

# 人口動態市区町村別統計へのベイズ統計の応用について

## (1) 標準化死亡比への応用

平子 哲夫\*<sup>1</sup> 佐伯 則英\*<sup>1</sup> 中田 正\*<sup>2</sup>

### I はじめに

保健・医療・福祉のように地域に密着して行われる行政の場合、保健所・市区町村といった比較的小さな行政単位で、その地域の特徴にあった政策立案が求められる。そうした政策立案及び執行過程においてはその地域に即した情報が不可欠になる。そのようなデータ需要に応えるため、市区町村レベルでの基礎資料として従来人口動態統計課では人口動態特殊報告保健所・市区町村別統計を、過去2回(昭和58年～62年、昭和63年～平成4年)刊行し、その中で年齢構成を考慮した加工統計として標準化死亡比<sup>(注1)</sup>、合計特殊出生率を算出してきた。しかしながら小地域間の比較や経年的な動向を標準化死亡比や合計特殊出生率で見ると、特に死亡数や出生数が少ない場合には、数値が大幅に上下し、その地域の死亡・出生の動向を把握することが困難であった。これは、標本数(死亡数や出生数)が少ないため、偶然変動の影響を受け、数値が不安定な動きを示すためである。平成5年～9年保健所・市区町村別統計を刊行するにあたり、ベイズ・モデルを適用することにより、出現数の少なさに起因する偶然性の影響を減少させ、地域間比較、経年比較に耐えうるより安定性の高い指標として標準化死亡比(ベイズ推定値)<sup>(注2)</sup>、合計特殊出生率(ベイズ推定値)を算出し、参考表として掲載した<sup>1)</sup>。本稿ではベイズ・モデルに基づくこうした指標の理論的背景、具体的な算出方法、精度、使用上の注

意を解説するが、今回標準化死亡比(ベイズ推定値)(以下ベイズ推定値という)、次回合計特殊出生率について解説する。

(注1) 標準化死亡比(SMR, *Standardized Mortality Ratio*)

$i$  市区町村における標準化死亡比を求める場合、

$d_i$ :  $i$  市区町村の死亡数

$p_i$ :  $i$  市区町村の年齢階級別人口

$Q$ : 基準人口集団における年齢階級の死亡率

$$SMR_i = \frac{d_i}{\sum_j Q \times p_i} \quad (\text{但し, } j \text{ 年齢階級})$$

となる。分母は市区町村の人口集団が基準人口における死亡率に従って死亡すると仮定した場合に期待される総死亡数である。通常100倍して表示することが多い。全国を基準とした場合、標準化死亡比が100より大きいと言うことはその市区町村の死亡状況は全国より悪いことを意味する。

(注2) 平成5年～9年人口動態保健所・市区町村別統計では従来の標準化死亡比・合計特殊出生率と区別するため、標準化死亡比(ベイズ推定値)、合計特殊出生率(ベイズ推定値)と表示している。

### II ベイズ推定値の理論的背景

(1) ベイズ推定の考え方

母数の  $\theta$  のとる値について、あらかじめ何ら

\*<sup>1</sup> 厚生省大臣官房統計情報部人口動態統計課課長補佐 \*<sup>2</sup> 同課長

かの情報が与えられて、それが確率分布の形で表されているとき、その分布を事前分布という。

$\theta$ の事前分布の密度関数を $p(\theta)$ とし、また、 $\theta$ を与えたときのデータ $X$ の確率密度関数を $f(X|\theta)$ とすると、データの観測値 $X=x$ が与えられたとき $\theta$ の条件付き密度はベイズの定理より

$p(\theta|x) = p(\theta) \cdot f(x|\theta) / \int p(\theta) \cdot f(x|\theta) d\theta$ で与えられる。これを $\theta$ の事後密度関数、その分布を事後分布という。

つまり、 $p(\theta|x) \propto p(\theta) \cdot f(x|\theta)$ となり、事前分布の情報とデータの観測値の情報から事後分布の情報が得られることとなる。

ベイズ統計学では、観測以前に我々にとって利用可能な事前情報（事前分布）を観測によって得られる標本情報（人口、死亡数）によって更新し（事後分布）、それを推定の基礎とするものである。

## (2) ベイズ推定値の算出方法

母集団の母数を $\theta_d$ （標準化死亡比）とし、事前分布としてガンマ分布 $\text{Gamma}(\alpha, \beta)$ を選択した。このとき、データの観測による市区町村の死亡数 $d$ にポアソン分布 $Po(d)$ を仮定すると、事後分布もガンマ分布 $\text{Gamma}(\alpha+d, \beta+e)$ となる。その期待値 $E$ 及び分散 $V$ は

$$E(\theta_d|d) = \frac{\alpha+d}{\beta+e}$$

$$V(\theta_d|d) = \frac{\alpha+d}{(\beta+e)^2}$$

となる。したがって、

$$\text{標準化死亡比 (ベイズ推定値)} = \frac{\alpha+d}{\beta+e}$$

ただし、 $d$ =死亡数、 $e$ =期待死亡数

$\alpha, \beta$ : 事前分布であるガンマ分布のパラメータ

この形から人口が大きい場合には通常の標準化死亡比の値に近づき、人口規模の小さい場合には地域の標準化死亡比の期待値に近づくことが理解できる。

## (注3) 事前分布の設定

ある市区町村の標準化死亡比の事前分布（ガンマ分布）の期待値及び分散は、その地域を含むより大きな地域（二次医療圏）の市区町村の標準化死亡比の期待値及び分散に等しいとして決定した（いわゆるモーメント法によっている）。今、ある二次医療圏を $K$ 、 $K$ 内の市区町村を $i$ とする。

