

科学研究費助成事業 研究成果報告書

平成 26 年 6 月 20 日現在

機関番号：32623

研究種目：基盤研究(B)

研究期間：2011～2013

課題番号：23300107

研究課題名(和文)健康危機管理のための突発的生起事象を検出する統計モデル

研究課題名(英文)Statistical models for detecting emerging outbreaks for health risk monitoring

研究代表者

丹後 俊郎 (TANGO, Toshiro)

昭和女子大学・生活機構研究科・客員教授

研究者番号：70124477

交付決定額(研究期間全体)：(直接経費) 13,600,000円、(間接経費) 4,080,000円

研究成果の概要(和文)：突発的生起事象(クラスター)の検出には、Kulldorff(1997)によって提案された空間スキャン統計量が広く使用されている。しかし、不規則な形状のクラスターを同定できないという問題点があった。一方、Tango-Takahashi(2005)が提案した方法は任意の形状のクラスターを柔軟に検出できるが、計算時間が膨大で小さなクラスターしか同定できない制限があった。本研究では、これらの問題点を改善するための研究を行った。その結果、クラスターの形状・サイズに制限がなく、計算時間も大幅に短縮でき、クラスターの相対危険度が増大するにしたがい検出力が大きくなり、実用上優れた性能を示すことできた。

研究成果の概要(英文)：The circular spatial scan statistic proposed by Kulldorff (1997) has been utilized to detect emerging outbreaks or clusters in many syndromic surveillance systems in USA. However, it cannot detect noncircular, irregularly shaped clusters. The flexible spatial scan statistic proposed by Tango and Takahashi (2005) has also been used for detecting irregularly shaped clusters. However, this method sets a feasible limitation of a maximum of 30 nearest neighbors for searching candidate clusters because of heavy computational load. In this study, we show a flexible spatial scan statistic implemented with a restricted likelihood ratio proposed by Tango (2008) to (1) eliminate the limitation of 30 nearest neighbors and (2) to have surprisingly much less computational time than the original flexible spatial scan statistic. As a side effect, it is shown to be able to detect clusters with any shape reasonably well as the relative risk of the cluster becomes large via Monte Carlo simulation.

研究分野：総合領域

科研費の分科・細目：情報学、統計科学

キーワード：疾病集積性 尤度比検定 ポアソン分布 バイオテロリズム

1. 研究開始当初の背景

2001年9月11日ニューヨーク市の世界貿易センターを襲った史上最大の国際テロ、それに引き続いて10月に発生した、「炭素菌を含んだ白い粉」が入った郵便物事件は世界中に衝撃を与えた。これらの事件をきっかけとして、バイオテロリズムやSARS、新型インフルエンザの勃発など、ヒトの健康を脅かす事件の勃発を疑わせる症候(syndrome)を早期に発見し事件の拡大を未然に防ぐための症候サーベイランス(syndromic surveillance)と早期発見のための方法論に関する研究が活発になってきている。そこでは対象とする地域から関連する症候の発生状況を定期的に収集し監視できる情報ネットワークシステムの構築と健康被害の勃発の兆候を示す何らかの突発的な事象が発生した場合に、それがいつ、どこで発生したのかを適切に検出できる統計的方法の開発が重要となる。

米国の多くの都市で運用されている症候サーベイランスに組み込まれている統計的方法はSaTScanというソフトウェアの名称で知られている疾病の集積地域を検出するために開発されたKulldorffのスキャン統計量である。しかし、この方法には幾つかの問題点が指摘されていた。我々はSaTScanの問題点の一つを改良し、検出できる集積地域の大きさに制限があるものの任意の形状の集積地域を柔軟に検出できるスキャン統計量を提案し(Tango and Takahashi, 2005)、そのソフトウェアFlexScan(2007)を公開した。その後、Kulldorffとの共同研究により一部の機能を時空間へ拡張したモデルを米国ニューヨーク市での症候サーベイランスデータに適用し、SaTScanでは検出できない時空間の集積領域を検出できる理論と具体的実現方法を開発した(Takahashi et al., 2008; Tango et al., 2010)。これらの方法は、既存の方法論より優れている、という点で、世界からある程度の注目を浴び、世界各地の疫学研究、症候サーベイランスで使用されてきた。Springer社からは、これまでに著者らが開発してきた方法を含めた「疾病の集積性の検出」に関するテキストの出版を依頼され、2010年3月に出版した。

