

# 位置情報を用いた疫学研究とその統計的方法

高橋 邦彦<sup>1</sup>・和泉 志津恵<sup>2</sup>・竹内 文乃<sup>3</sup>

(受付 2013 年 9 月 20 日；改訂 12 月 10 日；採択 12 月 16 日)

## 要 旨

近年、事象の地理的変動の評価や健康リスクの地域間比較を行うために、収集されるデータに位置情報を付加した空間データの利用が活発になってきている。疫学分野においても、疾病の発生状況の地理的な格差・変動の記述や種々の要因の地理的変動を考慮した検討などを扱う空間疫学研究、空間データをモデルに組み込んだ健康リスクの推定の試み、それらに関連する統計手法が注目されるようになってきている。本稿では空間データを用いたいくつかの疫学研究の事例を紹介し、その中で用いられる方法と研究の進め方について概説する。

キーワード：空間データ、空間疫学、疾病地図、疾病集積性、災害、放射線。

## 1. はじめに

今日、我々のまわりには様々な数字・データがあふれている。そのなかには系統的に収集されているデータもあるが、その解析などは意図されず、無秩序に収集されてしまっているものもある。たとえば災害時などのデータは乱雑なものも多いが、その中から有用な情報を取り出すことが災害の状況把握やその対策に役立つと考えられる。また新たに集められるものだけではなく、必要に応じてそれまでに報告されている別のデータも併せて利用することで、より有効な情報が得られることもある。なかでも事象の地理的変動や健康リスクの地域間比較を行うために、収集されるデータに位置情報を付加した空間データの利用が近年活発になってきている(矢島, 2012)。またデータから得られた情報や数値は現場や一般市民に広く伝えることが重要であるが、日頃からそれらのデータに接していない者にとっては、その数字を見ただけでは全体を把握することは難しい。そのため地域での空間的な分布の観察や、広域にまたがる大規模災害での状況把握においては GIS (geographic information system: 地理空間情報システム) の利用も効果的であるが、そこでも位置情報が必要となってくる (Lai et al., 2009)。実際、平成 24 年 (2012 年) に改正された災害対策基本法の中でも、災害応急対策ための情報の収集及び伝達等において災害応急対策責任者は「地理空間情報の活用に努めなければならない」「災害に関する情報を共有し、相互に連携して災害応急対策の実施に努めなければならない」と記されるなど、現実社会においても空間情報の利用の重要性が認識されるようになってきている。人間集団における疾病の頻度分布と決定要因についての研究とも定義される疫学研究においても、近年これらの空間データを用いた研究が進んできており、特に疾病の発生状況の地理的な格差・変動を記述したり、種々の要因の地理的変動を考慮した検討、さらにその予兆の早期発見などを扱う空

<sup>1</sup> 名古屋大学大学院 医学系研究科生物統計学分野：〒466-8550 愛知県名古屋市昭和区鶴舞町 65

<sup>2</sup> 大分大学 工学部知能情報システム工学科：〒870-1192 大分県大分市旦野原 700

<sup>3</sup> 国立環境研究所 環境健康研究センター：〒305-8506 茨城県つくば市小野川 16-2

表 1. クロウン病における全国の 2003 年度医療受給者証所持者数, 死亡数・年齢調整死亡率, ならびに推計された 2002 年総患者数.

	医療受給者 証所持者数 (2003 年度)	死亡数 (2003 年) *	年齢調整死亡率 (/100 万人) (2003 年) *	総患者数 (2003 年) **	受診間隔 90 日までの 者も含めた総患者数 (2002 年) **
男性	15,436	31	0.39	11,600	16,400
女性	6,904	29	0.34	4,300	5,400

\* 土井・横山編 (2006). 難病の死亡統計データブック

\*\* 横山・土井編 (2008). 平成 14 年患者調査による難病の受療状況データブック

間疫学研究やそのための統計解析手法も注目されるようになってきている (Waller and Gotway, 2004; Schabenberger and Gotway, 2005; Lawson, 2006; 丹後 他, 2007; Pfeiffer et al., 2008). しかし実際に空間データを用いた疫学研究を行ったことがある研究者はまだまだ少数であろう. 特にこれまで空間データや位置データを扱ったことがない者にとっては, 例えば緯度・経度などの情報をどう扱うべきか, またどういうデータが必要になるのかなどがわからず, それが空間データを用いた疫学研究を実施するにあたっての高いハードルになっているのではないかと思われる.

そこで本稿では空間データを用いたいくつかの疫学研究の事例を紹介し, その中で用いられる方法と研究の進め方について概説を行うことで, 実際の疫学研究のヒントを与えたい.

## 2. 疫学研究の事例

本節では空間データを用いた実際の疫学研究の事例として, 特定疾患(難病)医療受給者証所持者数の地域比較, 大分県内における交通事故発生状況の地域集積性の検討, 心肺機能停止傷病者救急搬送件数の時間集積性の検討, ならびに東日本における放射性セシウム曝露量の推計に関する研究を紹介しながら, そこで用いられるデータ, 疾病集積性の検定やモデルを用いた推計などの解析手法, 結果とその解釈について概説を行う.

### 2.1 特定疾患(難病)医療受給者証所持者数の地域比較

疾病の発生頻度や患者数の地域比較を行うことは, その疾病の特徴を捉えるうえで大変重要なことである. 難病における死亡数の地域比較については, 人口動態統計による死亡個票などの行政資料を用いた検討が行われてきた. 一方で難病において正確な患者数のデータを把握することは困難であり, これまで全国疫学調査, 患者調査をはじめとする標本調査による推計, 特定疾患医療受給者証所持者数(以下, 所持者数という)などが患者数の目安として用いられ, 検討が行われてきている. 例えば 2003 年度クロウン病に注目してみると概況は表 1 のようになっていた.

所持者数は, 現在, 厚生労働省衛生行政報告例において, 全国(対象疾患別, 性別, 年齢階級別)および都道府県別(対象疾患別, 性別)に集計されたデータが公開されており, 難病の各研究班においても, 患者数に代わる一つの目安として利用されている. 一方, 平成 13 年(2001 年)~15 年(2003 年)度においては地域保健・老人保健事業報告において, 保健所別の統計表が公開されている. ここでは平成 15 年(2003 年)度全国のクロウン病における所持者数を例として, 空間疫学的手法を適用した地域比較について検討する(土井 他, 2012).

#### 2.1.1 疾病地図

今回用いた所持者数は表 2 のようなデータになっている. なお当該年度のデータでは全国 566

表 2. 2003 年度特定疾患医療受給者証所持者数, 保健所, 性(女)・対象疾病別.

クローン病・女性							
保健所	総数	0-9 歳	10-19 歳	20-29 歳	...	60-69 歳	70 歳以上
00 全国	6,904	14	320	1,787	...	551	383
01 北海道	425	1	29	142		18	23
0110 札幌市	159	0	9	53		14	6
0136 小樽市	7	0	0	4		0	0
0137 市立函館	26	0	0	9		1	2
0138 旭川市	37	0	2	12		1	3
0151 江別	11	0	2	2		0	0
⋮							

保健所管轄(ただし京都市は1つの保健所として報告)での集計がされていた。

観察データに基づく統計解析を行うにあたっては、まずはそのデータの様子を視覚的に観察することが重要である。一般的なデータでは、最初にヒストグラムや散布図などによってデータの分布の様子を観察すべきであるように、疾病の発生などの空間データにおいてもその分布の様子を把握することは、最も基本的かつ重要なことである。

しかし地域比較を行う場合、単純に観測数のみを考えれば、その数は当然人口の多い地域ほど多くなる傾向がある。また疾病によってはその地域住民の性・年齢構成が影響を与えることもあり、単純に観測数やそれを人口あたりの観測数、また死亡であれば死亡率などだけで比較を行うことは適切ではない。そこで年齢などの影響を取り除くため標準化されたりリスク指標を用いた疾病地図がよく利用される。そのためには、まず基準集団(例えば日本全国など)の年齢階級別の発生割合が必要となるが、日本においては国勢調査による人口などが利用できる。表2のデータを解析する場合には、簡単のため日本全国を基準集団として1番目の年齢階級0-9歳で考えると

$$(\text{日本全国 0-9 歳女性での観測数}) / (\text{日本全国 0-9 歳女性の人口})$$

で0-9歳女性の人口あたり所持者数  $P_1$  が計算できる。これを各年齢階級(表2の場合、8階級)で求める。次に対象となる地域の年齢階級人口データを同様に入手する。今、地域  $i$  での年齢階級  $j$  の人口を  $n_j^{(i)}$  としたとき、年齢階級の数を  $J$  とすれば

$$e_i = \sum_{j=1}^J n_j^{(i)} P_j$$

を地域  $i$  における期待頻度(期待観測数)という。これは「各年齢階級で基準集団と同じ割合で所持者数がいたとすると、この地域では何人の所持者が観測がされると期待されるか」という値になることを意味し、その地域住民の年齢構成を考慮した値となっている。

地域  $i$  の観測数を  $d_i$  としたとき、この期待頻度  $e_i$  を用いて

$$(2.1) \quad SMR_i = d_i / e_i$$

を SMR (Standardized Morbidity Ratio) という(観測数が死亡数の場合には Standardized Mortality Ratio として標準化死亡比と呼ばれる。ここでは観測数が所持者数のため Morbidity を用いた)。この SMR を用いた疾病地図は従来からよく用いられているが、それと同時に SMR の指標としての問題点も論じられている。一番の問題は、SMR はその地域の人口の大きさの影響

