【原著論文】

地域間比較のためのリスク人口の推定方法 ー インフルエンザ定点報告数に関する分析 ー

片岡裕介・浅見泰司・多田有希・小坂 健

**A Risk Population Estimation Method for Interregional Comparison:
An analysis on the influenza data reported by the surveillance system**

Yusuke KATAOKA, Yasushi ASAMI, Yuki TADA and Ken OSAKA

Abstract: This study proposes a risk population estimation method based on the reported counts of patients of influenza in each area. A risk population estimation method is developed by regarding the sum of the weekly observations in a season following the binomial distribution, which enables estimations of population parameters and probability parameters through maximum likelihood method. The results show that the estimated infection probability is consistently correlated with population density, and the estimated population parameters in areas with high populations are obviously small in comparison with the proportion of population.

Keywords: 感染症サーベイランス (Epidemiological Surveillance of Infectious Diseases), 最尤法 (maximum likelihood method), インフルエンザ (influenza), 人口密度 (population density), 二項分布 (binomial distribution)

1. はじめに

地域事象に対して、地理的、空間的な視座から特徴やパターンを見出そうとする際には、地域相互を何らかの基準のもとで照合する過程が不可避である。しかしながら、現実には全ての観測値データが、集計された地域単位で基準化されているわけではなく、このことが事象解明に要する分析や適切な評価をおこなう上で致命的な問題となる。感染症を扱う本稿の契機もこの点にあった。

本稿で採りあげたインフルエンザは、流行が周期的に現れる感染症であり、国内においては、毎年

11月下旬から12月上旬頃に始まり、翌年の1～3月頃に患者数が増加、4～5月にかけて減少していくという流行のパターンを示す¹⁾。その個人や社会に与える影響の重大性から、インフルエンザは「感染症の予防及び感染症の患者に対する医療に関する法律」により、感染症発生動向調査における定点把握疾患に定められており（感染症定点サーベイランス）、新規患者の発生数は全国約5,000ヶ所の指定された医療機関（定点）から、保健所、都道府県を経て、国へ毎週報告されることになっている。

インフルエンザの流行を扱った研究を見てみると、Cliff et al. (1986) による空間的な側面に着目した流行分析から、Viboud et al. (2003) の気象学的手法を用いた流行予測に至るまで、その分析アプローチは多岐にわたる。日本を対象としたものとし

片岡：〒277-8568 千葉県柏市柏の葉 5-1-5
東京大学空間情報科学研究センター
Center for Spatial Information Science, University of Tokyo
E-mail ukataoka@csis.u-tokyo.ac.jp

て、中谷（1994）では空間的相互作用を考慮した理論疫学モデルによる分析がなされ、鈴木ほか（2003）では、GISを用いた患者情報や小中学校の休校情報の解析が行われた。

谷村（2003）や中谷ほか（2004）は、疾病の地域集積性や発生動向の特性を検出するためにGISを援用した空間疫学的アプローチの必要性を述べているが、ここでは特に以下の点について留意しておかなければならない。それは、特定の地域内の時系列的な推移のみに止まらず、地域間比較を伴う空間的な流行分析を念頭に置く際には、観測値の集計単位である各地域の定点の数および質が、分析結果に多大な影響を及ぼすという事実である。この点が考慮されていない分析は、その効果が著しく損なわれるばかりか、場合によっては結果から誤った見解が導かれるおそれさえある。

前出の感染症定点サーベイランスは、インフルエンザを始めとした感染症流行の早期把握やトレンド把握をその目的とするため、定点の選定に際しては、できる限り無作為によること、人口や医療機関の分布を勘案することが求められている。しかし、全国の感染症定点サーベイランスを対象とした一連の研究（村上ほか、1999、2003；橋本ほか、1999）のなかで、村上ほか（2003）は、保健所管轄人口にもとづいた定点の設置基準²⁾から見たときに、人口規模の大きい保健所については基準を下回るものが多いことなど、現状のサーベイランスの問題点を指摘した。また中谷（1994）は、インフルエンザの流行モデルを検討するにあたって、都道府県間の比較を行うために、各都道府県の総患者報告数を定点数で除した数を用いている。つまりは、定点報告数をもって単純に地域間を比較することは適切でないのである。