しかし、SaTScan, FlexScanを含めた現在の方法は、1) 集積地域がある程度以上大きいと実用上検出不可能となる、2) 時間空間的に成長していく勃発パターンの検出パワーが小さい、などの問題点が内在しており、更なる改良、あるいは新たな方法、の開発が課題であった。

2. 研究の目的

本研究では、特に、以下の2点に焦点を当てて研究を遂行する。

(1) 集積地域の大きさ(クラスターサイズ)に制限のない、任意の形状の集積地域を検出できる方法の開発: 現在のflexible scanが検

出できる実用上の最大クラスターサイズ(地域の数)は20-30程度で、それを超えると計算時間が膨大となり、短時間での解析が不可能である問題点が指摘されている。本研究では、この制限を取り外し、サイズに関係なく検出できる理論武装とその具体的アルゴリズムの開発を目的とする。

(2) 時空間的に成長している突発的事象の検出できる方法の開発: これまでの時空間的集積地域は時間的な成長は表現できても空間的な広がりを検出する性能は低かった。本研究では、「徐々に時間的、空間的に拡大する」突発事象のクラスターの検出力を改善できるoutbreak modelを内蔵したモデルの開発を目的とする。

3. 研究の方法

(1) 研究全体について

研究期間全体を通じ、理論的な統計量の検出と現実を反映した様々な勃発のシナリオに基づく理論的なシミュレーションを交互に繰り返し行いながら研究を遂行する。その際、シミュレーションで使用する時空間領域として東京都・神奈川県の113市区町村の実際の人口構成を利用する。各地域の地理的位置は、人口中心点を利用する。

(2) クラスターサイズの理論的研究

本研究に関しては、最近、Kulldorffのスキャン統計量の欠点の一つである「真の集積地域より数倍から数十倍大きい集積領域を検出してしまう」性質の改善を検討するためにscan統計量のエンジン部分である尤度比検定統計量を改良して、リスクの小さい領域を呑み込まないような制約付尤度比検定統計量を検討したところ、その良い意味での副作用として「計算時間が飛躍的に減少することが判明した」(Tango, 2009)。そこで、まずは、同様な方法がflexible scanにも適用可能かどうかをまず理論的に検討する研究から着手したい。それが可能であれば、次はそのアルゴリズムの検討に入りたい。これが実現できれば、クラスターサイズに関わらず計算時間と検出精度がともに大幅に向上されることが期待され、その可能性は高いと考えている。なお、本研究で検討・提案するスキャン統計量の評価には、下記に示す大規模なMonte Carloシミュレーションによる評価が必要になる。その際、SaTScanと異なり任意の形状の領域をスキャンするために相当な計算時間を要するが、効率的な計算プログラムの作成(S-plus, C++を予定)と計算速度の速いマシンを利用することで効率的に研究が遂行できると考えている。前者のプログラム作成の補助に研究補助者を予定し、後者のために計算速度の速いマシンの購入を予定している。

(3) Monte Carloシミュレーションによる検出力の検討

以下に記述する任意の形状をしたクラスターを人口的に作成し、それに対して、三つ

の方法、

- 1) Kulldorff's circular spatial scan statistic,
 - 2) Tango-Takahashi's flexible spatial scan statistic,
 - 3) 本研究で提案する制限付き尤度比を内蔵した flexible spatial scan statistic
- それぞれの検出力を Monte Carlo シミュレーションにより比較する。

研究対象全地域として、東京・神奈川
の地域(113 市区町村)を選択する。仮定する
クラスターの最大サイズは $K=50$ とするが、
Tango-Takahashi 法については、計算時間の
制限のため $K=20$ と設定する。

クラスターの形状については、次の2種
類を仮定した

A. 十字にクロスした形のクラスターA。
東京・神奈川区域の市区町村番号で表現す
ると $A=(25, 29, 30, 31, 32, 33, 38, 39, 40,$
 $41, 47, 49, 53, 72, 73, 74, 75, 76, 86, 93,$
 $92, 97)$ となり、この場合の真のクラス
ターサイズは $s^* = 22$ となる

B. 環状のクラスターB: $= (32, 56, 57, 58,$
 $59, 62, 63, 73, 74, 75, 76, 78, 81, 82, 86,$
 $93, 92)$ であり、この場合の真のクラス
ターサイズは $s^* = 17$ となる。

実際のシミュレーションでは、これらの2
種類の形状について大きさ(sample size)の
異なるクラスターを5個作成する。この2種
類のクラスターの形状を選択した主な理由
は、既存の方法である Kulldorff の方法と
Tango-Takahashi の方法では、これらの形状
のクラスターは正確に検出することは100%
できないからであり、提案する方法がどの程
度検出できるかに興味があるからである。