を受け、特に人口の少ない地域では SMR の値は不安定になり、人口の異なる市区町村間での地域比較などには適しているとはいえないことである。このことは、その意味する指標を統計学的推測の立場から眺めれば明らかである。いま、地域  $i$  の観測数を確率変数  $D_i$ 、その観測値を  $d_i$  とし、ある疾病の発生リスクを基準集団のリスクと比較することを考え、地域  $i$  の基準集団に対する相対リスク (relative risk, RR) を  $\theta_i$  とする。一般に死亡数をはじめ比較的稀な発生事象は Poisson 分布に従うと仮定されるので、地域  $i$  が基準集団と同じリスク、 $\theta_i = 1$ 、を持てば、地域  $i$  での観測数は基準集団のリスクから計算された期待頻度  $e_i$  を期待値とする Poisson 分布に従う。もし、地域  $i$  が基準集団よりリスクが大 ( $\theta_i > 1$ ) あるいは小 ( $\theta_i < 1$ ) であれば観測数の分布は期待頻度  $\theta_i e_i$  をもつ Poisson 分布

$$(2.2) \quad D_i \sim \text{Poisson}(\theta_i e_i) \quad (D_i \text{ と } D_{i'} (i \neq i') \text{ は独立})$$

と仮定できる。そのうえで、相対リスク  $(\theta_1, \dots, \theta_m)$  を未知の定数と考え、 $\theta_i$  の最尤推定量  $\hat{\theta}_i$  を求めたものが (2.1) 式となる。つまり疫学では相対リスクの最尤推定値を SMR と定義しているのである。このことから推定量  $\hat{\theta}_i$  の標準誤差は  $e_i$  に応じて変化することがわかり、市区町村単位など各地域の人口の大きさが大きく異なる場合にはその地域間比較には必ずしも適している指標とは言えなくなる。そのため相対リスク  $(\theta_1, \dots, \theta_m)$  をベイズ流に推定する方法がしばしば利用されている。代表的な Poisson-Gamma モデルでは

$$D_i \sim \text{Poisson}(\theta_i e_i), \quad \theta_i \sim \text{Gamma}(\alpha, \beta)$$

と考える。事前分布のパラメータ  $\alpha, \beta$  の推定には経験ベイズ法やフルベイズ法などが用いられる。また、さらに複雑なベイズモデルによるリスク指標の推定も行われている (Lawson, 2006, 2013)。

ここでは女性クローン病の所持者数について、日本全国を基準集団としたときの都道府県単位、保健所管轄単位での SMR の Poisson-Gamma モデルによる経験ベイズ推定値を求め、地図に示した (図 1)。また保健所管轄単位での解析において SMR のベイズ推定値が高い上位 10 保健所を表 3 に挙げた。なお保健所管轄の年齢階級別人口は、その管轄に含まれる市区町村の人口を合算したものをを用いた。表 3 のように単純に数値だけの表を見てもなかなか全体を把握できないが、地図に描くことで都道府県でも北海道ならびに九州地方が高い値をとっている様子が把握できる。さらに保健所管轄で細かく見ることで、例えば同じ北海道内でも地域差があることが観察できる。一方で、地図では面積の大きい地域が目立ってしまい、例えば東京周辺はかなり拡大しないと詳細がつかめない。またその階級の決め方や色分けの影響もあり、場合に

表 3. SMR (ベイズ推定値) の上位 10 保健所.

	観測数	期待頻度	SMR (ベイズ推定値) (%)
大分市保健所	45	24.72	148.06
旭川市保健所	37	19.78	146.21
福山市保健所	39	21.87	143.55
徳島保健所	43	24.73	143.27
八日市保健所	24	11.48	143.24
熊本市保健所	61	37.97	141.52
徳山環境保健所	26	13.29	141.30
伊集院保健所	15	5.74	139.87
北九州市保健所	83	54.69	139.22
出雲保健所	19	8.80	138.79

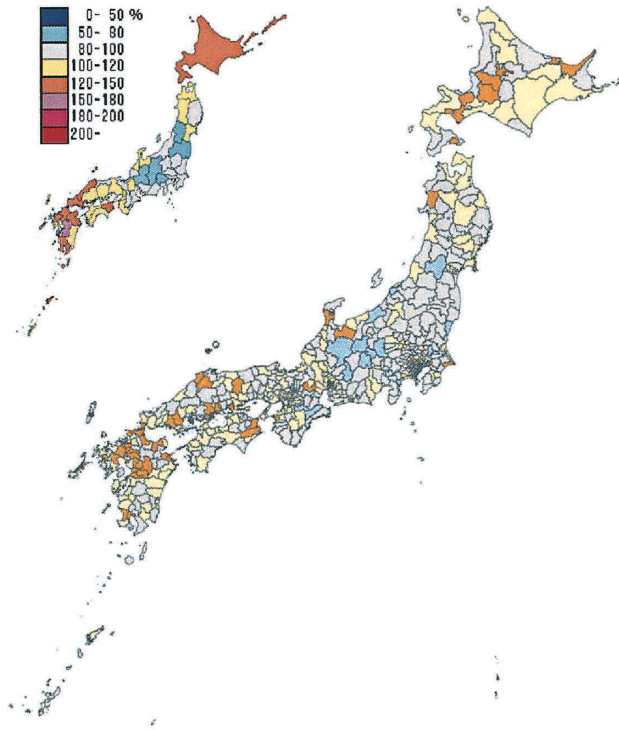


図 1. SMR (経験ベイズ推定値)の疾病地図：都道府県単位(左上)，保健所管轄単位(右)。

よっては誤った印象を与えてしまうこともあるため、結果をみる場合には注意深く観察や考察を行う必要がある。

### 2.1.2 考察

空間疫学の解析方法を用いた RR の推定とその視覚可である疾病地図によって所持者数の地域分布が異なることを見出すことができた。実際、他の疾患での検討も行ったところ、疾患毎に地域分布が異なる様子が観察できた。また、これらの疾病地図を観察した後、対象としている疾病のリスクの高い地域(もしくは低い地域)が、ある特定の地域に集中しているかどうかを検討する疾病集積性の検定を用いた解析を行うことができる(2.2 節参照)。実際、上記のデータに制限付尤度比に基づく flexible scan statistic (Tango and Takahashi, 2012) (以下、Flexible Scan 法という)を適用して解析を行ったところ、i) 久留米保健所を含む 29 保健所 ( $p = 0.0001$ )、ii) 札幌市保健所を含む 9 保健所 ( $p = 0.0004$ )、iii) 北九州市保健所を含む 7 保健所 ( $p = 0.0188$ )、の 3 つの有意な集積地域が検出された。

一方でここで用いたデータは所持者数であり、厳密な意味での患者数ではなく、それらの間には差異があることも考えられる。この点において、これらの結果の解釈の際には十分な注意が必要であろう。しかし近年、所持者数も年々増加しており、これらのデータを用いた地域比較は有用であると考えられる。今後、関連したデータを用いた、より高度な方法での解析なども検討しながら、これらのデータを有効に活用した解析を行い、その上で、実際の解析結果の解釈について臨床の専門家などと詳細な検討を行うことで、今後の難病研究推進に寄与できるものと考えられる。

## 2.2 大分県内における交通事故発生状況の地域集積性の検討

交通事故の原因解明や事故防止の対策を立てるうえで、交通事故の第1当事者(犯した過失が、ほかの当事者よりも相対的に重いと判断された当事者)の種別(自動車等, 自転車, 歩行者), 当事者の年齢や性別, 免許取得後の経過年数などの人的特徴とともに, 事故発生の日時, 地域, 天候, などの環境に関する特徴を明らかにすることは重要なことである. 全国の交通事故に関する情報は, 交通事故分析センター(ITARDA)により都道府県単位や政令指定都市単位のデータがいくつかの項目を掛け合わせた表形式で有償にて配布されている(イタルダ, 2013). 一方, いくつかの都道府県では市区町村単位でまとめられた交通統計も公開されている.

本節では特に大分県内における自動車等(原付以上, 以降自動車等と表記)による交通事故に注目し, 平成9年(1997年)~18年(2006年)に県内で発生した交通事故の中で, 自動車等乗車者の死亡数(交通統計では死者数)の地域比較を行う. データとしては和泉ら(2010)と同様に, 大分県内18市町村における調査期間10年間での事故発生市町村別の自動車等乗車者(全年齢)の死者数, 免許保有者数を用いる(表4). なお, ここでの死者数は事故発生後24時間以内での死亡のみがカウントされており, 免許保有者数(千人年)は小数点第1位において四捨五入されている. また地域 $i(i=1, 2, \dots, 18)$ の死者数を $d_i$ , 免許保有者数を $n_i$ とし, 同時期の県全体を基準集団として考え, 地域 $i$ における期待頻度 $e_i$ やRRを

$$e_i = n_i \times \frac{\sum_{i=1}^{18} d_i}{\sum_{i=1}^{18} n_i}, \quad RR_i = \frac{d_i}{e_i}$$

として計算し, 特に死者数が集中している地域があるか, 疾病集積性の検定による検討を行う.

### 2.2.1 疾病集積性の検定

疾病地図を観察すると, 対象としている疾病のリスクの高い地域(もしくは低い地域)がある特定の地域に集中しているのではないかと思われることがある. もしこの疾病の発生が集中

表4. 大分県内において発生した人身事故における自動車等(原付以上)乗車者の死者数, 免許保有者数, 免許保有者10万人対死者数(1997年~2006年).

