

**科学研究費助成事業 研究成果報告書**

平成 29 年 6 月 19 日現在

機関番号：32643

研究種目：基盤研究(B) (一般)

研究期間：2014～2016

課題番号：26280008

研究課題名(和文) ヒトの健康を脅かす稀少生起事象の集積性を検出する統計モデルに関する研究

研究課題名(英文) Statistical models for detecting clusters of rare events

研究代表者

丹後 俊郎 (Tango, Toshiro)

帝京大学・大学院公衆衛生学研究科・教授

研究者番号：70124477

交付決定額(研究期間全体)：(直接経費) 9,200,000円

研究成果の概要(和文)：本研究では、これまで申請者らが開発してきた人の健康を脅かす疾病集積性を検出するための統計モデルに関して、主に二つ、(1) 真の疾病の空間的集積性の完全同定確率の向上と、(2) 時空間的に成長する突発的事象をより良く検出できるスキャン統計量の更なる改良、に関する研究を行った。真の疾病集積エリアの同定確率が大きくなるような統計量の改良を試み、新たに発生しつつある局在化した疾病の勃発の時空間的な広がりを精度よく検出できる時空間スキャン統計量を検討した結果、実用上優れた性能を示すことができた。これらの性質を2006年に実施された北九州市の131の小学校の欠席者数のサーベイランスデータを利用して実証した。

研究成果の概要(英文)：In this study, we improved statistical models proposed by Tango et al. for detecting disease clusters and emerging disease outbreaks. Compared with other existing statistical models, Tango et al.'s models have shown a lot of good performance in actual epidemiological cluster detection but it has been said in the literature that (1) their models have low detection probability of the true cluster and (2) they have difficulty in correctly detecting the growth of space-time disease outbreaks. The first problem was improved by modifying a restricted likelihood ratio. To overcome the second problem we considered a improved space-time scan statistic for detecting localized emerging disease outbreaks which can detect outbreaks more timely and correctly. These properties are illustrated with data from the weekly surveillance of the number of absentees in primary schools in Kitakyushu-shi Japan, 2006.

研究分野：総合領域

キーワード：空間疫学 疾病集積性 環境統計

## 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、つまり、疾病の時間・地域集積性を適切に検出できる統計的方法の開発が重要となる。

実際、米国の幾つかの都市で運用されていた症候サーベイランスに当初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(circular spatial scan statistic)であった。ただ、この方法は症候の集積地域が文字通りの円状でないとうまく同定できないなど、幾つかの改善すべき問題点があった。申請者は平成16-18年度科学研究費補助金(基盤研究(B))を受け、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(flexible spatial scan statistic)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、平成20-22年度、平成23-25年度と科学研究費補助金(基盤研究(B))を受け、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., 2012)。

これらの方法は、世界からある程度の注目を浴び、ニューヨーク市をはじめとする世界各地の症候サーベイランスや、疫学研究などで使用されてきた。海外の出版社 Springer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年の3月に出版した。しかし、現在の方法は、既存の他の方法に比べて優れている点が多いものの、1) 真の疾病集積性の完全同定確率が期待したほどの大きさではない、2) 時空間的集積性の検出では、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い、

などの問題点が指摘されており更なる改良が望まれていた。

## 2. 研究の目的

本研究では、以下の2点に焦点を当てて研究を遂行する。

- (1) 真の疾病の空間的集積性の完全同定確率を大きくするための改良を行う。従来のスキャン統計量はそのエンジン部分にポアソン分布を仮定した尤度比検定統計量を用いていたが、この完全同定確率が大きくなるような統計量の修正を試みる。
- (2) 時空間的に成長する突発的事象を検出できる方法の開発を行う。これまでの時空間的集積地域は時間的な成長は検出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、「時間とともに空間的にも成長する」クラスターを精度よく検出できる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：

研究の計画と方法の概要は、研究期間全体を通じ、理論的な統計量の検討と現実を反映した様々な勃発のシナリオに基づく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。実データでの検証法については、2006年に実施された北九州市の131の小学校の欠席者数の週毎のサーベイランスデータを利用して検討する。

### (2) 空間的集積性の完全同定確率に関する研究：

真の疾病の空間的集積性を正確に同定できる確率が、集積性の強度(リスクの大きさ)が増大するにつれて、1.0に近づく方法を検討するとともに、突発事象の集積性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 統計量の開発を目指す。

