

## 死亡率曲面に対する平滑化適用基準の検討

ネギシ セイイチ ヨコヤマ ヒデアキ  
根岸 聖一\* 横山 英明\*

**目的** 年次・年齢階級別死亡率（死亡率曲面）に対して、死亡構造の特徴を明らかにするための平滑化を行った。用いた平滑化法は重み付き平均を繰返すものである。そこで、その適切な繰返し回数を判定する方法を考案した。幾つかの傷病別死亡率データにその方法を適用し、その適否を検討することが目的である。

**対象と方法** 対象とした死亡率は我が国の全悪性新生物、白血病、自動車交通事故である。用いた平滑化は単純な重み付き平均を繰返す方法である。繰返し回数が満たすべき第1の基準として、平滑化後の値が元の値からみて統計的に異ならないこと、すなわち、比率の信頼区間から外れない範囲で平滑化を行うこととした。具体的には次の手順によって適切な回数を判定した：①年次・年齢階級別死亡率の実測値に対して95%信頼区間を求める。②平滑化後の値が信頼区間から外れる割合が、5%を統計的に上まわらない回数を許容範囲内とする。

さらに、10回以上繰返しても信頼区間から外れる割合が5%以下の場合、第2の判定基準として平滑化の収束条件を設定し併用した。

**結果** 悪性新生物では、適切な平滑化の繰返し回数は3回であった。白血病では、10回繰返しても信頼区間を外れる割合が5%以内であったので、平滑化の収束条件をみる第2の基準を適用した。その結果5回が適切であると判定された。自動車交通事故では、10～14歳階級と15～19歳年齢階級の間は特異的であり、死亡率が急激に変化しているので、信頼区間の判定対象から除くことにした。その結果、適切な回数は5回であった。

**結論** 本報で用いた平滑化法は直観的で分かりやすく、また実行することも容易である。本報で示した2つの基準は、適切な繰返し回数を客観的に決定することができるので、死亡率の3次元的構造を把握するために有用と思われる。

**Key words** : 年次別・年齢階級別死亡率, 重み付き平均, 平滑化, 信頼区間

### I はじめに

死亡率は重要な保健衛生指標であり、その年齢・年次効果を分析することによって多くの知見が得られる。そのため色々な分析方法、例えば Age-Period-Cohort モデルやコホート死亡率の分析<sup>1)~6)</sup>などが行われている。それらは年次・年齢階級別死亡率データに対してモデルを想定し、年次効果・年齢効果・コホート効果を抽出しようとするものである。しかし、そこで使われているモデルが妥当かどうかを判断することは難しい。一方、何らかのモデルを想定せずとも、年次・年齢

と死亡率の関係がよく見通せるように工夫された表現方法を用いれば、それにより多くの知見が得られる<sup>7,8)</sup>。そのような立場から、死亡状況を3次元的に観察するグラフィックスシステムを作成した<sup>9)</sup>（以下、前報とする）。

ここでは死亡率は年次・年齢に関して滑らかであるとする。そのとき死亡率の実測値データに平滑化を行えば、それが持つ傾向や特徴はより明瞭になる。それで前報の死亡率観察システムでは平滑化を行うようにした<sup>9)</sup>。そこで用いた平滑化法は重み付き平均を繰返すものであり、死亡率曲面を観察しながら繰返し操作を続け、適切と思われる回数まで行うものであった。そのため客観的に妥当な平滑化回数をどのように定めたらよいかという問題が残されていた。すなわち、平滑化の回数が足りなければ当初の目的が果たせず、逆

\* 自治医科大学・医学情報学研究室  
連絡先：〒329-0434 栃木県河内郡南河内町祇園 2-36-3 自治医科大学 2号館ビル 2階  
自治医科大学医学情報学研究室 根岸聖一

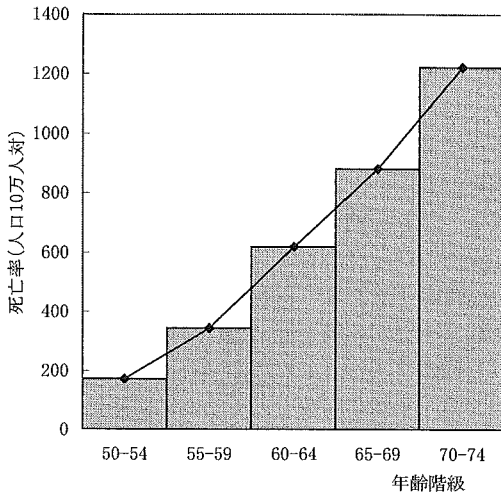
に多すぎると実態からかけ離れた姿を観察することになってしまう点である。そこで過不足のない平滑化を、直観によらずに客観的な基準に基づいて実現する方法が望まれる。そのための一方法を考案し、幾つかの典型的な例に適用し検討した結果を以下に報告する。