	死者数	免許保持者数(千人年)	免許保持者数10万人対死亡者数
大分市	99	2,834	3.5
別府市	14	718	1.9
中津市	36	509	7.1
日田市	43	467	9.2
佐伯市	38	508	7.5
臼杵市	20	280	7.1
津久見市	3	131	2.3
竹田市	11	170	6.5
豊後高田市	18	154	11.7
杵築市	21	199	10.6
宇佐市	52	379	13.7
豊後大野市	31	266	11.7
由布市	25	214	11.7
国東市	25	208	12.0
姫島村	0	14	0.0
日出町	18	163	11.0
九重町	19	70	27.1
玖珠町	12	115	10.4

(集積)しているとするれば、その地域になんらかの原因があるかもしれないし、その疾病が流行性のものであるかもしれない。もし疾病の集積が観察された場合には、集積地域を中心に調査を行い、原因を特定したり対策を講じることが必要となるだろう。しかしこれらの疾病地図だけから「どこかに集積しているか?それとも全体的にばらついているか?」、さらに集積しているとしても「どの範囲までか?」を視覚的に判断するだけでは説得力に欠け、また実際その判断も難しい場合も少なくない。ここに疾病集積性の有無を客観的に決定する分析方法が必要となる。このようなとき「イベントが対象地域内のどこかに集積しているか?」を統計的に検定を行う方法として空間集積性の検定が適用できる。

実際に空間集積性の有無を検出する場合には、研究目的、データの種類(個人ごとの発生情報のような個人単位の点データ、市区町村ごとの発生情報のような地域単位の集計データ)に応じてそれぞれ異なった検定方法が提案されており、それらを使い分ける必要がある(Kulldorff, 2006; Rogerson and Yamada, 2009; Tango, 2010)。空間集積性の検定は、焦点をあてた検定(Focused Test)、一般的な検定(General Test)の2つに分類することがBesag and Newell(1991)によって提案された。焦点をあてた検定は、原子力発電所やごみ焼却施設等の健康に影響を及ぼすと考えられる地点(固定発生源)を中心として、その周辺に空間集積性があるか否かの検定を行う方法である。この方法を用いた代表的な研究事例として、「イタリアにおける高周波無線曝露と白血病疾患との関連性」(Michelozzi et al., 2002)、「日本のごみ焼却施設周辺における週産期健康障害のクラスター」(Tango et al., 2004)などが挙げられる。検定手法としては、Stone's test(Stone, 1988)、Besag-Newell's test(Besag and Newell, 1991)、スコア検定(Waller et al., 1992; Tango, 1995, 2002)などがある。

一方、一般的な検定は、固定発生源を特定せずに、研究対象領域の中のどこかに空間集積性があるか否かの検討を行う方法である。この一般的な検定は、さらにGlobal Clustering Test(GCT)とCluster Detection Test(CDT)の2つに分けられる。GCTは空間集積性の有無を一般的に検定する方法であり、検定方法としては、点データに対しては、Cuzick-Edward's test(Cuzick and Edward, 1990)、Tango's test(Tango, 2007)が適用可能であり、集計データに対しては、Besag-Newell's test(Besag and Newell, 1991)、Tango's index(Tango, 2000)が適用可能である。この方法を用いた研究事例として、「英国における小児白血病・悪性リンパ腫の空間集積性の検討」(Tango, 2007)などがある。GCTでは統計的に有意な空間集積性があると判断されたとしても、それがどの地域であるかは把握することができない。一方CDTは、空間集積性の有無の検定と同時に空間集積性が認められる地域の位置を検出する方法である。この方法を用いた代表的な研究事例として、「米国北東部の乳がん死亡のクラスターの有無」(Kulldorff et al., 1997)、「英国での新型クロイツフェルト・ヤコブ病のクラスターの有無」(Cousens et al., 2001)などがある。

このCDTにはスキャン統計量を用いた検定がいくつか提案されている。例えば市区町村単位のデータを考える場合、集積(クラスター)とは1つもしくは複数の市区町村が連結してできる地域と考える。スキャン統計量による検定では、クラスターの候補となる連結した地域のひとつひとつをwindowと呼び、「クラスターが存在する」ということは「観測数が期待観測数に比べ、有意に高くなるwindowが存在する」と考えることができる。逆に「クラスターが存在しない」ということは「全てのwindowについて、その観測数は期待観測数とほぼ同じである」ということになる。前節のSMRの議論と同様、観測数がPoisson分布に従うというPoissonモデルを考える。あるwindow  $Z$  を考えて、 $Z$  に含まれる地域内のリスクが $\theta_Z$  であり、また $Z$  の外側の地域ではリスクが $\theta_{Z^c}$  であるとする。つまり、全体が $m$  個の地域に分割されているとき、地域 $i(i = 1, 2, \dots, m)$  の観測数 $D_i$  が確率変数としてそれぞれ独立に

$$D_i \sim \text{Poisson}(\theta_Z e_i) \quad (i \in Z); \quad D_i \sim \text{Poisson}(\theta_{Z^c} e_i) \quad (i \notin Z)$$

であると考え、クラスタの有無は

$$\text{帰無仮説(クラスタ無し)} \quad H_0: \theta_Z = \theta_{Z^c} \quad (\text{for } \forall Z)$$

$$\text{対立仮説(クラスタ有り)} \quad H_1: \theta_Z > \theta_{Z^c} \quad (\text{for } \exists Z)$$

という仮説検定問題を考えることになる。このとき、ひとつひとつの window  $Z$  に対して検定を繰り返すと検定の多重性の問題が発生してしまう。そこで Kulldorff and Nagarwalla (1995), Kulldorff (1997) は尤度比に基づく統計量  $\lambda(Z)$  を考え、ある全体集合  $\mathcal{Z}$  のすべての window  $Z (\in \mathcal{Z})$  の中から  $\lambda(Z)$  の値が最大のもの (most likely cluster; MLC) を探し、そのときの  $Z$  をクラスタの候補とする spatial scan statistic を提案した。このときの尤度比は

$$(2.3) \quad \lambda(Z) = \begin{cases} \left( \frac{d(Z)}{e(Z)} \right)^{d(Z)} \left( \frac{d(Z^c)}{e(Z^c)} \right)^{d(Z^c)}, & d(Z) > e(Z) \\ 1, & \text{その他} \end{cases}$$

となり、MLC  $Z^*$  は  $Z^* = \arg \max_{Z \in \mathcal{Z}} \lambda(Z)$  となる。ここで  $d(Z)$  は window  $Z$  内全体での観測数、 $e(Z)$  は期待頻度とする。MLC として同定された  $Z^*$  の有意性の判定については、モンテカルロ検定 (Dwass, 1957) が利用される。即ち、各地域の観測事象数を帰無仮説のもとで乱数によって  $B$  (たとえば 999 や 9,999) 組発生させ、各組の乱数に基づく検定統計量  $\lambda_b = \lambda_b(Z^*)$  ( $b = 1, 2, \dots, B$ ) を計算する。一方、実際の観測データに基づく検定統計量を  $\lambda^* = \lambda^*(Z^*)$  するとき、 $p$  値の近似値

$$\hat{p} = \frac{1 + \sum_{b=1}^B I(\lambda_b \geq \lambda^*)}{B + 1} \quad I(\cdot): \text{定義関数}$$

によって有意性の判定を行う。このとき、用いるスキャン統計量や window を構成する際のパラメータによって、帰無仮説のもとでの検定統計量の分布が変化し、 $p$  値も異なってくることに注意する必要がある。

スキャン統計量ではクラスタを探し出す (scan する) window  $Z$  の全体集合  $\mathcal{Z}$  のとり方の違いによってもいくつかの統計量が提案されている。Kulldorff and Nagarwalla (1995), Kulldorff (1997) は、同心円状に、ある限界まで地域を追加していく circular window の全体をとった circular scan statistic (Circular Scan 法) を提案した。 $Z_{ik}$  ( $k = 1, 2, \dots, K_i$ ) を地域  $i$  から近い順に、 $i$  自身を含む  $k$  個の地域からなる集合とする。ただし各  $i$  の座標はその地域の代表点 1 点 (市区町村役場の所在地や人口重心など) であらわすものとする。このとき Circular Scan 法では  $\mathcal{Z}$  の全体集合として

$$(2.4) \quad \mathcal{Z}_1 = \{Z_{ik} \mid 1 \leq i \leq m, 1 \leq k \leq K_i\}$$

を考える。 $K_i$  としてはクラスタに含まれる最大距離や人口、最大地域数などが用いられる。この方法は簡便であるが、明らかに円状のクラスタしか同定できない。そこで非円状のクラスタも同定できるよう Circular Scan 法を拡張したいくつかの方法が提案されてきている。たとえば、SA Scan 法 (Duczmal and Assunção, 2004), Upper Level Set (ULS) Scan 法 (Patil and Taillie, 2004), Flexible Scan 法 (Tango and Takahashi, 2005), Echelon Scan 法 (栗原・石岡, 2007) などが提案されてきている。これらの方法は非円状の window も同定できるようにしながら、また計算時間が大きくなりすぎないように工夫されている。例えば Flexible Scan 法では、制限された範囲内で非円状のクラスタを同定する方法として、まず地域  $i$  を中心として  $i$  自身を含み  $i$  から近い順に  $K$  個の地域からなる集合  $Z_{iK}$  を定める。この  $Z_{iK}$  から、 $i$  を含み、連結している部



分集合を考え、その全体  $\mathcal{Z}_2$  を考える。つまり  $Z_{ik}$  の中で  $i$  を含んで  $k$  個の地域からなる連結した window が  $j_{ik}$  個あるとすると、 $Z$  の全体集合は

$$(2.5) \quad \mathcal{Z}_2 = \{Z_{ik(j)} \mid 1 \leq i \leq m, 1 \leq k \leq K, 1 \leq j \leq j_{ik}\}$$

とあらわされる。このため地域の隣接情報(地域が連結している場合 1 となりそうでないとき 0)から定義される隣接行列が必要となる。一方、ULS Scan 法や Echelon Scan 法は距離以外の尺度を用いる。ULS Scan 法では、Flexible Scan 法と同様の隣接行列と事象発生確率を用いて、集積地域の候補となる window を形成する。たとえば地域  $i$  における免許保有者数あたりの死者数を

$$P'_i = (\text{地域 } i \text{ における観測された死者数}) / (\text{地域 } i \text{ における免許保有者数})$$

