

Surveillance (ISDS) の主催する会議 (Syndromic Surveillance Conference) などでも、サーベイランスに関する様々な発表・討論が行われている³⁾。

公衆衛生におけるサーベイランスとは米国の CDC (The Centers for Disease Control and Prevention; 疾病予防管理センター) では以下のように定義されている⁴⁾。

the ongoing, systematic collection, analysis, and interpretation of health data essential to the planning, implementation, and evaluation of public health practice, closely integrated with the timely dissemination of these data to those who need to know. The final link of the surveillance chain is the application of these data to prevention and control. A surveillance system includes a functional capacity for data collection, analysis, and dissemination linked to public health programs. (Thacker, 1994)

この定義のように、一般にサーベイランスの議論はそのデータの収集方法、システムから議論され、最終的には予防やそのコントロールまで包括して検討されるものである。しかし解析部分においては統計学が重要な役割を果たすことになる⁵⁾。バイオテロリズムのように突発的な症候の発生を発見するためには、日頃から関連の症状の発生状況を監視しておき、患者数が通常の状況に比べ突発的に集中した場合、それが重要なシグナルになっていると考えることができる。もちろん疾病によっては通常ではなかなか起きないもので、1件でも患者が発見されれば直ちに対応が必要なものもあるが、一般的には、日常的に似たような症状・疾患が少数ながらも起きてもおかしくないものも多い。このようなデータからシグナルを統計的に検出するため、いくつかの手法が用いられる。ある特定の地点における発生状況を追い続け、その変化を検出するための Process Control Charts などは代表的なものである。また、地域的な変化も観察するための方法としては、疾病地図を推定する手法なども適用される。

なかでも疾病集積性の検定 (disease clustering test) は、そのままサーベイランスの問題に適用可能であり、これらの解析の重要な手法のひとつとなっている。たとえば、ある地域において突発的な事象が発生したかどうかを検出するためには、日頃から定期的に観測される経時データを用いた伝統的な時間集積性の解析手法が利用できる。観測された疾病の頻度が通常期待される頻度比べて大きい現象が、ある時期に集中していれば、そこで何らかの事象が発生したと考えられ重要なシグナルとなる。この場合、疾病集積性の検定を行うことで集積 (cluster) の有無が判定できる。一方、ある時点における空間 (平面) 的な疾病の集積性の検討には空間集積性 (地域集積性) の解析手法が用いられる。この解析では、ある時点において、注目してい

る事象 (死亡など) が対象地域内のどこかに集積しているかどうかを判定することができる。なかでも集積の有無の判定と同時に集積地域の同定も行う Cluter Detection Test (CDT) は有用であり、とくに Kulldorff による spatial scan statistic⁶⁻⁷⁾ を用いる方法と、そのアプリケーションソフトウェア "SaTScan" は空間疫学の研究において広く利用されている。しかし最近、Kulldorff の方法の問題点として非円状の地域をうまく同定できないことが指摘され、それを改良した方法として Duczmal and Assunção⁸⁾, Patil and Taillie⁹⁾, Tango and Takahashi¹⁰⁾, Assunção *et al.*¹¹⁾, Kulldorff *et al.*¹²⁾ などいくつかの方法が提案されている。

さて、バイオテロリズムの発生など突発的な事象を早期発見するサーベイランスの目的としては、それが「いつから」発生したか、「どこで」発生したか、空間的・時間的の両面の検出が重要となる。この目的のために、疾病の空間・時間集積性 (space-time clustering) の検定を利用した方法がいくつか提案されているが、現時点で実際の解析に即利用できる形で提供されている手法は多くない。そこで Kulldorff¹³⁾ は空間集積性の検定のための spatial scan statistic に時間集積性の検出も含めた空間・時間スキャン統計量 (space-time scan statistic) を提案した。この方法により突発的な事象が「いつ」「どこで」発生したのかを検出し、その3次元的な cluster を同定することが可能となる。実際アメリカのいくつかのサーベイランスシステムにはこの方法が取り入れられ、日々解析が行われている。さらにこの改良として Takahashi *et al.*¹⁴⁾ による flexible space-time scan statistic も提案されている。本論では、これらの space-time scan statistic による検定に焦点をあて、サーベイランスのための解析について論じる。

2 後ろ向き研究と前向き研究

ある地点において継続して観測されたデータから、その発生の集積性を検出するためには、時間集積性の検定法が利用できる。また地域的な集積性も同時に検出するためには空間・時間集積性の検定法が利用できる。一般的にこれらの方法は過去に得られたデータから集積のあった (過去の) 集積性を検出していることになる。つまり後ろ向き (retrospective) の方法である。しかし、バイオテロリズムの発生の監視のようなサーベイランスを目的とした解析においては、すでに終結した集積性の発見よりも、むしろ現時点でも続いている集積を、発生時点から時間をあけずそれをいち早く発見・同定することが重要である。そのため Kulldorff¹³⁾ では、解析時点を含んだ「生きているクラスター (alive cluster)」を同定する前向き (prospective) の方法を提案している。この解析によって、解析時点において、まさに起こっている突発的な事象の発生が「いつから起きていたのか」を検出することができる。そのためサーベイランスの解析では、(目的に応じて) 短い間隔で定期的 (例えば毎日1回や毎週1回など) に解析するこ

とが求められるのである。

ところで、一般的な統計的検定の有意性の判定基準としては、 p 値が0.05や0.01などの値を用いることが多い。しかし、日々のサーベイランスにおいて5%の確率で起こるということは、 $1/0.05 = 20$ でほぼ20日に1回の頻度で起こっても不思議がないということになる。このように「〇〇日に1回の頻度よりも稀である」という期間を Recurrence Interval (RI) という¹⁵⁾。その考えから、毎日行われるサーベイランスにおいては、その有意性の判定基準として、ほぼ1年に1回の頻度よりも稀な現象 (RI=365日) に対応する $p = 1/365 = 0.0027$ がひとつの基準として用いられることもある。週単位のデータであれば、1年間の52週間後を基準として $p = 1/50 = 0.02$ などが適当であろう。

なお、発生直後の集積は一般的に (i) 空間的に集積地域は狭い地域、(ii) 時間的に発生時点は解析時点に近い最近、であることが想定されるであろう。また感染性の症候などは、いったん発生した後、徐々に近隣の地域へ広がっていくことなども考えられよう。このような空間的な広がりを把握するためのモニタリングもサーベイランスにとって重要な問題であるが、この目的のためにも空間集積性、もしくは空間・時間集積性の検定法などが適用可能となる。

3 サーベイランスのための Cluster Detection Test

時間的なデータを観察しながら集積性を検出する方法としては時間集積性の検定法が利用できる。一方で、サーベイランスを目的とした場合、その発生時点とともに発生地域の同定が重要となる。そこで Kulldorf¹³⁾ は平面上での疾病集積性の検定である spatial circular scan statistic に時間のデータを入れた cylindrical space-time scan statistic (SaTScan) を提案した。

いま、解析を行う対象地域 G が m 個の region (市区町村, counties, zip codes など) に分割されているものとし、各 region に対して、時点 $Y_1, Y_1 + 1, Y_1 + 2, \dots, Y_2$ のデータが存在するとする。この対象地域において、 $[Y_1, Y_2]$ の時点では集積が存在しない、つまり定期的に症候が観測されているということが帰無仮説となる。この帰無仮説のもとでは、 i 地区の時点 d における case の数 (観測数) N_{id} が互いに独立に期待値 μ_{id} の Poisson 分布

$$N_{id} \sim \text{Poisson} (\mu_{id})$$

に従うとし、その観測値を n_{id} とする。ただし、 μ_{id} は i 地区の人口に比例する値、または性・年齢などの共変量を調整した期待観測数とする。ここで、日々のサーベイランスを目的とした解析のため、解析時点を含んだ cluster のみを考える前向き研究として prospective space-time 解析を考える。つまり、時点 Y_2 のデータが得られた段階で解析

を行い、クラスターの時間 $[s, t]$ として常に $t = Y_2$ となる集積のみを考える。これによって同定されるクラスターの時間は $[s, Y_2]$ となる alive cluster になる。さらに、発生時点 s も解析時点 Y_2 からそれほど遠くない時点であると想定される。そこで、実際にデータは $[Y_1, Y_2]$ の全てを用いるが、発生時点はある時点 Y_T (Y_2 から T 時点前) 以降である、と考え、 $[s, t]$ として

$$[Y_2 - T + 1, Y_2], [Y_2 - T + 2, Y_2], \dots, [Y_2 - 1, Y_2], [Y_2, Y_2]$$

のいずれかであるような cluster を探しだすことになる。この方法が prospective scan 法 (図1) であり、このときの T を maximum temporal length という。

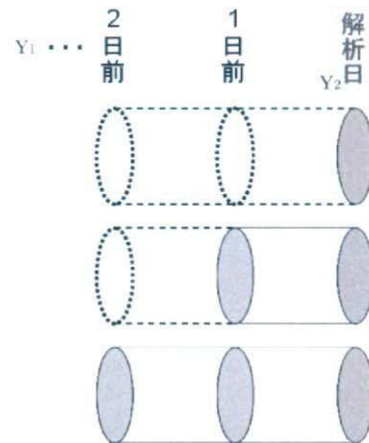


図1 prospective space-time 解析の模式図

さらに $Z_{ik} (k=1, 2, \dots, K)$ を region i から近い順に、 i 自身を含む k 個の region からなる集合とする。ただし各 i の座標はその region の代表点1点 (市区町村役場の所在地など) であらわすものとする。ここで K は予め定めしておく maximum geographical length とする。このとき cylindrical scan statistic では、底面を

$$Z_1 = \{Z_{ik} \mid 1 \leq i \leq m, 1 \leq k \leq K\}$$

および高さを

$$Y = \{[Y_2 - s + 1, Y_2] \mid 1 \leq s \leq T\}$$

とした全体集合 $W = Z_1 \times Y$ からの要素となる円柱状の window W を考える。

しかし、この方法は底面が円状の正円柱状の window だけを考えているため、平面上での circular scan 同様、非円状の地域を同定することは出来ない。そこで Takahashi *et al.*¹⁴⁾ は spatial flexible scan における window $Z \in Z_2$ を底面にもつ多角柱の window を scan することで非円状の地域の space-time の解析が可能となる flexible space-time

scan 法を提案した. まず region 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 Z_2 を考える. つまり Z_{ik} の中で i を含んで k 個の region からなる連結したものが j_{ik} 個あるとすると, 底面を

$$Z_2 = \{Z_{ik(j)} \mid 1 \leq i \leq m, 1 \leq k \leq K, 1 \leq j \leq j_{ik}\}$$

とする $W = Z_2 \times \mathcal{V}$ からの要素となる window W を考える.

