظور فرمول‌بندی فرضیات علت‌شناسی، ارزیابی خطر و تعیین نواحی دارای خطر بالا که نیازمند جلب توجه و اعمال مداخله هستند، تهیه نقشه دقیقی از خطر بیماری در نواحی به منظور اختصاص بهتر منابع و ساخت اطلس بیماری‌ها.

تاکنون مطالعات موردی متعددی انجام گرفته است که در آن از کاربردها و مدل‌های نقشه‌بندی بیماری‌ها استفاده گردیده است. با این وجود جز در چند مطالعه اندک و به صورت محدود، روشهای متعدد تحلیل نقشه‌های بیماری‌ها با هم مقایسه نشده و شرایط کلی آنها مورد بررسی قرار نگرفته است. با توجه به اهمیت موضوع و کاربرد رو به گسترش نقشه‌بندی بیماری‌ها، در مطالعه حاضر مهمترین مدل‌های نقشه‌بندی (مدل‌های حداکثر درستی، هموارسازی، بیز خطی، بیز تجربی، بیز کامل، آمیخته و چندسطحی) معرفی گردیده و مبنای تئوری، کاربردها، محاسن و معایب هر یک به اختصار بیان گردیده و از جنبه‌های مختلف مورد مقایسه قرار گرفته‌اند.

۱ مقدمه

نقشه‌بندی بیماری شامل مجموعه‌ای از روش‌های آماری است که به تهیه نقشه‌های دقیق براساس برآوردهایی از میزان‌های بروز، شیوع و مرگ و میر منجر می‌شوند تا کاربران بتوانند با اطمینان توزیع جغرافیایی بیماری موردنظر را برآورد کنند [۱، ۲].

نمایش و تحلیل نقشه‌های بیماری‌ها اکنون به‌عنوان یک ابزار اساسی در تحلیل‌های بهداشت عمومی در نظر گرفته می‌شود. قدیمی‌ترین مثال نقشه‌بندی بیماری‌ها، نقشه آدرس قربانیان وبا برحسب فاصله از منابع آبی در لندن است که در سال ۱۸۵۴ توسط جان اسنوارائه شده است [۳]. در سالیان اخیر موضوع نقشه‌بندی بیماری‌ها به شکل قابل ملاحظه‌ای گسترش یافته است. این گسترش منجر به استفاده بیشتر از ابزارهای آمار فضایی در تحلیل داده‌ها گردیده است [۲].

نقشه‌ها روشی کارا و منحصر به فرد برای پاسخگویی به پرسشهایی درباره مکان و میزان بیماری ایفا می‌کنند. پاسخ به این پرسش‌ها منجر به تحلیل جغرافیایی بیماری می‌گردد. تحلیل تغییرات جغرافیایی میزان‌های بیماری‌ها دارای کاربردهای بسیاری است. این کاربردها تنها محدود به فرمول‌بندی و ارزیابی فرضیات علت‌شناسی نیست بلکه در حوزه اختصاص منابع یا ارزیابی خطر در بهداشت عمومی نیز کاربرد دارد. اهداف اصلی نقشه بندی بیماری عبارتند از:

- ۱- تشریح تغییرات فضایی در شیوع بیماری به منظور فرمول‌بندی فرضیات علت‌شناسی.
- ۲- ارزیابی خطر و تعیین نواحی دارای خطر بالا که نیازمند جلب توجه و اعمال مداخله هستند.
- ۳- تهیه نقشه دقیقی از خطر بیماری در نواحی به منظور اختصاص بهتر منابع.
- ۴- ساخت اطلس بیماری‌ها [۱، ۲، ۴، ۵].

دو حالت رایج در مطالعات توزیع جغرافیایی بیماری‌ها وجود دارد: ۱- ناحیه‌ای تعریف می‌گردد و مکان موارد یک بیماری خاص، درون این ناحیه در یک دوره زمانی ثابت، ثبت می‌گردد. این مکانها معمولاً آدرسهای محل سکونت هستند. پس از مشخص شدن این آدرسها، ادامه کار با تحلیل مستقیم مکان موارد امکان‌پذیر است. این حالت را تحلیل وقایع موردی^۱ می‌نامیم. اغلب این تحلیل نیازمند استفاده از مدل‌های روند نقطه‌ای و متدولوژی مربوط است [۶، ۷، ۸]. ۲- در اغلب موارد به دلیل لزوم محرمانه بودن برخی از مسائل در پزشکی، امکان دستیابی به داده‌ها در مقیاس و سطح بالا وجود ندارد و باید دسته‌بندی مجدد برای تحلیل تعداد موارد درون نواحی کوچک مطالعه انجام شود. این نواحی کوچک، مناطقی اختیاری مانند شهرستان، شهر، منطقه، ناحیه و محله هستند که معمولاً برای اهداف اجرایی و مدیریتی تعریف می‌شوند. تحلیل این نوع داده‌ها که شامل تعداد موارد در نواحی کوچک هستند را تحلیل تعداد موارد در نواحی^۲ می‌نامیم. تعداد موارد درون یک ناحیه از تجمیع همه

¹case-event analysis

²tract count analysis

موارد درون آن ناحیه به دست می آید. بدین ترتیب مکان‌های جداگانه‌ی موارد از دست رفته و اطلاعات تعداد موارد را تنها می‌توان از مکان ناحیه استخراج کرد. اغلب نقطه میانی ناحیه، برای معرفی مکان ناحیه در نظر گرفته می‌شود. در یک ناحیه انتخاب شده برای مطالعه، m ناحیه را در نظر می‌گیریم. مقادیر درون m ناحیه را با y_i ($i = 1, 2, \dots, m$) نشان می‌دهیم. تاکنون مطالعات موردی متعددی انجام گرفته است که در آن از کاربردها و مدل‌های نقشه‌بندی بیماری‌ها استفاده گردیده است. به عنوان نمونه می‌توان به نقشه‌بندی سندروم مرگ سریع نوزادان، سرطان لب در اسکاتلند، مرگ و میر کودکان در میسوری، سرطان در فرانسه و سرطان ریه در آلمان اشاره کرد [۹ - ۱۵]. با این وجود جز در چند مطالعه اندک و به صورت محدود، روشهای متعدد تحلیل نقشه‌های بیماری‌ها با هم مقایسه نشده و شرایط کلی آنها مورد بررسی قرار نگرفته است. با توجه به اهمیت موضوع و کاربرد رو به گسترش نقشه‌بندی بیماری‌ها، در این مقاله مهمترین مدل‌های نقشه‌بندی معرفی و مبانی تئوری، کاربردها، محاسن و معایب هریک به اختصار بیان می‌گردد.