とし、これを事象発生確率とおく。地域  $i$  を事象発生確率  $P'_i$  の降順に並べ替えたものを  $i'$  ( $1 \leq i' \leq m$ ) とおく。隣接行列の行、列を地域の事象発生確率  $P'_i$  の降順に並び替え、得られた行列の左上から  $l \times l$  ( $1 \leq l \leq L$ ) 行列を取り出す。ただし  $L$  は作成される window の最大サイズとする。このとき、この部分行列に含まれる地域の事象発生確率は  $P'_i$  以上となる。 $(l-1) \times (l-1)$  行列までで生成された window 全体を  $\mathcal{Z}_{3,l-1}$  とし、そこに含まれる長さ  $k$  ( $k = 1, 2, \dots, K_{l-1}$ ) の window を  $Z'_{kj}$  ( $j = 1, 2, \dots, j_{l-1,k}$ ) とする。 $i' = l$  としたとき、地域  $i'$  があるひとつの  $Z'_{kj}$  とだけ連結している場合には、 $Z'_{kj} \cup i'$  を新たな window とし、いずれの  $Z'_{kj}$  とも連結していない場合には  $i'$  自身を window に加える。ただし複数の  $Z'_{kj}$  と連結する場合には、 $i'$  と連結したもののなかで長さが最大となる window だけを候補に加える。また  $i'$  によって、複数  $Z'_{kj}$  が連結し 1 つの連結した地域となる場合には、これを新たな window とし、その部分集合は新たな window には加えないものとする。 $P'_i \geq P'_L$  までこれを繰り返して与えられたときの  $Z'_{kj}$  の全体は

$$(2.6) \quad \mathcal{Z}_3 = \mathcal{Z}_{3,L} = \{Z'_{kj} \mid 1 \leq k \leq K_L, 1 \leq j \leq j_{L,k}\}$$

とあらわされ、これを Upper Level Set という。なお Echelon Scan 法では、ULS Scan 法と同様に、地域の隣接情報と事象発生確率を用いて、エシェロンデンドログラムを描き、それを用いて、集積地域の候補となる window を形成する。

また尤度関数  $\lambda(Z)$  に制約を設けた制約付き尤度比検定 (Restricted Likelihood Ratio Test) も提案されている (Tango, 2008; Tango and Takahashi, 2012)。このときの制約付き尤度比は

$$(2.7) \quad \lambda'(Z) = \begin{cases} \lambda(Z) \cdot I(p_i < \alpha_1), & d(Z) > e(Z) \\ 1, & \text{その他} \end{cases}$$

とあらわされる。ここで  $p_i$  は帰無仮説  $\theta_i = 1$ 、対立仮説  $\theta_i \neq 1$  に対する片側  $p$  値であり、mid- $p$  値として以下の式で与えられる。

$$p_i = \Pr\{D_i \geq d_i + 1 \mid D_i \sim \text{Poisson}(e_i)\} + \frac{1}{2} \Pr\{D_i = d_i \mid D_i \sim \text{Poisson}(e_i)\}.$$

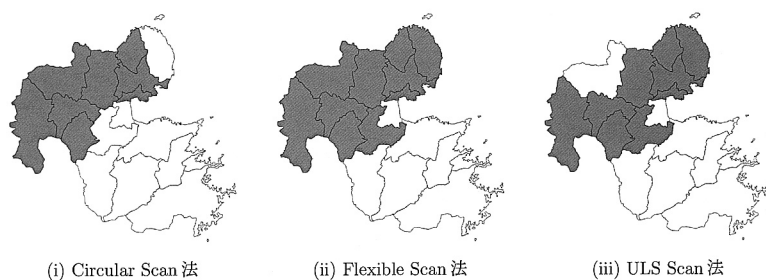
また  $\alpha_1$  は各地域に対しての有意水準として設定されるが、ひとつの目安として  $\alpha_1 = 0.20$  が推奨されている (Tango, 2008)。

### 2.2.2 解析結果

ここでは大分県内における交通事故での自動車等乗車者の死者数について、スキャン統計量を用いた CDT である Circular Scan 法、Flexible Scan 法、ULS Scan 法の 3 つの方法による疾病集積性の検定による解析結果を比較する。Circular Scan 法や Flexible Scan 法を適用する際には、FleXScan V3.1 (Takahashi et al., 2012) を利用し、ULS Scan 法を適用する際には統計ソ

表 5. 検出された MLC の比較.

$K$	Circular Scan 法			Flexible Scan 法			ULS Scan 法		
	18	15	12	18	15	12	18	15	12
大分市									
別府市									
中津市	x	x	x	x	x	x	x	x	
日田市	x	x	x	x	x	x	x	x	x
佐伯市				x			x	x	
臼杵市				x	x		x	x	
津久見市									
竹田市				x	x		x	x	
豊後高田市	x	x	x	x	x	x	x	x	x
杵築市	x	x	x	x	x	x	x	x	x
宇佐市	x	x	x	x	x	x	x	x	x
豊後大野市				x	x		x	x	
由布市				x	x	x	x	x	x
国東市				x	x	x	x	x	x
姫島村									
日出町	x	x	x	x	x	x	x	x	x
九重町	x	x	x	x	x	x	x	x	x
玖珠町	x	x	x	x	x	x	x	x	x
MLC における									
観測数	219	219	219	369	331	269	369	369	233
期待頻度	134.77	134.77	134.77	242.66	209.36	162.43	242.66	242.66	129.07
RR	1.6	1.6	1.6	1.5	1.6	1.7	1.5	1.5	1.8
$\log \lambda(Z^*)$	33.15	33.15	33.15	69.19	61.96	49.07	69.19	69.19	50.62
$p$ 値	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001	0.001

図 2. Circular Scan 法, Flexible Scan 法, ULS Scan 法 ( $K=12$ ) によって検出された MLC の比較 (大分県の自動車等事故による死亡).

フトウェアの R において作成したアプリケーションを利用した. 最大 window サイズ (Circular Scan 法の  $K_i$ , Flexible Scan 法の  $K$ , ULS Scan 法の  $L$ ) には 12, 15, および総地域数の 18 を選び, spatial scan statistic  $\lambda(Z)$  を統計量として用い, モンテカルロ検定のための繰り返し数を 999 回とし, 対数尤度比統計量が最も大きくなるものを MLC とした (表 5, 図 2). MLC のうち  $p < 0.05$  となるものを有意な集積地域とし, MLC 以外にも有意なクラスタがあった場合には secondary cluster としてそれらも同定した. 表 5 では, MLC に含まれる市町村に x 印を記し, MLC における観測数, 期待頻度, 相対リスク (RR), 対数尤度比統計量 ( $\log \lambda(Z^*)$ ),  $p$  値を示した.

解析の結果、最大 window サイズの値に係わらず、ULS Scan 法が最も大きな統計量をもつ MLC を検出していた。最大 window サイズの値を 18 としたとき、Flexible Scan 法による MLC は ULS Scan 法のものと同じとなった。一方、Circular Scan 法は、最大 window サイズの値に係わらず、同じ地域を含む MLC を検出し、その地域に含まれる市町村数は 8 と、他の 2 方法と比べて最も少なかった。

さらに、secondary cluster を探してみると、Circular Scan 法では、最大 window サイズの値に係わらず、{由布市、竹田市、豊後大野市} (検定統計量 6.62,  $p$  値=0.011, RR=1.6) であった。これらの地域は、最大 window サイズの値が 15 と 18 のときに Flexible Scan 法、ULS Scan 法によって検出された MLC に含まれる。一方、Flexible Scan 法、ULS Scan 法では secondary cluster は検出されなかった。

### 2.2.3 考察

CDT の 3 つの方法にて解析した結果、大分県における自動車等事故の発生に地域差があることが分かった。3 つの方法にて検出された MLC の共通地域には、北部の中津市 (RR=1.1)、豊後高田市 (RR=1.8)、宇佐市 (RR=2.1)、西部の日田市 (RR=1.4)、九重町 (RR=4.1)、玖珠町 (RR=1.6)、由布市 (RR=1.8)、中部の杵築市 (RR=1.6)、日出町 (RR=1.7) が含まれた。自動車等事故の発生が南部以外の地域に集中していたことが示唆される。ただし本研究における死者数の位置情報は、事故発生地点を表しているが、免許保持者数はその住所地をデータとして用いている。したがって本解析では、各地域の住民が自身が住む地域内で発生した事故で死亡すると仮定していることになる。

解析で検出された MLC は、3 つの方法間において異なり、最大 window サイズの値によっても異なることが示された。最大 window サイズの値として、解析者の経験則から、全地域数の  $2/3$  ( $K = 12$ ) から  $1$  ( $K = 18$ ) までの間の値を選択したが、標準的な値はこれまで特に示されていない。さらに、実際のデータ解析においては、真となるクラスタが不明であり、結果の解釈の際には十分な注意が必要であろう。

なお前述のとおり、本データでの死者数は事故発生後 24 時間以内での死亡のみがカウントされているが、負傷者の中には、交通事故による負傷が原因で 24 時間以後に死亡する者もいる。そこで死者数と同様に、大分県の自動車等事故による負傷者数の集積性の検定を行ったところ、Circular Scan 法は最大 window サイズの値に係わらず、{大分市、別府市、中津市、日田市、竹田市、由布市、日出町、九重町、玖珠町} が MLC (RR=1.09) として検出された。一方、Flexible Scan 法や ULS Scan 法では、最大 window サイズの値に係わらず、{大分市、別府市、中津市、杵築市、宇佐市、由布市、日出町、九重町} が MLC (RR=1.11) が検出された。これら負傷者数において検出された MLC は、死者数のものと類似していた。

近年、高齢者が関与する交通事故が全国的に増加しており、このような地域別の交通事故データや関連したデータを活用して、地域の特徴を明らかにすることが大切である。今後の交通事故の方策づくりにおいて、これらの地域比較の結果が役立つものと期待される。