ここで window W における観測数を $N(W)$ とし, その観測値を $n(W)$, さらに W が cluster でないときの期待観測数を $\mu(W)$ とする. cluster の有無の検定として

$$\text{帰無仮説: } E(N(W)) = \mu(W) \text{ for } \forall W \in \mathcal{W}$$

$$\text{対立仮説: } E(N(W)) > \mu(W) \text{ for } \exists W \in \mathcal{W}$$

を考え, window W が cluster となる対立仮説のもとでの尤度を $L(W)$ とし, 帰無仮説のもとでの尤度を L_0 とする. 尤度比統計量

$$\lambda(W) = \begin{cases} \frac{L(W)}{L_0} = \left(\frac{n(W)}{\mu(W)} \right)^{n(W)} \left(\frac{\mu(W)}{n(W)} \right)^{n(W)} & (n(W) > \mu(W)) \\ 1 & (\text{その他}) \end{cases}$$

を考え

$$\lambda^* = \lambda(W^*) = \max_{W \in \mathcal{W}} \lambda(W)$$

のように最大尤度比 λ^* をとる window W^* を most likely cluster (MLC) とし, これを cluster の候補と考える. ここで, この MLC が統計的に有意な集積性をもつかどうかの評価が必要となる. そのため帰無仮説のもとでの $\max_{W \in \mathcal{W}} \lambda(W)$ の分布を使って有意性を見るが, 一般的には Monte Carlo 法を利用して数値的に求めた p 値によってその有意性が検討される. その結果有意となれば, W^* が有意な cluster となり, その地域, その期間に疾病が集積していたと判定できる.

4 米国 Massachusetts 州東部のサーベイランスデータ

このデータは Harvard Vanguard Medical Associate によって利用されているデータで, zip code ごとにまとめられた電子的な医療記録として毎日集計され¹⁶⁻¹⁷⁾ SaTScan を利用して毎日解析が行われている. ここでは2005年8月の呼吸器疾患 (respiratory) の発生について, Kulldorff の cylindrical scan 法 (SaTScan) と Takahashi *et al.* の flexible scan 法 (FleXScan) によって解析を行った. 解析に利用したデータ, パラメータは以下のとおりである.

• データ:

2005年8月における respiratory の発生

Massachusetts 州東部にある $m = 385$ の zip codes からなる地区

• 解析:

解析モデル: Poisson モデル

2005年8月7日から30日まで毎日解析

scan する window の平面における最大長 (Maximum geographical length): $K = 20$ area

scan する window の最大時間 (Maximum temporal length): $T = 7$ 日

Monte Carlo シミュレーションの回数: 999

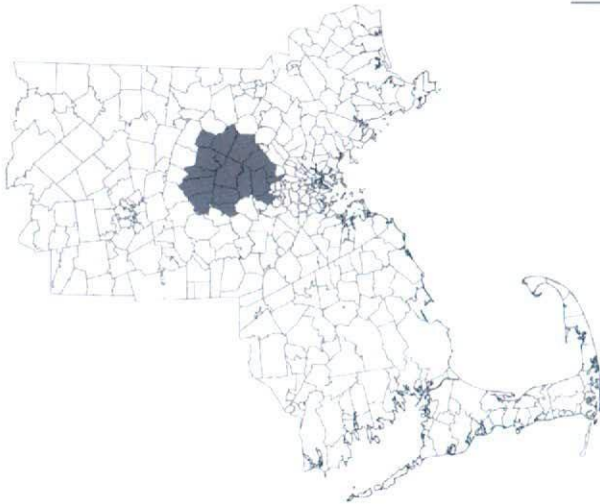
有意と判断する recurrence interval: 6ヶ月以上 (すなわち $p < 0.0054$)

なお, 期待観測数は同地域における過去1年分以上のデータを用い, 性別, 年齢以外にも季節, 月, 曜日などの影響も調整されたものを利用している.

この解析によって, 8月12日~15日の解析において有意な集積が検出された (図2). まず8月12日において, SaTScan, FleXScan とも $RI=2.7$ 年 ($p=0.001$) の集積が検出された. SaTScan においては18地区からなる地域で8月11日~12日の2日間に集積が同定された. 一方, FleXScan では12地区からなる地域が同定され, 微妙に同定された地域が異なっていた (11地区は共通). 8月13日, 14日においては SaTScan は12日と同じ地域を同定したが, それらの RI は短く有意とはならなかった. 一方, FleXScan では13地区からなる有意な集積地域が同定された. 8月15日ではそれぞれ有意な集積地域が同定されたがその集積の時間が異なった (SaTScan では5日間, FleXScan では6日間). このとき FleXScan では集積地域がまた少し変化したが SaTScan では集積地域の変化の様子は観察されなかった.

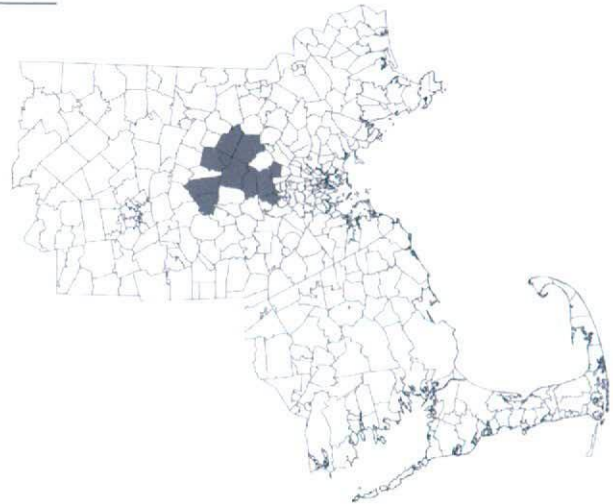
5 北九州市内小学校欠席数によるサーベイランス

次に, 小学校の欠席児童数によるサーベイランス解析について検討を行った. このデータは2006年度福岡県北九州市内の132小学校 (全児童数52,189名) について, 毎週水曜日 (小学校の休校日は除く40週) にその日の欠席者数が報告されたものである. ここでは特に目立つ欠席増減がなかった4~9月 (18週) の小学校別平均欠席児童数を各小学校のベースライン (期待欠席数) として用い, 10月11日から12月冬休み前まで毎週解析を行った. 解析には FleXScan (maximum length: $K=15$, $T=2$) とし, その有意性の判定として $p < 0.02$ ($RI \approx 1$ 年) を用いた. その結果, 11月15日までの解析 (10月11日, 18日, 25日, 11月1日, 8日, 15日) では有意な集積は検出されなかった. しかし, 11月22日に有意な欠席数の集積が検出され, その後, 冬休み前の12月20日までの5週間にわたり集積が認められた (表1, 図3). 特に12月20日においてはその欠席の集積地域が小さい範囲になっている様子が観察された.

Aug. 12

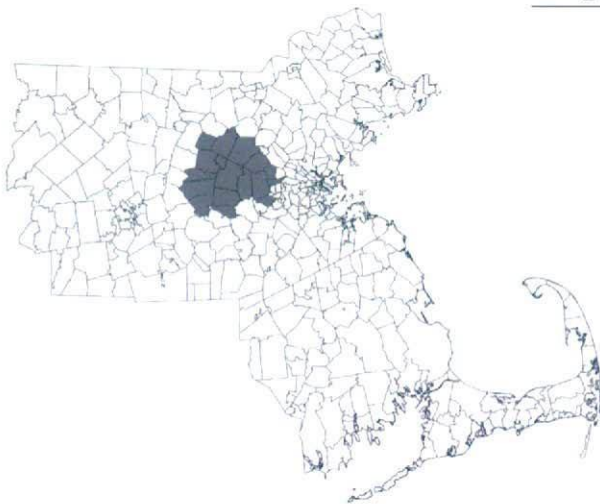
SaTScan

Cluster period: Aug.11-12
 RI=2.7 years ($p = 0.001$)
 No. of cases: 51, RR: 2.25
 No. of zip codes: 18



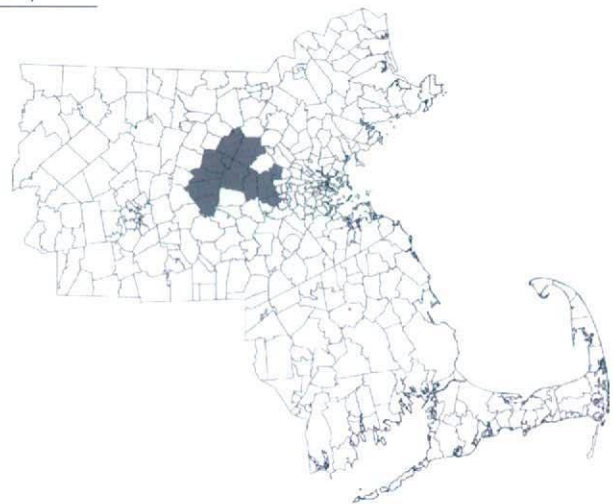
FleXScan

Cluster period: Aug.11-12
 RI=2.7 years ($p = 0.001$)
 No. of cases: 42, RR: 3.37
 No. of zip codes: 12

Aug. 13 / 14

SaTScan

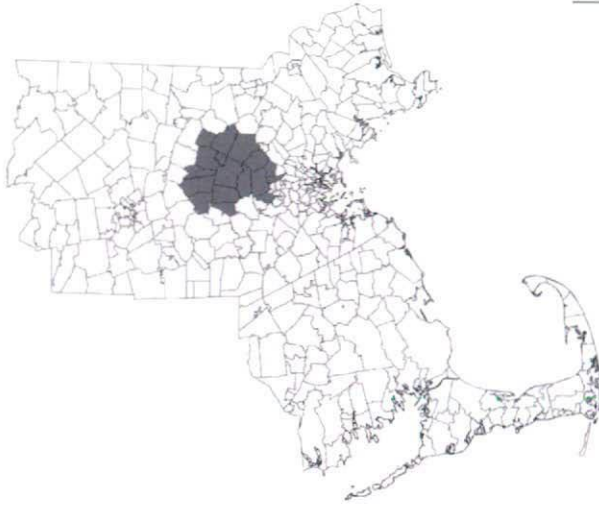
Cluster period: Aug.11-13 / 11-14
 RI=91 days ($p = 0.011$) / 30 days ($p = 0.033$)
 No. of cases: 55 / 59, RR: 2.31 / 2.19
 No. of zip codes: 18