そこで本稿では、地域内の定点の設置状況が必ずしも基準を満たしてはいないという現状の条件下で、地域間を比較するために最低限必要となる各地域のリスク人口の推定方法について論じる。ここで言う「リスク人口」とは、地域のなかで、既に免疫を持っているなどの理由により感染する可能性のない人を除いた、潜在的に感染するおそれのある人の

集合を指す。これは実際の定点の圏域人口とは数値的には異なるものの、各地域で比例するかたちで圏域人口に対応する。そして、その「リスク人口」の中から実際に感染する確率をここでは「罹患率」と呼ぶことにする³⁾。

以上をふまえ、各地域のリスク人口、および各シーズンの罹患率を推定するにあたり、最尤法を用いた推定モデルを提案する。続いてその適用事例として、3シーズンにわたるインフルエンザの定点報告数を対象とし⁴⁾、地域の特性を加味した推定方法により得られる結果について考察する。

2. リスク人口推定モデルの提案

本節では、インフルエンザという流行に周期性がある事象に対し、各シーズンの発生数の総計を用いて、各々の地域のリスク人口を最尤法により推定するモデルを提案する。推定に用いた最尤法は、周知のように母数推定方法として一般的に用いられる手法である。なかでも二項分布を扱った例をあげると、Myung（2003）では、確率項に複数の関数型を用いた推定結果とともに、最小二乗法による場合との比較を交えて紹介されている。

まずモデル化の前提として、対象とする地域数は十分大きく、また地域間で罹患率に差はないとする。これは、シーズン中の流行のピーク時には地域差が生じることを認めるにせよ、シーズン全体でみたときに各地域の罹患率は同等であることを意味する。そこで、各地域 i ($i=1,2,\dots,M$)における定点報告数のシーズン j ($j=1,2,\dots,T$)を通した総計を $x_{i,j}$ で表す。ここで、発症する確率である罹患率が地域によらず一定であるときに $x_{i,j}$ は、地域 i におけるリスク人口 n_i 、シーズン j における罹患率 p_j の二項分布に従う（Dawson-Saunders and Trapp, 1994）。さらに、地域数 M が十分大きいという条件のもとでは、この二項分布は近似的に正規分布に従うことから、 $x_{i,j}$ は以下の式(1)のような、平均 $n_i p_j$ 、分散 $n_i p_j (1-p_j)$ の確率密度関数 $f(x)$ で表される。

$$f(x_{i,j}; n_i p_j, n_i p_j (1-p_j)) = \frac{1}{\sqrt{2\pi \cdot n_i p_j (1-p_j)}} \exp \left\{ -\frac{(x_{i,j} - n_i p_j)^2}{2n_i p_j (1-p_j)} \right\} \quad (1)$$

式(1)の対数尤度関数 L を求めると,

$$\begin{aligned} \ln L &= \ln \prod_i^M \prod_j^T f(x_{i,j}; n_i p_j, n_i p_j (1-p_j)) \\ &= \ln \prod_i^M \prod_j^T \frac{1}{\sqrt{2\pi \cdot n_i p_j (1-p_j)}} \\ &\quad \times \exp \left\{ -\frac{(x_{i,j} - n_i p_j)^2}{2n_i p_j (1-p_j)} \right\} \\ &= \sum_i^M \sum_j^T \ln \frac{1}{\sqrt{2\pi \cdot n_i p_j (1-p_j)}} \\ &\quad - \sum_i^M \sum_j^T \frac{(x_{i,j} - n_i p_j)^2}{2n_i p_j (1-p_j)} \end{aligned} \quad (2)$$

となるので, 式(2)で n_i と p_j ($j=1, 2, \dots, T$) について, 以下のように偏微分をおこない, 各々の最尤値を求める.

$$\begin{cases} \frac{\partial \ln L}{\partial n_i} = 0 \end{cases} \quad (3)$$

$$\begin{cases} \frac{\partial \ln L}{\partial p_j} = 0 \quad (j=1, 2, \dots, T) \end{cases} \quad (4)$$

まず式(3)より,

$$\frac{\partial \ln L}{\partial n_i} = \frac{T}{2n_i} + \sum_j^T \frac{x_{i,j}^2 - n_i^2 p_j^2}{2n_i^2 p_j (1-p_j)} = 0 \quad (5)$$

となり, これを n_i について解くと,

$$n_i = \frac{-T + \sqrt{T^2 + 4 \cdot \sum_j^T \frac{x_{i,j}^2}{p_j (1-p_j)} \cdot \sum_j^T \frac{p_j}{1-p_j}}}{2 \cdot \sum_j^T \frac{p_j}{1-p_j}} \quad (n_i > 0) \quad (6)$$

が得られる.