## 2.3 心肺機能停止傷病者救急搬送件数の時間集積性の検討

近年、わが国では救急需要が増加し、2008 年の救急出場件数は約 510 万件で、1998 年からの 10 年間で約 38% 増加していた。その中で毎年約 10 万件が心肺機能停止傷病者の搬送となっている。日本では、平成 15 年 (2003 年) 「救急業務高度化推進検討会」によるウツイン様式導入の提言をもとに、平成 17 年 (2005 年) 1 月より全国の消防本部における全ての心肺機能停止傷病者のウツイン様式に基づくデータ収集が開始され、救急蘇生統計として公表されている。ウツイン様式の調査項目では、病院搬送前的心肺機能停止傷病者に対する救命活動による蘇生、

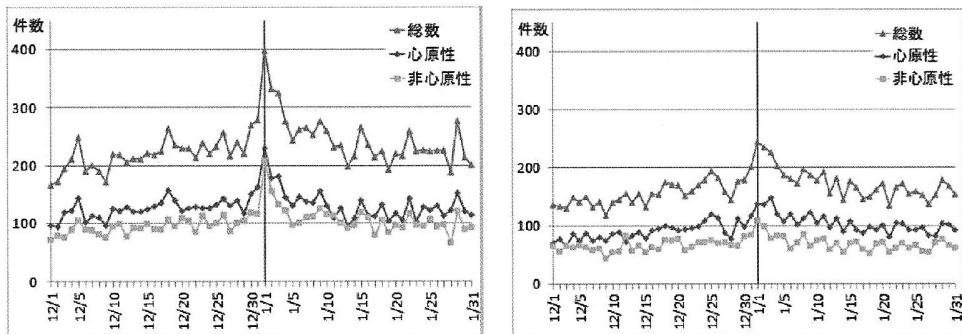


図 3. 2005 年 12 月～2006 年 1 月における日単位での心肺機能停止傷病者救急搬送件数。  
左：男性，右：女性。

1 ヶ月後の生存率や回復程度など，救命処置の成績に関する項目に重点が置かれている．その一方で，心肺機能停止傷病者が発生するリスク要因に関しては十分に検討されていないのが実情である．

一般に心肺機能停止は冬季に多いことが知られており，平成 21 年(2009 年)度第 2 回救急統計活用検討会の集計や新潟市における検討などでも明らかになっている．しかし具体的にどの時期にどの程度発生しているかなど，その発生時期についての詳細な分析や厳密な統計学的検討は十分に行われていない．発生日などの時間でも直線で見れば位置をもった空間データとして捉えることができ，これまでの平面上と同様の解析手法が適用できる．そこでここではウツイン様式に基づき集計され消防庁救急企画室から公開されている救急蘇生統計データを用い，特に心肺機能停止傷病者発生が多い冬季 12～1 月の 2 ヶ月間に注目し，その期間内において発生が集中している時期があるかどうかを検討する(高橋 他, 2012)．

### 2.3.1 対象データ

2005～2006 年の日本全国において救急搬送された心肺機能停止傷病者を対象として，匿名化・電子化されたうえで消防庁救急企画室から提供されているウツイン様式に基づく救急蘇生統計データを用いた．本研究では特にその発生の特徴をつかむことを目的とし，発生日，都道府県，性別，年齢，心原性・非心原性の別，および覚知の時刻の項目を用いた．データは年毎のファイルに分けられ提供されているが，入力されている発生日の年が収録されている年と異なっていたものについては，収録されているファイルの年に合わせた．また発生日が未入力のものについては解析から除外した．以上により対象解析例数は 2005 年が 102,738 件(うち発生日不明 2 件)，2006 年が 105,943 件(うち発生日不明 4 件)であった．その中で 2005 年 12 月～2006 年 1 月の 2 ヶ月間においては，男性 14,442 件，女性 10,216 件であった．そのうち心原性の件数は男性 8,177 件，女性 6,020 件，非心原性は男性 6,265 件，女性 4,196 件であった．この 2 ヶ月間における男女別日毎の発生状況は図 3 のようになった．

### 2.3.2 時間集積性の検定

上記 12～1 月の 2 ヶ月について日単位の発生件数を用いて，特定の期間に統計的に有意な時間集積性が確かめられるか，男女別の総数，ならびに心原性・非心原性別に集積性の検定を行った．ここでは第  $i$  日の搬送件数を確率変数  $D_i$ ，その観測値を  $d_i$  とし(2.2)式と同様の Poisson モデル

表 6. 2005 年 12 月～2006 年 1 月で検出された心肺機能停止傷病者救急搬送件数の有意な集積期間.

		全国 総件数	1日あたり 期待件数	検出された 集積期間	日数	期間内 件数	$p$ 値	RR
総数	男性	14,442	232.94	12/30-1/10	12	3,439	0.0001	1.23
	女性	10,216	164.77	12/30-1/11	13	2,587	0.0001	1.21
心原性	男性	8,177	131.89	12/30-1/10	12	1,930	0.0001	1.22
	女性	6,020	97.10	12/29-1/11	14	1,652	0.0001	1.22
非心原性	男性	6,265	101.05	1/1-1/3	3	495	0.0001	1.63
	女性	4,196	67.68	12/30-1/5	7	614	0.0001	1.30

$$D_i \sim \text{Poisson}(\theta_i e_i)$$

に従うと仮定した。このときこの2ヶ月内に集積期間がないことを帰無仮説とし、1日あたり期待件数は、当該2ヶ月間の搬送件数の総数を日数(62日)で割った数、すなわち  $e_i = \sum_{i=1}^{62} d_i/62$  ( $i = 1, 2, \dots, 62$ ) を用いることで、発生日という(時間的)位置データと、各日の発生件数だけで解析を行うことができることになる。検定には2.2節で述べた疾病集積性の検定法である Kulldorff (1997) のスキャン統計量を用いた。ただしここでは直線上の window  $Z$  を考えるだけになるので、circular scan statistic (Circular Scan 法) と flexible scan statistic (Flexible Scan 法) は同じになる。検定に必要なパラメータとして、検出される集積期間の最大期間(maximum temporal length)を15日間とし、モンテカルロ検定のための繰り返し数を9,999回、有意水準5%として検定を行った。最も  $p$  値が小さい期間を MLC とし、それ以外にも  $p < 0.05$  となる有意な集積期間があったものは、secondary cluster として  $p$  値の小さい順に挙げた。この解析には FlexScan Version 3.1 を用いた。

結果は表6のようになった。まず総数での結果をみると、男性で2005年12月30日～2006年1月10日の12日間が有意な集積期間( $p = 0.0001$ )として検出された。12～1月の2ヶ月間を基準とした集積期間内の発生の相対リスク(RR)は1.23であった。同様に女性については、2005年12月30日～2006年1月11日の13日間が有意な集積期間( $p = 0.0001$ , RR=1.21)として検出された。心原性・非心原性別の検定においても、いずれの場合でも有意な時間的集積が検出された。男女問わず、心原性、非心原性とも、それぞれ最も有意な集積期間として検出されたのは1月1日を含む年末年始時期であった。特に非心原性で検出された集積期間では、相対リスクが1.63(男性)など、心原性のみの場合の年末年始時期の集積期間の相対リスク1.22よりもさらに高いリスクをもつ期間が検出されていた。

### 2.3.3 考察

時間集積性の検定による解析を行ったところ、男女別、心原性・非心原性の別のいずれの場合でも1月1日を含む年末年始時期に最も有意な集積期間が検出された。ここでは詳細は割愛したが、2006年12月～2007年1月、2007年12月～2008年1月においても同様に1月1日を含む年末年始に有意な集積が検出された。またこの結果は人口の多い都道府県だけの影響ではなく、全国的に心肺機能停止傷病者の救急搬送件数がこの時期に増加していると認められた。さらに検出された集積期間内とそれ以外で搬送された傷病者の年齢構成を比較しても大きな違いが見られなかった。このことから特定の年代だけが年末年始に増加しているのではなく、どの年代でも増加していると考えられる。今回用いた時間集積性の検定による検討では、集積の原因を明らかにすることは困難であるが、今後の対策や原因を検討する上で有益な情報になると

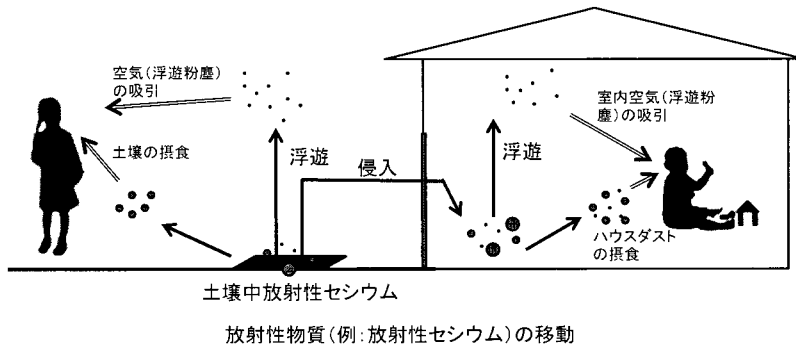


図 4. 放射性セシウムのヒト取り込み経路の想定。

考えられる。

#### 2.4 東日本における放射性セシウム曝露量の推計

2011年3月11日に発生した東日本大震災後の福島第一原子力発電所の事故により大気環境中に漏出した放射性核種は、主に福島県の一部地域で高い汚染を生じている。一方、同21日からの関東地方での降雨によって、茨城県南部から千葉県北西部にかけて放射性核種の湿性沈着が起こり、放射線空間線量が高い地域が点在している。特に千葉県柏市周辺では、2011年11月に局所的に45万Bq/kgの土壤が検出されるなど、人口の多い地域に高線量区域があり、住民の曝露評価及び除染方法の検討、除染効果の定量など課題が山積である。

