

感染症サーベイランスにおける患者発生予測に関する統計的解析

北山 秋雄¹⁾ 市村 博¹⁾ 太田原 美作雄¹⁾ 飯田 恭子²⁾

A Statistical Analysis on the Forecasting of Numbers of Patients in the Surveillance of Infectious Diseases

Akio KITAYAMA, Hiroshi ICHIMURA, Misao OHTAWARA
and Kyoko IIDA

I はじめに

1981年7月に厚生省の「感染症サーベイランス」と呼ばれる全国的な監視体制が発足して、今年で5年目を迎えた。その間、全国から報告されるデータの精度も向上し、対象疾病の患者発生動向や流行の実態が徐々に解明されている。一方、各都道府県に於いても、社会的環境や地理的条件等を考慮に入れて、地方の特徴を活かした解析システムが開発されている^{1),2)}。しかし、患者発生数を定量化して将来予測のモデル関数を作成し、データ解析に利用している地方機関はほとんどない。将来予測は、衛生行政担当者が予防措置を講ずる際、最も適切な意志決定をするための貴重な情報を提供する。そこで、本研究では、千葉県に於ける過去4年間のデータを時系列解析 (Time Series Analysis) し、患者発生数の予測を試みたところ、若干の知見を得たので報告する。

II 方法

(1)資料：①千葉県感染症情報 (1982~1985)、②「感染症サーベイランス情報 (1982~1985) 週報」(厚生省)
(2)処理機器：本体SORD(株)M343SX, 主な周辺機器、40MBハードディスク, プリンタ (PT-350), 1MB拡張メモリー等, 使用言語, 日本語BASIC十一部アセンブリ, (3)データ解析法：自己回帰分析 (Auto Regressive Analysis), アルゴリズムの概要³⁾は注)に記した通りである。

III 結果

表1, は自己回帰分析を使い、千葉県における主な対象疾病の一医療定点当りの患者発生数を解析して得られたモデル関数である。Lag (遅れ)=20に於ける最適項数(m)を算出したところ、最小の手足口病ではm=4, 最大のインフルエンザ様疾患ではm=7を示した。通常、

表1. 千葉県に於ける疾病別自己回帰分析による解析結果

疾病名	最適AR (m)	トレンド	モデル関数	相関係数
麻疹様疾患	AR (2)	$Y(t)=0.601+0.358*t$	$y(t)=0.538*y(t-1)+0.398*y(t-2)+0.064$	0.917
水痘	AR (3)	$Y(t)=3.219-0.76 *t$	$y(t)=0.606*y(t-1)+0.438*y(t-2)-0.143*y(t-3)+0.099$	0.898
手足口病	AR (3)	$Y(t)=0.072+0.005*t$	$y(t)=0.947*y(t-1)+0.168*y(t-2)-0.215*y(t-3)+0.1$	0.949
インフルエンザ様疾患	AR (3)	$Y(t)=9.198-0.044*t$	$y(t)=1.123*y(t-1)-0.257*y(t-2)-0.144*y(t-3)+0.278$	0.879

1) 千葉県衛生研究所

2) 愛国学園短期大学

(1986年9月30日受理)

観測値と予測値の相関があまり低くならない程度に項数を少なくする。今回は、Lag=3に於ける最適項数を算出した。相関係数(r)の低いインフルエンザ様疾患ではr=0.879, 高い手足口病ではr=0.949であった。こ

これらの値はLag=20に於ける相関係数と比較しても最大0.08低かったにすぎない。

表2. は昭和57年～昭和60年までの4年間の全国に於ける麻疹様疾患一医療定点当りの患者発生数を、自己回帰分析を用いて解析して得られたモデル関数である。Lag=20に於ける最適項数はm=15、相関係数はr=0.991であった。また、Lag=3に於ける相関係数はr=

0.986の高い値を示した。

図1～4. は本県に於ける昭和60年の一医療定点当りの患者発生数と表1. のモデル関数によって得られた理論値の折線グラフである。スケール (Scale) の決め方によって平滑化 (Smoothing) に違いが生じるが、麻疹様疾患や水痘では比較的週別の増減が顕著にみらる。しかし、何れの疾病も年間の発生パターンが観測値と理

表2. 全国に於ける麻疹様疾患の自己回帰分析による解析結果

疾病名	最適AR (m)	トレンド	モデル関数	相関係数
麻疹様疾患	AR (3)	$Y(t)=0.730-0.001*t$	$y(t)=1.010*y(t-1)+0.275*y(t-2)-0.308*y(t-3)+0.023$	0.986