FleXScan

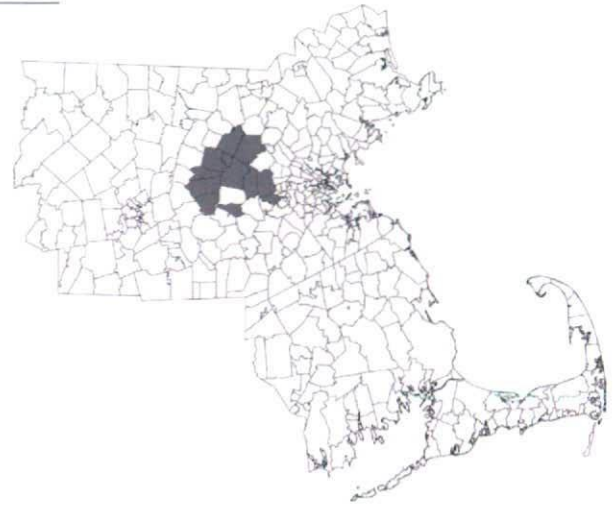
Cluster period: Aug.11-13 / 11-14
 RI=333 days ($p = 0.003$) / 250 days ($p = 0.004$)
 No. of cases: 46 / 49, RR: 3.08 / 2.89
 No. of zip codes: 13

Aug. 15



SaTScan

Cluster period: Aug.11-15
 RI=200 days ($p = 0.005$)
 No. of cases: 82, RR: 2.00
 No. of zip codes: 18



FleXScan

Cluster period: Aug.10-15
 RI=1.4 years ($p = 0.002$)
 No. of cases: 72, RR: 2.40
 No. of zip codes: 12

図2: Massachusetts 州における respiratory の発生の集積

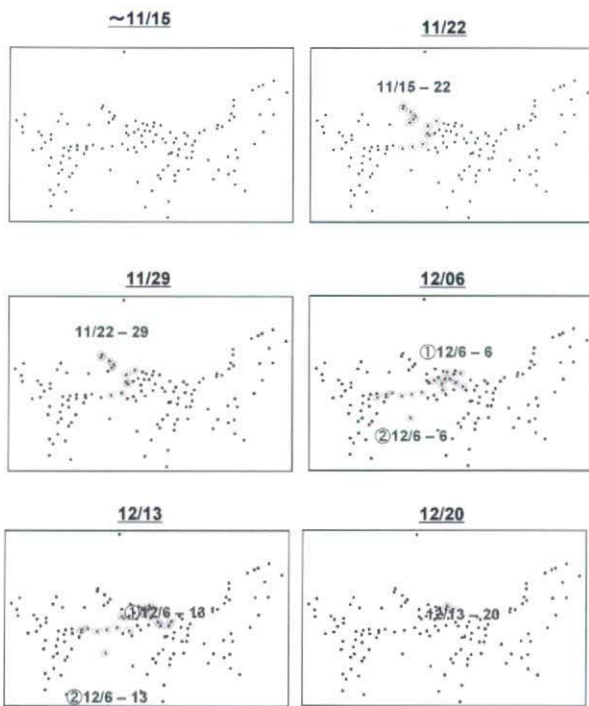


図3 北九州市小学校欠席サーベイランスで検出されたクラスター (10月~12月). 点は小学校の座標を表わす.

表1 北九州市内小学校欠席数によるサーベイランス解析の結果 (10~12月の各週)

解析日	校数	cluster period	cases	expectd	p-value
11月22日	13校	11月15~22日	283	191.4	0.001
11月29日	11校	11月22~29日	238	164.9	0.001
12月6日	9校	12月6~6日	141	84.5	0.002
	7校	12月6~6日	84	45.8	0.006
12月13日	11校	12月6~13日	305	225.4	0.001
	10校	12月6~13日	203	142.3	0.006
12月20日	2校	12月13~20日	80	44.3	0.012

もちろん, この解析は欠席数のみからの検討であり, その欠席理由がなんであるかは特定できない. しかし, 事後的に得られたこの期間の北九州市内の定点観測での感染性腸炎の発生 (図4) と見比べると, この結果がうまく符合していることがうかがえ, この欠席は感染性腸炎のものであったのではないかと推測される.

なお, 1月以降の解析においては3月初旬まで有意な集積が検出された. これはインフルエンザの流行によるものと推測される.

6 サーベイランス解析における解釈

統計学的に有意な集積性が検出された場合でも, サーベイランスにおいてはいくつかの注意が必要となる. たとえば先に述べたように, $p < 0.05$ となったものが直ちに問題として考えることはせず, そのデータに応じた RI などを考えることになる. データが日単位のものなのか, 週単位

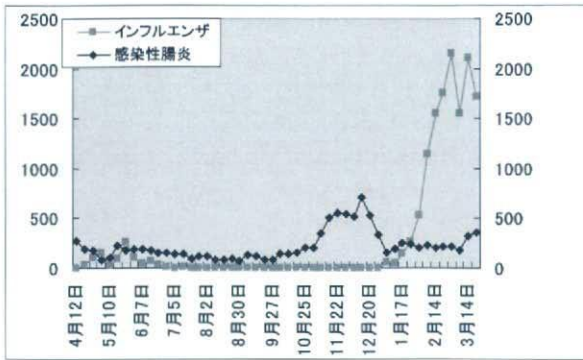


図4 北九州市内定点観測によるインフルエンザと感染性単核炎の発生報告件数

のものなのかでそのシグナルの有意性の判断が違ってくることになる。

また、単に有意性だけみればよいわけではない。表2は、先ほどの例と同じく Massachusetts において2005年8月に発疹 (Rash) の発生を解析した結果 $p < 0.0054$ となったものである。この解析では SaTScan, FleXScan と同じ結果が得られ、8月7日～10日の解析で、1箇所の zip code だけが同定され、その有意性はいずれも強いものであった ($p = 0.001$)。ここから「この突発的な発生は8月2日であり、そこから10日までおきていた」と解釈されそうである。しかし、この結果をよく眺めると、8月7日の解析では、8月2日～7日の6日間が同定され、その観測数は7件であった。次に8月7日の解析でも集積性が検出されたが、そこで同定されたクラスターは8月2日～8日の7日間で観測数は7件のままである。つまり、8月8日にはこの地区で疾病発生は起きていないことが読み取れる。さらに、9日、10日には、期待観測数が同じであるにも係わらず、観測数は減少している。つまりこの場合、8

月7日の解析で検出されたクラスターは検討しなくてはいいけないが、8日～10日の有意なクラスターはそれほど問題にならないと考えられよう。

また、同定されたクラスターの情報 (症候, 場所, 期間など) を検討し、このクラスターが本当に問題となるかどうかを検討する必要がある。例えば発生した曜日の問題 (週明けは患者が多い) や特定のイベントがあった日など、様々な検討がされる。これは統計学者だけではなく、疫学や医学, 保健医療, 行政, データ解析など, それぞれの専門家がチームとなって検討されるべきものである。

さらに、北九州市内の小学校欠席児童数による解析では、特定の疾患などの情報はなく、理由を問わず学校欠席のデータから集積が検出されるだけであり、その欠席理由までは断定できない。むしろこの結果で検出された集積地域で「何か異常がおきているのではないか」、「感染性の疾患が発生しているのではないか」という疑いが出たのであって、実際のサーベイランスとしては、同定された小学校の詳細調査を即時に行うことで、その原因究明が必要となるであろう。しかし、定常的に観測し続けるなかで何らかの健康危機事象の発生をいち早く見つけ出すという観点では、この小学校欠席数によるサーベイランスも有用であると考えられる。

本論のはじめに述べたように、近年、サーベイランスのための研究は重要となってきた。その解析手法として空間疫学における手法は有用なツールとなっている。サーベイランスを目的とした手法の開発もさらに進んでいくことになるであろう。しかし、同じくはじめに述べたように、サーベイランスの検討は解析手法だけではなく、データの収集から解析, それを行うシステム, さらに対策までを総合的に考える必要もある。わが国においても健康危機管理のひとつとしてこれらのサーベイランスがますます重要な課題になってくると考えられる。

表2: Detected outbreaks of Rash based on daily syndromic surveillance data in eastern Massachusetts during August 1-30, 2005.

Day	zip codes	cluster period	cases	expectd	R.I. (p-value)
Aug.07	01951	Aug.02-07	7	0.0427	2.7years (0.001)
Aug.08	01951	Aug.02-08	7	0.0545	2.7years (0.001)
Aug.09	01951	Aug.03-09	6	0.0545	2.7years (0.001)
Aug.10	01951	Aug.04-10	5	0.0545	2.7years (0.001)

参考文献

- 1) 丹後俊郎, 横山徹爾, 高橋邦彦. 空間疫学への招待. 東京: 朝倉書店; 2007.
- 2) Brookmeyer R, Stroup DF, editors. Monitoring the health of populations. New York: Oxford University Press, Inc.; 2004.
- 3) International Society for Disease Surveillance. <http://www.syndromic.org/>
- 4) Lawson AB, Kleinman K, editors. Spatial & syndromic surveillance for public health. Chichester: John Wiley & Sons; 2005.
- 5) Wilson AG, Wilson GD, Olwell DH, editors. Statistical methods in counterterrorism. New York: Springer; 2006.
- 6) Kulldorff M, Nagarwalla N. Spatial disease clusters: detection and inference. *Statistics in Medicine* 1995; 14: 799-810.
- 7) Kulldorff M. A spatial scan statistic. *Communications in Statistics: Theory and Methods* 1997; 26: 1481-1496.
- 8) Duczmal L, Assunção R. A simulated annealing strategy for the detection of arbitrarily shaped spatial clusters. *Computational Statistics & Data Analysis* 2004; 45: 269-286.
- 9) Patil GP, Taillie C. Upper level set scan statistic for detecting arbitrarily shaped hotspots. *Environmental and Ecological Statistics* 2004; 11: 183-197.
- 10) Tango T, Takahashi K. A flexibly shaped spatial scan statistic for detecting clusters. *International Journal of Health Geographics* 2005; 4: 11.
- 11) Assunção R, Costa M, Tavares A, Ferreira S. Fast detection of arbitrarily shaped disease clusters. *Statistics in Medicine* 2006; 25: 723-742.
- 12) Kulldorff M, Huang L, Pickle L, Duczmal L. An elliptic spatial scan statistic *Statistics in Medicine* 2006; 25: 3929-3943.
- 13) Kulldorff M. Prospective time periodic geographical disease surveillance using a scan statistic. *Journal of the Royal Statistical Society, Series A* 2001; 164: 61-72.
- 14) Takahashi K, Kulldorff M, Tango T, Yih K. A flexibly shaped space-time scan statistic for disease outbreak detection and monitoring. *International Journal of Health Geographics* 2008; 7: 14.
- 15) Kleinman K, Lazarus R, Platt R. A generalized linear mixed models approach for detecting incident clusters of disease in small areas, with an application to biological terrorism. *American Journal of Epidemiology* 2004; 159:217-224.
- 16) Lazarus R, Kleinman K, Dashevsky I, DeMaria A, Platt R. Using automated medical records for rapid identification of illness syndromes (syndromic surveillance): The example of lower respiratory infection. *BMC Public Health* 2001; 1: 9.
- 17) Lazarus R, Kleinman K, Dashevsky I, Adams C, Kludt P, DeMaria A, Platt R. Use of automated ambulatory-care encounter records for detection of acute illness clusters, including potential bioterrorism events. *Emerging Infectious Diseases* 2002; 8(8): 753-760.