ここでは、放射線関連の数多くの取り組みの中から、第4次航空機モニタリングによって測定された放射線の放射線量(空間データ)から、居住地域ごとの事故後1年間での内部曝露量および外部曝露量の推計を試みた研究を取り上げる(環境省, 2012)。居住地域から放射線被ばく量を推計するといっても、同一地域に居住していても、個人によって居住環境や生活習慣、行動様式には大きな違いがある。しかし、個人ごとの曝露量を測定するのは実質的に不可能であるため、ここでは地域ごとに放射線量を測った空間データとサンプリング調査による実測データに基づき、モンテカルロシミュレーションの手法を用いて確率的な誤差を与えた放射線曝露量の評価を行った。曝露が住環境や生活習慣と関連する場合、疫学研究のための空間的な曝露評価が重要になる。曝露量の推計は、年齢ごとの放射線被ばくに対する感受性を考慮し、0歳、1-6歳、7-15歳、16歳以上の4群に分けて実施することとした。また、推計を行うエリアは、福島県以外は都道府県単位で、福島県に関しては市町村単位とすることとした。

ここでは、飲食物、土壤由来の粉じんの吸入、屋外大気、屋内空気、ハウスダストの吸入からの曝露を内部曝露(図4)、屋内および屋外の空間線量から推計される曝露を外部曝露とする。また、半減期が短く、曝露が事故後初期に限定されたヨウ素による影響は考慮せず、放射性セシウムによる曝露のみを評価対象とした。また、高線量地域である千葉県柏市及びその周辺地域で2012年度に実施された住民調査では、少数の個人を対象として内部曝露・外部曝露が詳細に測定されたが、その中で食事由来の内部被ばく量(7名を対象に陰膳法で月1~2回程度継続的に実施)は非常に寄与が小さいことが示唆された(国立環境研究所, 2013)。また、水道局の調査により水道に含まれる放射性物質は事故後ほどなくして検出限界以下となっているため、ここでは飲食物および飲料水由来の放射線曝露は評価の対象から外すこととした。

### 2.4.1 対象データ

ここでは、空間データであるメッシュ単位で測定された GIS データと、個人の属性(年齢階級や在住都道府県)ごとに規定されるそれ以外のデータを利用する。空間データも研究利用申請によって利用が可能であり、それ以外のデータは公表データとなっている。

- 外部空間線量評価のためのデータ

事故後、文部科学省が定期的に航空機を用いて放射性核種の土壌沈着量を評価する取り組みを実施している。ここでは、東日本を対象に 2011 年 11 月 5 日に実施された第 4 次航空機モニタリングデータを利用する。当該データは、核種ごとに 25m メッシュの GIS データとして利用することができる。

- 屋外滞在時間割合

年齢階級ごとの屋外滞在時間に関しては、公表されている調査結果を利用する (United States Environmental Protection Agency (USEPA), 2011b)。その結果、0 歳 (1 歳未満) の屋外滞在時間の 5% 点 が 0.05 時間/日、50% 点 が 0.70 時間/日、1-6 歳の 5% 点 が 0.08 時間/日、50% 点 が 2.10 時間/日、7-15 歳の 5% 点 が 0.01 時間/日、50% 点 が 3.00 時間/日、16 歳以上の 5% 点 が 0.02 時間/日、50% 点 が 4.3 時間/日となっている。

- 建物種類比

地方によって、木造および非木造の割合が異なることから、公表されている都道府県ごとの木造・非木造の構成比を利用した (総務省統計局, 2008)。土壌沈着した放射性物質の影響は、土壌(地上)からの距離によって減衰するため、特に人口密集地であれば居住階数も考慮すべきであるが、現段階では利用可能なデータがなく、検討をしていない。

- 遮蔽係数

木造および非木造の建物がどの程度放射線を遮る効果があるかは、遮蔽係数として定量、公表されている (International Atomic Energy Agency (IAEA), 2000)。外部放射線量は木造住宅では 40% (0.4)、非木造住宅では 20% (0.2) にまで遮蔽される。

### 2.4.2 放射線曝露量のモデル推計

#### (1) 外部被ばく

外部被ばくによる被ばく線量は、 $L$  を都道府県または市町村 (福島県のみ)、 $Y$  を年齢カテゴリー、 $i$  を放射線核種 (Cs134 または Cs137) として、次の式で推計することとした。

$$(2.8) \quad E_{\text{ext}}(L, Y, i) = C(L, i) \cdot \Phi(i) \cdot [\tau(Y) + (1 - \tau(Y)) \cdot b \cdot B]$$

ここで  $C$  は積算沈着量であり、

$$C(L, i) = \int_0^t C_t(L, i) \cdot e^{\lambda^{(i)} \cdot t} dt + \int_t^1 C_t(L, i) \cdot e^{-\lambda^{(i)} \cdot t} dt$$

とする。ここでは 2011 年 3 月 12 日から 2012 年 3 月 12 日までの 1 年間の被ばく量の推計を目的としているため、積分区間(年)は 0 から 1 とする。 $t$  は、文部科学省による第 4 次航空機モニタリングデータ採取年月日である 2011 年 11 月 5 日が積分区間における 239 日目 (1 年間の 65.3% 時点) に相当することから 0.653 となる。 $C_t$  は、データ採取日の各地点の放射性セシウムの土壌沈着量 ( $\text{Bq m}^{-2}$ ) である。このデータ採取日の土壌沈着量を基準に崩壊定数  $\lambda$  (Cs134 で 0.3460, Cs137 で 0.0231) を含めた放射線変の方程式を用いて、放射性セシウムの物理的半減期による減衰のみを時間に伴う変動として考慮して 1 年分に相当する区間での定積分で 2011 年 3 月 12 日から 2012 年 3 月 12 日までの 1 年間の積算沈着量を計算した。 $\tau$  は年齢階級ごとの屋外滞在時間割合、 $\Phi$  は線量換算係数 (Cs134 で  $4.7 \times 10^{-5} \text{ mSv}/(\text{year} \cdot \text{Bq m}^{-2})$ 、Cs137 で

$1.8 \times 10^{-5} \text{ mSv}/(\text{year} \cdot \text{Bq m}^{-2})$ )で(IAEA, 2000),  $B$ は都道府県ごとの建物種類比,  $b$ は建物の種類ごとの遮蔽係数とする. 屋外滞在時間割合  $\tau$ については, 年齢階級ごとの屋外滞在時間の5%点と50%点が調査結果(USEPA, 2011b)と一致するような対数正規分布からのモンテカルロサンプリングによって確率的変動を与え, 各市町村・都道府県の第4次航空機モニタリングデータ採取地点分だけ乱数を発生させ, 各地点の積算沈着量にかけ合わせた.

#### (2) 内部被ばく(飲食物以外)

ここでは, 地表面に吸着した放射性セシウムがヒト体内に侵入する経路としては図4のように土壌やハウスダストの経口摂取や屋内外の浮遊微粒粉じんの吸入摂取を想定した. これまでも被ばく線量の推定は様々な機関で試みられているが, ハウスダストや室内空気に関して確率的な推計を行った研究はほとんどない. 以下に経路別内部被ばく線量推計のための式を示す.

##### ● 土壌( $s$ )経由

$$(2.9) \quad E_{\text{int},s}(L, Y, i) = \frac{1}{\rho} M_2 \cdot IR_s(Y) \cdot K(i) \cdot 10^6 \cdot C(L, i)$$

##### ● ハウスダスト( $d$ )経由

$$(2.10) \quad E_{\text{int},d}(L, Y, i) = M_2 \cdot IR_d(Y) \cdot K(i) \cdot A \cdot 10^6 \cdot C(L, i)$$

##### ● 室内空気経由(ina)

$$(2.11) \quad E_{\text{int},\text{ina}}(L, Y, i) = M_1 \cdot IR_{\text{ina}}(Y) \cdot K(i) [1 - \tau(Y)] \cdot A \cdot 10^6 \cdot C(L, i)$$

##### ● 屋外大気経由(outa)

$$(2.12) \quad E_{\text{int},\text{outa}}(L, Y, i) = \frac{1}{\rho} M_1 \cdot C_{\text{outa}} \cdot IR_{\text{outa}}(Y) \cdot K(i) \cdot \tau(Y) \cdot 10^6 \cdot C(L, i)$$

ここで, 係数および係数を定めるための事前情報について示す.  $\tau$ は屋外滞在時間割合を表し, 前述外部被ばくと同じ情報を用いた.  $M_2$ は濃縮係数であり, 国立環境研究所が千葉・茨城地域で実施した実測データに基づいて土壌で1.1, ハウスダストで1.5と設定した. また,  $\rho$ は土壌密度であり一律に  $65 \text{ kg m}^{-2}$ ,  $M_1$ は空気中浮遊微粒子であり一律に2とし(環境省, 2012),  $K(i)$ は核種  $i$ の経口摂取による実効線量係数で, Cs134で  $1.9 \times 10^{-8} \text{ (Sv Bq}^{-1})$ , Cs137で  $1.3 \times 10^{-8} \text{ (Sv Bq}^{-1})$ とした(The International Commission on Radiological Protection (ICRP), 1996). さらに,  $C_{\text{outa}}$ は屋外大気中の粉じん濃度であり, 環境省大気汚染物質広域監視システム「そらまめ君」による観測データから  $0.10 \times 10^{-6} \text{ (kg m}^{-3})$ とした. ここまでの係数はいずれも定数として確率的変動は考えなかった. 一方,  $A$ は屋外の物質が屋内に入り込む割合を表す侵入係数で, 前述と同様の国立環境研究所による実測データに基づいて中央値は  $0.03 \text{ (Bq kg}^{-1}) \text{ (Bq m}^{-2})^{-1}$ , 95%値は  $0.18 \text{ (Bq kg}^{-1}) \text{ (Bq m}^{-2})^{-1}$ となる対数正規分布をとるものとした. また,  $IR_j$ は  $j$ の摂取量 ( $\text{kg year}^{-1}$ ) および呼吸率 ( $\text{m}^3 \text{ year}^{-1}$ ) で,  $j$ は土壌( $s$ ), ハウスダスト( $d$ ), 室内空気(ina), 屋外大気(outa)の4種であり, 米国環境省の調査結果に基づいて中央値と95%点が表7に一致するような対数正規分布をとるものとした(USEPA, 2011a). 係数  $\tau$ ,  $A$ ,  $IR_j$ については, 上記事前情報で得られたパーセント点や分散に関する情報から正規分布または対数正規分布を利用して, モンテカルロサンプリングによって確率的変動を与えるものとした.