我々が最初提案した flexible scan 統計量(Tango and Takahashi, 2005)は検出できる実用上の最大クラスターサイズが25から30程度で、それを超えると計算時間が膨大(例：一か月以上)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、最近の研究により、尤度比検定統計量に基づく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」し、その結果として、クラスターサイズの制限がなくなること示した(Tango and Takahashi, 2012)。本年度の課題についても、この尤度比検定統計量の更なる改良により解決可能と考えられるので、まずは、理論的に検討する研究から着手し、次にアルゴリズムの検討に入る。

(3) 時空間的広がりを検出するモデルの検討：この研究テーマは過去に何度も挑戦してなかなかうまくいかなかったが、最近の研究(Tango, 2008, Tango and Takahashi, 2012)によりその大きなヒントが与えられた。申請者らの提案した flexible spatial scan は、空間集積性の検出においては、広く利用されている circular spatial scan 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キャン法の検討を行う。circular spatial scan では、ある地域(円柱の底面)を固定したとき、底面×時間(長さ)のシリンダー(円柱)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、一日前から起こっているのか、あるいは二日前から起こっているのか、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である。一方、flexible spatial scan では集積地域の形状は任意で制限はないので底面を非円状の多角柱からなるウィンドウをスキャンすることで現実的な時空間領域の同定を可能にする。しかし、これらの方法は、勃発の発生した地域が時間とともに拡大していくという状況をうまく同定することができなかった。ところが、最近の研究成果により、1) 観測された事象数の異常性を判定する期待頻度の計算と、2) ポアソン分布のバラツキを超えた過分散を適切に処理することにより「時空間的に徐々に拡大する」領域の同定方法に関する「手がかり」がつかめたところである。ここで提案する方法は、既存の方法、(特に、Kulldorff の時空間スキャン統計量、Takahashi-Tango 時空間スキャン統計量)とシミュレーションでその性能を比較する。

#### 4. 研究成果

(1) 空間的集積性の同定確率に関する研究  
まず、任意の空間領域(ウィンドウ)W に対し観測ケース数を  $n(W)$ 、期待ケース数を  $\mu(W)$  と置くと、空間集積性検出のための統計量として制限付き尤度(Tango, 2008)

$$\frac{n(W)/\mu(W)}{n(W^c)/\mu(W^c)} \times \prod_i I(n(W)/\mu(W) > n(W^c)/\mu(W^c)) \times \prod_i I(p_i < \alpha_i)$$

を修正する。ここで、 $p_i$  は各地域(region)毎の疾病集積性の有意差検定の片側 midp 値であり  $\alpha_i$  は事前に指定された各地域の有意水準である。ここで、 $\alpha_i = 1$  とすれば、この尤度は制約なしの通常の尤度比検定統計量である。 $\alpha_i$  の大きさをいろいろ変えて、クラスターの検出の状況を検討した結果、次の結果が得られた。

$\alpha_i = 0.10 \sim 0.20$  と設定すると集積リスクの大きいものの、小さいサイズのクラスターの検出力が高い、

$\alpha_i = 0.20 \sim 0.30$  と設定すると、中程度の集積リスクでクラスターサイズも中程度の検出力が高い

$\alpha_i = 0.30 \sim 0.40$  と設定すると、小さい集積リスクでクラスターサイズが大きいクラスターの検出力が高い

そこで、シミュレーションにより、様々な形状、リスクの大きさを持つクラスターを生成させる。その際に、クラスターのリスクの大きさ(集積性の強度)を次のステップで計算する。まず、帰無仮説の下で期待される期待ケース数を  $E_i$  としたときのポアソン乱数、 $N_i \sim \text{Poisson}(E_i)$

を発生させ、帰無仮説の下での観測ケース数とする。次に、事前に与えられたポアソン上側確率  $Q_i$ (小さいほどリスクが大きい)に対して

$Q_i = \Pr\{N_i \geq M_i \mid N_i \sim \text{Poisson}(E_i)\}$  となるポアソン乱数  $M_i$  を発生させる。つまり、クラスターに属する地域  $i$  の観測ケース数として  $N_i + M_i$  を仮定する。また、シミュレーション全体の検出力については、Tango and Takahashi (2005)で提案された2変量検出力分布  $P(l, s)$  (Bivariate power distribution)

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

の形でまとめる。したがって、2次元座標で  $(l, s) = (s^*, s^*)$  (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。その結果、代表的な結果の例をリスクの小さい方から大きい方へと、0.05, 0.01, 0.001, 0.0001 の4段階で、真のクラスターサイズより2個地域数が多く検出する場合について、示す次のようになる。