$Q$ : 全国の年齢階級別死亡率

$pi$ :  $i$ 市区町村の性・年齢階級別人口

$di$ :  $i$ 市区町村の性別死亡数

$ei$ :  $i$ 市区町村の期待死亡数

$MK$ :  $K$ 二次医療圏内の市区町村の性別標準化死亡比の期待値

$VK$ :  $K$ 二次医療圏内の市区町村の性別標準化死亡比の分散

とすると

$$MK = \sum_{i=1}^K SMRi \times wi$$

$$VK = \sum_{i=1}^K wi \times (SMRi - MK)^2$$

但し、

$$ei = \sum_j Q \times pi \quad wi = \frac{ei}{\sum ei} \quad SMRi = \frac{di}{ei}$$

( $j$  年齢階級)

ここで事前分布に仮定したガンマ分布は、期待値 $=\alpha/\beta$ 、分散 $=\alpha/\beta^2$ であるので、

$$MK = \frac{\alpha}{\beta} \quad VK = \frac{\alpha}{\beta^2}$$

これらを $\alpha, \beta$ について解くと

$$\alpha = \frac{MK^2}{VK} \quad \beta = \frac{MK}{VK}$$

となる。この $\alpha, \beta$ が二次医療圏の市区町村に対する性別の標準化死亡比の事前分布（ガンマ分布）のパラメータとなる。

実際の計算において人口は平成7年国勢調査の人口を5倍して使用、死亡数は平成5年～9年のデータをまとめて使用した。

## (注4) 事前分布に二次医療圏を採用した理由

二次医療圏は、概ね30万人程度の人口規模を持つ日常生活圏として地理的条件、社会的条件を考慮し医療計画を設定するために全国を覆うように定められている。従って死亡の発生に関してある程度同一性が考えられる。また事前分

布に県，市郡，二次医療圏の3通りにて算出した際には，二次医療圏を事前分布にした場合が最も分散が小さく，地域内の均一性を支持する結果となっている<sup>3)</sup>。ただし，保健所管轄区域であるとか，行政区域ではなく人口規模や地理的状况を考えて事前分布を設定することも考えられるが，全国的に統一的な状況観察を行うためには人口・死亡のデータが簡単に入手できることが実務上重要である。そのため参考表の算出にあたっては，当該市区町村を含む二次医療圏の数値に基づき，ベイズ・モデルを適用し，市区町村の数値を推定している。

### Ⅲ 標準化死亡比とベイズ推定値の比較

#### (1) 精度

市区町村の人口規模と全死因の標準化死亡比の標準誤差率(推計値に対する標準誤差の割合)との関係を図1に示している。

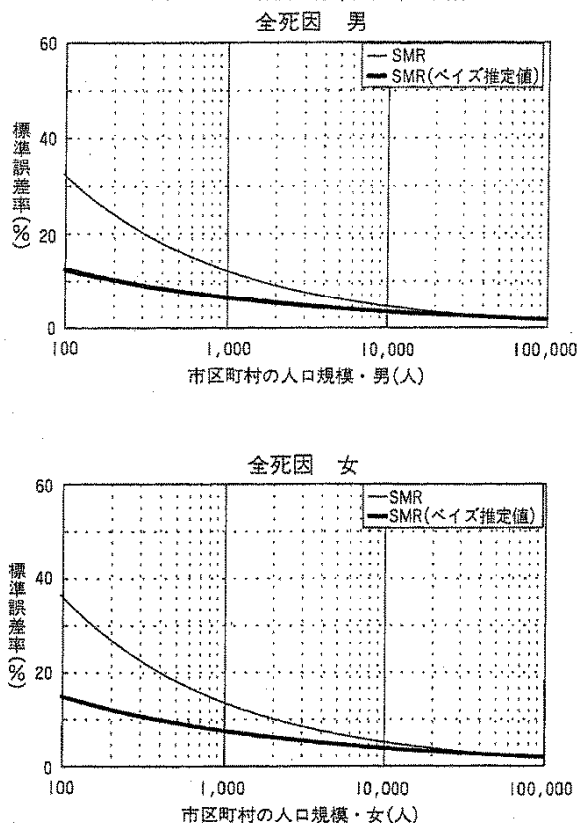
図1を見ると，標準化死亡比(ベイズ推定値)は，全死因の場合人口規模が2～3万人以上である大きな市区町村では，誤差の値は標準化死亡比，ベイズ推定値ともにほとんど精度は変わらない。逆に2～3万人以下の人口規模の市区町村では標準化死亡比よりもベイズ推定値のほうが標準誤差率が小さく，人口規模が小さくなるほどその差は大きくなっていることが読みとれる。