### 2.4.3 結果

都道府県ごとまたは市町村ごとの空間線量データから各種パラメータを用いて推定した年間被ばく量の推計値を階級別に色分けして図5に示す. その結果, 福島県, 群馬県, 栃木県を除く東日本の大部分が  $0.2 \text{ mSv year}^{-1}$  以下という結果になった. また, 比較的高線量な区域が点



表 7. 米国環境省調査による年齢階級ごとの各物質の体内取り込み係数. i) 土壌(年摂取量), ii) ハウスダスト(年摂取量), iii) 大気(年呼吸量).

i) 土壌			ii) ハウスダスト			iii) 大気		
$IR_s$	kg year <sup>-1</sup>		$IR_d$	kg year <sup>-1</sup>		$IR_{ina}, IR_{outa}$	m <sup>3</sup> year <sup>-1</sup>	
年齢	平均	95%点	年齢	平均	95%点	年齢	平均	95%点
0-1 歳	0.011	0.073	0-1 歳	0.011	0.037	0-1 歳	1,971	3,358
1-6 歳	0.018	0.073	1-6 歳	0.022	0.037	1-6 歳	3,687	5,037
7-15 歳	0.018	0.073	7-15 歳	0.018	0.037	7-15 歳	5,548	7,300
16 歳以上	0.007	0.073	16 歳以上	0.011	0.037	16 歳以上	5,950	9,125

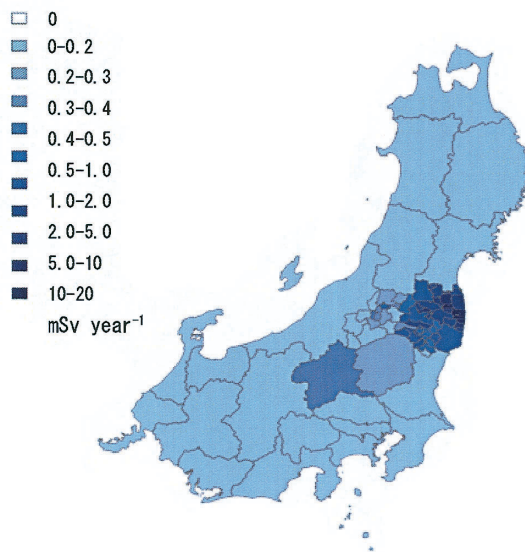


図 5. モデル推計された外部被ばく線量の 50%点(1-6 歳).

在するとされる柏市を抽出し、福島市と比較した。ここでは 1-6 歳の年齢階級に限定して福島市、柏市を図 6 に示す。

なお、今回のような 1 年分の被ばく量の推計では、0-1 歳の年齢階級については生まれ月にこの階級に含まれる期間が大きく変わること、7-15 歳は小中学生、16 歳以上は高校生以上の幅広い年齢を含んだ結果であるため、今回は 1-6 歳の年齢階級の結果のみを示したが、年齢階級による推定の傾向の違いは見られなかった。

#### 2.4.4 考察

今回は、空間データとモンテカルロサンプリングの手法を用いて、東日本での年間被ばく量の推計値の算出を試みた。福島県内では、平成 23 年(2011 年)度にガラスバッジ等を用いた外部被ばく線量の実測が行われていたため、表 8 のように今回の推計結果と実測データとの比較検証することができた。今回は、実測データが公表されていたいくつかの市町村について、実測値とモデル推定値を比較した結果、実測値は平均値しか公表されていない場合が多いが、曝露量が対数正規分布であることを考慮すると、平均値は中央値より高いと考えられる。モデル推定値は C 市では同程度で、C 市以外はモデル推定値が高めの値であったが、すべて数倍程度

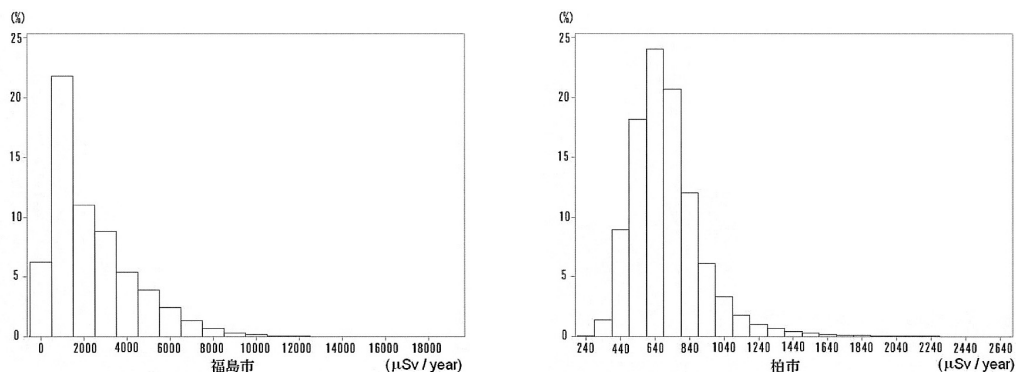


図 6. モデル推計された外部被ばく線量の分布(1-6 歳). 左図: 福島市, 右図: 柏市.

表 8. 福島県市町村のモデル推定値と住民の実測値との比較( $\text{mSv year}^{-1}$ ).

市町村	モデル推定値 (1-6 歳)		実測値の範囲	
	50%点	95%点	平均値	95%点
A 市	1.7	6.3	0.4-1	4-6
B 市	1.2	4.3	0.44	
C 市	1.6	4.4	1.3	
D 市	2.1	8.7	0.6-0.8	1.4-1.6
E 市	4.8	8.0	1.6	

の差であった。外部被ばく線量に関してはモデル推定値は実測値とよく一致しているといえる。

本節で述べた結果の限界としては、沈着量が少ない地域における第4次航空機モニタリングデータには航空機測定による検出限界があるため、値の確からしさを今後さらに検証する必要がある。本節では、アウトカムとの関連の疫学的評価には至っていないが、原発事故直後にしか測定することのできなかつた空間データを用いて今後の疫学研究の土台となるべき曝露評価を実施した。なお、本節で示した結果は、国立環境研究所による報告書「東日本大震災後の災害環境研究の成果」の一部である(国立環境研究所, 2013)。

東日本大震災および原発事故前の時期にあたる2011年1月から全国15地域を対象とした大規模出生コホート研究「エコチル調査」が始まっており、今回の曝露評価を行った東日本地域の中でも福島県全域、宮城県、神奈川県、千葉県、山梨県、長野県、富山県の一部地域がリクルート対象地域となっている。エコチル調査対象者では妊娠・出産の経過が調べられ、今後の追跡調査では子どもの発達発育、各種疾患発症情報が集められていく予定である。その際には、空間疫学的に本事例での曝露評価と疾患発症の関連が突合されることが期待される。

### 3. おわりに

本稿では、収集されるデータに位置情報が付加される空間データを用いた疫学研究について、いくつかの研究事例を通して、そこで用いられるデータ、解析手法、結果の解釈等について詳細を行った。空間データとしては、1点1点の緯度・経度を指定するものや、メッシュデータのように面積をもったデータを取り扱うものもある。空間データを用いた疫学研究の代表的なものとしては、2.1, 2.2節のような市区町村単位やその結合した地域での地域比較などがある。

これは、そもそも位置データ以外の用いられるデータが、市区町村単位など行政単位で集計されていることから、それに合わせて位置データも市区町村をそれぞれ1点で代表させて扱っている。一方でこのような行政区分にとらわれず、時間的に不変な空間単位での観測や解析のため、2.4節のようにメッシュデータも利用されるようになってきている。日本においても、国勢調査や気候などの各種統計が規則正しい格子状に定められた標準メッシュ体系で集計され、公開されるようになってきている。福島第一原子力発電所事故での放射線モニタリングなどにおいては、モニタリングポストのような点データとともにメッシュデータも利用されている。

疫学研究では地域に重点が置かれることが多いが、ここで紹介した統計手法の応用範囲は幅広く、同様に「位置」が定まるものであればさまざまな応用が可能である。例えば、ゲノム研究における遺伝子座を記述する HapMap のような遺伝子地図などの位置データを用いれば、疾患との関連遺伝子の検出なども可能である (Nishiyama et al., 2011)。また、2.3 節での救急搬送件数の時間集積性検出のように、時間データだけでも1次元空間データとして平面上の2次元データと同様な解析手法を適用することもできるが、一方で平面上の位置に時間も付加したものを3次元データとして扱うこともある。実際、感染症の発生やバイオテロによる異常を早期検出するための症候サーベイランスの解析においては、「いつ」「どこで」異常が発生したかを検出することが重要であり、そのような場合には時間-空間集積性の検出として3次元空間での集積性の検定も利用できる (Lawson and Kleinman, 2005; 丹後 他, 2007; Takahashi et al., 2008; Robertson et al., 2010)。

本稿では、著者らが取り組んだいくつかの疫学研究を取り上げ、そこで用いられた統計的方法の詳解を行った。最近ではこのような空間データを用いた疫学研究も活発となり、それとともに新たな解析手法も提案されてきている。今後、より複雑な構造をもつ空間データの利用も可能になると考えられ、それを解析する統計的方法の更なる発展とともに、疫学分野における統計家の貢献が期待される。