$Q_i = 0.05$  の場合

$$P(s^*, s^*) = 0.059$$

$$P(s^*+1, s^*) = 0.238$$

$$P(s^*+2, s^*) = 0.496$$

$Q_i = 0.01$  の場合

$$P(s^*, s^*) = 0.107$$

$$P(s^*+1, s^*) = 0.367$$

$$P(s^*+2, s^*) = 0.638$$

$Q_i = 0.001$  の場合

$$P(s^*, s^*) = 0.245$$

$$P(s^*+1, s^*) = 0.589$$

$P(s^{*+2}, s^*) = 0.816$   
 $Q_i = 0.0001$  の場合  
 $P(s^*, s^*) = 0.503$   
 $P(s^{*+1}, s^*) = 0.738$   
 $P(s^{*+2}, s^*) = 0.971$

となる。これは、当初想定された、真の疾病の空間的集積性を正確に同定できる確率が、集積性の強度（リスクの大きさ）が増大するにつれて、1.0 に近づく現象を示している。これらの研究成果の一部は、第 37 回国際バイオメトリクス学会の招待講演（2014、フィレンツェ、イタリア）で発表した。その一部は神奈川県乳がん死亡集積性の解析（Katayama, 2014）に応用された。

## （2）時空間的広がりを検出するモデルの検討

時空間的広がりを検出するための一般モデルとして、地域  $i$ 、時点  $t$  で観測される観測ケース数は無視できない時点間の過分散を考慮して、負の二項分布 NB に従い、

$$H_0: N_{it} \sim \text{NB}(\mu_{it}, \phi_{it}),$$

$$H_1: N_{it} \sim \text{NB}(\theta_{it}\mu_{it}, \phi_{it})$$

を考える。ここで、 $(\mu_{it}, \phi_{it})$  は負の二項分布のパラメータで、 $\theta_{it}$  は地域  $i$  の観測時点  $t$  での未知の疾病相対リスクである。ここで、時点  $t_p - u$  で勃発が始まったアウトブレイクモデルとして、

$$\theta_{it} = h(\tau + \beta_W(t - t_p + u)), \text{ if } (i, t) \in W = Z \times I_{\{u\}} \\ 1, \text{ otherwise}$$

ここで  $h(\cdot)$  は時点  $t = t_p - u$  での相対危険であり、その時点を超えて開始されたアウトブレイクの速度（傾き）は  $\beta_W h'(\tau)$  となるモデルである。つまり、上記のモデル（仮説）はすべての探索可能なドメイン  $W$  上での単純な次の仮説検定に帰着される：

$$H_0: \beta_W = 0, \quad H_1: \beta_W > 0,$$

このモデルを組み込むことにより、時空間的広がりを Kulldorff らの時空間スキャン統計量（2001）、Takahashi-Tango ら（2008）の時空間スキャン統計量（以下では「これまでの時空間モデル」と呼ぶ）よりも、感度良く、しかも、精度よく検出可能なことが予想される。ここでは、詳細な検定統計量の導出は省略するが、その性能を 2006 年に実施された北九州市の 131 の小学校の欠席者数の週毎のサーベイランスデータを利用して Kulldorff らのモデル（時空間スキャン統計量）と比較検証した。その主な結果は次の通り：

これまでの時空間モデルでは type I エラー、すなわち、偽陽性（間違っただけの発生）の確率が名義水準  $\alpha$  の 10 倍となることも稀ではないが、提案するモデルではほぼ名義の有意水準  $\alpha$  の付近であった。

アウトブレイク地域での帰無仮説の下での期待ケース数が小さいとき、これまでの時空間モデルでは検出力が有意水準  $\alpha$  より低下するという現象が起きるが、提案モデルではそのような現象は起こらない。つまり、従来のスキャン統計量は正しい検定の要求事項を満足していない、ことが分かる。

通常の見出力に関して、これまでの時空間モデルは提案モデルより高い値を示すこともしばしばであるが、正しい勃発地域の検出確率では、これまでの時空間モデルは提案モデルよりかなり低いことがわかった。つまり、従来のモデルでは non-outbreak の周辺地域を飲み込んだ形で実際より膨れ上がった地域を検出していたことが分かる。

また、より重要な問題として、これまでの時空間モデルでは、疾病の発生が減少傾向にある地域群をアウトブレイク地域として検出することもまれではなかったが、提案するモデルではそのような現象は観察されていない。