特集：地域診断・症候サーベイランスに向けた空間疫学の新展開

健康危機管理と小学校欠席状況サーベイランス

郡山一明¹⁾²⁾，片岡裕介³⁾，竹中ゆかり²⁾，浅見泰司³⁾，高橋邦彦⁴⁾，丹後俊郎⁴⁾

¹⁾ 北九州市危機管理参与，²⁾ 救急救命九州研修所，

³⁾ 東京大学空間情報科学研究センター，⁴⁾ 国立保健医療科学院技術評価部

Health Crisis Management and Absentee Surveillance in Elementary School

Kazuaki KOHRIYAMA¹⁾²⁾，Yusuke KATAOKA³⁾，Yukari TAKENAKA²⁾，
Yasushi ASAMI³⁾，Kunihiko TAKAHASHI⁴⁾，Toshiro TANGO⁴⁾

¹⁾ Supervisor of Crisis Management in Kitakyushu City，²⁾ Emergency Life Saving Training Academy of Kyushu，

³⁾ Center for Spatial Information Science，University of Tokyo，

⁴⁾ Department of Technology Assessment and Biostatistics，National Institute of Public Health

抄録

健康危機管理における小学校欠席状況サーベイランスの有用性について解説した。欠席状況をサーベイランスシステムとして使用するためには、日常的に存在する小学校在籍者数及び欠席状況の学校間較差を解消する必要がある。この較差を解消するために標準化欠席比という考え方を導入した。本方法を用いることで、実際に起きた地域の健康危機事態のアウトブレイクを早期に把握することが可能であった。また、GISを用いることでインフルエンザ流行期の欠席状況には早期より空間集積性が見出された。健康危機管理におけるサーベイランスの目的は、危機拡大を抑制するための予防行動強化を起こすきっかけを得ることにある。小学校欠席状況サーベイランスは、健康危機事態の早期把握に有用であり、空間-時間集積性検定を応用することでさらに有用性が高まると考えられる。

キーワード： 健康危機管理，サーベイランス，欠席率，GIS，空間-時間集積性

Abstract

The utility of the elementary school absentee surveillance in the health-crisis-management was explained. As a surveillance system, it is necessary to dissolve interscholastic difference of the elementary school enrollment and the absence situation that exists daily to use the absence situation. To dissolve this difference, we introduced the idea of standardized absentee ratio. The outbreak of the health crisis in the region that actually occurred was able to be detected by using this method at the early stage. Moreover, at the early stage, the space cluster was found to the absence situation of the influenza fashion period by using GIS. The purpose of surveillance in a health-crisis-management is to obtain the chance to cause the preventive action to control the expansion of crisis. We consider that the elementary school absentee surveillance is useful for detect at the early stage of the health-crisis-situation. Moreover, we consider that the utility of surveillance rises further because it applies spatial-time scan statistic.

Keywords: health crisis management, surveillance, absentee ratio, GIS, spatial-time clustering

〒807-0874 福岡県北九州市八幡西区大浦3-8-1

救急救命九州研修所

3-8-1 Oura, Yahata-nishi-ku, Kitakyushu, Fukuoka, 807-0874 Japan.

Email : kooriyama@fasd.or.jp

1 はじめに

危機は自然と文明の共同産物である。特定地域の危機にすぎなかった風土病が物流や交通路の発達によって全世界の日常的な危機となったように、人間生活の発展・文明化は絶えず新たな危機を生み出していく。また、都市化、すなわち生活空間の密集化は単一危機で大量被災者を生む上に、都市システムはそれぞれの機能が関連して構築されているがゆえに、次々とシステムを破綻して被災の場を増やし、複合危機を形成していくという構造的な脆弱性を持っている。このような現代社会では危機管理、とりわけ健康危機の兆候を把握するサーベイランスの重要性は益々大きくなってきている。

現在実施されている感染症をはじめとする健康危機に関するサーベイランスの多くは、予め原因リストを作成しておき、その数の変化を感知する方法が用いられている。この方法では原因判明に至るまでに検査を含めて一定の時間がかかることは避けられず、原因が判明した時には既に危機が拡大していることが少なくない。そもそも新興感染症をはじめとして、原因リストが準備されていない危機には対応できない。危機のアウトブレイクという用語は曖昧に使われがちであるが、その真意は次の2つの場合に集約できると考えられる。すなわち、危機の要因が地域内で、①一定の量を越えた場合、②空間的拡大をしていく場合、である。この観点にたてば、現在、わが国で実施されている健康危機に関するサーベイランスシステムは迅速性に乏しく、また数の増減は空間的拡大を反映しないので、危機のアウトブレイク把握には不十分である。

健康危機管理におけるサーベイランスの目的は、危機の拡大を抑制するための予防行動強化を起こすきっかけを得ることにある。求められる最初の条件は、原因の正確性ではなく危機兆候把握のスピードであり、そのためには人間の社会生活の中から健康危機に関係する現象の突然の変化を検知する prospective なシステムを構築することが望まれる。具体的には定まった地点で日常的に観察されている現象の中から、危機に関連した変化を把握するサーベイランスシステムである。

小学校は地域毎に相当数が存在し、毎日一定数の児童が継続的に登校してくる。そして在籍児童が健康被害に陥った場合には、その原因が感染症であれ化学物質であれ、欠席という形で反映される。

我々は小学校のこの特性を健康危機兆候把握に応用できないかと考え調査・研究を進めてきた¹⁾。

本稿ではこれまでの成果の一部を示し、健康危機管理システムとしての小学校欠席状況サーベイランスの可能性について解説する。

2 小学校欠席状況サーベイランスに期待される特徴とそのための準備

小学校は校区によって地域性が限定さえているうえに在

籍している年齢層と数が一定である。すなわち、欠席状況の解析にあたっては、年齢調製の必要がなく、在籍者数を分母として欠席率を算出することで健康危機に陥っている状況(危機の大きさ)を異なる小学校間で数学的に比較できるという大きな利点を持っている。このことは、空間疫学の方法²⁾を用いることで異常地域の集積性を検出でき、経時的に解析することで空間的拡大も検出できるという極めて魅力的な可能性にも繋がっている。

欠席率を学校間比較に使用するためには、平常時の欠席率はどの程度なのか、在籍者数による影響はないのか、平常時の欠席状況に学校間較差はないのか、欠席率の異常をどのように定義するのか、などについて検討しなければならない。そこで、我々はまずこれらについて北九州市を対象として、市内の公立小学校のデータを収集して検討した。北九州市は九州北部に位置する面積485.55Km²、人口1,000,559名(平成16年6月1日)の政令指定都市である。調査時点での市立小学校総数は132校、総児童数は52,339名であった。

132小学校の在籍者数分布を示す。301名から400名の小学校が最も多かったが、全体の36%は300名以下であり100名以下の学校も8校存在した(図1)。

このような在籍者数分布の元で、市内で健康危機が発生していなかった15週間の各小学校の欠席率を定期的に週1回調べた。132校の欠席率は2.0±1.2%であった。縦軸に各週のそれぞれの学校の欠席率、横軸に小学校在籍者数をグラフにプロットすると、在籍者数が少ない小学校において欠席率のばらつきが大きくなっていることが分かった(図2下段)。これは、在籍者数が少ない小学校では欠席者数の増減の影響が大きく反映されて不安定になるからと考えられる。さらに規模が等しい小学校においても欠席率に大きな較差がある場合があることが分かった(図2上段)。

この結果から、各小学校の欠席状況を比較するためには、欠席率を単純に比較するのではなく何らかの標準化を図る必要があることが示された。

標準化の方法として、各学校別に健康危機事態が起きていない期間を基準として、この間の平均欠席率を期待欠席率(E)とし、実際に観測された欠席率を(O)として、O/E比をとったものを標準化欠席比とした。当然のことながら標準化欠席比を用いることで学校間較差はなくなり(図3上段)、在籍者数によるばらつきも小さくできた(図3下段)。

健康危機が発生していない時期の学校の標準化欠席比がm(平均)+2σ(標準偏差)を超えた割合は0.03であったので、m+2σをカットオフ値としてカットオフ値以上の場合を「異常」と定義した。すなわち、カットオフ値はp<0.03に相当する。

3 実際の事例による検証

本方法の有用性を過去の健康危機事態を用いて検証した。用いたのは1996年に埼玉県越生市で起きた集団下

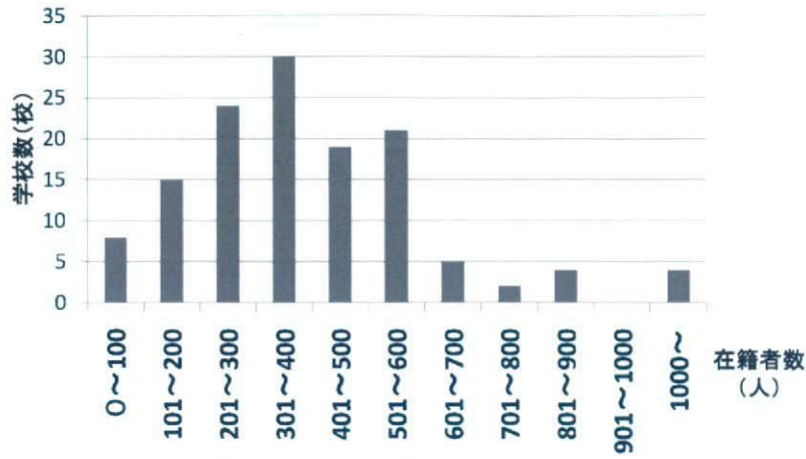


図1 北九州市小学校の在籍者数別学校数

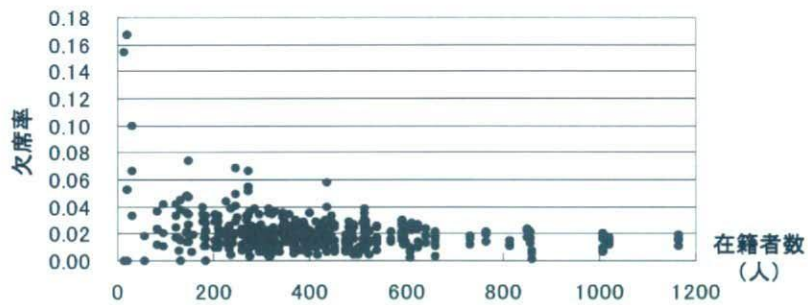
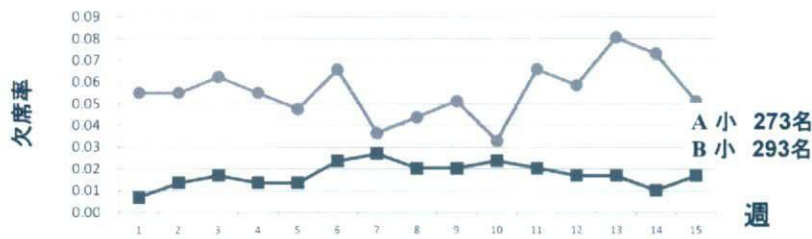