۲ مدل‌های نقشه‌بندی بیماری‌ها

۲-۱ مدل حداکثر درست‌نمایی (نسبت مرگ و میر استاندارد شده (SMR))

برای ارزیابی وضعیت یک ناحیه از نظر میزانهای بروز و ابتلای بیماری، به عنوان گام نخست مناسب است که ارزیابی کنیم چه میزان بروزی برای آن بیماری در آن ناحیه موردانتظار است، سپس موارد مشاهده شده را با بروز موردانتظار مقایسه کنیم. این رویکرد برای تحلیل مقادیر درون نواحی استفاده می‌شود.

نسبت مقادیر مشاهده شده به مقادیر موردانتظار مرگ و میر، نسبت مرگ و میر استاندارد شده^۳ (SMR) نامیده می‌شود. در مطالعات مربوط به ابتلا و بروز بیماری، از نسبت مقادیر مشاهده شده به مقادیر موردانتظار بروز بیماری استفاده می‌کنیم که نسبت بروز استاندارد شده^۴ (SIR) نامیده می‌شود. این نسبت برآورد خطر نسبی درون هر ناحیه است [۱، ۲].

برای به‌کارگیری این روش، نقشه به n ناحیه مجاور نامتداخل تقسیم می‌شود. O_i تعداد رویداد موردنظر مشاهده شده در ناحیه i ام را نشان می‌دهد. ($i = 1, 2, \dots, n$)

³Standardized Mortality/Morbidity Ratio

⁴Standardized Incidence Ratio

تعداد رویداد موردانتظار در ناحیه i ام با (E_i) نشان داده می‌شود که طی دوره مطالعه ثابت و معلوم فرض شده و از حاصل ضرب نسبت جمعیت هر ناحیه به جمعیت کل در کل موارد مشاهده شده به دست می‌آید. خطر نسبی بیماری در ناحیه i ام نیز با θ_i نشان داده می‌شود. در نقشه‌بندی بیماری، فرض می‌شود که تعداد رویدادها، بین نواحی، مستقل از هم بوده و از توزیع پواسون با میانگین $E_i\theta_i$ پیروی می‌کند. بنابراین درست نمایی O_i از رابطه زیر به دست می‌آید:

$$L(\theta) = \prod_{i=1}^n \frac{\exp(-E_i\theta_i)}{O_i!} \{E_i\theta_i\}^{O_i} = \prod_{i=1}^n \text{pois}(O_i; E_i\theta_i) \propto \prod_{i=1}^n \theta_i^{O_i} \cdot \exp \left\{ - \sum_{i=1}^n E_i\theta_i \right\}$$

لگاریتم تابع درست‌نمایی نیز عبارت است از:

$$l(\theta) = \ln L(\theta) \propto \sum_{i=1}^n O_i \ln \theta_i - \sum_{i=1}^n E_i \cdot \theta_i$$

بدین ترتیب برآورد حداکثر درست‌نمایی عبات است از:

$$\hat{\theta} = SMR_i = \frac{O_i}{E_i}$$

گرچه استفاده از SMR و SIRها در نقشه‌بندی بیماری‌ها بسیار رایج است اما نقاط ضعفی در استفاده از این شاخص‌ها وجود دارد.

۱- SMR و SIRها مبتنی بر برآورد نسبتی هستند و بنابراین اندک تغییراتی در مقادیر موردانتظار می‌توانند منجر به تغییرات بزرگی در برآورد گردد.

۲- هنگامی که مقدار موردانتظار صفر (یا نزدیک به صفر) باشد، SMR و SIR برای هر مقدار مشاهده شده مثبت، وجود نداشته یا بسیار بزرگ خواهد بود.

۳- SMR و SIRهای صفر تفاوت‌های موجود در مقادیر موردانتظار را منعکس نمی‌کنند.

۴- واریانس SMR و SIR برابر $\frac{\theta_i}{E_i}$ و متناسب با عکس مقدار موردانتظار بوده و در نواحی با جمعیت بزرگ، دارای مقادیر کوچک و در نواحی با جمعیت کم، دارای مقدار بزرگ است. این مسئله تصمیم‌گیری براساس آن را مشکل می‌کند.

۵- SMR و SIR تلاشی برای نشان دادن ساختار پایه‌ای داده‌ها انجام نمی‌دهد و دقیقاً یک برآورد اشباع از خطر نسبی است و بنابراین استفاده از آن مقرون به صرفه نیست.

از آنجایی که این برآوردگر دارای عیوب مذکور است می‌توان از مدل‌های دیگر برای غلبه بر این مشکلات استفاده کرد.

۲-۲ مدل‌های هموارسازی

رویکردی برای تحلیل اکتشافی داده‌ها که اغلب در مسائل تک متغیره به کار می‌رود و در جهت هموارسازی نوسانات و اختلالات تلاش می‌کند [۱۶]. رایج‌ترین روش در این کلاس، روش هموارسازی کرنل نادارایا - واتسون است که هنگام استفاده برای SMR، میانگین وزنی

$$\hat{\theta} = \sum_{j \neq i} w_j \varphi \text{ (ها) است:}$$

که در آن وزن‌ها از رابطه $w_j = \frac{k((\varphi_i - \varphi_j)/h)}{\sum_i k((\varphi_i - \varphi_j)/h)}$ به دست آمده و توابعی از مقادیر

همسایگی‌ها هستند و K تابع چگالی احتمال متقارن با میانگین صفر است و h برای مینیمم کردن برخی از معیارهای نکویی برازش انتخاب می‌شود. رویکردهای وزن‌دهی دیگری نیز وجود دارند اما ویژگی‌های آماری این برآوردگر را ندارند. ثابت هموارسازی h را می‌توان با استفاده از cross-validation انتخاب کرد.

رویکرد دیگر هموارسازی، هموارسازی جداگانه صورت و مخرج SMR کسر است. یک روش توسط کلسال و دیگل برای برآورد خطر نسبی مکان و آدرس رویدادها پیشنهاد شده است [۱۷، ۱۸]. این روش برای داده‌های مکان رویدادهای مورد و شاهد به کار می‌رود که مکان مشاهدات براساس مکان جور نشده‌اند و از cross-validation حداقل مربعات برای به دست آوردن ثابت هموارسازی مشترک برای برآوردگر به دست آمده استفاده می‌شود.

