

令和 3 年 6 月 8 日現在

機関番号：12602

研究種目：基盤研究(C) (一般)

研究期間：2017～2020

課題番号：17K00046

研究課題名(和文)空間医学データにおける統計的複数クラスタ検出法の開発と実践

研究課題名(英文)Development of a multiple cluster detection method for spatial medical data and applications

研究代表者

高橋 邦彦(Takahashi, Kunihiko)

東京医科歯科大学・M&Dデータ科学センター・教授

研究者番号：50323259

交付決定額(研究期間全体)：(直接経費) 3,500,000円

研究成果の概要(和文)：本研究では、位置情報を含む空間的な医療データを用いて、これまで十分に検討されていなかった疾患発生に関連する複数のクラスタを検出し、その総合的な有意性を評価する新しい統計手法を提案した。この手法を実データやシミュレーション研究に適用したところ、従来の手法では不可能であった複数のクラスタの検出と評価を高い精度で行うことができることがわかった。

研究成果の学術的意義や社会的意義

これまで十分な議論がされていなかった空間データにおける複数クラスタの検出において、適切なクラスタ数、各クラスタの位置を推定し、その上で複数クラスタ全体としての統計的有意性が評価できるという一連の新たな解析手法が提案でき、学術的意義は大きい。また本研究では現実に即した問題設定での議論を踏まえており、特にクラスタ検出は感染症をはじめとする疾病発生のサーベイランスでも活用できる方法で、解析ソフトの開発を進められており、国内外の研究者、実務者に向けた社会的意義も大きい研究である。

研究成果の概要(英文)：We proposed a new statistical method for detecting multiple clusters related to disease occurrence using spatial medical data including location information, and evaluating their overall significance, which has not been sufficiently studied yet. By applying this method to real data and simulation studies, we found that the method can detect and evaluate multiple clusters with high accuracy, which has not been possible with conventional methods.

研究分野：生物統計学

キーワード：空間疫学 スキャン統計量 多重検定 医学データ

1. 研究開始当初の背景

近年、疫学研究や臨床研究を含めた幅広い医学研究分野において、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。地図上における市区町村や緯度経度などの2次元空間データに限らず、時間を点とした1次元空間データ、時間+空間や医療画像診断での3次元空間データなど様々なものが位置情報として用いられる。また遺伝子解析における遺伝子間の関係や、薬剤疫学における薬剤間や副作用間の関連への適用や検討も行われている。このような医学空間データに対して、それぞれの検討目的に応じた解析が適用されるが、位置の細分化が詳細になればなるほどデータ量も大きく、検討する課題も複雑なものになる。そのためビッグデータの枠組みなども含め、様々な観点から、より適切かつ効率的な統計手法を開発することが生物統計学分野の重要な課題となっている。

そのひとつとして、注目する疾病発生と医学的な意味で関連が強い位置・領域を客観的に同定することを目的とした「統計的クラスタ検出」の問題がある。通常、疾病発生との関連の評価については、それぞれの点で独立に、各点ごとに観測されるデータに基づき、統計的に定量可能な指標によって行われることが多い。しかし、一般に疾病発生との関連は近距離の点では類似性を持つと考えられ、疾病発生と強い関連をもつクラスタ(近距離範囲の隣接した点からなる集積領域)が対象空間内に存在することが想定される。

このようなとき、適切なクラスタの位置を推定するとともに、個々の点ではなく、クラスタとして統合された関連性の指標による評価が必要となる。このことは、近距離座標間の相関が考慮されるとともに、個々の点での独立な評価よりも、より強い関連を評価することが可能になることにつながる。また結果の解釈においても、独立な点の集合ではなく、クラスタ領域として解釈することで、実際上の議論・検討においてもメリットが多い。例えばクラスタとして検出された地域・時間の共通性を探ることで、疾病発生の要因を探りあてることができ、集中発生する感染症などを早期発見すること症候サーベイランスなどにも活用されている。

ここで大きな問題となるのがクラスタ候補の選択とその有意性評価である。先行研究などを参考に事前にクラスタ領域候補をいくつか設定し、それぞれを検定により評価する方法も用いられるが、その方法では、検出されるクラスタ領域は恣意的なものになってしまう。また複雑なネットワーク構造を予め絞りこむことなどは困難なため、それらの問題に対応するためには事前にクラスタ候補を制限せず、より一般的かつ柔軟に適切なクラスタを検出できる統計的クラスタ検出法が必要となってくる。