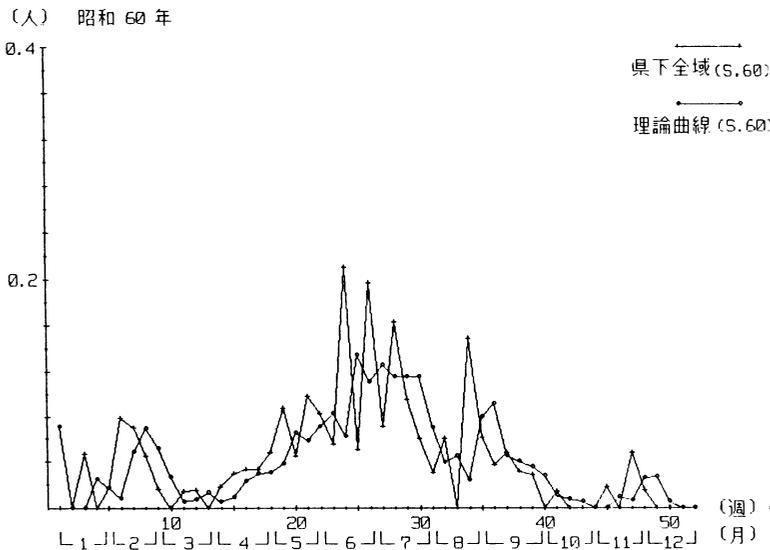


図1. 麻疹様疾患 年別1定点当りの調査状況

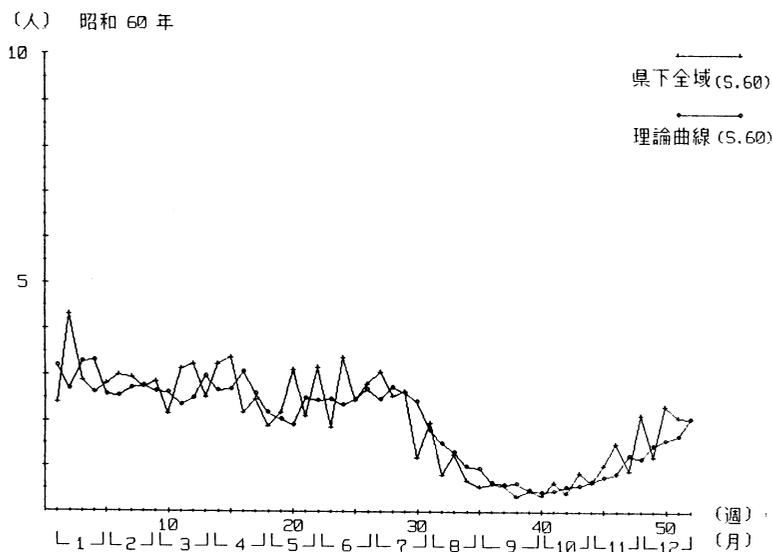


図2. 水痘 年別1定点当りの調査状況

論値ではほぼ一致している。図1, の麻疹様疾患は本県では, 昭和59年に大流行したが, 昭和60年は一部の地域を除いて発生数が少なかった。図2, の水痘も一部の地域を除いて発生数が少なかった。例年, 水痘は夏から秋にかけて減少する傾向にあるが, 過去4年間の年間発生パターンに大きな変化がみられないことから, 比較的環境的变化に影響されない, 常在性の感染症と考えられる。

図3, の手足口病は昭和60年に大きな流行がみられた。患者の発生時期も早く, 5月中旬より急増し, 6月~7月にかけてプラトーに達し, 以後急速に減少した。病原ウイルスはコックサッキーA16型であった。図4, のインフルエンザ様疾患は, 過去4年間で最も大きな流行であった。昭和59年~60年の冬期はB型株が, 昭和60年~61年の冬期はA型香港株が流行した。