۳-۲ مدل‌های بیز خطی

مبتنی بر توابع خطی از SMR هستند. بکارگیری این روش‌ها در نقشه‌بندی بیماری‌ها، نخستین بار توسط مارشال بیان شده است [۱۹، ۱۱]. ثابت شده است که برآورد بیز خطی $\{\theta_i\}$ براساس SMR (φ_i) عبارت است از:

$$\hat{\theta} = E_{\Theta}\{\theta_i\} + \frac{cov(\theta_i, \varphi_i)}{var_{\Theta}(\varphi_i)}[\varphi_i - E_{\Theta}\{\theta_i\}]$$

که در آن $E_{\Theta}\{\theta_i\}$ امید θ_i در توزیع پیشین است. این برآوردگر بهترین برآوردگر به فرم $\hat{\theta}_i = \alpha_i + \beta \varphi_i$ است. این معادله منجر به برآوردگر به فرم

$$\hat{\theta}_i = m_i + \frac{v_i}{\frac{m_i}{v_i} + v_i} \{\varphi_i - m_i\}$$

می‌شود که m_i و v_i به ترتیب میانگین و واریانس پیشین خطر نسبی اند. بنابراین اگر بتوان میانگین و واریانس پیشین θ_i را برآورد کرد، به سادگی می‌توان برآورد پسین آن را به دست آورد.

برخی از ویژگیهای این روش عبارتند از:

- ۱- اگر E_i بزرگ باشد آنگاه برآورد به سمت SMR میل می‌کند اما اگر E_i کوچک باشد برآورد به سمت میانگین پیشین میل می‌کند.
 - ۲- این برآوردگرها اصل درستنمایی را نقض کرده و به همین دلیل از منظر تئوری سؤال برانگیزند. اصل درستنمای مشخص می‌کند که اگر دو درستنمایی متناسب با هم بوده و ثابت تناسب وابسته به پارامتر مورد نظر نباشد، آنگاه نتایج به دست آمده باید مشابه باشد.
 - ۳- این رویکرد تنها مقادیر برآوردها را ارائه می‌دهد و اقدامی در جهت محاسبه فاصله اطمینان و انجام آزمون فرض صورت نمی‌پذیرد.
- مارشال دو نوع از این برآوردگر را پیشنهاد کرده است:
برآورد کلی: میانگین و واریانس ثابت فرض شده و از روش گشتاوری به دست می‌آیند.

$$\hat{v} = \frac{1}{\sum E_i} \sum E_i \cdot \{\varphi_i - \hat{m}\}^2 - \frac{\hat{m}}{\sum E_i}, \quad \hat{m} = \frac{\sum O_i}{\sum E_i}$$

با توجه به اینکه که ممکن است مقدار برآورد v منفی باشد، اگر $\hat{v} < 0$ باشد، $\hat{\theta} = \hat{m}$ در نظر گرفته می‌شود.

برآورد موضعی: برآورد میانگین و واریانس پیشین صرفاً با استفاده از نواحی مجاور به دست

$$\hat{v}_i = \frac{1}{\sum_{j \sim i} E_j} \sum_{j \sim i} E_j \cdot \{\varphi_j - \hat{m}_j\}^2 - \frac{\hat{m}_j}{\sum_{j \sim i} E_j} \quad \text{و} \quad \hat{m} = \frac{\sum_{j \sim i} O_i}{\sum_{j \sim i} E_i}$$

مارشال انحراف معیار این برآوردگرها را با استفاده از روش دلتا به دست آورده است. در نظر گرفتن واریانس توزیع پسین $(var(\theta_i|O_i))$ نسبت به واریانس مقادیر برآورد شده $(var(\hat{\theta}_i))$ ارجح است.

۴-۲ مدل‌های بیز تجربی

مشابه روشهای بیزی بوده اما توزیع پیشین خطرات نسبی از داده‌های مشاهده شده برآورد می‌شود [۲۰، ۲۱]. این روشها فرض می‌کنند که $\{\theta_i\}$ از توزیع خاصی مانند $g(\theta)$ پیروی می‌کنند که با مجموعه‌ای از پارامترها مشخص می‌شوند (Ω) . این پارامترها را می‌توان

از داده‌ها و با حداکثر نمودن درست‌نمایی حاشیه‌ای $l(\Omega) = \int f(x|\theta)g(\theta|\Omega)d\theta$ برآورد کرد که در آن $f(x|\theta)$ توزیع توأم موارد مشاهده شده است. حداکثر نمودن را می‌توان از طریق الگوریتم EM که بی‌نیاز از انتگرال‌گیری است یا از طریق تقریبات تحلیلی برای انتگرال از قبیل تقریب لاپلاس انجام داد. مهم‌ترین مدل‌هایی که در چارچوب مدل‌های بیز تجربی قرار دارند مدل‌های گاما - پواسون و لگ نرمال هستند که توسط کلایتون و کالدور معرفی گردیده‌اند [۱۰].

مدل گاما - پواسون

در این مدل، فرض بر این است که تعداد مرگ‌ها در نواحی دو به دو مستقل بوده و از توزیع پواسون پیروی می‌کنند ($O_i \sim poisson(E_i\theta_i)$). توزیع پیشین برای خطرات نسبی نیز $gamma(a, b)$ در نظر گرفته می‌شود. برای به دست آوردن توزیع پسین گاما، توزیع پیشین گاما برای خطرات نسبی، با درست‌نمایی پواسون ترکیب می‌شود. بدین ترتیب خطر نسبی دارای توزیع پسین $Gamma(a + O_i, b + E_i)$ با میانگین $E[\theta_i|O_i, a, b] = \frac{a + O_i}{b + E_i} = w_i SMR_i + (1 - w_i) \frac{a}{b}$ است که در آن $w_i = \frac{E_i}{b + E_i}$. بنابراین میانگین پسین برای λ امین ناحیه، میانگین وزنی SMR برای λ امین ناحیه و خطر نسبی بیماری در نقشه کلی است که وزن به طور معکوس با واریانس SMR در ارتباط است. اگر مقادیر مشاهده شده و مقادیر موردانتظار هر دو بزرگ باشند آنگاه برآوردگر به سمت SMR میل می‌کند. اما اگر مقادیر مشاهده شده و مقادیر موردانتظار هر دو کوچک باشند آنگاه به سمت میانگین پیشین $(\frac{O_i}{E_i})$ میل می‌کند [۱۰، ۲۰، ۲۱]. با وجود این که مقادیر برآورد شده را می‌توان از برآوردگر بیز خطی کامل به دست آورد، مدل گامات پواسون دارای چند حسن است:

- ۱- این مدل پسین کامل را برآورد می‌کند که می‌تواند برای محاسبه فواصل اطمینان و انجام آزمون‌های فرض استفاده شود درحالی‌که روش بیز خطی چنین قابلیت را ندارد.
 - ۲- حتی اگر توزیع واقعی خطر نسبی گاما باشد مدل گاما - پواسون میانگین و واریانس این توزیع را از طریق حداکثر درست‌نمایی برآورد می‌کند که نسبت به روش گشتاوری استفاده شده در مدل‌های بیز خطی برتر و ارجح است [۱۰، ۲۰، ۲۱].
- عیب اساسی این روش این است که: این مدل قادر به در نظر گرفتن همبستگی فضایی نیست [۲۰].