研究代表者らは、空間疫学分野を中心に、クラスタの位置や大きさを予め想定しない状況における疾病発生と関連の強いクラスタ領域をシグナルとして検出する統計的方法の検討と実践を行ってきた。特にスキャン統計量を用いて、疾病発生が集中している地域・時間をクラスタとして検出し、その有意性を統計的検定に基づいて評価する疾病集積性の検定法として flexible scan statisti を提案し、続けて、その拡張や評価について統計的な視点から様々な検討を行ってきた。またアプリケーション FleXScan の開発も行い、国内外の疫学研究や GIS 分野で利用されている。さらに難病患者発生の空間的集積性、救急搬送件数の時間的集積性、疾病関連遺伝子研究などに適用した、実践的な研究も実施してきている。

しかし、申請者らの提案法をはじめ、これまでのクラスタの有意性評価は、対象空間内に単一クラスタのみを仮定するモデルに基づく統計的検定となっている。そのため複数のクラスタの検出としては、通常、単一クラスタの検出を繰り返し用い、複数クラスタであることを考慮した上で、各クラスタについて個々に有意性(p値)評価を行っている。つまり複数クラスタ全体としての検出や評価は十分に行われていない状況である。最近のビッグデータをはじめ、従来よりも広い対象空間での解析においてはむしろ複数のクラスタの存在が考えられ、そのためには統計的な検出と評価法が必要となる。

2. 研究の目的

本研究では、医学分野における医学空間データを用いた統計的クラスタ検出において複数クラスタの検出と検定による評価法について汎用性のある統一的な方法論の開発を目的とする。特に空間疫学の事例において、事前にクラスタ候補が制限されていない状況でのスキャン統計量を利用した集積性の検定法の拡張について検討を行う。また対象空間内に存在する複数クラスタの数を適切に推定するための新たな情報量規準を開発し、さらに関連性の有意性評価指標の視点から適切な検出法の検討を行い、複数クラスタ研究と評価のための総合的な評価手順の開発と提案を行う。

3. 研究の方法

まず、空間疫学における時間集積性の検定での複数クラスタ検出法の理論構築に焦点をあて、従来の単一クラスタを仮定したスキャン統計量に基づくクラスタ検出法(Kulldorff, 1997; Tango and Takahashi, 2012, 他)のモデルを拡張し、複数クラスタのモデルを構築した。具体的にはクラスタの個数 K を固定したもとで、Jung(2009)が提案する一般化線形モデルを参考に、よ

り一般化されたもとでの K クラスタモデルの定式化を行った。構築された K クラスタモデルにおいて、最適なクラスタ数を選択するための規準指標を開発した。具体的にはクラスタの候補選択に確率構造を導入し、そのうえで周辺尤度に基づく情報量規準を構築した。その情報量規準にもとづく検定統計量を開発し、時間軸上での複数クラスタの存在を対立仮説とした場合の p 値を算出する手順を提案した。この精度について、シミュレーション実験および実際のデータに適用し、従来法との比較を行った。

さらに、この方法を拡張し、平面空間上における複数クラスタの検出と評価法についても検討を行い、その精度についてもシミュレーション実験および実際のデータに適用した評価を行った。