## 参 考 文 献

- Besag, J. E. and Newell, J. (1991). The detection of cluster in rare disease, *Journal of the Royal Statistical Society, Series A*, **154**, 143–155.
- Cousens, S., Smith, P. G., Ward, H., Everington, D., Knight, P. S. G., Zeidler, M., Stewart, G., Smith-Bathgate, E. A. B., Macleod, M. A., Mackenzie, J. and Will, R. G. (2001). Geographical distribution of variant Creutzfeldt-Jakob disease in Great Britain, 1994–2000, *Lancet*, **357**, 1002–1007.
- Cuzick, J. and Edwards, R. (1990). Spatial clustering for inhomogeneous populations (with discuss), *Journal of the Royal Statistical Society, Series B*, **52**, 73–104.
- 土井由利子, 横山徹爾, 高橋邦彦, 金谷泰宏 (2012). 難病患者の頻度の推計方法及び疫学的特徴の把握方法の開発, 厚生労働科学研究費補助金(難治性疾患克服研究事業)「今後の難病対策のあり方に関する研究」(研究代表者 林 謙治), 平成 23 年度総括・分担研究報告書, 31–46.
- Duczmal, L. and Assunção, R. (2004). A simulated annealing strategy for the detection of arbitrarily shaped spatial clusters, *Computational Statistics & Data Analysis*, **45**, 269–286.
- Dwass, M. (1957). Modified randomization test for nonparametric hypotheses, *Annals of Mathematical Statistics*, **28**, 181–187.
- International Atomic Energy Agency (2000). *Generic Procedures for Assessment and Response during a Radiological Emergency*(IAEA-TECDOC-1162), International Atomic Energy Agency (IAEA), Vienna, Austria.
- イタルダ (2013). 平成 24 年全国市区町村別交通事故死者数, ITARDA INFORMATION 交通事故分析レ

- ポート, **101**, 1–12, 公益財団法人交通事故総合分析センター, 東京, <http://www.itarda.or.jp/>.
- 和泉志津恵, 大嶋美登子, 小畑経史, 西村靖史, 原 恭彦, 矢島潤平(2010). 自動車等(原付以上)の交通事故の特徴, 大分県交通安全協会受託研究「交通事故防止に向けた統計的データ分析—高齢者事故の背景要因の探索—」(研究代表者 和泉志津恵), 平成 20・21 年度研究報告書, 23–43, 大分県交通安全問題研究会, 大分.
- 環境省(2012). 平成 24 年度災害・放射能と環境に関する委託研究報告書, 環境省, 東京.
- 国立環境研究所(2013). 東日本大震災後の災害環境研究の成果, 独立行政法人国立環境研究所, 茨城.
- Kulldorff, M. (1997). A spatial scan statistic, *Communications in Statistics: Theory and Methods*, **26**, 1481–1496.
- Kulldorff, M. (2006). Tests of spatial randomness adjusted for an inhomogeneity: A general framework, *Journal of the American Statistical Association*, **101**(475), 1289–1305.
- Kulldorff, M. and Nagarwalla, N. (1995). Spatial disease clusters: Detection and inference, *Statistics in Medicine*, **14**, 799–810.
- 栗原考次, 石岡文生(2007). 空間データの階層構造による分類とその応用, *日本統計学会誌*, **37**(1), 113–132.
- Lai, P. C., So, F. M. and Chan, K. W. (2009). *Spatial Epidemiological Approaches in Disease Mapping and Analysis*, CRC Press, Boca Raton.
- Lawson, A. B. (2006). *Statistical Methods in Spatial Epidemiology*, 2nd ed., John Wiley & Sons, Chichester.
- Lawson, A. B. (2013). *Bayesian Disease Mapping: Hierarchical Modeling in Spatial Epidemiology*, 2nd ed., Chapman & Hall/CRC, Boca Raton.
- Lawson, A. B. and Kleinman, K. (eds.) (2005). *Spatial and Syndromic Surveillance for Public Health*, John Wiley & Sons, Chichester.
- Michelozzi, P., Capon, A., Kirchmayer, U., Forastiere, F., Biggeri, A., Barca, A. and Perucci, C. A. (2002). Adult and childhood leukemia near a high-power radio station in Rome, Italy, *American Journal of Epidemiology*, **155**, 1096–1103.
- Nishiyama, T., Takahashi, K., Tango, T., Pinto, D., Scherer, S. W., Takami, S. and Kishino, H. (2011). A scan statistic to extract causal gene clusters from case-control genome-wide rare CNV data, *BMC Bioinformatics*, **12**, 205.
- Patil, G. P. and Taillie, C. (2004). Upper level set scan statistic for detecting arbitrarily shaped hotspots, *Environmental and Ecological Statistics*, **11**, 183–197.
- Pfeiffer, D., Robinson, T., Stevenson, M., Stevens, K., Rogers, D. and Clements, A. (2008). *Spatial Analysis in Epidemiology*, Oxford University Press, New York.
- Robertson, C., Nelson, T. A., MacNab, Y. C. and Lawson, A. B. (2010). Review of methods for space-time disease surveillance, *Spatial and Spatio-temporal Epidemiology*, **1**, 105–116.
- Rogerson, P. and Yamada, I. (2009). *Statistical Detection and Surveillance of Geographic Clusters*, Chapman & Hall/CRC, Boca Raton.
- Schabenberger, O. and Gotway, C. A. (2005). *Statistical Methods for Spatial Data Analysis*, Chapman & Hall/CRC, Boca Raton.
- 総務省統計局(2008). 平成 20 年住宅・土地統計調査, 総務省統計局, 東京.
- Stone, R. A. (1988). Investigations of excess environmental risks around putative sources: Statistical problems and a proposed test, *Statistics in Medicine*, **7**, 649–660.
- Takahashi, K., Kulldorff, M., Tango, T. and Yih, K. (2008). A flexibly shaped space-time scan statistic for disease outbreak detection and monitoring, *International Journal of Health Geographics*, **7**, 14.
- 高橋邦彦, 武村真治, 長谷川 学, 金谷泰宏, 齋藤大蔵, 阪本敏久(2012). わが国における救急蘇生統計を用いた冬季の心肺機能停止傷病者救急搬送件数の時間的集積性の検出, *日本臨床救急医学会雑誌*, **15**(5), 652–661.

- Takahashi, K., Yokoyama, T. and Tango, T. (2012). *FlexScan: Software for the Flexible Scan Statistics*, Version 3.1, National Institute of Public Health, Saitama, Japan, <http://www.niph.go.jp/soshiki/gijutsu/download/flexscan/>.
- Tango, T. (1995). A class of tests for detecting “general” and “focused” clustering of rare diseases, *Statistics in Medicine*, **14**, 2323–2334.
- Tango, T. (2000). A test for spatial disease clustering adjusted for multiple testing, *Statistics in Medicine*, **19**, 191–204.
- Tango, T. (2002). Score tests for detecting excess risks around putative sources, *Statistics in Medicine*, **21**, 497–514.
- Tango, T. (2007). A class of multiplicity adjusted tests for spatial clustering based on case-control point data, *Biometrics*, **63**, 119–127.
- Tango, T. (2008). A spatial scan statistic with a restricted likelihood ratio, *Japanese Journal of Biometrics*, **29**(2), 75–95.
- Tango, T. (2010). *Statistical Methods for Disease Clustering*, Springer, New York.
- Tango, T. and Takahashi, K. (2005). A flexibly shaped spatial scan statistic for detecting clusters, *International Journal of Health Geographics*, **4**, 11.
- Tango, T. and Takahashi, K. (2012). A flexible spatial scan statistic with a restricted likelihood ratio for detecting disease clusters, *Statistics in Medicine*, **31**(30), 4207–4218.
- Tango, T., Fujita, T., Tanihata, T., Minowa, M., Doi, Y., Kato, N., Kunikane, S., Uchiyama, I., Tanaka, M. and Uehata, T. (2004). Risk of adverse reproductive outcomes associated with proximity to municipal solid waste incinerators with high dioxin emission levels in Japan, *Journal of Epidemiology*, **14**, 83–93.
- 丹後俊郎, 横山徹爾, 高橋邦彦 (2007). 『空間疫学への招待』, 朝倉書店, 東京.
- The International Commission on Radiological Protection (1996). *Age-dependent Doses to Members of the Public from Intake of Radionuclides (Part 5), Compilation of Ingestion and Inhalation Dose Coefficients*, ICRP Publication 72.26:1, Oxford, New York.
- United States Environmental Protection Agency (2011a). *Preliminary Report on Operational Guidelines Developed for Use in Emergency Preparedness and Response to a Radiological Dispersal Device Incident*, United States Environmental Protection Agency (USEPA), Washington, D.C.
- United States Environmental Protection Agency (2011b). *Exposure Factors Handbook 2011 Edition*, United States Environmental Protection Agency (USEPA), Washington, D.C.
- Waller, L. A. and Gotway, C. A. (2004). *Applied Spatial Statistics for Public Health Data*, John Wiley & Sons, New Jersey.
- Waller, L. A., Turnbull, B. W., Clark, L. C. and Nasca, P. (1992). Chronic disease surveillance and testing of clustering of disease and exposure: Application to leukemia incidence and TCE-contaminated dumpsites in upstate New York, *Environmetrics*, **3**, 281–300.
- 矢島美寛 (2012). 「特集 時空間統計解析：新たな分野横断的展開」について, *統計数理*, **60**(1), 1–2.

## Epidemiologic Studies with Location Information and Their Statistical Methods

Kunihiko Takahashi<sup>1</sup>, Shizue Izumi<sup>2</sup> and Ayano Takeuchi<sup>3</sup>

<sup>1</sup>Department of Biostatistics, Nagoya University Graduate School of Medicine

<sup>2</sup>Department of Computer Science and Intelligent Systems, Faculty of Engineering, Oita University

<sup>3</sup>Center for Environmental Health Sciences, National Institute for Environmental Studies

In recent years, spatial data consisting of observed data with location information has been more frequently used to assess geographical variation of events and perform inter-regional comparisons of health risk. In epidemiology, spatial epidemiological studies dealing with descriptions of the geographical gap of incidence of disease and considering the geographical variation of various factors, attempts at spatial modelling for health risk estimation, and related statistical methods have attracted attention. This paper introduces some examples of epidemiological studies using spatial data, and outlines how to proceed with such research and use statistical methods.