## II 資料と方法

年齢階級別死亡率は人口動態統計<sup>10)</sup>の値を用いた。ただし、5歳年齢階級別死亡率に基づいて等高線地図を描くと等高線が折線的で粗さが目立つこと、また、年次に関しては1年毎であることから、さらに1歳毎の死亡率を線形補間によって求めた。すなわち、各5歳年齢階級別死亡率は、図1に示すように、その階級の中央年齢の死亡率であるとし、他の年齢は線形補間した値を用いた。また、悪性新生物と白血病の85歳以上の死亡率は便宜的に87歳の死亡率とし、自動車交通事故の80歳以上は82歳の死亡率とした。こうして悪性新生物と白血病に関しては0歳～85歳までの、自動車交通事故に関しては0歳～80歳までの1歳毎の死亡率を推定した。また、年次に関しては、悪性新生物と白血病では1958年～1994年の37年間分について、自動車交通事故では1963年～1994年の32年間分を対象にした。このようにして年次×年齢のセルに配される死亡率のデータセットを作成した。

用いた平滑化法は、対象セルとその最近接4セ

図1 年齢階級別死亡率から年齢別死亡率への補間  
悪性新生物死亡率（男1993年）の例



ルとの重み付き平均を繰り返すものである<sup>9)</sup>。その重み係数の与え方を図2に示す。ただし、0歳や85歳のようにデータの端では、欠けたセルの重みは中心セルの重みに加えることにした。これらは前報<sup>9)</sup>と同じである。

平滑化法に対する適切な繰り返し回数を求めるために次のように考えた。この平滑化法では繰り返し回数が増すほど平滑化の効果は大きくなる。一方、繰り返し数が多くなると、平滑化後の値が実測値から大きく離れて、その食い違いが無視し得ない場合も起こる。したがって実測値からのズレをある程度の範囲内に抑えることが、適切な平滑化のための条件として必要である。

その条件を具体的に示す量として信頼区間に基づく基準を設定した。すなわち、元の実測値である“年次×5歳年齢階級”別死亡率に対して、それぞれの95%信頼区間を求め、平滑化後の値がそれから外れる割合は5%付近に留まることを条件とした。

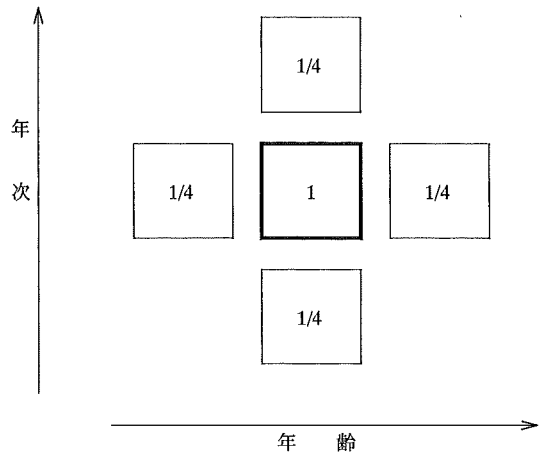
具体的には、適切な平滑化回数を求める手順は以下の通りである：

- ① 実測値であるj年次、5歳年齢階級iの死亡率  $m_{ij}^{(0)}$  の95%信頼限界を求める (Nは分母人口)：

$$m_{ij}^{(0)} \pm 1.96 \sqrt{\frac{m_{ij}^{(0)}(1 - m_{ij}^{(0)})}{N}} \quad (1)$$

- ② n回平滑化後の値  $m_{ij}^{(n)}$  が、上式の信頼限

図2 重み付き平滑化におけるセル位置と重み係数（相対値）



界から外れるセルの割合  $\hat{P}_{out}(n)$  を求める。

③ セル数  $K$  が有限なので  $P_{out}$  の95%信頼限界を求める：

$$P_{out(\pm)} = 0.05 \pm 1.96 \sqrt{\frac{0.05 \times 0.95}{K}} \quad (2)$$

④  $\hat{P}_{out}(n)$  が(2)式の範囲内であれば、そのときの回数は許容範囲内であるとする。

⑤ また、 $\hat{P}_{out}(n)$  は離散的に変化するの、 $n$  回目の  $\hat{P}_{out}(n)$  が④の許容範囲を多少超えても、 $(\hat{P}_{out}(n) - P_{out(+)})$  の大きさが、 $(P_{out(+)} - \hat{P}_{out}(n-1))$  と同程度の時には、 $n$  回も許容範囲とする。そして  $(n-1)$  回の等高線地図に粗さが目立つならば、 $n$  回を適切な回数として採用する(後の図4)。