(注5) 標準誤差とは，推定値の標準偏差のことである。推定値を中心としてその前後に2倍づつの幅をとれば，その範囲内に約95%の確率で真の値が存在すると考えてよい。

一般に死亡数の発生が少ない地域ほど誤差は大きくなる。ある地域において死亡数が100であれば標準化死亡比の標準誤差率は10%となる。つまりその地域の標準化死亡比は，偶然変動によって生じる誤差を考慮に入れると80～120の間に約95%の確率で真の値が存在すると考えられる。

市区町村の死亡数の分布がポアソン分布に従うと仮定すれば，標準化死亡比については，

図1 人口規模と標準誤差率の関係



(注3) で定義した記号を用いると

$$\text{標準誤差} = \sqrt{\frac{di}{ei^2}} \quad \text{標準誤差率} = \frac{1}{\sqrt{di}} \times 100 (\%)$$

として算出される<sup>2)</sup>。またベイズ推定値については

$$\text{標準誤差} = \sqrt{\frac{\alpha + di}{(\beta + ei)^2}} \quad \text{標準誤差率} = \frac{1}{\sqrt{\alpha + di}} \times 100 (\%)$$

として算出される。

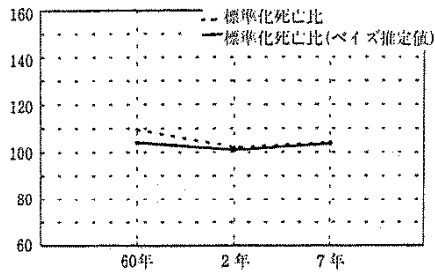
#### (2) 年次比較

昭和60年，平成2年，7年を中心にした5年間のデータより算出した標準化死亡比(ベイズ推定値)と標準化死亡比を図2-1，2-2に示している。人口規模の小さな市区町村，大きな市区町村をそれぞれ5つずつ例示としてあげている。人口規模が小さな市区町村では標準化死亡比の場合上下への変動が大きく年次の比較が困難であるのに対し，ベイズ推定値においては極端な数値の変動がなく，年次比較が可能とな

図2-1 ベイズ推定値と標準化死亡比の比較（全死因・男）

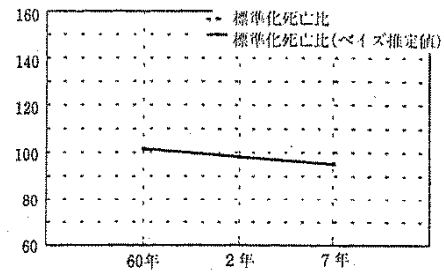
<人口規模の小さな市区町村>

A村(7年人口:600)

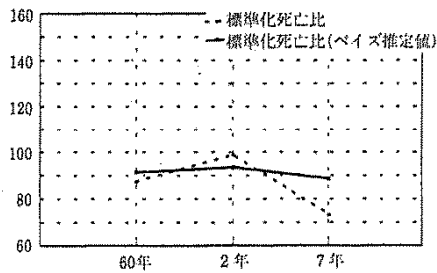


<人口規模の大きな市区町村>

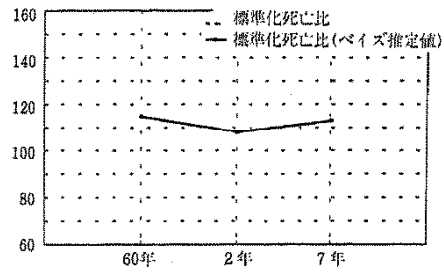
F市(7年人口:170,000)



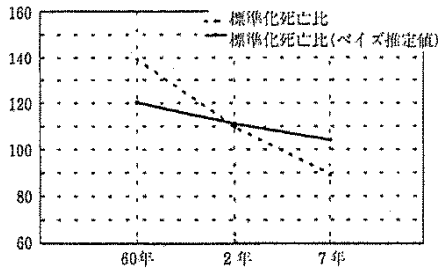
B村(7年人口:800)



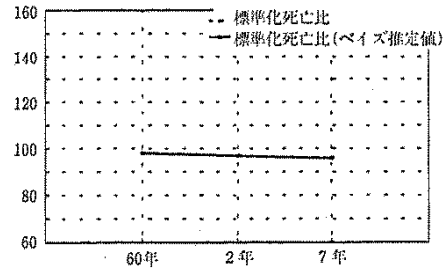
G市(7年人口:137,000)



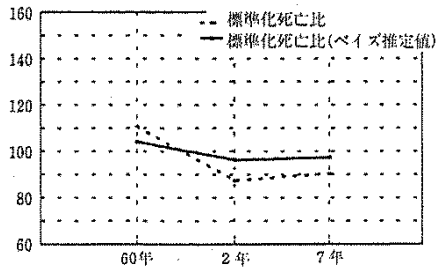
C村(7年人口:700)