シミュレーションデータの発生法
シミュレーションは次に示すステップにし
たがって行う。

Step 1. ランダムな観測ケース数の発生

それぞれの地域 i ($=1, \dots, m=113$) に、
ポアソン分布(期待値 E_i)にしたがう非ケ
ースのランダムな観測数 n_{0i} を発生させる。
ここで各地域の人口を w_i と置くと、

$$E_i = n_0 w_i / w_i$$

で計算される。ここで、 n_0 はクラスターが
存在しないという仮定の下で、事前に指定さ
れる期待ケース数であり、

$$n_0 = E_i$$

である。 n_0 の値としては、

$$100, 200, \dots, 500$$

と5通りを設定する。

Step 2. 観測ケース数の発生:

事前に指定されたクラスターに属する地域
については、次式で定義される「クラスター
の強さ」を表現する Q

$Q = \Pr\{X \geq n_{1i} \mid X \sim \text{Poisson}(E_i)\}$
に相当するクラスターによる観測ケース数
 n_{1i} を発生させる。 Q の値としては、

$$Q = 0.05, 0.01, 0.001$$

の3通りを設定する。

Step 3. 各地域の観測ケース数 n_i の計算

Step 1 と Step 2 で観測されたケース数の
合計を各地域で観測されるケース数とする。
つまり、クラスターに属さない地域は $n_i=n_{0i}$
であり、クラスターに属する地域は

$$n_i = n_{0i} + n_{1i}$$

と計算する。

Step 4. Step 1-Step 3 を1000回繰り返す。

上記に記述したように、クラスターの形状
が2種類、サイズが5種類、 n_0 の値が3種類、
 Q の値が3種類、合計で、 $2 \times 5 \times 3 \times 3 = 90$
通りのクラスタリングシナリオに基づいて、
それぞれの spatial scan statistic の性能
を比較する。これらのシミュレーションの結果
は Tango and Takahashi (2005) で提案され
た2変量検出力分布 $P(I, s)$ (Bivariate power
distribution)

$P(I, s) =$ 検出されたクラスターの大きさ
(地域の数)が I で、その中で真のクラス
ターに含まれる地域数 s ($s \leq s^*$) となる
確率

の形でまとめる。したがって、2次元座標で
 $(I, s) = (s^*, s^*)$ (仮定した真のクラス
ターの大きさ)の周辺で分布していれば、その方
法の検出力が高いということになる。

(4) 時空間的に成長している突発的事象の
検出できる方法の開発研究

我々が提案した FlexScan は、空間集積性
の検出においては、広く利用されている
SaTScan よりも任意の形状のエリア(集積
地)が同定できる点で利点がある。この利点
を生かして、時間軸の加わった時空間領域
においても精度よく同定できる時空間領域
のスクリーン法の検討を行う。時間的変動のモ
デルとしては、ある地域(多角形)を固定したと
き、底面×時間(長さ)からなるウィンドウを
利用して、稀少生起事象の勃発の兆候が現時
点で起きたのか、一日前から起きているの
か、あるいは二日前から起きているのか、
など現在から過去にさかのぼって、最も可
能性の高い時空間領域を同定する方法が基本
と考えている。ただ、これでは、空間的な広
がりを検出するには不十分なので時間の増
加にともない近隣地域にも拡大できるモデ
ルを導入することで対処する。

最終年度には、これらの成果を取り入れた
ソフトウェア(仮称: Disease Mapping
System)の試作システムの開発を行う予定で
ある。

4. 研究成果

(1) クラスターサイズの理論的研究

まず、突発事象が起きた地域の大きさ(ク
ラスターサイズ)に関わらず短時間で、かつ、
任意の形状の集積地域を正しく検出できる、
真の flexible scan 統計量の理論的かつシミ

ュレーションによる性能評価の研究を行った。現在の flexible scan 統計量は検出できる実用上の最大クラスターサイズは、地域の構造により 20 から 30 程度で、それを超えると計算時間が膨大となり、大きなクラスターの存在が疑われる状況下では事実上適用が不可能となる問題が指摘されてきた。しかし、最近の研究で、circular scan 統計量の欠点の一つである「真の集積地域より数倍から数十倍大きい集積領域を検出してしまう」性質の改善を検討するために scan 統計量のエンジン部分である尤度比検定統計量を改良して、リスクの小さい領域を呑み込まないような制約付き尤度比検定統計量を検討したところ、その良い意味での副作用として「計算時間が飛躍的に減少する」ことが判明した。