図2 欠席率：学校による欠席率較差（上段），在籍者数の関係（下段）

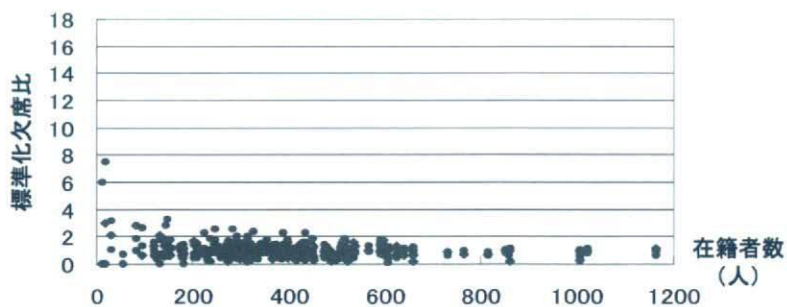
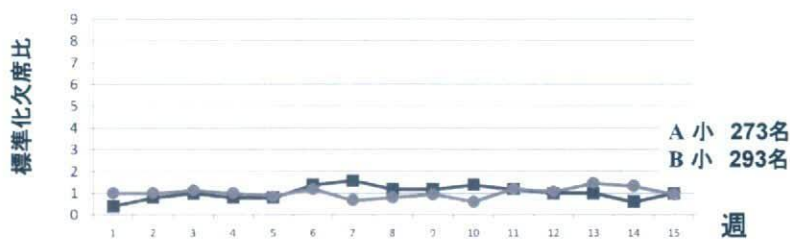


図3 標準化欠席比：学校間較差（上段），在籍者数との関係（下段）

痢症である。

[事態概要]³⁾⁴⁾

1996年(平成8年)6月10日, 埼玉県越生町で下痢・腹痛等の症状で小中学校の児童生徒が多数欠席しているとの一報が入り, 同日から調査したところ町内の住民に同様の患者が多数いることが分かった。6月17日, 有症者の検便からクリプトスポリジウム(原虫)が検出され, その後, 同町の水源, 水道水からも検出されたため, 水道水を介したクリプトスポリジウムによる集団下痢症であることが判明した。

[標準化欠席比による解析]

埼玉県越生町と隣接する毛呂山町, 鳩山町に所在する10小学校の欠席状況について, 6月3日から6日までの4日間を基準として17日までの標準化欠席比を10小学校全体と学校別に算出した(図4)。10小学校全体の標準化欠席比は6月10日から異常となった。学校別にそれぞれの

標準化欠席比とカットオフ値を検討したところ, 越生町2小学校が6月7日の時点から異常となり, その後も継続して上昇していた。隣接する毛呂山町の毛呂山小は標準化欠席比の大きさは小さいものの10日から異常を示し, そのピークは越生町の2小学校の欠席状況のピークと同様であった(図5)。

[解説]

標準化欠席比を用いることで, 越生町周辺に何らかの健康危機事態が起きていることを早期に把握できた。本方法では毛呂山小でも欠席状況の異常が観察された。報告書には越生町以外での健康危機についての記載はないが, 毛呂山町の水源になっている阿諏訪川でもクリプトスポリジウムが確認されたことを記載しており, 毛呂山小の欠席状況のピークが越生町の欠席状況と一致することを考えれば毛呂山町でも少ないながら健康危機を来たしていた可能性も否定できない。

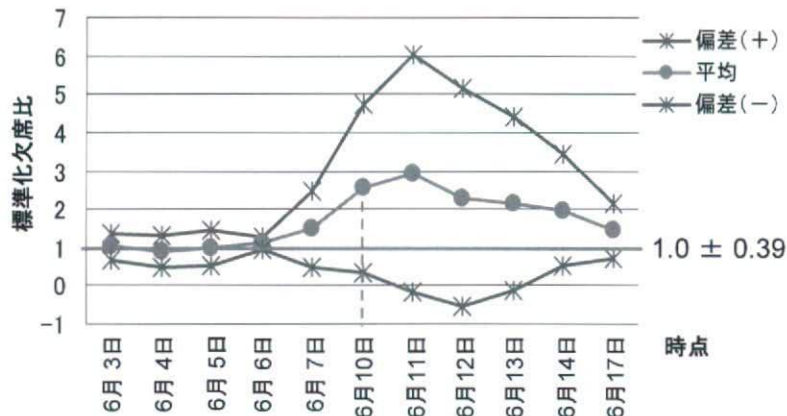


図4 越生町近隣の10小学校の標準化欠席比の推移

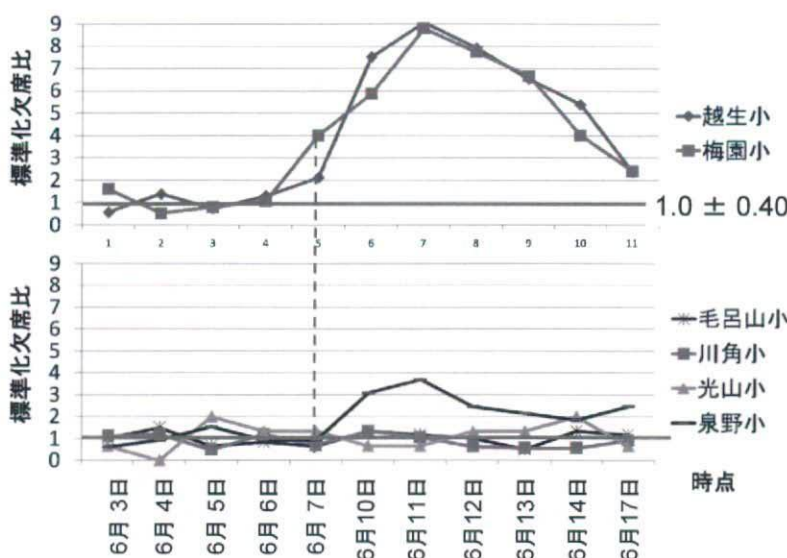


図5 標準化欠席比の推移:越生町(上段), 毛呂山町(下段)

小学校欠席状況サーベイランスは健康危機兆候把握に有効であった。

4 GISを用いた地域感染症流行への応用

2005年の北九州市において市立の132小学校の欠席状況を1月11日から3月22日まで定期的に週1回調べた。この年度の市内全体の平常時標準化欠席比は 1.0 ± 0.57 であった。市内全体の標準化欠席比の変化を示す(図6)。

定点観測ではこの時期に地域で流行していたのはインフルエンザのみであり、小学校の欠席状況はインフルエンザの流行によるものが殆どと考えられた。実際、インフルエンザ定点値と標準化欠席比の相関係数は0.85であり、両者には強い相関が認められた(図7)。市内全体の標準化欠席比がカットオフ値を超えて異常となったのは1月25日であった。定点観測値がインフルエンザの注意報基準値である10を超えたのは2月1日であった。

学校別にみると1月11日の時点で標準化欠席比が自校のカットオフ値を超えたものが13校あり、18日の時点で超えたものが50校と経時的に増加していった。

それぞれの学校で観察された標準化欠席比をもとに、クリギング法によって空間補間を行い、GISを用いて地図上に視覚化した。1月11日の時点で欠席状況が異常な地域に集積性があることが示され(図8)、18日には一気に拡大していた(図9)。

[解説]

クリギング法は元来、任意の地点のボーリング調査により鉱産物の埋蔵量を過大評価せずに推計するために開発された代表的な空間補間法のひとつである。この方法では座標上の事象を連続的な局地的な変数とみなしたうえで、局地的な空間変動は①全体的傾向、②空間的に相関しているローカルな残差、③ランダムなノイズ、の3つの成分から構成されるとする⁹⁾。この考え方は現在、気象学、農学、