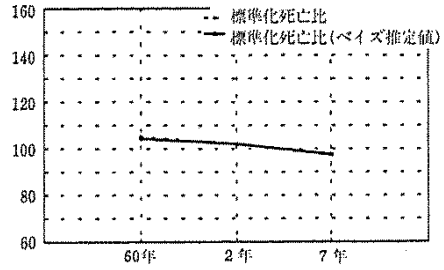
H区(7年人口:142,000)



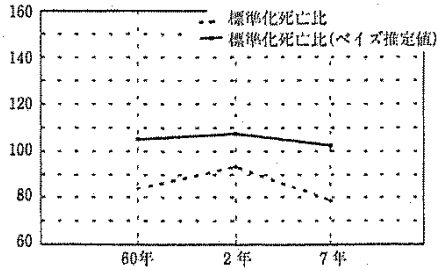
D村(7年人口:800)



I区(7年人口:121,000)



E村(7年人口:800)



J区(7年人口:117,000)

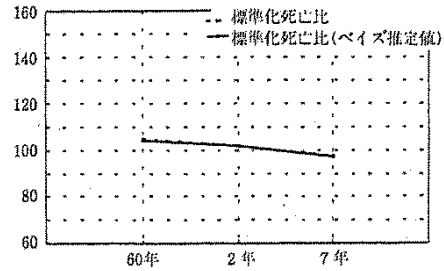
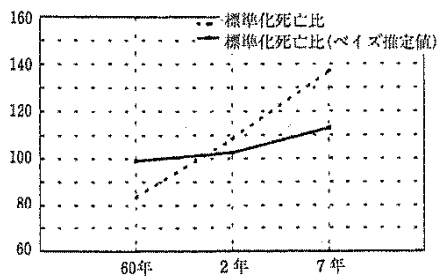
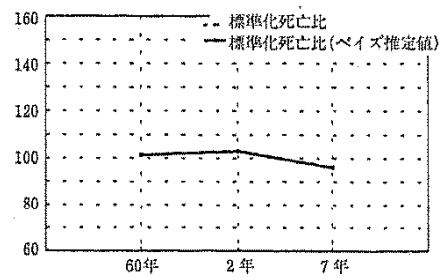