これらの結果の一部の研究成果は国際会議 GEOMED 2015 の招待講演（2015、フローレンス、イタリア）で発表した。また、国際的に著名な雑誌 *Statistics in Medicine*（2016）に掲載された。

本研究により、突発的生起事象の時空間上での勃発の兆候（症候の増加、広がり、あるいは局所的集積性）が精度よく検出できる統計モデルの開発が期待され、健康危機管理対策を推進するための有用なツールの提供が期待できる。

## < 引用文献 >

Tango T and Takahashi K. A flexibly shaped spatial scan statistic for detecting clusters. *International Journal of Health Geographics* 2005, 4:11

Tango T. A spatial scan statistic with a restricted likelihood ratio. *Japanese Journal of Biometrics* 2008, 29(2):75 - 95.

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

Tango T. *Statistical Methods for Disease Clustering*, *Statistics for Biology and Health Series*, Springer, 2010.

Tango T and Takahashi K. A flexible spatial scan statistic with a restricted likelihood ratio for detecting disease clusters. *Statistics in Medicine* 2012, 31(30): 4207 - 4218

## 5 . 主な発表論文等

### 〔雑誌論文〕(計 4 件)

Manabe T, Yamaoka K, Tango T, Binh NG, Co, DX, Tuan ND, Izumi S, Takasaki J, Chau NQ and Kudo K. Chronological, geographical, and seasonal trends of human cases of avian influenza A (H5N1) in Vietnam, 2003–2014: A Spatial Analysis. *BMC Infectious Diseases* 2016; **16**: 64. 査読有.

Tango T. On the recent debate on the space-time scan statistic for prospective surveillance. *Statistics in Medicine* 2016; **35**: 1927-1928. 査読有.

Tango T. On the repeated measures designs and sample sizes for randomized controlled trials. *Biostatistics* 2016;**17**: 334-349. 査読有.

Katayama K, Yokoyama K, Yako-Suketomo H, Okamoto N, Tango T, Inaba Y. Breast Cancer Clustering in Kanagawa, Japan: A Geographic Analysis. *Asian Pac J Cancer Prev*. 2014, **15**: 455-460. 査読有

### 〔学会発表〕(計 4 件)

Yamaoka K, Watanabe M, Watanabe J, Adachi M, Nemoto A, Tango T. A cluster randomized controlled trial to examine the effects of lifestyle education program for adolescents. The XXVIIIth International Biometric Conference, July 10-15, 2016, the Victoria Conference Centre, Victoria, British Columbia, Canada.

Tango T. On the recent debate on SaTScan for prospective surveillance. (Invited Speaker).

GEOMED 2015 (the IX international, interdisciplinary conference on spatial statistics, geographical epidemiology), September 10-12, 2015, Florence, Italy.

渡辺純子、渡辺満利子、山岡和枝、安達

美佐、根本明日香、丹後俊郎. 中学生のライフスタイル改善による愁訴減少の効果の評価 クラスタ-無作為化比較試験. 第 26 回日本疫学会学術総会、2016 年 1 月 21 - 23 日、米子コンベンションセンター、米子市。

Tango T. Statistical methods for disease clustering. The XXXVIIth International Biometric Conference (Invited Speaker). July 6-11, 2014, Firenze, Italy.

### 〔図書〕(計 5 件)

Tango T. Repeated measures design with generalized linear mixed models for randomized controlled trials. CRC Press, 2017.

丹後俊郎. 新版メタアナリシス入門 エビデンスの統合を目指す統計手法、朝倉書店、2016.

丹後俊郎、継時的繰り返し測定デザイン - 治療効果を評価する混合効果モデルとその周辺. 朝倉書店、2015.

山岡和枝、安達美佐、渡辺満利子、丹後俊郎. ライフスタイル改善の実践と評価 生活習慣病発症・重症化の予防に向けて. 朝倉書店、2015.

Saeki H. and Tango T. Statistical Inference for Non-inferiority of a Diagnostic Procedure Compared to an Alternative Procedure Based on the Difference in Correlated Proportions from Multiple Raters. In Kees van Montfort, Johan Oud, Wendimagegn Ghidye (Eds.) *Developments in Statistical Evaluation of Clinical Trials*, p119-137, Springer, 2014.

## 6 . 研究組織

### (1) 研究代表者

丹後俊郎 (TANGO Toshiro)

帝京大学・大学院公衆衛生学研究科・教授  
研究者番号：70124477