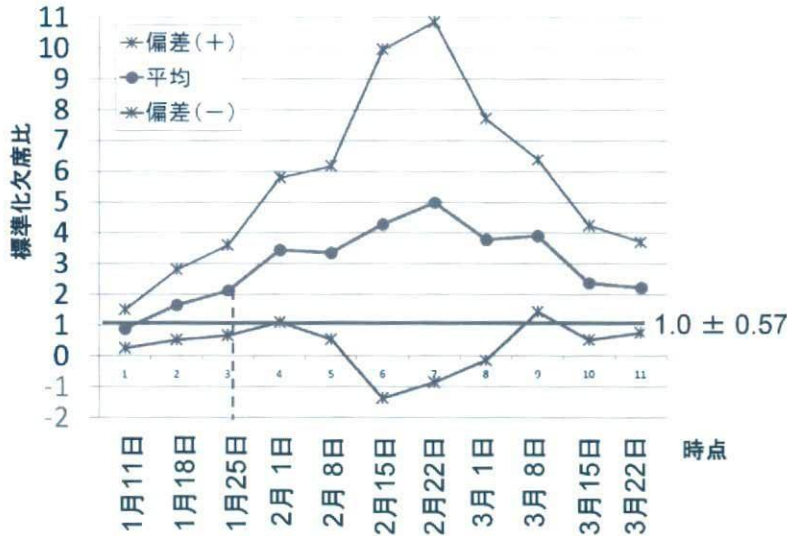


図6 2005年インフルエンザ流行期の北九州公立小学校標準化欠席比の推移

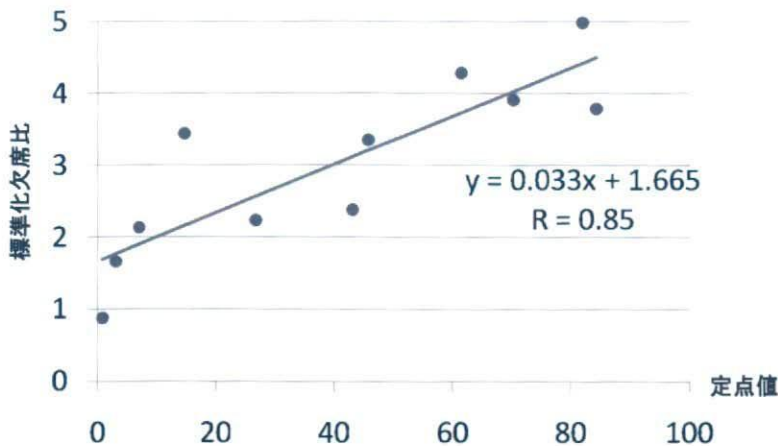


図7 北九州市のインフルエンザ定点値と小学校標準化欠席比の相関

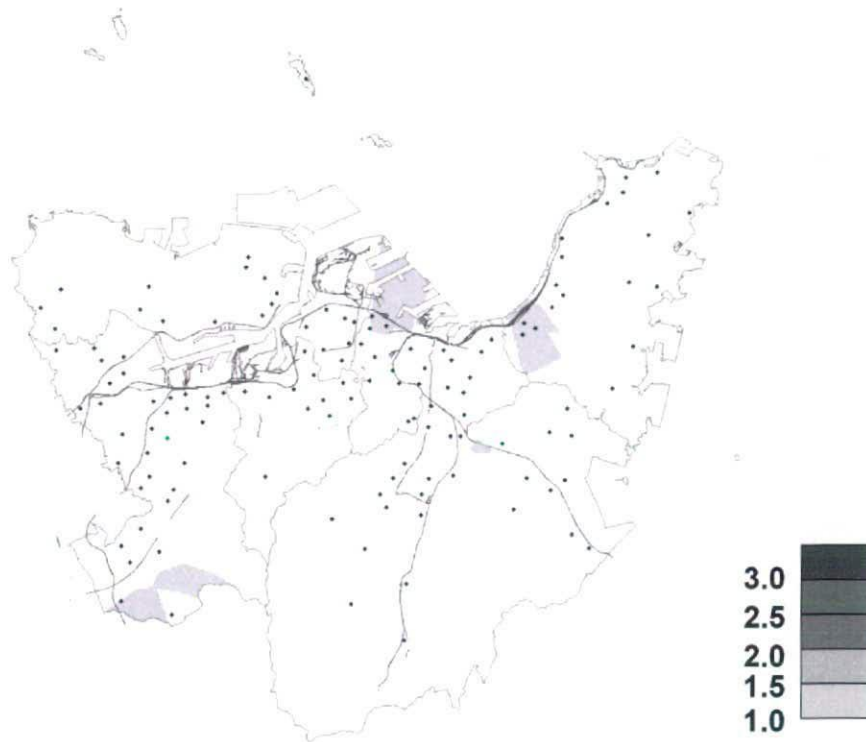


図8 カットオフ値による異常欠席状況小学校の空間集積：1月11日

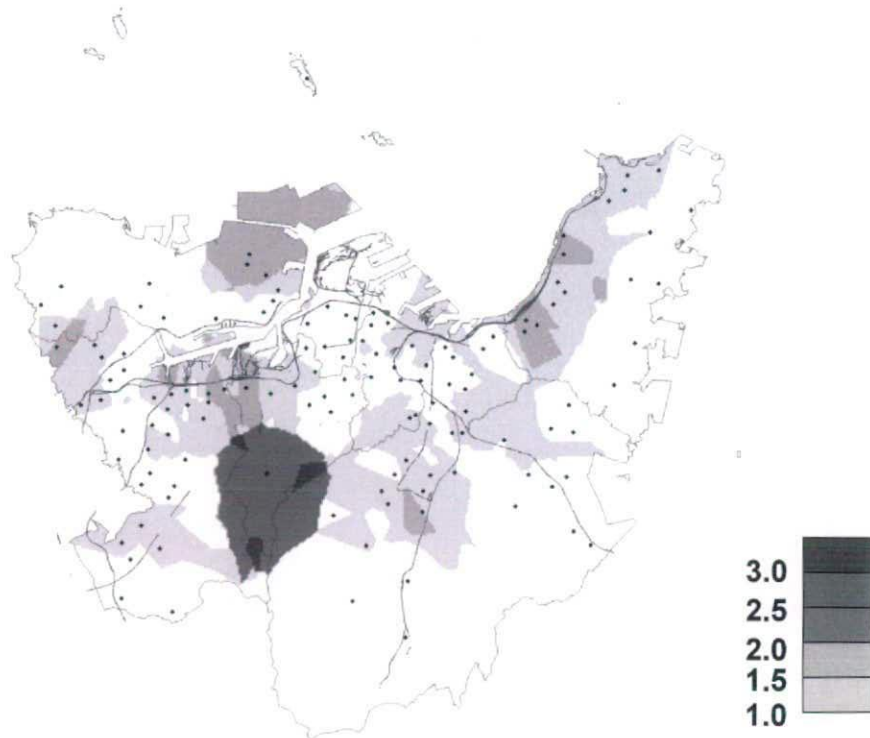


図9 カットオフ値による異常欠席状況小学校の空間集積：1月18日

森林学等に広く応用されている。同様に、地域の欠席状況について空間的な多寡を過大評価せずに評価できるであろう。GIS と併用すれば欠席状況の空間的集積性を視覚化して表示できる。

一般に欠席状況に空間的集積性があることは、集積地域の学童が同一の曝露源に接したか、ヒト-ヒト伝播を来たず感染症のアウトブレイクが起きているかの2つの場合に限られる。図8のように異常の集積性が散在して認められる場合はヒト-ヒト伝播を来たず感染症のアウトブレイクが強く疑われる。この結果はインフルエンザの地域発生と拡大状況を反映したものと考えると矛盾はない。この場合、1月11日の時点でインフルエンザのアウトブレイクの可能性を指摘することが可能である。さらにこの結果で注目すべきは、わずか1週間での欠席異常状況の拡大であろう。健康危機にとって兆候把握のサーベイランスの迅速性がいかに重要であるか、その必要性を如実に示している。

これらの結果から、小学校欠席状況サーベイランスは感染症の地域アウトブレイク把握に有効であり、GIS を応用することによりさらに有利性を発展できると考えられる。

5 今後の展開とそのための検討課題

我々はこれまでの研究を通じて、小学校欠席状況サーベイランスは一定の工夫を加えれば、健康危機管理のためのサーベイランスとして非常に有用かつ実用的なものになりえると考えている。過去の事例でも危機発生初期に当該地域の小学校欠席率が上昇していたものも少なくない⁶⁻⁸⁾。しかしながらその殆どは「事態を把握した後に振り返ってみれば初期に小学校欠席率が上昇していた」というretrospectiveな解析にとどまっている。

一方、学校の欠席状況を危機把握のためにprospectiveに収集しようという試みもある。アメリカNorth Carolina州Mecklenburgでは、社会における疾患アウトブレイクの早期発見、バイオテロリズムの早期把握を目的として地域の救急医療機関や中毒センターの患者来院状況とともに、188の公立学校の欠席率を収集するMedical On-Line Surveillance Tool (MOST) というシステムを構築している⁹⁾。ある地点における事象が、その周辺地区の事象を受けて互いに相関関係が発生していないかについては、クリギングをはじめとする空間的自己相関の分析手法があるが、MOSTでも単に欠席率を表示しているのみで、その欠席状況をふまえた考察方法については未だ確立されていない。

欠席状況のような時間的データを観察しつつ空間的集積性を検出するには、現時点でも続いている集積性をいち早く「生きているクラスター」として同定するprospective

な解析方法が必要となる¹⁰⁾。これについては、近年、集積期間と地域について集積性の有無を検定するflexible spatial scan等の方法¹¹⁾¹²⁾が開発されており、欠席状況サーベイランスへの活用が期待される。

欠席状況サーベイランスで最も難しいのは「異常」を定義するカットオフ値の決め方であろう。カットオフ値を高く設定すればspecificityは上がるだろうが危機の兆候把握は遅れるに違いない。低すぎれば兆候把握は早くなるが、false positiveも多くなる。

今後はカットオフ値と欠席者の症候の集積性とあわせて検討し、さらにGISと併用することで、欠席状況をふまえた考察方法を開発し、本システムの実用化に向けて研究を進めていく予定である。

- 1) 浅見泰司. 地理及び社会状況を加味した地域分析法の開発に関する研究 平成16年度～18年度 厚生労働科学研究総合研究報告書.
- 2) 丹後俊郎, 横山徹爾, 高橋邦彦. 空間疫学への招待. 東京: 朝倉書店; 2007.
- 3) 埼玉県衛生部. クリプトスポリジウムによる集団下痢症-越生町集団下痢症発生事件: 1997.
- 4) 坂保三継. クリプトスポリジウムとその水系感染症 水情報 1996; 16(3): 8-11.
- 5) 張 長春. 空間データ分析. 東京: 古今書店; 2001.
- 6) 堺市学童集団下痢症対策本部編: 堺市学童集団下痢症報告書. 1997.
- 7) 横浜市衛生局感染症・難病対策課. 市立文庫小学校での下痢・嘔吐者の集団発生について (2005/05/26) 記者発表資料 2005.
- 8) 国立感染症研究所. 小学校における風疹集団発生の調査-岡山市. 病原微生物検出情報(IASR)2006; 24: 58-59.
- 9) <http://www.co.mecklenburg.nc.us/Departments/Health+Department/Top+News/Pandemic+Flu.htm>
- 10) Kulldorf M. Prospective time periodic geographical disease surveillance using a scan statistic. Journal of the Royal Statistical Society, Series A, 2001;164:61-72.
- 11) Tango T and Takahashi K. A flexibly shaped spatial scan statistic for detecting clusters. International Journal of Health Geographics, 2005;4:11.
- 12) Takahashi K, Kulldorf M, Tango T and Yih K. A flexible space-time scan statistic for disease outbreak detection and monitoring. Advances in Disease Surveillance.2007;2:70.

A Scan Statistic based on Anscombe's Variance Stabilization Transformation

Kunihiko Takahashi, Toshiro Tango

Department of Technology Assessment and Biostatistics
National Institute of Public Health, Japan

OBJECTIVE

This paper proposes a new scan statistic which detects disease clusters more accurately than that based on the likelihood ratio.