なお、死亡数が5未満の時には2項分布の正規近似が不適当なので、(1)式の信頼限界は2項分布に立ち戻ってF分布による方法で求めた<sup>1)</sup>。

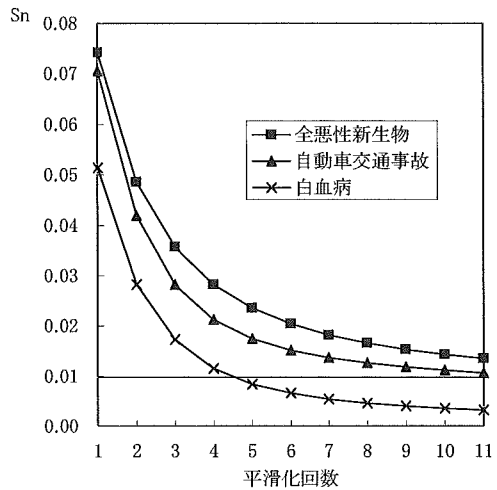
繰り返し回数  $n$  が少ない段階で信頼区間から外れるセルの付近は、死亡率が急激に変化すると場所なので、死亡率観察システムにそれを表示するように機能を追加した。すなわち等高線図上に、信頼区間の下限よりも低い値となったセルは○記号で表し、逆に信頼区間の上限を越えたセルは●記号で表すようにした。

後の白血病でみるように、死亡率が全年齢階級にわたってかなり小さいケースでは、平滑化を10回以上繰り返しても、信頼区間を外れるセルの割合  $\hat{P}_{out}(n)$  が(2)式の範囲内に留まることがある。すなわち、視覚的に観察すると平滑化の程度は十分であると思われても、平滑化を更に多数回繰り返さなければ  $\hat{P}_{out}(n)$  が5%付近に達しないケースがある。そうしたケースに対処するため、第2の基準として、平滑化の収束の程度を測る量を導入した。すなわち次の(3)式による量  $S_n$  を評価することにした。

$$S_n = \frac{\sum |m_{ij}^{(n)} - m_{ij}^{(n-1)}|}{\sum (\text{信頼区間})_{ij}} \quad (n \geq 1) \quad (3)$$

平滑化が十分になれば  $m_{ij}^{(n)}$  は収束して分子は小さくなる。上式は、信頼区間に対する各回毎の平滑化による相対的な変化の大きさを表すものである。この  $S_n$  が0.01以下になれば、信頼区間幅の1%以下の変化となるので、収束は十分であると判断することにした。

図3  $S_n$  の値による平滑化の収束の程度



### III 結 果

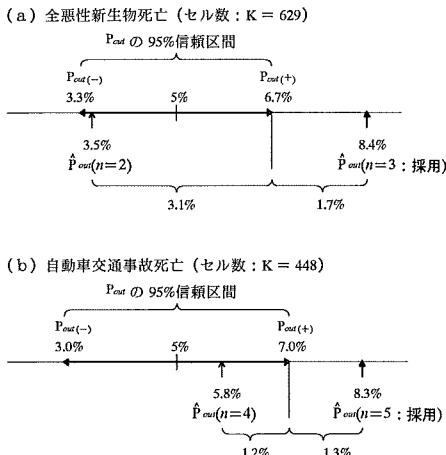
死亡率が高い例として全悪性新生物を、死亡率が比較的低い例として白血病を取り上げた。そして後述するように、社会制度的な要因が死亡率に影響を与えていると思われる例として、自動車交通事故を取り上げた。なお、本報で扱う死亡率はすべて男性についての人口10万人対の値である。

#### 1. 全悪性新生物

95%信頼区間を外れる割合  $\hat{P}_{out}(n)$  は、平滑化2回で  $\hat{P}_{out}(2) = 3.5\%$  であり、 $\hat{P}_{out}(3) = 8.4\%$ 、 $\hat{P}_{out}(4) = 13.5\%$  となった。元データの年次×年齢階級の全セル数は  $K = 37 \times 17 = 629$  なので、(2)式による  $P_{out}$  の95%信頼区間は  $P_{out(-)} = 3.3\% \sim P_{out(+)} = 6.7\%$  である。また、平滑化の収束度を表す  $S_n$  は  $S_2 = 0.049$ 、 $S_3 = 0.036$ 、 $S_4 = 0.028$  の各値であった(図3)。