4. 研究成果

時間集積性の事例として、日本における心肺機能停止傷病者の日別搬送件数のデータに適用し評価を行った（図 1）。

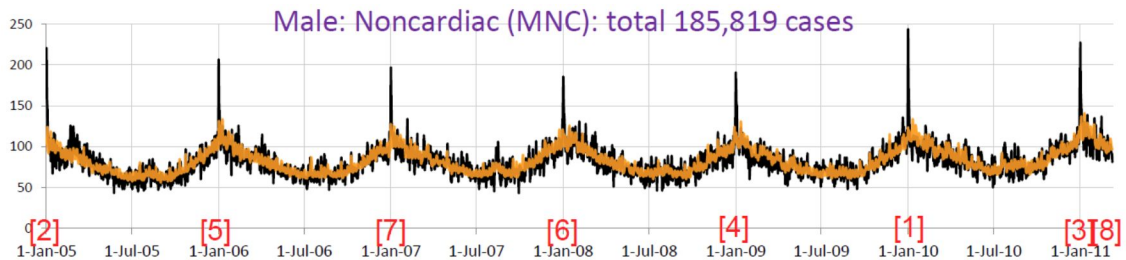


図 1. 男性非心原性の心肺機能停止傷病者数の日別搬送件数（黒線）、期待頻度（オレンジ）と検出されたクラスタ位置

実データの解析において検出されたクラスタは、従来法と同じであった。一方、従来法では評価不能であった複数クラスタの存在を仮定した包括的な統計的有意性を本方法により評価することができた。また本提案での情報量規準を用いることで、存在するクラスタ個数・位置を適切に推定・評価できていることが確認できた（図 2）。さらにシミュレーション実験により、提案法が複数クラスタを精度よく検出できることが確かめられた。本成果について国内外での学会発表とともに、国際的な学術専門誌に掲載された。

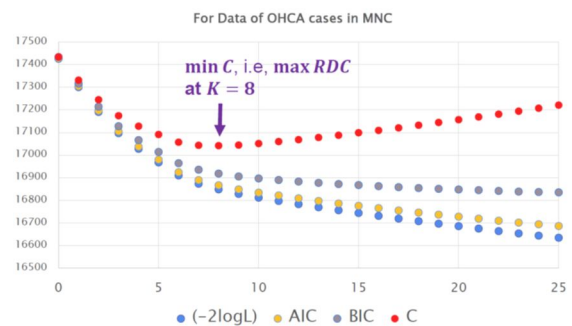


図 2. クラスタ個数選択のための情報量規準の挙動

つぎに上記の時間集積性の検定における複数クラスタ検出のための方法を平面空間上のデータに適用するための拡張を行った。時間集積性の検定と同様に複数クラスタのモデル化、それに基づく最適なクラスタ数推定のための情報量規準の提案、さらに包括的な複数クラスタ存在の統計的有意性の評価方法を開発した。

事例としてイングランドにおける COPD 入院患者数の地域データへの適用したところ、今回の提案した新たな情報量規準とそれに基づく統計量を用いる提案法によって平面空間内に存在する複数クラスタの数を的確に定めることが確かめられた。また詳細なシミュレーション実験による数値的評価を行い、平面上におけるクラスタ検出においても、今回提案する方法によって、従来の方法ではできなかった複数クラスタを精度よく検出できることが確かめられた。（図 3）この成果についても国内外での学会発表とともに、国際的な学術専門誌に掲載することができた。