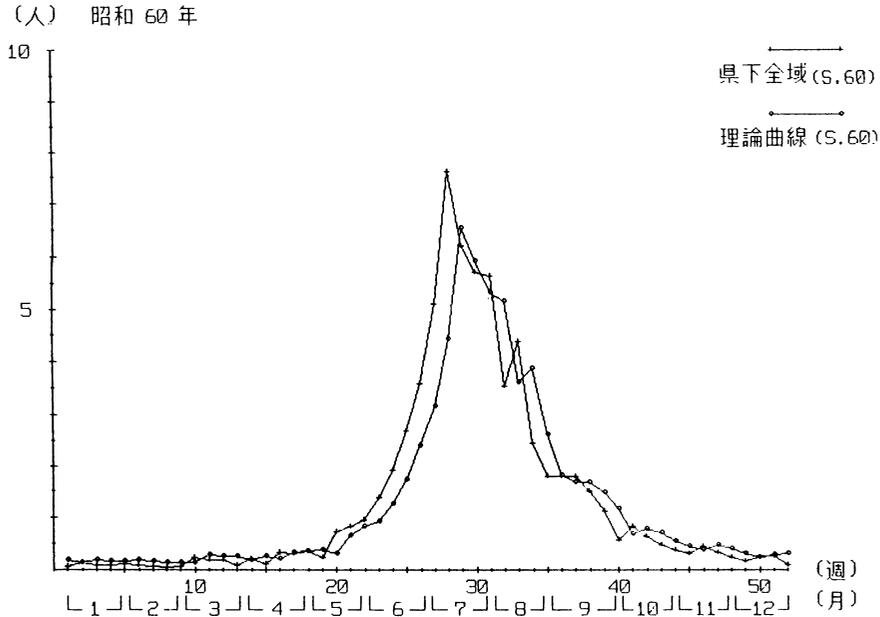


図3. 手足口病 年別1定点当りの調査状況

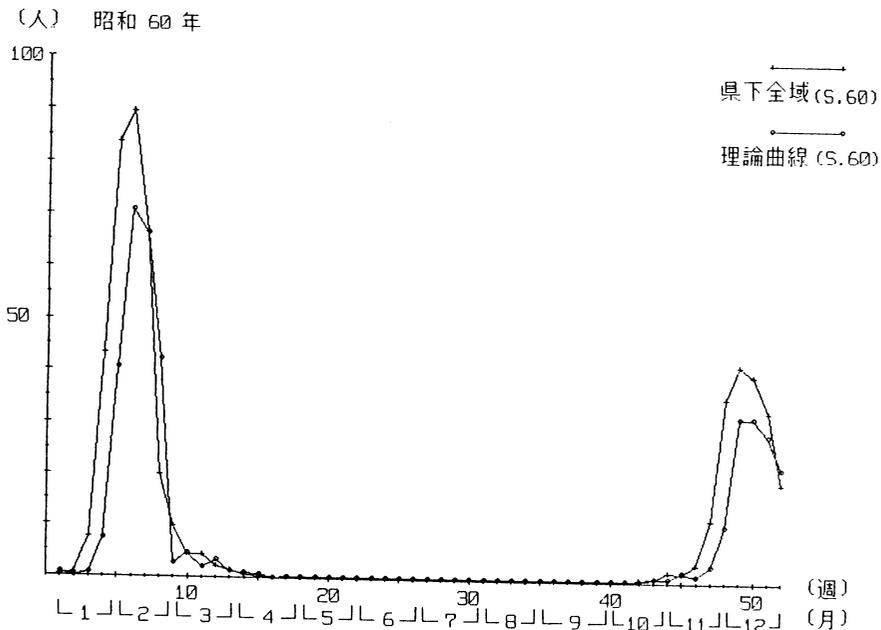


図4. インフルエンザ様疾患 年別1定点当りの調査状況

IV 考察

時系列解析のひとつである自己回帰分析は、定常時系列のデータ（ある時点での系列値がそれ以前の系列値に従属し、しかもその依存のしかたが各時点で同じであるデータを意味し、一般的には時系列データから傾向変動（Trend）を除去したデータ）の特性の解析にしばしば用いられる。近年、物理学的現象だけでなく経済指標の推移⁴⁾や大気汚染物質の日内変動⁵⁾のような社会現象の動向を予測する方法としても応用されている。自己回帰分析は、観測値に対する予測だけでなく、時系列データの時間的依存特性も抽出するという利点がある。この他にも、移動平均法や指数平滑法、スペクトル解析法⁶⁾等、その対象や目的によって様々な時系列解析法がある。通常、移動平均法や指数平滑法は、平滑化の際、項数のとり方や定数 ρ の決め方によって時間的依存特性が変化するという問題があり、スペクトル解析法⁶⁾は、定常時系列の中の複数の周期成分を各周期に分離する点では卓れているが、元来、周期成分の大きさ（Power Spector）を問題にしているため、例えば水痘や突発性発疹のような通年的に患者発生数に大きな変化がみられず、しかもパターンも恒常化している場合には、解析能が必ずしも十分に発揮されない。最近、コンピュータの普及や Cooley, J. W. and Tuckey, J. W⁷⁾等による高速フーリエ変換法（FFT : Fast Fourier Transform）の開発により、比較的容易にデータ処理が可能になったとはいえ、感染症の将来予測にはスペクトル解析は必ずしも適当ではないように思われる。