この方法論を flexible scan に適用して、次に示す、大規模な Monte Carlo シミュレーション研究を実施したところ、クラスターサイズに関わらず計算時間と検出精度がともに大幅に向上され、検出性能が従来の方法より優れていることを明らかにできた。

(2) シミュレーションによる検出力の検討

推定された 2 変量検出力の分布の一例として、十字にクロスした形のクラスター A (クラスターサイズ=22) を検出する場合の提案する方法の結果の一つを図 1、Kulldorff の circular spatial scan statistic の結果の一つを図 2 に示した。この図からも、本研究で提案した制限付き尤度比を内蔵する flexible spatial scan statistic が kulldorff の方法に比較して、明らかに、目的とするクラスターを検出できていることがわかる。シミュレーション評価の結果、提案する方法は、他の方法に比較して、期待ケース数が大きくなるにつれ、また、クラスターの強さを表現する指標 Q の値が小さくなるにつれて、より正確にクラスターを検出していることが確認された。これらの結果の特徴は、次の 3 点にまとめられる。

(A) 現在の flexiblespatial scan に存在した検出可能なクラスターサイズの制限 (現実的な計算時間を考慮した最大値 30) がない、

(B) 計算時間が少なくとも、既存の方法の 1000 分の 1 程度に短縮できている、更に、

(C) Monte Carlo シミュレーションにより、クラスターの相対危険度が大きくなるにつれて、どのような形状をもつクラスターでも、その検出確率が 1.0 に近くなることが示された。

これらの結果の一部の研究成果は国際的に著名な雑誌 Statistics in Medicine に掲載された。

(3) 時空間的に成長する突発的事象の検出法

本研究では、インフルエンザの感染のように、突発的事象が集積している地域が時間的に成長・拡大する様々なモデル化を検討した。中でも、現実的でありかつ単純なモデルとして、単位時間当たりの突発的事象の広がり

flexScan with restricted LRF (k=113, r=0.2)																									
s	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	total	
1	0	0																						0	
2		0																						0	
3			0																					0	
4				0																				0	
5					0																			0	
6						0																		0	
7							0																	2	
8								0																0	
9									0															0	
10										0														0	
11											0													0	
12												0												0	
13													0											1	
14														0										0	
15															0									0	
16																0								0	
17																	0							0	
18																		0						0	
19																			0					0	
20																				0				0	
21																					0			0	
22																						0	59	59	
23																						1	178	179	
24																							258	258	
25																							211	211	
26																							133	133	
27																							87	87	
28																							35	35	
29																							19	19	
30																							12	12	
31																							3	3	
32																							1	1	
total	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	998	1000

図 1 . 制限付き尤度比を内蔵する Flexible scan statistic の 2 変量検出力。この例では、99%はクラスター A を同定できている (s=22 のラインに集中) 状況が理解できる。

Scan (k=50)																								
s	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	total
1	0	0																						0
2		0																						0
3			0																					0
4				0																				0
5					0																			0
6						0																		0
7							0																	0
8								0																0
9									0															0
10										0														0
11											0													0
12												0												0
13													0											0
14														0										0
15															0									0
16																0								0
17																	0							0
18																		0						0
19																			0					0
20																				0				0
21																					0			0
22																						0	203	203
23																							8	8
24																							3	3
25																							4	4
26																							1	1
27																							81	81
28																							3	3
29																							7	7
30																							3	3
31																							2	2
32																							4	4
33																							7	7
34																							1	1
35																							10	10
36																							3	3
37																							4	4
38																							2	2
39																							2	2
40																							2	2
41																							133	133
42																							1	1
43																							72	72
44																							35	35
45																							7	7
46																							20	20
47																							8	8
48																							34	34
49																							13	13
50																							3	3
total	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1000

図 2 . Kulldorff の circular spatial scan Statistic の 2 変量検出力。この例ではクラスター A をきちんと同定できている場合は全くなく (s=22 のラインに一つもない) クラスター A の一部を同定できているにすぎないことがわかる

隣接地域のみ、かつ、突発的事象の終息の時間的変動も隣接地域から終息していく、という制限付き併合法・終息法を導入した。この方法をシミュレーションで検討した結果、当初の研究計画で想定していた広がり方をある程度は捉えることは可能となったが、広がり方を柔軟に同定するまでには至らなかった

た。この問題に関しては、今回の研究期間内には残念ながら、十分な成果を出せなかったが、今後の研究課題としたい。

一方で、これまでの研究成果を反映した方法を組み込んだソフトウェアの試作システムを開発した。このシステムを発展させることで健康危機管理対策を推進するための有用なツールとして期待される。

5. 主な発表論文等

〔雑誌論文〕(計 8 件)

Hida, E. and Tango T. Three-arm non-inferiority trials with a prespecified margin for inference of the difference in the proportions of binary endpoints. *Journal of Biopharmaceutical Statistics*, 2013; **23**: 774-789, 査読有.