مدل لگ نرمال

توزیع پیشین گاما برای خطر نسبی گرچه از منظر ریاضیاتی مناسب است، می‌تواند محدود کننده باشد زیرا ورود کووریت به آن مشکل بوده و امکان در نظر گرفتن همبستگی فضایی در نواحی نزدیکتر وجود ندارد [۱]. مدل لگ نرمال برای خطرات نسبی بیماری انعطاف پذیرتر از مدل گاما - پوآسون است:

$$y_i \sim \text{poisson}(E_i \theta_i)$$

$$\log \theta_i = \alpha + v_i$$

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

۵-۲ مدل بیز کامل (BYM)

در این مدل فرض بر این است که تعداد مرگ‌ها در هر منطقه دوه‌دو مستقل بوده و از توزیع پوآسون پیروی می‌کنند. این مدل توسط کلایتون و کالدور مطرح شده و توسط بیسگ، یورک و مولی بسط داده شد [۱۰، ۲۲]. فرمول‌بندی این مدل به شرح زیر است:

$$y_i \sim \text{Poisson}(E_i \theta_i)$$

$$\log \theta_i = \alpha + u_i + v_i$$

در این مدل پارامتر خطر نسبی بیماری به سه مؤلفه تجزیه می‌شود:

۱- α : مؤلفه روند که سطح کلی خطر نسبی است.

۲- u_i (بیش پراکندگی فضایی (ناهمگنی همبسته فضایی)): منطقی است که نواحی نزدیک هم، خطرات نسبی مشابه هم داشته باشند. برای در نظر گرفتن این امکان، متغیر تصادفی u_i که با دیگر $\{u_i\}$ ها همبسته است در مدل وارد می‌شود. برای این مؤلفه، ساختار همبستگی فضایی استفاده می‌شود که برآورد خطر نسبی در هر ناحیه وابسته به نواحی همسایه است. مدل خودبازگشتی شرطی پیشنهاد شده توسط بیسگ و همکاران عبارت است از:

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

$$\text{که در آن: } \bar{u}_i = \frac{1}{\sum_j w_{ij}} \sum_j u_j w_{ij} \text{ و } \tau_i^2 = \frac{\tau_u^2}{\sum_j w_{ij}}$$

اگر i و j مجاور باشند، $w_{ij} = 1$ و در غیر این صورت $w_{ij} = 0$ است.

۳- v_i (بیش پراکندگی غیرفضایی (ناهمگنی ناهمبسته فضایی)): با فرمول‌بندی مدل برای بیش پراکندگی همبسته فضایی، واریانس وابسته به تعداد همسایگی‌هاست و استقلال به خوبی تعریف نمی‌شود. برای حل این مشکل، می‌توان عبارت دیگری (v_i) را معرفی کرد که پارامتر بیش پراکندگی ناهمبسته استاندارد است. توزیع پیشین برای این پارامتر عبارت است از: $v_i \sim N(0, \tau_v^2)$ پارامترهای τ_u^2 و τ_v^2 تغییرپذیری u و v را کنترل می‌کنند. در تحلیل بیز کامل باید برای این پارامترها توزیع‌های پیشین تعیین گردد. برناردینی توزیع‌های پیشین گاما را برای این پارامترها پیشنهاد داده است [۲۲، ۲۳].

۲-۶ مدل‌های آمیخته

می‌توان یک مدل ساخت که توزیع حاشیه‌ای مقادیر، توزیعی آمیخته شامل مجموعه‌ای از اجزاء گسسته در نظر گرفته شود. این رویکرد کاملاً از مدل‌های قبلی - که در آن پراکندگی و تغییرات در خطر با مخلوطی از چند جزء مدل‌بندی می‌شود - متفاوت است. این مدل K سطح متفاوت از خطر را در نظر می‌گیرد $\{\theta_1, \theta_2, \dots, \theta_k\}$ و مقادیر مشاهده شده را از ناحیه‌ای با خطر θ_j و احتمال p_j فرض می‌کند. بدین ترتیب تابع درست‌نمایی آن عبارت است از:

$$L = \prod_{i=1}^n \sum_{j=1}^k p_j \text{pois}(O_i, E_i, \theta_j)$$

در این حالت جواب به فرم بسته برای خطرات نسبی وجود ندارد اما الگوریتم‌های قابل اعتمادی مانند VEM برای این هدف وجود دارند [۲۴]. این الگوریتم‌ها برآوردهای حداکثر درست‌نمایی برای $\{\theta_i\}$ و $\{p_j\}$ را ارائه می‌دهند و همچنین می‌توانند برای طبقه‌بندی هر ناحیه با یک خطر استفاده شوند. یک راه دیگر استفاده از امید ریاضی پسین

$$\hat{\theta}_i = \frac{\sum_{j=1}^k \hat{\theta}_j \hat{p}_j \text{Pois}(O_i, E_i, \hat{\theta}_j)}{\sum_{j=1}^k \hat{p}_j \text{Pois}(O_i, E_i, \hat{\theta}_j)}$$

کلارک (۲۰۰۲) فرض می‌شود که لگاریتم خطر نسبی می‌تواند به سه مؤلفه جمع‌پذیر

تفکیک شود. یک مؤلفه ثابت نشان دهنده‌ی ناهمگنی ناهمبسته (v) که بیش‌پراکندگی در یک ناحیه خاص را اندازه‌گیری می‌کند و دو مؤلفه آمیخته (u, φ) که نشان دهنده‌ی جنبه‌های متفاوت همبستگی فضایی هستند. این دو مؤلفه ترکیبی از یک مؤلفه همبستگی فضایی و یک مؤلفه که گامهای گسسته را مدل‌بندی می‌کند می‌باشند [۲۵].