図2-2 ベイズ推定値と標準化死亡比の比較（全死因・女）

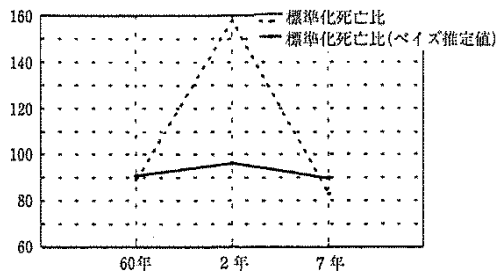
＜人口規模の小さな市区町村＞  
A村（7年人口：600）



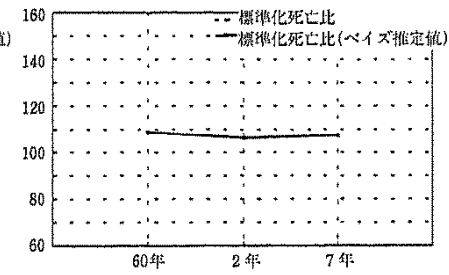
＜人口規模の大きな市区町村＞  
F市（7年人口：190,000）



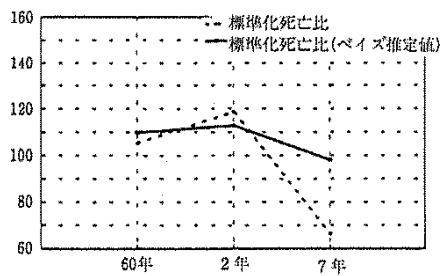
B村（7年人口：600）



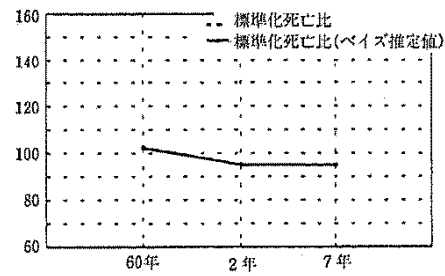
G市（7年人口：161,000）



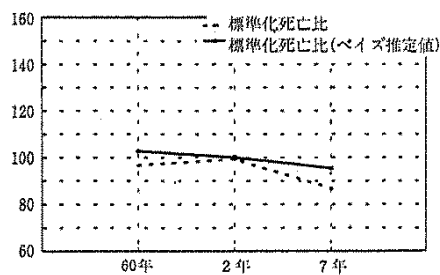
C村（7年人口：800）



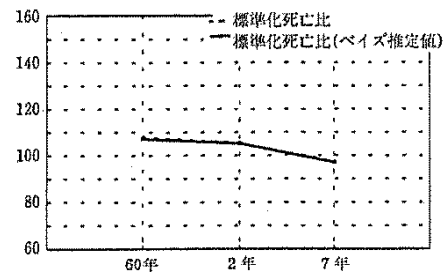
H区（7年人口：153,000）



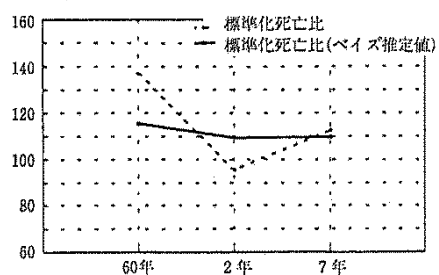
D村（7年人口：700）



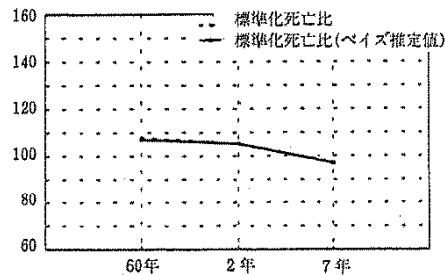
I区（7年人口：129,000）



E村（7年人口：800）



J区（7年人口：122,000）

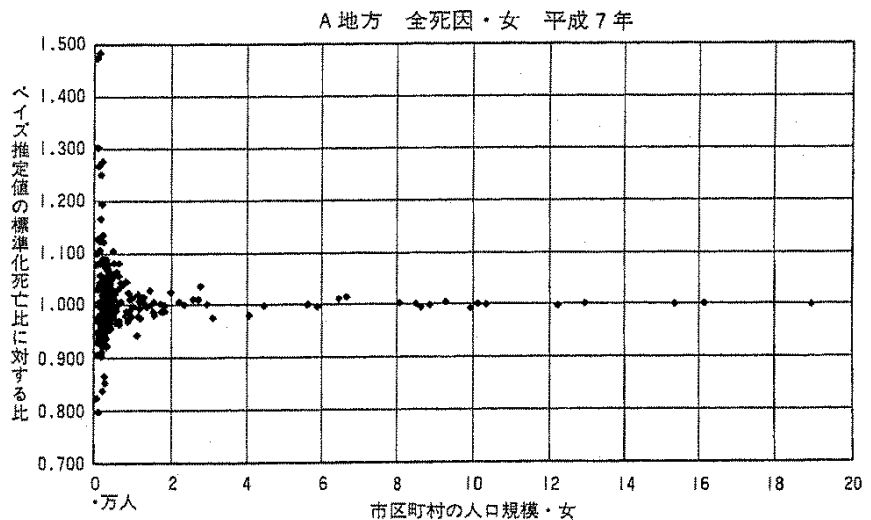
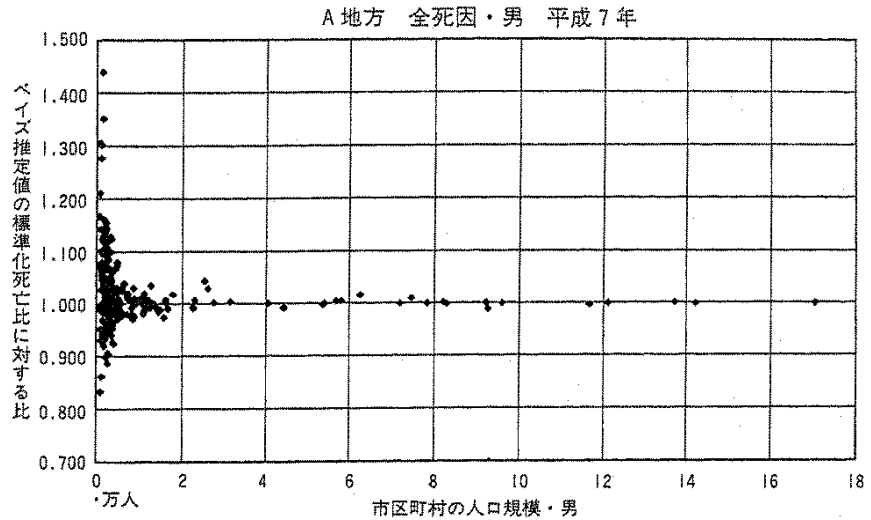


っている。これに対し、人口規模の大きな市区町村ではどちらの値を使用しても値に変化はなく、ベイズ推定値は人口規模の小さな市区町村にとって有用な方法であることが理解できる。

### (3) 人口規模との関係

A地方の全死因・男の人口規模とベイズ推定値の標準化死亡比に対する比をみたものを図3に示している。人口規模が小さいほど割合のばらつきは大きく、人口規模が2～3万人以上ではほとんど比が1である。つまり、全死因の場合人口規模が2～3万人を境にして、それ以下の市区町村ではベイズ推定値と標準化死亡比の値が異なる。

図3 人口規模とベイズ推定値の標準化死亡比に対する比の関係



### (4) 疾病地図での比較

四国においてベイズ推定値と標準化死亡比につ

いて描いた疾病地図を図4-1～図4-2に示している。両者それぞれの四国における平均値、標準誤差を用いて、平均値を中心にその前後0.5×標準誤差、1.5×標準誤差を加えたところで値を切り分け、5段階に表示した。そのため両者を比較するとベイズ推定値においては標準誤差が小さくなっているため、階級の値が異なることに注意を要する。