BACKGROUND

The circular spatial scan statistic proposed by Kulldorff and Nagarwalla [1] has been widely used along with SaTScan software for cluster detection. To detect arbitrarily shaped clusters which cannot be detected by the circular scan statistic, Duczmal and Assunção [2] and Tango and Takahashi [3] have proposed different scan statistics. All of these tests are based on maximizing the likelihood ratio statistic $\lambda(Z)$ for each window Z . However, Tango and Takahashi [3] have shown examples in which Duczmal and Assunção's procedure detected quite large and peculiar shaped clusters that had the largest likelihood ratio λ among the three scan statistics applied. It cast a doubt on the validity of the model selection based on maximizing $\lambda(Z)$.

METHODS

One of reasons for detecting undesirable clusters is that $\lambda(Z)$ is derived only from the observed number of cases $n(Z)$ and the expected number $\mu(Z)$ under the null hypothesis H_0 of no clustering. $\lambda(Z)$ ignores the variability of the relative risks of regions included in Z . Then we propose an alternative scan statistic that can take such variability into account.

Assume that, under H_0 , the observed number of cases X_i is a Poisson random variable with expected value μ_i in each region $i = 1, 2, \dots, m$. Then, let us apply Anscombe[4]'s variance stabilization transformation:

$$Y_i = 2\sqrt{X_i + (3/8)} - 2\sqrt{\mu_i + (1/8)},$$

Table. Bivariate power distributions $P(l, s) \times 1000$ of the likelihood ratio statistic λ and the proposed statistic T using the flexible scanning method ($K = 15$) for the hot-spot cluster **A** with $s^* = 3$ regions, where l is the length of significant MLC, s is the number of regions identified out of the assumed true cluster (see details [3]). Total number of cases is set to be 500 in the entire $m = 113$ regions (total number of population is 19,803,618), and relative risk in the hot-spot is set to be 3.0. Nominal α -level is set as 0.05 and 1000 trials are carried out. Lines of $s = 0, 1, 2$ whose all the cells have zero power are not shown. The mark "*" is the powers of accurate detection.

likelihood ratio statistic λ (traditional power = 1000/1000)																
Length l	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Total
Include $s = 3$			*468	179	123	86	78	40	19	5	2	0	0	0	0	1000
proposed statistic T (traditional power = 1000/1000)																
Length l	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Total
Include $s = 3$			*891	73	18	9	5	4	0	0	0	0	0	0	0	1000

where $E(Y_i) = 0$ and $\text{Var}(Y_i) = 1$ under H_0 . For any Z , let $\bar{y}(Z)$ and $\bar{y}(Z^c)$ be the means of y_i within Z and outside, respectively. Then, we propose a new scan statistic T as

$$T = \max_{Z \in \mathcal{Z}} \{ \bar{y}(Z) - \bar{y}(Z^c) / \sqrt{(1/l(Z)) + (1/l(Z^c))} \},$$

where $l(\cdot)$ denotes the number of regions included therein. The window Z^* which attains the maximum T is defined as the most likely cluster (MLC). In the same manner as Kulldorff's scan statistic, Monte Carlo testing is required for the distribution of T under H_0 .

RESULTS

Several scenarios of simulation were used to illustrate the proposed test statistic T with scanning methods of the circular [1] and the flexible [3]. The bivariate power distribution proposed by [3] shows that T has shorter tails and, consequently, better ability of pinpointing the assumed hot-spot cluster compared with the likelihood ratio.

CONCLUSIONS

The proposed scan statistic can detect disease clusters more accurately than that based on the likelihood ratio.

REFERENCES

- [1] Kulldorff M, Nagarwalla N. Spatial disease clusters: detection and inference. *Statistics in Medicine* 1995; **14**:799-810.
- [2] Duczmal L, Assunção R. A simulated annealing strategy for the detection of arbitrarily shaped spatial clusters. *Computational Statistics & Data Analysis* 2004; **45**:269-286.
- [3] Tango T, Takahashi T. A flexibly shaped spatial scan statistic for detecting clusters. *International Journal of Health Geographics* 2005; **4**:11.
- [4] Anscombe F. J. The transformation of Poisson, binomial and negative-binomial data. *Biometrika* 1948; **35**:246-254.

A Spatial Scan Statistic Scanning Only the Regions with Elevated Risk

Toshiro Tango

Department of Technology Assessment and Biostatistics
National Institute of Public Health, Japan

OBJECTIVE

To propose a new spatial scan statistic that has higher ability of pinpointing the true cluster.

BACKGROUND

The spatial scan statistic proposed by Kulldorff[1] has been applied to a wide variety of epidemiological studies and also to disease surveillance for the detection of disease clusters along with SaTScan software. However, it does not seem to be well recognized that maximizing the likelihood ratio tends to detect the most likely cluster much larger than the true cluster by swallowing neighbouring regions with non-elevated risk, Tango and Takahashi[2] have shown such typical examples which casted a doubt on the validity of the model selection based on maximizing the likelihood ratio.

METHODS

To avoid such undesirable phenomena, we propose the following modified likelihood ratio test statistic by taking individual region's risk into account:

$$\lambda_T(\mathbf{Z}) = \sup_{\mathbf{Z} \in \mathcal{Z}} \left(\frac{n(\mathbf{Z})}{\xi(\mathbf{Z})} \right)^{n(\mathbf{Z})} \left(\frac{n - n(\mathbf{Z})}{n - \xi(\mathbf{Z})} \right)^{n - n(\mathbf{Z})} \cdot I \left(\frac{n(\mathbf{Z})}{\xi(\mathbf{Z})} > \frac{n - n(\mathbf{Z})}{n - \xi(\mathbf{Z})} \right) \prod_{i \in \mathbf{Z}} I(p_i < \alpha_1)$$

where where $n()$ and $\xi()$ denote the random number of cases and the null expected number of cases within the specified window \mathbf{Z} , respectively, p_i is the p -value for the i -th region's risk and α_1 is the pre-specified individual region's significance level. Individual region's significance level α_1 is set equal to the significance level α_0 for detecting MLC.

RESULTS

The proposed circular spatial scan statistic was, via Monte Carlo simulations, shown to have quite high ability of pinpointing the true circular cluster assumed in the simulation (ex., see Table 1).

CONCLUSIONS

Our result suggests that, if other spatial scan statistics for detecting arbitrary shaped clusters adopt the proposed likelihood ratio test statistic, their performance is expected to be improved.

REFERENCES

- [1] Kulldorff M. A spatial scan statistic. *Communications in Statistics: Theory and Methods* 1997; **26**, 1481-1496.
- [2] Tango T, Takahashi T, A flexibly shaped spatial scan statistic for detecting clusters. *International Journal of Health Geographics* 2005; **4**:11.

Table 1. Simulated bivariate power distributions $P(l, s) \times 1000$ of Kulldorff's circular spatial scan statistic and the proposed circular spatial scan statistic for the hot-spot circular cluster $\mathbf{A} = \{14, 15, 20\}$ with length $l = 3$ and relative risk $\theta = 3.0$ that was assumed in the areas of Tokyo Metropolis and Kanagawa prefecture in Japan (113 regions). Total number of cases was set as 235. Two kinds of α -levels were set as $\alpha_0 = \alpha_1 = 0.05$ and 1000 trials were carried out. The power of accurate detection ($l = s = 3$) was indicated by "*".

Kulldorff's circular scan					Proposed circular scan				
Length l	Include s hot-spot regions				Length l	Include s hot-spot regions			
	0	1	2	3		0	1	2	3
1	1	0			1	1	0		
2	0	0	0		2	1	0	0	
3	0	0	0	*672	3	0	0	0	*845
4	0	0	0	147	4	0	0	0	38
5	0	0	0	38					
6	0	0	0	26					
7-10	0	0	0	15					
11-20	0	0	0	50					
21-47	0	0	0	14					
Total	1	0	0	962	Total	2	0	0	883
	usual power=0.963					usual power=0.885			

A Comparison of SaTScan and FleXScan for Outbreak Detection and Monitoring

– An Application to Data of Absentees in Primary School –

Kunihiko Takahashi, Toshiro Tango

*Department of Technology Assessment and Biostatistics
National Institute of Public Health, Japan*

OBJECTIVE

This paper describes a comparison between two statistics – SaTScan and FleXScan, applying to a data of absentees in primary school in Japan.

BACKGROUND

Early detection of disease outbreaks enables public health officials to implement disease control and prevention measures at the earliest possible time. A time periodic geographical disease surveillance system based on a cylindrical space-time scan statistic proposed Kulldorff(2001) has been used extensively for disease surveillance along with the SaTScan software. Recently, Takahashi *et al.*(2008) proposed a flexibly shaped space-time scan statistic (“FleXScan” hereafter) to detect arbitrarily shaped disease outbreaks, and the statistic is well suited for detecting and monitoring disease outbreaks in irregularly shaped areas. They have suggested that we may use both statistics for disease outbreak detection, but for different purposes. Then it is important to know their performances.

In many cities such as New York City, Boston, real-time, geographic, early outbreak detection system have been implemented, and the SaTScan is used for the analysis. On the other hand, in Japan, there is not yet such a system which includes statistical analysis using cluster detection tests, and its implementation is demanded now.

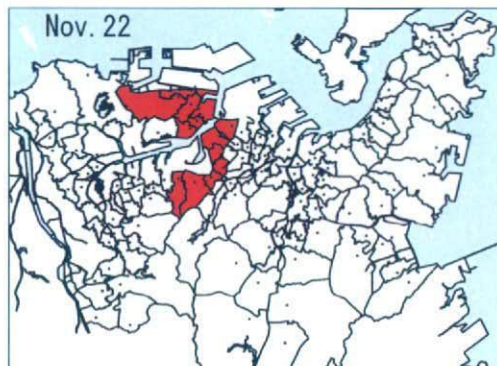
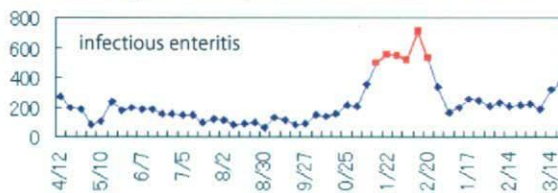
METHODS

We applied a prospective space-time analysis using the SaTScan and the FleXScan to a surveillance data of number of absentees in primary school in Kita-kyushu, Japan. The data came from weekly surveillance of the absentees on every Wednesday in 132 primary schools in Kita-kyushu, during April 2006 to March 2007. Total number of school children was 52,189. We set $K = 15$ as the maximum length of the geographical window, and the maximum temporal length to be $T = 2$ weeks. The expected number (baseline) of absentees were calculated as the averages during April to September for each school, and the analysis started from October 11. The number of replications for the Monte Carlo procedure was set to $B = 999$ and its significance level was set to $p < 0.02$ corresponding to the recurrence interval of 50 weeks, i.e., 1 year.