مدل آمیخته عبارت است از: $\log \theta_i = \alpha + v_i + p_i u_i + (1 - p_i) \varphi_i$

این مدل دارای حالات خاصی برحسب مقادیر p_i است. اگر $p_i = 1$ این مدل مشابه مدل BYM است و اگر $p_i = 0$ باشد این مدل تبدیل به مدل گام خالص می‌شود. برای پارامترهای این مدل، توزیع‌های پیشین تعیین می‌گردد.

۷-۲ مدل‌های چندسطحی

روش دیگر نقشه‌بندی بیماری‌ها فرم لگ خطی را به کار می‌گیرد. $\theta_i = \exp\{\beta_0 + v_i\}$ که عبارت تصادفی دارای توزیع نرمال با میانگین صفر است $v_i \sim N(0, \sigma_v^2)$ و σ_v^2 واریانس اثرات تصادفی v است. این مدل را می‌توان برحسب تعداد موارد نیز نوشت (به جای میزانها):

$$y_i \sim \text{Poisson}(\mu_i), \quad \log(\mu_i) = \log(E_i) + \beta_0 + v_i$$

در حالت کلی مدل‌های چندسطحی به داده‌هایی برازش می‌یابند که دارای سطوح خوشه‌بندی در ساختار خود هستند. در نقشه‌بندی بیماری‌ها و کاربردهای جغرافیایی، این سطوح، سطوح متفاوت جغرافیایی هستند [۲۶]. به عنوان مثال بخش‌های شهرستان‌ها، شهرستان‌های استان‌ها و استان‌های کشور. سپس برای هر سطح جغرافیایی می‌توان اثرات تصادفی دارای توزیع نرمال را برازش داد. به عنوان مثال برای شهرستان‌های استان‌ها، می‌توان برازش زیر را داشت:

$$y_{ij} \sim \text{Poisson}(\mu_{ij}), \quad \log(\mu_{ij}) = \log(E_{ij}) + \beta_0 + v_{ij} + u_{ij}$$

که اثرات تصادفی شهرستان‌ها و استان‌ها دارای توزیع نرمال هستند $v_j \sim N(0, \sigma_v^2)$ و $u_{ij} \sim N(0, \sigma_u^2)$

مدل‌های چندسطحی با پاسخ پواسون را می‌توان هم با رویکردهای فراوانی‌گرایانه و هم با رویکردهای بیزی برازش داد. رویکردهای فراوانی‌گرایانه معمولاً نیازمند استفاده از تقریب هستند. به عنوان مثال نرم‌افزار MLwiN از روشهای شبه‌درست‌نمایی استفاده می‌کند که نیازمند استفاده از تقریب‌های سری‌های تیلور برای تبدیل مسئله به حالتی است که بتوان با استفاده از الگوریتم حداقل مربعات مکرر عمومی برازش را انجام داد [۲۷، ۲۸، ۲۹، ۳۰].

دیگر رویکردهای رایج فراوانی‌گرایانه شامل تقریبات لاپلاس و مربع نرمال است. برآورد بیزی شامل بسط مدل‌ها به مدل‌هایی است که شامل اطلاعات پیشین بوده و از برآوردهای MCMC برای برازش استفاده می‌کنند [۳۱، ۳۲].

۳ مقایسه مدل‌ها

با وجود مطالب و مقالات فراوان درخصوص طراحی مدل‌های نقشه‌بندی بیماری‌ها، توجه اندکی بر ارزیابی و مقایسه مدل‌ها به‌ویژه با استفاده از معیارهای نکویی برازش مدل‌ها معطوف شده است. لائسون و همکاران (۲۰۰۰) در مطالعه‌ای با استفاده از شبیه‌سازی داده‌ها از برخی از مدل‌های ممکن، به ارزیابی چگونگی پوشش پراکندگی فضایی توسط هر کدام از مدل‌ها و بررسی شایستگی نسبی هر کدام از مدل‌های برازش‌یافته با استفاده از چندین معیار نکویی برازش پرداختند [۲]. ریچاردسون (۲۰۰۴) یک مطالعه شبیه‌سازی جامع را برای روشن نمودن میزان هموارسازی خطرات در مدل‌های بیزی نقشه‌بندی بیماری‌ها به‌منظور ارزیابی عملکرد آنها در تشخیص نواحی با خطر بالای واقعی انجام داد [۳۳]. مارشال (۱۹۹۱) روشهای بیز تجربی و برخی روش‌های ابتدایی بیز کامل در نقشه‌بندی بیماری‌ها را مرور و مورد مقایسه قرار داد (۴۳). بیتل (۲۰۰۰) درخصوص رویکردهای ناپارامتری و مبتنی بر مدل بحث نمود [۳۵]. لائسون (۱۹۹۹) و ریچاردسون (۲۰۰۳) استفاده‌ی مدل‌های فضایی در کاربردهای اپیدمیولوژیک را مورد بحث قرار دادند [۱۴، ۳۶]. ماندا و لیلاند (۲۰۰۷) روش‌های فراوانی‌گرایانه (حداکثر درستمایی) را با مدل‌های فضایی بیزی مقایسه کرده‌اند [۳۷]. برناردینی و مونتومولی مقایسه‌ای بین مدل‌های بیز کامل و بیز تجربی و مدل‌هایی که با پراکندگی بیماری‌ها سروکار دارند انجام داده‌اند [۳۸]. بست (۲۰۰۵) جدیدترین مدل‌های بیزی که در نقشه‌بندی بیماری‌ها از مدل‌های بیز کامل استفاده می‌کنند را خلاصه کرده است [۳۹]. کلایتون و کالدور مقایساتی درباره مدل‌های بیز تجربی و بیز کامل انجام داده‌اند [۱۰]. تلاش‌های دیگری برای ارزیابی حساسیت مدل بیز کامل نیز انجام شده است که این ارزیابی‌ها نیز تنها محدود به دامنه محدودی از مدل‌های بیزی است [۲۶، ۴۰]. براساس این مطالعات نتایج زیر حاصل شده است.