地図を比較するとベイズ推定値においてはより地域性が明確に描出されていることが見て取れる。図4-1を観察してみると、標準化死亡比では人口規模が小さな市区町村が多い山間部に、極端な値を示す白で塗られた地域がちりばめられている。また周囲の市区町村と比較して全く

異なる死亡状況が頻繁に生じるとは考えにくい。が、階級の分布が周囲の市区町村とは必ずしも滑らかな分布を形成しておらず、地理的關係も理解しがたい。これに対しベイズ推定値においては、標準化死亡比と比較して周囲の市区町村から滑らかに階級が移行していることが観察される。

## IV 結果の考察

平成5～9年人口動態保健所・市区町村別統計において、性別・全死因の標準化死亡比（ベイズ推定値）が市区町村別に算出されたものが掲載されている<sup>1)</sup>。報告書には算出された数値

図4-1 四国地方・全死因・男 平成5～9年

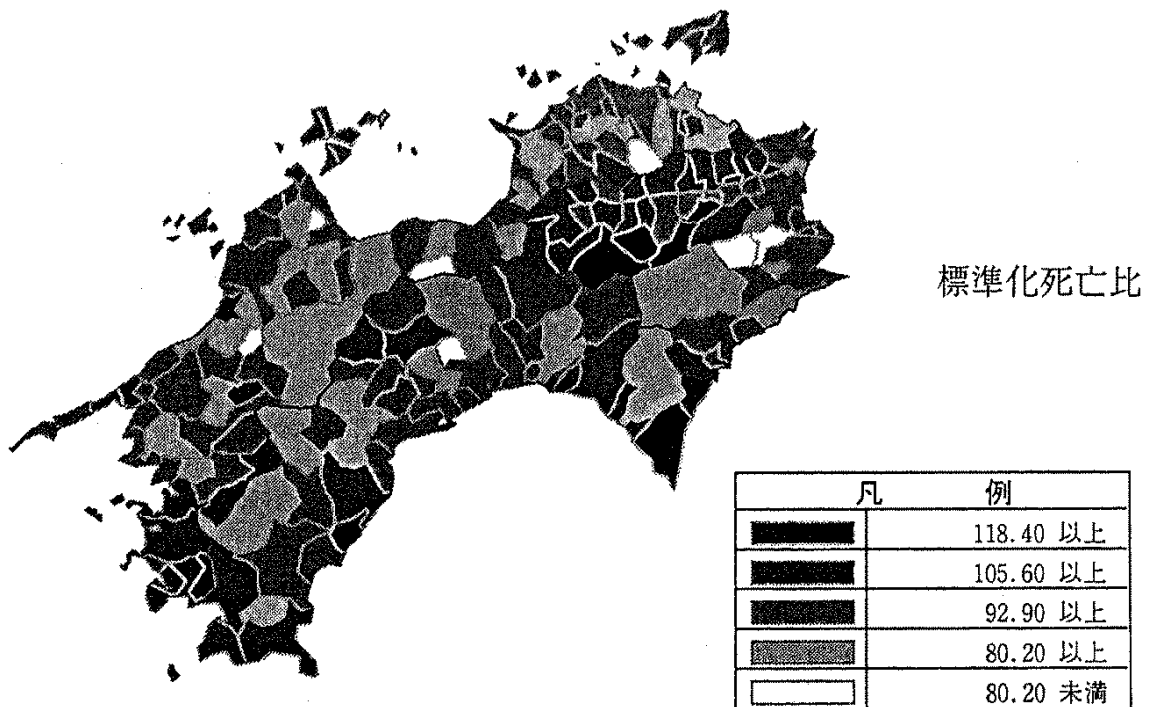
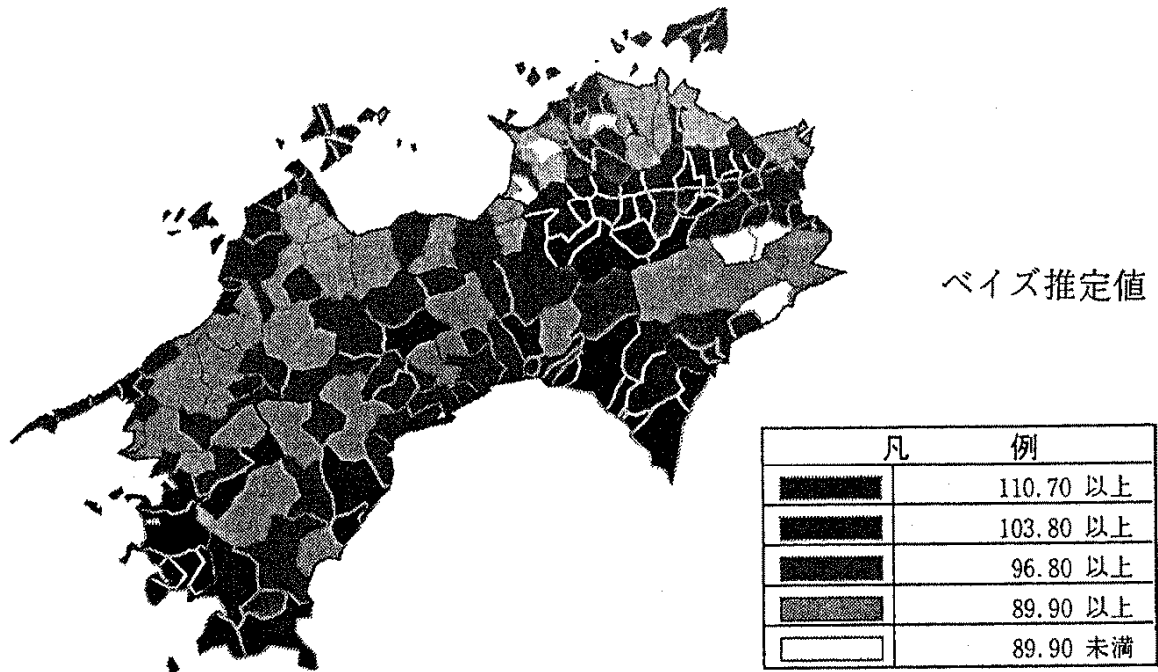
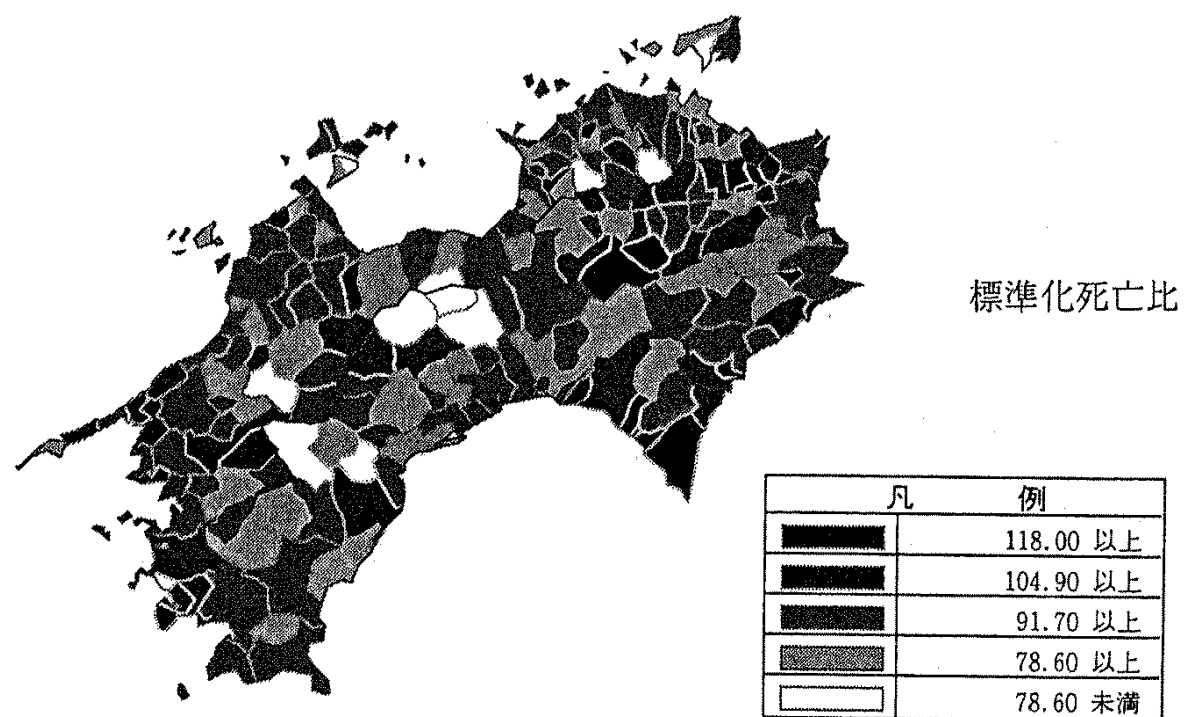
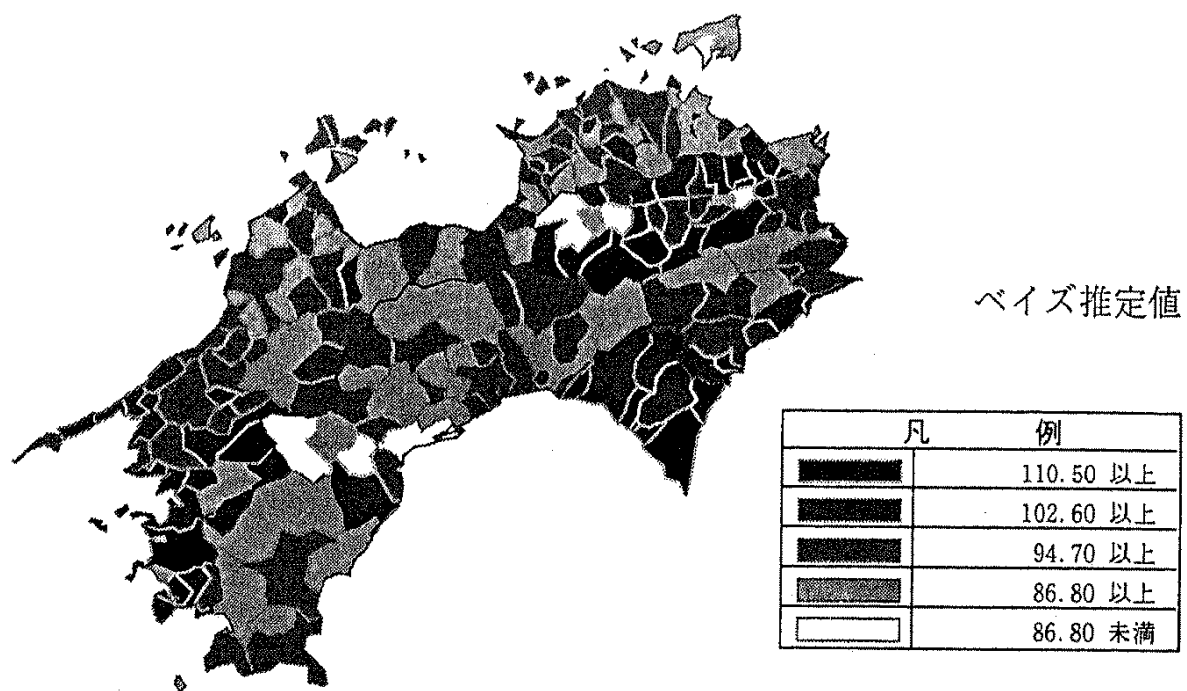