RESULTS

For the analysis during October 11 to November 15, both tests did not detect any signals. On November 22, both tests detected a cluster of 2 weeks (i.e., from November 15 to 22) with $p = 0.001$. The SaTScan detected a circular cluster of 11 schools, on the other hand, the FleXScan detected a non-circular cluster of 13 schools. The secondary cluster was not detected on the day. From November 22 to December 13, significant clusters were detected by both tests. Their results were similar but detected areas were slightly different. On December 20, both tests detected the same cluster, which was a small cluster of only 2 schools.

At the same periods, an outbreak of infectious enteritis has been observed in the area from a sentinel medical system, which consists of some hospitals and clinics. It seems that the outbreak was detected by our analysis of absentees in school.



CONCLUSIONS

The FleXScan is well suited for detecting and monitoring disease outbreaks. And the data of absentees in primary school is useful for detecting outbreaks.

REFERENCES

Takahashi K, Kulldorff M, Tango T, Yih K. A flexibly shaped space-time scan statistic for disease outbreak detection and monitoring. *International Journal of Health Geographics* 2008; 7:14.

A New Space-Time Scan Statistic for Timely Outbreak Detection Taking Overdispersion Into Account

Toshiro Tango and Kunihiko Takahashi

Department of Technology Assessment and Biostatistics
National Institute of Public Health, Japan

OBJECTIVE

To propose a new space-time scan statistic taking overdispersion into account for accurate and timely detection of disease outbreaks.

BACKGROUND

As a major analytical method for outbreak detection, a software SaTScan of Kulldorff's space-time scan statistic (2001) has been implemented in many syndromic surveillance systems. Since it is based on circular windows in space, it cannot detect non-circular areas. Takahashi *et al.* (2008) proposed a flexible space-time scan statistic with capability of detecting non-circular areas, which is implemented in a software FlexScan. It seems to us, however, that the detection of *Most Likely Cluster* defined in these space-time scan statistics are not always appropriate for the purpose of timely detection of a localized disease outbreak. Furthermore, these scan statistics fail to take overdispersion into account and thus the rate of erroneous false alarm will increase.

METHODS

We assume that, under the null hypothesis of no outbreaks, the number of cases $N_i(t)$ in the region i ($i = 1, \dots, m$) at time t is a Negative Binomial random variable $NB(\mu_i(t), \phi_i(t))$ where the expected baseline covariate-adjusted values $\mu_i(t)$ and the over-dispersion parameter $\phi_i(t)$ can be estimated from the baseline data. Since we are only interested in detecting outbreaks that are alive (active) at the current time t_P , we only consider 'alive' outbreaks that are present in the following T time intervals: $I_u = [t_P - u, t_P]$, $u = 0, \dots, T - 1$ where T is a pre-specified maximum temporal length of the outbreak. In this situation, we shall propose the following *outbreak model*:

$$N_i(t) \sim \begin{cases} NB(\theta_{in}(t)\mu_i(t), \phi_i(t)), & \text{if } i \in Z \text{ and } t \in I_u \\ NB(\theta_{out}(t)\mu_i(t), \phi_i(t)), & \text{otherwise} \end{cases}$$

where it is assumed that

$$\begin{aligned} H_0 &: \theta_{in}(t) = \theta_{out}(t) = 1 \\ H_1 &: \theta_{in}(t) > \theta_{out}(t) = 1 \end{aligned}$$

where the relative risk $\theta_{in}(t)$ could be exponentially increasing in time under H_1 : $\theta_{in}(t) = \exp\left(\beta \frac{t-t_p+u+1}{u+1}\right)$, for $t_p - u \leq t \leq t_p$. The spatial window Z , time interval I_u and the slope $\hat{\beta}$ for

which the likelihood ratio is maximized identifies the *Most Likely Outbreak*.

APPLICATION

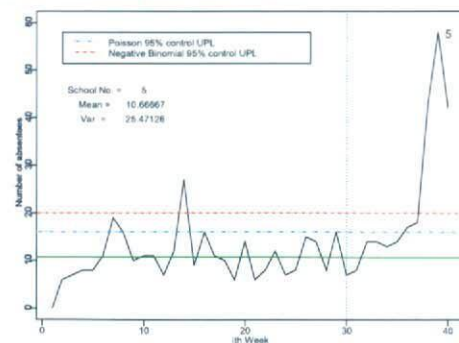
We applied two kinds of the proposed space-time scan statistic, one with Kulldorff's circular window and the other with Tango and Takahashi's non-circular window, to data from weekly surveillance of the absentees in 132 primary school in Kita kyushu shi, during April 2006 to March 2007, Japan. The primary purpose of the application is (a) the timely detection of the *Most Likely Outbreak* of the absentee during the influenza season (December to February) and (b) the comparison with the results from existing two space-time scans, SaTScan and FlexScan. Total number of school childrens were 52,189. Baseline period was considered to be from April to September 2006 and the analysis period was from October to February 2007. Maximum temporal length was set as $T = 2$.

RESULTS AND CONCLUSIONS

Detailed results will be presented at the conference. These results suggest that the proposed space-time scan statistic is better suited for accurate and timely detection of disease outbreaks compared with existing space-time scan statistics.

REFERENCES

- [1] Kulldorff M. Prospective time periodic geographical disease surveillance using a scan statistic. *Journal of the Royal Statistical Society, Series A*, 2001; **164**: 61-72.
- [2] Takahashi K, Kulldorff M, Tango T and Yih K. A flexibly shaped space-time scan statistic for disease outbreak detection and monitoring. *International Journal of Health Geographics* 2008; **7**:14.



Cluster Detection Comparison in Syndromic Surveillance

C. Goranson^{1,3}, Takahashi K.², Tango T.², A. Cajigal¹, M. Paladini¹, E. L. Murray¹,
T. Nguyen¹, K. Konty¹, F. Hardisty³

New York City Department of Health and Mental Hygiene(NYCDOHMH), NY, USA¹, National Institute of Public Health, Wako-shi, Saitama, Japan², The Pennsylvania State University³, PA, USA

OBJECTIVE

To use the New York City Department of Health and Mental Hygiene's (NYC DOHMH) emergency department (ED) syndromic surveillance data to evaluate FleXScan's flexible scan statistic and compare it to results from the SaTScan circular scan. A second objective is to improve cluster detection in by improving geographic characteristics of the input files.

BACKGROUND

The NYC DOHMH collects data daily from 50 of 61 (82%) EDs in NYC representing 94% of all ED visits (avg daily visits ~10,000). The information collected includes the date and time of visit, age, sex, home zip code and chief complaint of each patient. Observations are assigned to syndromes based on the chief complaint field and are analyzed using SaTScan to identify statistically significant clusters of syndromes at the zip code and hospital level [1]. SaTScan employs a circular spatial scan statistic and clusters that are not circular in nature may be more difficult to detect. FlexScan employs a flexible scan statistic using an adjacency matrix design [2][3].

METHODS

Counts of syndrome visits were aggregated at the zip code level for 2005. FleXScan's flexible scan and SaTScan's circular scan were analyzed by comparing the most likely cluster (primary cluster) identified; the secondary clusters identified; location and area of identified cluster; P-value and relative risk. Both projected and unprojected coordinate systems were used to identify sensitivity in clusters to changes in measurement and coordinate systems. Improving the FleXScan matrix file provided a method for capturing area connectivity where bridges, tunnels, or subway lines existed between them. This was not possible to do in SaTScan. ZIP code area centroids were weighted to reflect the underlying population distribution of the areas. Both FleXScan and SaTScan were run again using the reweighted centroids.

RESULTS

FleXScan and SaTScan both detected similar, overlapping areas in three of the time periods investigated. Non-circular clusters with a high relative risk were detected by FleXScan's flexible scan, but this was not detected by SaTScan (Figure 1). However, known clusters were detected at a more significant p-value by SaTScan than FleXScan ($p=0.002$ vs. $p=0.179$). Weighting ZIP code centroids based on

population and improving the connectivity matrix changed results; over a one week period p-values increased 50% of the time, decreased 36% of the time, and stayed the same 14% of the time when weighted centroids were employed. The differences were most prominent where unweighted centroids had not been representative of underlying population distributions in the areas.

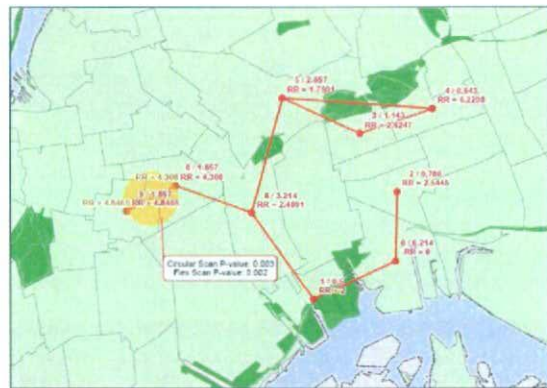


Figure 1 – Circular scan significant cluster (orange) and flexible scan significant cluster (red). Relative risk is also shown.

CONCLUSIONS

Flexible and circular scans are complementary tools that each have their strengths and should be used together. SaTScan provides a useful method for detecting clusters more circular in nature; FleXScan is one approach to better identify non-circular clusters by employing a flexible spatial scan statistic. Used together they may provide the best alternative to characterizing an outbreak.

Updating and improving the quality of area centroid files through population weighting, and the creation of accurate matrix files may improve the accuracy of syndromic surveillance. Additional analyses should be conducted to provide a statistical basis to comparisons between the methods. A second comparison between other confirmed outbreaks will attempt to identify the sensitivity and specificity of each under different circumstances.

REFERENCES

- [1] Heffernan R, Mostashari F, Das D, Karpati A, Kulldorff M, Weiss D, Syndromic surveillance in public health practice, New York City. *Emerging Infectious Diseases*. 2004 May;10(5):858-64.

[2] Tango T, Takahashi K, A flexibly shaped spatial scan statistic for detecting clusters. *International Journal of Health Geographics*. 2005 May 18;4(1):11.

[3] Takahashi K, Kulldorff M, Tango T and Yih K. A flexibly shaped space-time scan statistic for disease outbreak detection and monitoring. *International Journal of Health Geographics* 2008, 7:14

This project funded under NSF / EAPSI Award # 0812980.