الف) مقایسه مدل‌های بیزی و غیربیزی: با وجود تشابه برآورد پارامترها در دو مدل، برآورد واریانس اثرات تصادفی در رویکرد فراوانی‌گرایانه عموماً کمتر از روش بیزی است. همچنین اجرای روش بیزی آسان‌تر است و برآوردهای قابل اطمینان‌تری ارائه می‌دهند میزان‌های بروز بیماری‌ها که با استفاده از روش بیزی به‌دست می‌آیند دقیق‌تر (دارای فاصله اطمینان

کوتاه‌تر) هستند [۴۱، ۴۲]. ثابت شده است که روش‌های بی‌زی مبادله‌ای بین آریبی و کاهش واریانس برآوردها ایجاد می‌کنند. ثابت شده است که به‌ویژه هنگامی که حجم نمونه کم است این روشها ویژگی‌های خوبی برحسب کمینه‌کردن زیان مربع خطا دارند [۱۴]. کاهش واریانس از طریق اقتباس اطلاعات منتج از ساختار سلسله مراتبی در نظر گرفته شده، به‌دست می‌آید. اثر انقباض^۵ نیز وابسته به ساختار پیشین مفروض است. در نتیجه انتظار داریم که پیشنهادهای فضایی متفاوت منجر به انقباض‌های متفاوت گردند و بدین ترتیب ویژگی‌های مطلوب و مورد نظر برای برآوردها وابسته به هدف غایی برآورد است [۱۴].

ب) مدل‌های هموارسازی: هموارسازی با حذف نوسانات، امکان مشاهده الگوهای ذاتی و واقعی فضایی را فراهم می‌کند و برآوردهای دقیق‌تر و فواصل اطمینان کوتاه‌تری نسبت به حالت عدم هموارسازی را نتیجه می‌دهد. در حالت کلی، اثر هموارسازی کاهش تعداد واحدها در بالاترین و پایین‌ترین گروهها و بدین ترتیب افزایش تعداد واحدها در گروههای میانی است. در حالت کلی این مدلها عملکرد مطلوبی نداشته و استفاده از هموارساز کرنل توصیه نمی‌شود اما این روش در برآورد لگاریتم خطر نسبی عملکرد بهتری دارد [۲]. در صورت وجود خودهمبستگی هموارساز کرنل عملکردی ضعیف دارد [۴۳].

ج) مدل‌های بی‌ز خطی: در بین روش‌های بی‌ز خطی، مدل کلی مارشال نتایج بهتری را ارائه می‌دهد گرچه خودهمبستگی در باقیمانده‌ها را به‌جا می‌گذارد. همچنین این مدلها رفتاری مشابه مدل گاما - پواسون از خود نشان می‌دهند [۲].

د) مقایسه مدل‌های بی‌ز تجربی و بی‌ز کامل: برآوردهای ارائه شده توسط مدل بی‌ز تجربی، بنیابین برآوردهای مدل‌های SMR و بی‌ز کامل هستند [۴۳]. مدل بی‌ز تجربی از این لحاظ مورد انتقاد است که برآوردهای تولید شده بسیار زیاد به سمت میانگین کلی منقبض‌اند. این مسئله در توزیع مجموعه‌ی برآوردها منعکس است و پراکندگی کمتری در توزیع برآوردها نسبت به برآوردهای SMR وجود دارد. به‌همین دلیل برآوردهای بی‌ز تجربی مقید^۶ معرفی گردیده‌اند که برآوردهایی تولید می‌کنند که پراکندگی توأمشان مشابه داده‌های خام است [۴۴]. مدل بی‌ز کامل برتر از مدل بی‌ز تجربی است زیرا عدم قطعیت در پارامترهای مدل را در نظر می‌گیرد درحالی‌که برآوردهای مدل بی‌ز تجربی، برآوردها را نسبت به برآوردهای نقطه‌ای پارامترهای مدل شرطی می‌کند. بنابراین ممکن است برآوردهای مدل بی‌ز تجربی صحت کمتری داشته باشند. اما از سوی دیگر روش بی‌ز کامل نیازمند برنامه‌های رایانه‌ای پیچیده‌تر بوده و اجرای آن مشکل‌تر است [۳۹]. با این وجود مدل بی‌ز تجربی برای ارزیابی

⁵shrinkage⁶constrained

اولیه و مقدماتی خطر بیماری در سیستم‌های اطلاعات جغرافیایی^۷ - که تعداد زیادی داده مدیریت می‌شوند و الگوریتم مورد استفاده باید سریع و آسان باشد - مفید است. سپس اگر تحلیل اکتشافی پراکندگی و تغییرات قابل توجهی در خطرات نواحی را نشان دهد، برآورد بیز کامل را می‌توان استفاده نمود [۴۵]. در برآورد خطر نسبی نیز با وجود تفاوت برآورد خطر نسبی تحت دو روش، رتبه خطر نسبی در کل ناحیه نقشه‌بندی مشابه است [۱۰]. بررسی مجلات مهم اپیدمیولوژیک در دهه ۲۰۰۰-۱۹۹۰ نشان داد که مدل BYM تنها مدل فضایی بیز کامل مورد استفاده در کاربردهای منتشرشده‌ی نقشه‌بندی بیماری‌ها بوده است. مقایسات انجام شده در مطالعه اخیر و مطالعات دیگر نشان می‌دهد که این مدل هنوز هم یک مدل مناسب در نقشه‌بندی بیماری‌های نواحی کوچک است اما مدل‌های فضایی بیزی دیگر که در سالیان اخیر پیشنهاد شده‌اند، در برخی کاربردها توانایی بهبود برآوردها را دارند [۳۹]. مدل BYM بالاترین ΔBIC را در بین مدل‌های BYM، گاما-پوآسون و مدل‌های آمیخته نتیجه می‌دهد [۲].