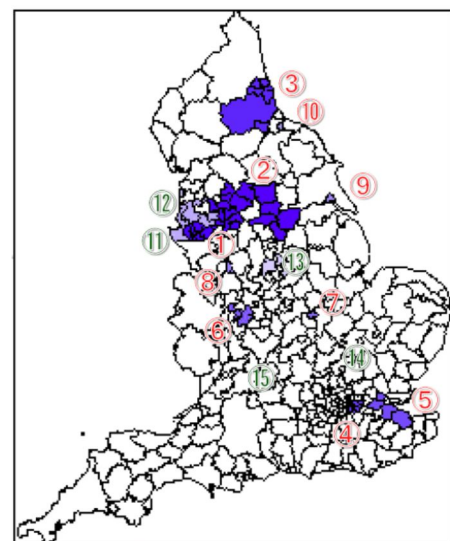


図 3. 空間集積性における複数クラスタの検出の比較

5. 主な発表論文等

〔雑誌論文〕 計6件（うち査読付論文 5件/うち国際共著 2件/うちオープンアクセス 3件）

1. 著者名 Takahashi Kunihiko, Shimadzu Hideyasu	4. 巻 13
2. 論文標題 Multiple-cluster detection test for purely temporal disease clustering: Integration of scan statistics and generalized linear models	5. 発行年 2018年
3. 雑誌名 PLOS ONE	6. 最初と最後の頁 e0207821
掲載論文のDOI（デジタルオブジェクト識別子） 10.1371/journal.pone.0207821	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 該当する
1. 著者名 Kuriki Satoshi, Takahashi Kunihiko, Hara Hisayuki	4. 巻 1
2. 論文標題 Multiplicity adjustment for temporal and spatial scan statistics using Markov property	5. 発行年 2018年
3. 雑誌名 Japanese Journal of Statistics and Data Science	6. 最初と最後の頁 191～213
掲載論文のDOI（デジタルオブジェクト識別子） 10.1007/s42081-018-0007-5	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 Horisaki Ken, Takahashi Kunihiko, Ito Hidemi, Matsui Shigeyuki	4. 巻 28
2. 論文標題 A Dose-Response Meta-analysis of Coffee Consumption and Colorectal Cancer Risk in the Japanese Population: Application of a Cubic-Spline Model	5. 発行年 2018年
3. 雑誌名 Journal of Epidemiology	6. 最初と最後の頁 503～509
掲載論文のDOI（デジタルオブジェクト識別子） 10.2188/jea.JE20170201	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -
1. 著者名 Anzai Tatsuhiko, Takahashi Kunihiko, Watanabe Michiko	4. 巻 73
2. 論文標題 Adverse reaction reports of neuroleptic malignant syndrome induced by atypical antipsychotic agents in the Japanese Adverse Drug Event Report (JADER) database	5. 発行年 2019年
3. 雑誌名 Psychiatry and Clinical Neurosciences	6. 最初と最後の頁 27～33
掲載論文のDOI（デジタルオブジェクト識別子） 10.1111/pcn.12793	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Takahashi Kunihiko, Shimadzu Hideyasu	4. 巻 19
2. 論文標題 Detecting multiple spatial disease clusters: information criterion and scan statistic approach	5. 発行年 2020年
3. 雑誌名 International Journal of Health Geographics	6. 最初と最後の頁 33
掲載論文のDOI (デジタルオブジェクト識別子) 10.1186/s12942-020-00228-y	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 該当する

1. 著者名 高橋邦彦	4. 巻 276
2. 論文標題 保健医療のための空間データサイエンス	5. 発行年 2021年
3. 雑誌名 医学のあゆみ	6. 最初と最後の頁 843-848
掲載論文のDOI (デジタルオブジェクト識別子) なし	査読の有無 無
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計7件 (うち招待講演 1件 / うち国際学会 3件)

1. 発表者名 Takahashi K, Shimadzu H.
2. 発表標題 Detection and evaluation of multiple clusters in spatial epidemiology
3. 学会等名 3rd International Conference on Econometrics and Statistics (EcoSta 2019) (招待講演) (国際学会)
4. 発表年 2019年

1. 発表者名 高橋邦彦, 島津秀康.
2. 発表標題 空間スキャン統計量を用いた複数クラスタの検出と評価
3. 学会等名 2019年度統計関連学会連合大会
4. 発表年 2019年

1. 発表者名 Takahashi K, Shimadzu H
2. 発表標題 Detecting multiple spatial-clusters by scan statistics
3. 学会等名 XXIXth International Biometric Conference (国際学会)
4. 発表年 2018年

1. 発表者名 Otani T, Takahashi K, Takeuchi A, Asami M
2. 発表標題 Comparison of Spatial Interpolation Methods based on Exposure Assessments of Air Pollutants: A Case Study on Nuclear Substances in Fukushima
3. 学会等名 12th International Conference on Health Informatics (国際学会)
4. 発表年 2019年

1. 発表者名 立森久照, 高橋邦彦, 加藤直広, 水流聡子, 熊倉陽介, 小高真美
2. 発表標題 GISを用いた地域の特徴の可視化の取り組み - 川崎市みまもりサービスの効率化に向けた活動 -
3. 学会等名 第2回日本臨床知識学会学術集会
4. 発表年 2018年

1. 発表者名 堀崎健, 高橋邦彦, 伊藤秀美, 松井茂之
2. 発表標題 日本人におけるコーヒー摂取量と大腸がん発生リスクの用量反応関係メタアナリシス
3. 学会等名 第28回日本疫学会学術総会
4. 発表年 2018年

1. 発表者名 大谷隆浩, 高橋邦彦
2. 発表標題 複数の疾病集積性を解析するためのRパッケージ“multicluster”の開発と評価
3. 学会等名 第31回日本疫学会学術総会
4. 発表年 2021年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関		
英国	Loughborough University		