Adachi M, Yamaoka K, Watanabe M, Nishikawa M, Kobayashi I, Hida E, Tango T. Effects of lifestyle education program for type 2 diabetes patients in clinics: a cluster randomized controlled trial. *BMC Public Health*, 2013, **13**: 467, 査読有.

Tango T. and Takahashi K. A flexible spatial scan statistic with a restricted likelihood. *Statistics in Medicine* 2012, **31**: 4207-4218, 査読有.

Yamaoka K. and Tango T. Effects of lifestyle modification on metabolic syndrome: a systematic review and meta-analysis. *BMC Medicine*, 2012; **10**: 138, 査読有.

Tango T., Takahashi K., Kohriyama K. A space-time scan statistic for detecting emerging outbreaks. *Biometrics* 2011; **67**: 106-115, 査読有.

Hida, E. and Tango T. On the three-arm non-inferiority trial including a placebo with a prespecified margin. *Statistics in Medicine* 2011; **30**: 224-231, 査読有.

Nishiyama T, Takahashi K., Tango T., Pinto D, Scherer SW, Takami S, Kishino H. A scan statistic to extract causal gene clusters from case-control genome-wide rare CNV data. *BMC Bioinformatics* 2011; **12**: 205, 査読有.

Saeki H. and Tango T. Non-inferiority test and confidence interval for the difference in correlated proportions in diagnostic procedures based on multiple raters. *Statistics in Medicine* 2011; **30**: 3313-3327, 査読有.

〔学会発表〕(計 7 件)

丹後俊郎、山岡和枝、高橋邦彦. 空間疫学の研究、教育、実務のための統計手法と統計ソフトウェアの開発、第 24 日日本疫学会学術総会、2014 年 1 月 24-25 日(宮城県).

Tango T. and Takahashi K. A flexible scan statistic with a restricted likelihood ratio

for detecting disease clusters. 2013 年度統計関連学会連合大会、2013 年 9 月 8-11 日(大阪府).

佐伯浩之、丹後俊郎、汪金芳、複数の読影者による対応のあるクラスターデータの割合の差の信頼区間、2013 年度統計関連学会連合大会、2013 年 9 月 8-11 日(大阪府).
安達美砂、山岡和枝、渡辺満利子、丹後俊郎. クラスター無作為化比較試験による 2 型糖尿病のための生活習慣改善プログラムの介入効果の評価、第 23 回日本疫学会学術集会、2013 年 1 月 24-26 日(大阪府).

Matsuda A., Matsuda T., Nishimoto H., Yamaoka K. and Tango T. Effectiveness of social supports on quality of life in breast cancer patients: a meta-analysis of randomized controlled trials. The 19th annual conference of the International Society for Quality of Life Research, October 24-27, 2012, Budapest, Hungary.

Tango T. and Takahashi K. FleXScan with a restricted likelihood ratio: a spatial scan statistic for irregularly shaped clusters. The 32nd annual conference of the international Society of Clinical Biostatistics. August 23rd 2011, Ottawa, Canada.

丹後俊郎、高橋邦彦. On a flexible scan statistic with a restricted likelihood ratio. 日本計量生物学会年会、2011 年 6 月 2 日(大阪府)講演予稿集

〔図書〕(計 3 件)

古川俊之監修、丹後俊郎著：医学への統計学、第 3 版、朝倉書店、2013、304 頁

丹後俊郎、山岡和枝、高木晴良著：新版ロジスティック回帰分析、朝倉書店、2013、296 頁

Saeki H. and Tango T. Statistical inference for non-inferiority of the difference in correlated proportions in diagnostics procedures based on multiple raters. In Kees van Montfort et al. (Eds.) *Developments in Statistical Evaluation of Clinical Trials*, Springer, Chapter 7: 119-137, 2013.

6. 研究組織

(1) 研究代表者

丹後 俊郎 (TANGO Toshiro)

昭和女子大学・生活機構研究科・客員教授
研究者番号：70124477

(2) 研究分担者：なし

(3) 連携研究者

高橋 邦彦 (TAKAHASHI Kunihiko)
名古屋大学・医学系研究科・准教授
研究者番号：50323259