同じく式(4)より,

$$\begin{aligned} \frac{\partial \ln L}{\partial p_j} &= -\frac{(1-2p_j)M}{2p_j(1-p_j)} \\ &\quad - \sum_i^M \frac{(n_i p_j - x_{i,j})(n_i p_j + (1-2p_j)x_{i,j})}{2n_i p_j^2 (1-p_j)^2} = 0 \end{aligned} \quad (7)$$

となる.

ここで, 式(7)に式(6)を代入することで, 以下の式(8)が導かれる.

$$\begin{aligned} &-\frac{(1-2p_j)M}{2p_j(1-p_j)} - \sum_i^M \frac{(n_i p_j - x_{i,j})(n_i p_j + (1-2p_j)x_{i,j})}{2n_i p_j^2 (1-p_j)^2} = 0 \\ \Leftrightarrow &\frac{(1-2p_j)M}{2p_j(1-p_j)} - \frac{1}{2(1-p_j)^2} \\ &\times \sum_i^M \frac{-T + \sqrt{T^2 + 4 \cdot \sum_j^T \frac{x_{i,j}^2}{p_j (1-p_j)} \cdot \sum_j^T \frac{p_j}{1-p_j}}}{2 \cdot \sum_j^T \frac{p_j}{1-p_j}} \\ &+ \frac{1}{(1-p_j)^2} \sum_i^M x_{i,j} + \frac{1-2p_j}{2p_j^2 (1-p_j)^2} \\ &\times \sum_i^M \frac{2 \cdot \sum_j^T \frac{p_j}{1-p_j}}{-T + \sqrt{T^2 + 4 \cdot \sum_j^T \frac{x_{i,j}^2}{p_j (1-p_j)} \cdot \sum_j^T \frac{p_j}{1-p_j}}} \cdot x_{i,j}^2 = 0 \end{aligned} \quad (8)$$

よって式(8)の p_j に関して, $j=1, 2, \dots, T$ で与えられる連立方程式の解として各 p_j が求まる.

さらに, $p_1, p_2, \dots, p_T$ の値を式(6)に代入すれば, 最終的に各地域のリスク人口 n_i が得られる.

3. 推定モデルの適用事例

前節で提案されたリスク人口推定モデルの適用事例として, インフルエンザ定点報告数のデータを用いた各推定値の導出とともに, 地域特性を考慮した推定方法について考察する.

3.1. インフルエンザ定点報告数について

使用したデータは1999年から2002年までのインフルエンザに関する毎週の定点報告数である. この期間の定点報告数の推移を図1に示す.

図1 定点報告数の推移 (1999～2002)

始めに毎週の定点報告数を、分析上の地域単位とした二次医療圏⁵⁾で集計した後に、各シーズンの定点報告数の総計（以下、シーズン総報告数と呼ぶ）を求めた。ただし、各シーズン期間の範囲の設定については、流行が終息に向かい次シーズンの発生が始まったとされる付近での、全国の定点報告数が最小となる時点をもってシーズンの区切りとした。

ここで2節の推定モデルでは、対象とする地域群について、罹患率がほぼ一定である状況を仮定している。そこで、実際の場合での地域差を把握するために、各医療圏のシーズン総報告数の全国比を求め、各シーズンについて見てみることにする。

図2 シーズン総報告数の全国比（一部）

図2は、全国360（2001年の時点）の医療圏のなかの50ヶ所（医療圏の番号が先頭から数えて101番から150番に相当）におけるシーズン総報告数の全国比を表している。同じ医療圏で全国比について比較すると、少数の医療圏を除きシーズンにはさほど依存せず、ほぼ一定している傾向がみえる。同様のことを全医療圏に対しても言うために、各医療圏でのシーズン間におけるバラツキとして、3シーズンの全国比の最大値と最小値との差をとった「範囲」を求めた。その頻度分布（階級幅は、0.2）を図3に示す。