ه) مدل‌های آمیخته: مدل BYM و مدل آمیخته دارای ویژگی‌های خوب و مطلوبی در مدل‌بندی یک بیماری هستند. ویژگی اختصاصی مدل آمیخته، توانایی اقتباس اطلاعات هم از نواحی مجاور و هم از نواحی غیرمجاور است. این ویژگی منجر به اربیی کمتر برآوردهای نقطه‌ای خطرات، نسبت به مدل‌های دیگر برای خطرات بالا می‌گردد. مدل BYM گرایش به بیش‌هموارسازی برآوردهای نقطه‌ای دارد اما در طبقه‌بندی کلی نواحی، اندکی بهتر از مدل آمیخته است. مدل‌های آمیخته حتی هنگامی که خطرات نسبی واقعی آمیخته باشند، خطرات نسبی را به خوبی پوشش نمی‌دهند [۳۹]. این مدل‌ها در تحلیل مقادیر خوب به خوبی عمل می‌کنند اما این تحلیل‌ها معمولاً با مدل‌های BYM و گاما بهتر انجام می‌شوند [۲]. چندین مدل برای تشریح توزیع فضایی میزان‌ها معرفی شده‌اند. این مدل‌ها، مدل‌های پوآسون آمیخته^۸ بوده که اثرات فضایی اختصاصی نواحی را وارد مدل می‌کنند. این اثرات می‌توانند معرف عوامل محیطی باشند و برخی اوقات همبستگی فضایی را نیز نشان دهند [۱۰، ۲۴، ۳۸، ۴۶]. بسطی از مدل پوآسون توسط لوسون و کلارک (۲۰۰۲) پیشنهاد شده است که شامل مخلوطی از اجزاء خودبازگشتی شرطی نرمال و غیرنرمال است که جزء غیرنرمال به منظور برقرار کردن گام‌های گسسته^۹ در ساختار خطر برای اجتناب از بیش‌هموارسازی^{۱۰} خطرات، در نظر گرفته شده است [۲۳].

⁷ Geographical Information Systems

⁸ Mixed Poisson Models

⁹ discrete jumps

¹⁰ over smoothing

و) مدل‌بندی توأم: اگر داده‌های چند بیماری یا رویداد بهداشتی که الگوهای جغرافیایی مشترکی دارند در دسترس باشد مدل‌بندی توأم خطرات می‌تواند منجر به بهبود برآوردها نسبت به تحلیل جداگانه‌ی هر رویداد گردد. حتی در صورتی که یک رویداد به‌عنوان رویداد هدف مدنظر باشد، ورود داده‌های دیگر رویدادهای مرتبط در دسترس در تحلیل مفید است. در حالت وجود دو رویداد، مدل جزء مشترک ^{11}SC ، اندکی بهتر از مدل BYM چندمتغیره عمل می‌کند. این مسئله ناشی از آن است که مدل $MVBYM$ همبستگی بین دو رویداد را در همه نواحی ثابت در نظر می‌گیرد در حالی که مدل SC امکان تغییر و متفاوت بودن تابع مشترک در نواحی مختلف را فراهم می‌کند [۳۹]. در سالیان اخیر مطالعات فراوانی در خصوص مدل‌بندی توأم بیماری‌ها به‌عنوان یکی از مهم‌ترین مباحث در نقشه‌بندی بیماری‌ها انجام شده که منجر به معرفی و بسط چندین مدل در این خصوص شده است [۴۲، ۴۷، ۴۸، ۴۹، ۵۰].

۴ نتیجه‌گیری

مدل‌های متعددی برای نقشه‌بندی بیماری‌ها مطرح شده است که هر کدام دارای مبانی تئوری، مزایا و معایب و نقاط قوت و ضعف خاص خود بوده و در مواردی خاص نسبت به دیگر مدل‌ها ارجح‌اند. در این مقاله مجموعه‌ای از مهم‌ترین مدل‌های نقشه‌بندی بیماری‌ها تشریح و مورد مقایسه قرار گرفته و تفاوتها و شباهت‌های آنها مورد بررسی قرار گرفت اما از آنجا که در سالیان اخیر تلاش‌های فراوانی در جهت رفع نقاط ضعف مدل‌های موجود و گسترش و بسط بیش از پیش و روزافزون این مدل‌ها به‌عمل آمده است، لزوم مقایسه‌ی دامنه‌ی گسترده‌تری از مدل‌ها، هم از منظر تئوری و هم از منظر نتایج در یک مطالعه شبیه‌سازی جامع ضروری به‌نظر می‌رسد. ضمن آنکه در این مقاله تنها در خصوص نقشه‌بندی یک بیماری بحث گردیده و با توجه به اهمیت و کاربرد فراوان مدل‌های نقشه‌بندی توأم و چند متغیره‌ی بیماری‌ها، مطالعه و مقایسه‌ی این مدل‌ها نیز می‌تواند مفید حائز اهمیت باشد.

¹¹shared component model

مراجع

- [1] Lawson A.B, Browne W.J and Rodeiro C.L. (2003). Disease Mapping with WinBUGS and MLwiN. John Wiley&Sons.
- [2] Lawson A.B, Biggeri A.B, Boehning D, Lesaffre E, Viel J.F, Clark A, Schlattmann P, Divin F. (2000). Disease mapping models: an empirical evaluation, STATISTICS IN MEDICINE. 19:2217-2241.
- [3] Snow, J. (1854). On the Mode of Communication of Cholera (2nd edn). London: Churchill Livingstone.
- [4] Pickle L, Mungiole M, Jones G, and White A. (1999). Exploring spatial patterns of mortality: the new atlas of United States mortality. Statistics in Medicine 18, 3211-3220.
- [5] Cancer E-Atlas (2008) [<http://www.ncin.org.uk/eatlas/>].
- [6] Congdon P. (2003). Applied Bayesian Modelling. Clinchester: John Wiley&Sons, Ltd.
- [7] Lawson, A. B. (2001). Statistical Methods in Spatial Epidemiology. Chichester: John Wiley & Sons, Ltd.
- [8] Elliott P, Wakefield J. C, Best N. G, and Briggs D. J. (Eds) (2000). Spatial Epidemiology: Methods and Applications. London: Oxford University Press.
- [9] Cressie N, Chan NH.(1989). Spatial modelling of regional variables. Journal of the American Statistical Association; 84:393-401.
- [10] Clayton D, Kaldor J. (1987). Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. Biometrics; 43:671-681.
- [11] Marshall R. (1991). Mapping disease and mortality rates using empirical Bayes estimators. Applied Statistics; 40: 283-294.
- [12] Tsutakawa RK, Shoop GL, Marienfeld C. (1985). Empirical Bayes estimation of cancer mortality ranges. Statistics in Medicine; 4:201-212.
- [13] Mollie A, Richardson S. (1991). Empirical Bayes estimates of cancer