図4-2 四国地方・全死因・女 平成5～9年





と市区町村別地図が掲載されている。ここでは結果の詳細については紹介せず、考察のみ行う。

市区町村別の地図であるので、県内の地域間の分析が可能になっており、同一県内においても死亡状況のばらつきがあることが観察される。全国を統一した基準で市区町村別に段階分けしており、死亡状況に関し全国で比較した場合どのあたりの位置づけであるかが市区町村別地図として表示されている。ベイズ推定値の場合高位・低位に人口規模の小さな市区町村がほとんどを占めるということはない。今回の算出した値が高位・低位30位に含まれる市区町村の中で、死亡数が100未満(標準化死亡比の標準誤差率が10%)の市区町村が占める割合は、ベイズ推定値では全死因・男16.7%, 全死因・女18.3%であるのに対し、標準化死亡比では全死因・男53.3%, 全死因・女47.5%となる。つまり標準化死亡比では相当数が偶然に起こり得る誤差の影響を受け、算出された数値が極端な値となっていると解釈できる。

また、特に今回の結果を評価するにあたり、平成5年に起きた北海道南西沖地震、平成7年に起きた阪神・淡路大震災の与えた影響を考慮に入れるべきである。

## V ベイズ推定値の使用上の注意

### (1) 死亡数が0の場合

ベイズ推定値は、市区町村において死亡数が0の場合であっても、潜在的な死亡状況を示す値として数値が表章されており、基本的には値は0にはならない。但し二次医療圏で死亡数が0の場合にはパラメータの推定ができない。

### (2) ベイズ推定値の算出の際の仮定

今回のベイズ推定値は前述のようにいくつかの仮定に基づいて算出されている方法である。その中でも特に二次医療圏において死亡の状況が同じであるという仮定の適合性の問題がある。

個々の地域に特異な状況があれば、個々の事例において別途検討することが必要である。今回の算出は、全国ベースの算出を目的とし、仮定がおおむね成立する場合に人口規模が小さな市区町村の値を安定化させるものであることに留意すべきである。

## VI おわりに

平成5年～9年保健所・市区町村別統計について掲載された標準化死亡比(ベイズ推定値)について解説した。ベイズ推定値は小地域の分析方法として有用である。その特性が十分に理解され、各地域における保健・医療・福祉等の施策の推進、評価に利用されることを願うものである。

### 謝辞

本報告書の作成にあたり、多くの方々からの助言を賜った。特に鈴木雪夫先生(東京大学名誉教授、多摩大学大学院教授)、丹後俊郎先生(国立公衆衛生院)にベイズ統計学についてご指導いただいた。ここに深甚なる感謝の意を表すものである。

### 文献

- 1) 厚生省統計情報部 平成5～9年人口動態保健所・市区町村別統計
- 2) 鈴木雪夫, 福富利夫 小地域における死亡指標 厚生指標 1980; Vol27; 3-19.
- 3) 鈴木雪夫 平成元年度厚生科学研究「小地域における死亡指標—ベイズ統計学からのアプローチ」研究会報告書
- 4) 丹後俊郎 死亡指標の経験的ベイズ推定量について—疾病地図への適用—, 応用統計学 1988; 17: 81-96.
- 5) 丹後俊郎 疾病地図と疾病集積性—疾病指標の正しい解釈をめざして— J. Natl. Inst. Public Health 1999; 48: 84-93.