図3より医療圏全体で見ても、各シーズンの全国比に差がない地域が圧倒的に多くを占めることがわかる。つまり、流行の時期や罹患率についてはシー

ズン間の差がある一方で、シーズン総報告数については、各定点の圏域人口に大きく依存していることが推測されよう。

図3 シーズン総報告数全国比の範囲の頻度分布

3.2. 対象地域の分類

前項では、各シーズンにおける罹患率が、大多数の地域でほぼ同程度と見なせることを示した。続いて本項では、流行に多大な影響を及ぼす地域特性に基づいた対象地域の分類をおこなう。

インフルエンザは、飛沫感染とともに飛沫核感染によっても伝播することから⁶⁾、地域の人口密度が感染状況に及ぼす影響が大きいと言われる。そこで各二次医療圏の、2000年国勢調査による人口と総面積を用いた人口密度により、全国の医療圏を複数の地域ブロックに分類してみる。その分類法としては、Chiu et al. (2001)により提案され、大規模データでは非常に有効とされているTwo-Stepクラスター分析を採用する。この方法では、BIC (Bayesian Information Criteria) のような情報量基準を用いることで、最適となるクラスター数についても求めることが可能となる。

図4は、全国の各医療圏における人口密度に基づき、Two-Stepクラスター分析を実行した結果として、クラスター数とBICの関係について示したものとなる。BICが最小となるクラスター数より、全国の二次医療圏は図5のように3つのグループに分類される⁷⁾。

分類された各クラスターとしては、概して、東京

図4 クラスター数と BIC

都区部の都心近郊、および大阪府の堺市以南の泉州地域が最も人口密度が高いグループとなり、それらの周辺地域が2番目に、そしてその他の地域が3番目のグループとなるといった地理的分布を呈している。また、人口の多い大阪市や名古屋市が、人口密度が低いクラスターに分類されている理由は、二次医療圏の各圏域面積が影響していることによる。

図5 地域の分類

3.3. リスク人口および罹患率の推定

前項までの考察をふまえ、分類されたそれぞれの地域ブロックで推定モデルを適用し、リスク人口と罹患率の推定値を求める。

報告数が3シーズン分あることから、2節の式(8)は3元非線形連立方程式となる。そこで、まず各シー

ズンの罹患率 p_1, p_2, p_3 の数値解を、Trust-Region Dogleg 法 (Conn et al., 2000) を用いて求めた。その結果、各ブロックについて得られた罹患率と、全医療圏を分類せずに一括して推定した場合のものを併せて以下の図6に示す (クラスターの番号は図5にならう)。

図6より、各ブロックの罹患率の3シーズン間の推移から、罹患率の推移が図1の報告数の推移と対応していることと、人口密度が高いブロックほど罹患率の推定値が高い傾向があることが言える。また、ブロック間で見られる罹患率の差異については、人口密度の高い地域の感染率が高いという、現状の理解を裏付ける結果として解釈できよう。それぞれの罹患率の推定値をみると、現実のそれとは幾分高い印象を与えているが、前述のとおりここでいう罹患率とは、潜在的に感染するおそれのある人たちが感染する確率であるため、通常の罹患率とは若干性質が異なるものである。

図6 ブロック別の罹患率 (推定値) の推移

さらに、得られた p_1, p_2, p_3 を式(6)に代入すれば、全ての二次医療圏の報告数のリスク人口が得られる。

図7は得られた推定値の頻度分布を表し、また図8は推定値の地理的分布について図示したものとなる。

両図において、階級幅は原則的に2,500人とした。まず、図7では定点の圏域人口は広範囲にわたって分布すること、そして図8からは、概して中山間地

域に、推定値の小さい地域が比較的多く分布していることが理解される。

実際には、定点報告数とはインフルエンザの発病者のうち医療機関で受診した患者の数であることを

図7 リスク人口推定値の頻度分布

考えれば、ここで推定されるリスク人口は本来の意味での医療機関の圏域人口よりも少なく見積もられるはずである。よって、それらの対応が明らかにされた資料を用いることで、より現状を反映した人口の推定値を得ることが可能となるであろう。また、インフルエンザには、複数の型が存在して流行に影響を及ぼすとされるが、確率項を含んだモデル化などによる対処も今後は検討できると考える。