- mortality rates using spatial models. *Statistics in Medicine*; 10:95-112.
- [14] Lawson AB, Biggeri AB, Boehning D, Lesaffre E, Viel J-F, Bertollini R. (1999). *Disease Mapping and Risk Assessment for Public Health*. Wiley: New York.
- [15] Lawson AB, Cressie NAC. (2000). Spatial statistical methods in environmental epidemiology. In *Handbook of Statistics: Bioenvironmental and Public Health Statistics*, vol 18, 357-396
- [16] Simonof JS. (1996) *Smoothing Methods in Statistics*. Springer: New York.
- [17] Kelsall J, Diggle P. (1995). Non-parametric estimation of spatial variation in relative risk. *Statistics in Medicine*; 14:2335-2342.
- [18] Kelsall J, Diggle P. (1998). Spatial variation in risk of disease: a non-parametric binary regression approach. *Journal of the Royal Statistical Society, Series C*; 47(4):559-573.
- [19] Lee PM. (1997). *Bayesian Statistics: An Introduction*, 2nd ed. Arnold.
- [20] Wolpert R, Ickstadt K. (1998). Poisson/gamma random field models for spatial statistics. *Biometrika*; 85:251-267.
- [21] Carlin B, Louis T. (1996). *Bayes and Empirical Bayes Methods for Data Analysis*. Chapman and Hall.
- [22] Besag J, York J, Mollie A. (1991). Bayesian image restoration with two applications in spatial statistics. *Annals of the Institute of Statistical Mathematics*; 43:1-21.
- [23] Bernardinelli, L., D. G. Clayton, C. Montomoli. (1995b). Bayesian estimates of disease maps: how important are priors? *Statistics in Medicine* 14, 2411-2431.
- [24] Bohning D, Schlattmann P, Lindsay. (1992). CA.MAN*Computer Assisted Analysis of Mixtures: Statistical Algorithms. *Biometrics*; 48: 283-303.
- [25] Lawson, A. B. and A. Clark. (2002). Spatial mixture relative risk models applied to disease mapping. *Statistics in Medicine* 21, 359-70.
- [26] Goldstein, H. (1995). *Multilevel Statistical Models* (2nd ed). London:

- Edward Arnold.
- [27] Goldstein, H. (1986). Multilevel mixed linear model analysis using iterative generalized least squares. *Biometrika* 73, 43-56.
- [28] Goldstein, H. (1991). Nonlinear multilevel models with an application to binary response data. *Biometrika* 78, 45-51.
- [29] Goldstein, H. and J. Rasbash (1996). Improved approximations for multilevel models with binary response. *Journal of the Royal Statistical Society* 159, 505-513.
- [30] Rasbash J, Browne W, Goldstein W, Yang M, Plewis I, Healy M, Woodhouse G, Draper D, Langford I and Lewis T (2000). *A User's Guide to MLwiN*. London: Institute of Education.
- [31] Raudenbush, S, Yang M, and Yosef M (2000). Maximum likelihood for generalized linear models with nested random effects via high-order, multivariate Laplace approximation. *Journal of Computational and Graphical Statistics* 9, 141-157.
- [32] Rabe-Hesketh, S, Toulopoulou T, and Murray R (2001). Multilevel modeling of cognitive function in schizophrenic patients and their first degree relatives. *Multivariate Behavioural Research* 36, 279-298.
- [33] Richardson S, Thomson A, Best NG, Elliott P. (2004). Interpreting posterior relative risk estimates in disease mapping studies. *Environmental Health Perspectives*; 112: 1016-25.
- [34] Marshall RJ. (1991). A review of methods for the statistical analysis of spatial patterns of disease. *Journal of the Royal Statistical Society Series A (Statistics in Society)*; 154: 421-41.
- [35] Bithell JF. A. (2000). Classification of disease mapping methods. *Statistics in Medicine*; 19: 2203-15.
- [36] Richardson S. (2003). Spatial models in epidemiological applications. In Green PJ, Hjort NL, Richardson S eds. *Highly structured stochastic systems*. Oxford: Oxford University Press, 237-59.
- [37] Manda SOM, Leyland A. (2007). An empirical comparison of maximum likelihood and Bayesian estimation methods for multivariate

- spatial disease modelling. *South African Statistical Journal*, 41:1-21.
- [38] Bernardinelli, L. and Montomoli, C. (1992). Empirical bayes versus fully bayesian analysis of geographical variation in disease risk. *Statistics in Medicine*, 11: 983-1007.
- [39] Best, N, Richardson, S. and Thomson, A. (2005). A comparison of Bayesian spatial models for disease mapping. *Statistical Methods in Medical Research* 14(1), 35-59.
- [40] Best N, Waller L, Thomas A, Conlon E, Arnold R. (1998). Bayesian models for spatially correlated disease and exposure data. In *Bayesian Statistics 6*, Bernardo J et al. (eds). Oxford University Press: London.
- [41] Downing A, Forman D, Gilthorpe M.S, Edwards K.L, Manda S.O. (2008). Joint disease mapping using six cancers in the Yorkshire region of England. *International Journal of Health Geographic*. 7:41
- [42] Assunção RM, Castro MSM. (2004). Multiple cancer sites incidence rates estimation using a multivariate Bayesian model. *International Journal of Epidemiology*. 33:508-516.
- [43] Empirical and Full Bayes estimators for disease mapping V. Gomez-Rubio. A. Lopez-Quilez.
- [44] Devine, O.J. and Louis, T.A. (1994). A constrained empirical bayes estimator for incidence rates in areas with small populations. *Statistics in Medicine*, 13:1119-1133.
- [45] Aylin, P., Maheswaran, R., Wakefield, J., Cockings, S., Jarup, L., Arnold, R., Wheeler, G. and Elliot, P. (1999). A National Facility for Small Area Disease Mapping and Rapid Initial Assessment of Apparent Disease Clusters Around a Point Source: The U.K. Small Area Health Statistics Unit. *Journal of Public Health Medicine*, 21(3):289-298.
- [46] Inskip H. Standardization methods. (1998). In *Encyclopedia of Biostatistics*, Vol. VI, Armitage P, Corton T (eds). Wiley: New York. 4237-4250.
- [47] Knorr-Held L, Best N. (2001). A shared component model for detecting joint and selective clustering of two diseases. *Journal of the Royal*

- Statistical Society Series A (Statistics in Society); 164: 73-86.
- [48] Dabney, A. R. and Wakefield, J. C. (2005). Issues in the mapping of two diseases. *Statistical Methods in Medical Research* 14, 83-112.
- [49] Held L, Nat_rio I, Fenton S. E, Rue H, and Becker N. (2005). Towards joint disease mapping. *Statistical Methods in Medical Research* 14, 61-82.
- [50] Dreassi E. (2007). Polytomous Disease Mapping to Detect Uncommon Risk Factors for Related Diseases, *Biometrical Journal* 49, 4, 520-529
DOI: 10.1002/bimj.2006