表1は、週単位の時系列データから得られた疾病別の自己回帰モデルである。このような自己回帰モデルを作成するには2つの問題がある。ひとつは項数(m)の取り方であり、他のひとつはパラメータ係数の決め方である。通常、Lagは全データNの1/10以下にとり、最適な項数(m)とパラメータ係数を得る。項数を多く取ればそれだけ原データに対するあてはまりは良くなるが、その後に観測される値に対する予測能力が落ち、しかも実用性に欠けることから、観測値と予測値の相関があまり低くない程度に項数を決定することが望ましい。経験的には項数(m)が4以下、相関係数(r)が0.8以上なければ予測モデルとしての実用性に欠ける。

図1~4は対象疾病中、患者発生パターンに特徴がみられる麻疹様疾患、水痘、手足口病、インフルエンザ様疾患について、週別の観測値と理論値を折線グラフで示した。麻疹様疾患は昭和53年に定期予防接種の対象疾病になり、それ以前は隔年毎に流行がみられたパターン

が変化し、最近では昭和59年に大きな流行が起きた。年間の患者発生パターンは、予防接種によって大きな山がなくなったが、凡そ5月~6月をピークとして不規則な増減を繰り返す。このように、予防接種対象疾病の多くは流行年を除けば年間を通して大きな山を持たず、不規則な増減を繰り返す。手足口病は、6月下旬から7月上旬にかけて一峰性の山を形成する。類似の疾病としてヘルパンギーナがある。水痘は、毎年、年間の発生パターンに大きな変化のみられない疾病である。突発性発疹もこの範疇に入る。インフルエンザ様疾患は、2月と12月に大きな二峰性の山を形成する。他の疾病に比べて比較的規則的な患者発生パターンがみられ、また発生規模も大きい。これら4疾病は、各疾病とも予測値が観測値とよく近似していることがわかる。尚、予測値が観測値より一週遅く示されるのは、自己回帰モデルは後のほうの予測値には先に観測された観測値（最適な項数(m)即ちm週前までの観測値）が関係するからである。

以上、4疾病について自己回帰分析（Auto Regressive Analysis）によるデータ解析を試みた。今後、このような結果がどのような環境条件下で成立しているか、様々な要因—気象要因や社会経済的要因、予防接種等—について定量化し、併せて構造的な分析を試みたい。
注）自己回帰モデルのパラメータ推定アルゴリズムは、大概次の通りである。

(1) データの定常性を得るため、元データをトレンドからの偏差として求める。

$$y_t = Y_t - \hat{Y}_t \quad (\text{但し } \hat{Y}_t \text{ はトレンドによる推定値})$$

(2) 有限のデータによって共分散関数を求める。

$$\hat{r}_y(\ell) = \frac{1}{T} \sum_{t=1}^{T-\ell} y(t-\ell)y(t)$$

(3) 1の値を予測誤差が最小になるように決定する。

（計算手順省略）

VI 文献

- 1) 石田名香雄（1983）：東北地方における感染症サーベイランス評価方法とスタイル，東北六県防疫月報，81，10-17.
- 2) 北山秋雄，市村博，太田原美作雄（1986）：パーソナルコンピュータによる千葉県感染症情報解析—麻疹様疾患を例にして—，第24回千葉県公衆衛生学会発表.
- 3) 尾崎タイヨ（1985）：計量モデル分析と数値算法，CBS出版
- 4) 円山由次郎（1970）：需要予測と時系列分析，日本

生産性本部.

- 5) Takakatsu Inoue (1986) : Regression analysis of nitrogen Oxide Concentration, Atmosphere Environment, 20, 1, 71-85.
- 6) 日野幹雄 (1977) : スペクトル解析法, 朝倉出版
- 7) Cooley, J. W. and Tuckey, J. W (1965) : An algorithm for the machine Calculation of Complex Fourier series, Mathematics of Computation, 19, 297-301.