また、ここで得た罹患率を用いた報告数の期待値の推定量と、実際の報告数との間で適合度検定を行った結果、尤度比はシーズンが前のものから順に、 2.20×10^4 , 6.66×10^4 , 4.72×10^4 となった。このとき有意水準 0.01 のもとで $\chi^2(359, 0.01) = 4.24 \times 10^2$ であることから、各シーズンの何れに場合においても両者が等しいとする帰無仮説は棄却される。全体

図8 全国二次医療圏のリスク人口推定値

としてはある程度の対応が認められるにせよ、両者が異なるというこの検定結果について、次に挙げるような問題が考えられる。1) 分類された地域群の中でも、実際には罹患率が一定ではなく、その影響が無視できない。2) シーズン総報告数全国比のバラツキが大きい地域が、全体に及ぼす影響が大きい。

また、3.2.での考察と関連するが、地域をさらに細分化してみた場合を検討してみたところ、含まれる地域数が極端に少ないクラスターが発生し、計算上不安定な解を得たこともあり、適合性に関する問題が改善する傾向は見えなかった。

3.4. 人口と推定されたリスク人口との関係

さらに下の図9は、都道府県人口が多い順に6都道府県を選択し、各都道府県人口が占める全国比と、同じく、本稿で得られた都道府県単位のリスク人口の推定値の全国比との比較を表したものである。

図9 都道府県人口とリスク人口推定値との比較

図9から、人口が多い都道府県においては、リスク人口の推定値の全国比は、人口の全国比と比べて小さい傾向があることが確認できる。つまりこの結果は、人口が多い都道府県において現行基準を下回るものがあるとした、村上ほか（2003）によって導かれた結論と概ね合致するものと言えよう。

さらに、「推定人口」と「医療圏人口」の相関係数をとった結果、0.775という一見すれば高い相関関係が示されるが、図9のように仔細を見てもわかる通り、二次医療圏人口は地域間を比較する基準として十分でないと判断できる。

4. おわりに

本稿では、地域間の比較分析に資するための、集計単位で基準化されていないインフルエンザ定点報告数を用いた、リスク人口の推定方法に関して論じた。リスク人口推定モデルの提案、および全国二次医療圏を対象とした適用事例を扱った本稿により得られた成果は、主に以下の通りである。各医療圏のリスク人口と各シーズンの罹患率からなる二項分布から、最尤法を用いたリスク人口推定モデルを提案した。また、3シーズンにわたるインフルエンザの定点報告数を用いて、リスク人口と罹患率の最尤値を数値的に得た。さらに、流行に多大な影響を及ぼすような地域特性を考慮した場合の方法について検討を行った。

現状のサーベイランスにとって、地域間の相互比較を伴った流行状況の把握が可能となることの効果は非常に大きい。具体的には、流行が先んじて発生した他地域の動向を参考にした流行予測や、また各地域の推定人口の一定割合を水準として用いるなどして、流行の発生を早期に感知し、アラートを発する時期を判断する指針の作成にも貢献できると考えられる。

なお、本稿でも問題とされたように、発生率が一定となるような地域の選定が非常に重要である。本稿では、まず全対象地域の罹患率に関する検討を行ったうえで、さらに地域を分類するという方法をとった。推定された罹患率にもとづく報告数と、観測データの適合度からみたときに、分類された地域内での罹患率が全て等しくなるほどの効果は得られなかったが、これについては定点報告数を扱う問題の困難性を考慮しつつ今後の課題としたい。

謝辞

本研究をおこなうにあたり、国立保健医療科学院技術評価部丹後俊郎部長、および国立感染症研究所感染症情報センター谷口清州感染症対策計画室長、大日康史主任研究官、安井良則主任研究官より貴重なコメントを頂いた。

本研究は、厚生労働科学研究費補助金（がん予防等健康科学総合研究事業）「地域における健康危機

情報の伝達、管理及び活用に関する研究」(代表：下田智久)の支援を受けた。

記して謝意を表する。

補注

- 1) 国立感染症研究所感染症情報センター HP <<http://idsc.nih.go.jp/disease/influenza/index.html>>より
- 2) 『感染症発生動向調査実施要項 1999』では、インフルエンザ定点のうち内科定点の場合には、保健所管轄人口が、「7.5 万人未満で定点数は 1, 7.5 万人以上 12.5 万人未満で定点数は 2, それ以上では定点数は $3 + (\text{人口} - 12.5 \text{ 万人}) / 10 \text{ 万人}$ 」と定められている。
- 3) 日本疫学会 (1996) においては、罹患率は、「ある人口集団における、ある一定の観察期間における疾病の発症頻度の率」と定義されている。
- 4) 分析の対象とした期間内において、若干の定点数の変動や、定点となる医療機関の変更の可能性はあるが、本稿の推定モデルでは定点について全く変動がないことを前提にしている。ただ、3.3. の適合度検定では、結果として定点の変動が多少影響していることが推測される。
- 5) 二次医療圏とは、「医療法に基づき策定される医療計画の単位となる区域のひとつであり、特殊な医療を除く一般の医療需要で主として病院における入院医療を提供する体制の確保を図る区域」とされる。一般に、二次医療圏の人口は 2,30 万人から数十万人単位となっているが、実際はかなりの地域差があるとみられる。
- 6) 中谷 (1994) によると、飛沫感染とは、患者から周囲の人々へ直接的に大きな飛沫を介して引き起こされる感染を指し、一方で飛沫核感染とは、閉鎖環境において長時間空気中に浮遊した小さい飛沫によって引き起こされる感染を指す。
- 7) 図 4 ではクラスター数が 6 のときにも、BIC が極小値をとっている。なお、6 つのグループに分類してみた場合では例外的に、人口密度が高い地域になるに従って、罹患率の推定値が必ずしも高くはならない。このような事情が BIC

の値の変化にも関係しているとも考えられる。

参考文献

- 鈴木宏・坂井貴胤・齋藤玲子・古俣修・佐藤勇 (2003) GIS (地理情報システム) によるインフルエンザ感染症の疫学解析. 「インフルエンザ」, 14 (1), 35-41.
- 谷村晋 (2003) 空間疫学アプローチは疾病対策にどのように役立つか. 「日本熱帯医学会雑誌」, 31, 237-241.
- 中谷友樹 (1994), インフルエンザの時・空間的流行モデル - 1988 ~ 1989 年におけるわが国の流行を事例として -. 「人文地理」, 46, 254-273.
- 中谷友樹・谷村晋・二瓶直子・堀越洋一 (2004), 『保健医療のための GIS』, 古今書院.
- 日本疫学会 (1996), 『疫学 - 基礎から学ぶために -』, 南江堂.
- 橋本修二・村上義孝・谷口清洲・永井正規 (1999) 感染症発生動向調査における全国年間罹患数推計のための定点設計. 「日本公衆衛生雑誌」, 46, 1068-1077.
- 村上義孝・橋本修二・谷口清洲・永井正規 (1999) 感染症発生動向調査における定点配置の現状評価. 「日本公衆衛生雑誌」, 46, 1060-1068.
- 村上義孝・橋本修二・谷口清洲・小坂健・瀧上博司・永井正規 (2003) 感染症法施行後における感染症発生動向調査の定点配置状況. 「日本公衆衛生雑誌」, 50, 732-738.
- Chiu, T., Fang, J., Wang, Y., and Jeris, C. (2001) A Robust and Scalable Clustering Algorithm for Mixed Type Attributes in Large Database Environment. *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 263-268.
- Cliff, A.D., Haggett, P. and Ord, J.K. (1986) *Spatial Aspects of influenza epidemics*. London: Pion
- Conn, A.R., Gould, N.I.M. and Toint, P.L. (2000) *Trust-Region Methods*. Philadelphia: SIAM.
- Dawson-Saunders, B. and Trapp, R.G. (1994) *Basic and Clinical Biostatistics second edition*. Norwalk: Appleton & Lange.
- Myung, I.J. (2003) Tutorial on maximum likelihood estimation. *Journal of Mathematical Psychology*, 47, 90-100.
- Viboud, C., Boëlle, P.Y., Carrat, F., Valleron, A.J. and Flahault, A. (2003) Prediction of the Spread of Influenza Epidemics by the Method of Analogues. *American Journal of Epidemiology*, 158, 996-1006.