以上の95%信頼区間による結果からすると、許容されるのは2回までと言うことになる。しかし、2回では  $S_2$  の値、および、等高線図の滑らかさの程度からみてもやや不十分であった。そこで図4(a)に示すように⑤の基準、すなわち、 $\hat{P}_{out}(n)$  と  $P_{out(\pm)}$  との関係も考慮して判断するケースとした。上限値と  $P_{out(+)} = 6.7\%$  と  $\hat{P}_{out}(3) = 8.4\%$  の差1.7%は、 $\hat{P}_{out}(2) = 3.5\%$  との差3.2%よりも小さいので、 $n = 3$  回も許容範囲とみなすことにして、適切な繰り返し回数は3回と判定した。以上の結果は図6に示す通りである(図5は  $n = 0$  の

図4 P<sub>out</sub> の95%信頼区間と  $\hat{P}_{out}(n)$  との関係 (白血病は第2の基準 S<sub>n</sub> で決定されるので省略)



元データの図)。

高齢者層において死亡率が増加傾向にあることは、図5に示した元データからも窺えるが、図6によって次の特徴が一層明瞭になる：①40歳台以降から加齢に伴って急増する②その加齢に伴う死亡率の増加傾向は年次が進むにつれて著しくなっ

ている。

### 2. 白血病

元データの等高線図である図7からわかるように、白血病は死亡率が低いために年次変動が相対的に大きく、等高線は入り組んでいる。このケースでは、平滑化を10回繰り返しても信頼区間を外れるセルの割合は  $\hat{P}_{out}(10) = 2.7\%$  に過ぎず、5%を越える(5.1%)には27回も要した。そこで、(3)式の平滑化の収束条件 S<sub>n</sub> を第2の指標として用いることにした。その結果5回で S<sub>5</sub> = 0.008 となったので(図3)、平滑化回数は5回が適切であると判定した。

その結果を図8に示す。これより白血病死亡率は以下の特徴を持つことが非常に観察しやすくなった：①1970年以降、高齢層で増加しており、1990年代に入ると80歳台では30に達している。②1960年～1970年では0～10歳の低年齢で死亡率が3～5と他の年代に比してやや高い。③全体的に死亡率は低く(最大値が1994年の85歳以上階級の39.5)加齢に伴う上昇は比較的緩やかである。

### 3. 自動車交通事故

図9に示すのは、1回の平滑化を行った等高線

図5 年齢×年次についての悪性新生物死亡率 (10万人対：男)

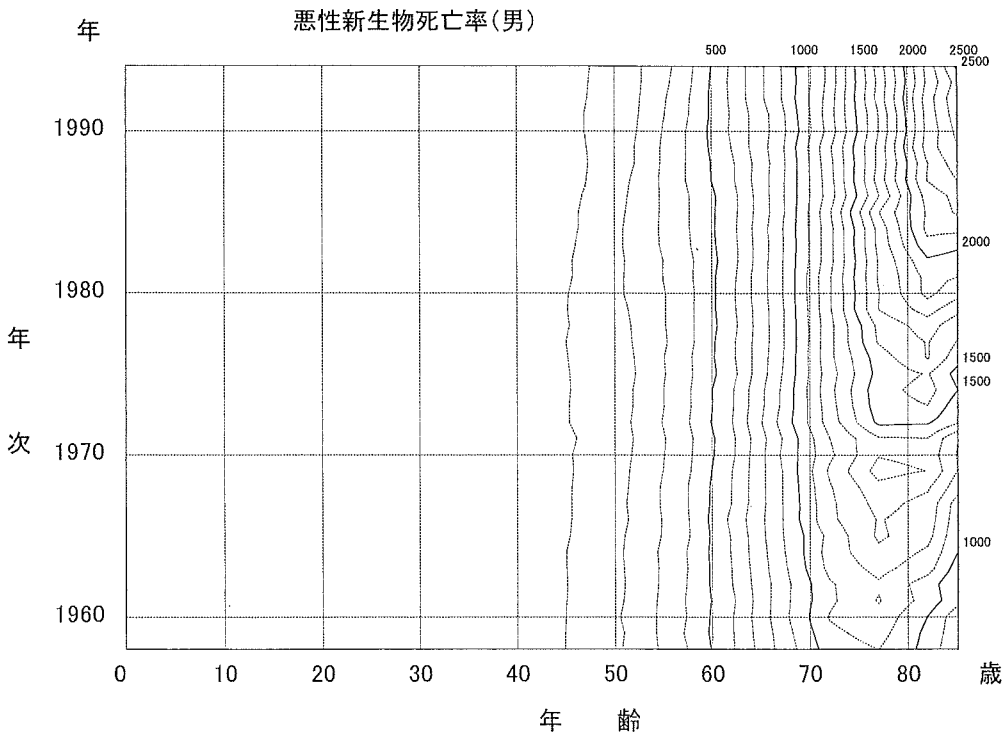


図6 年齢×年次についての悪性新生物死亡率の等高線図 (10万人対：男)  
 —重み付き平均の線り返し回数3回—  
 (●○は95%信頼限界以上・以下を表す)

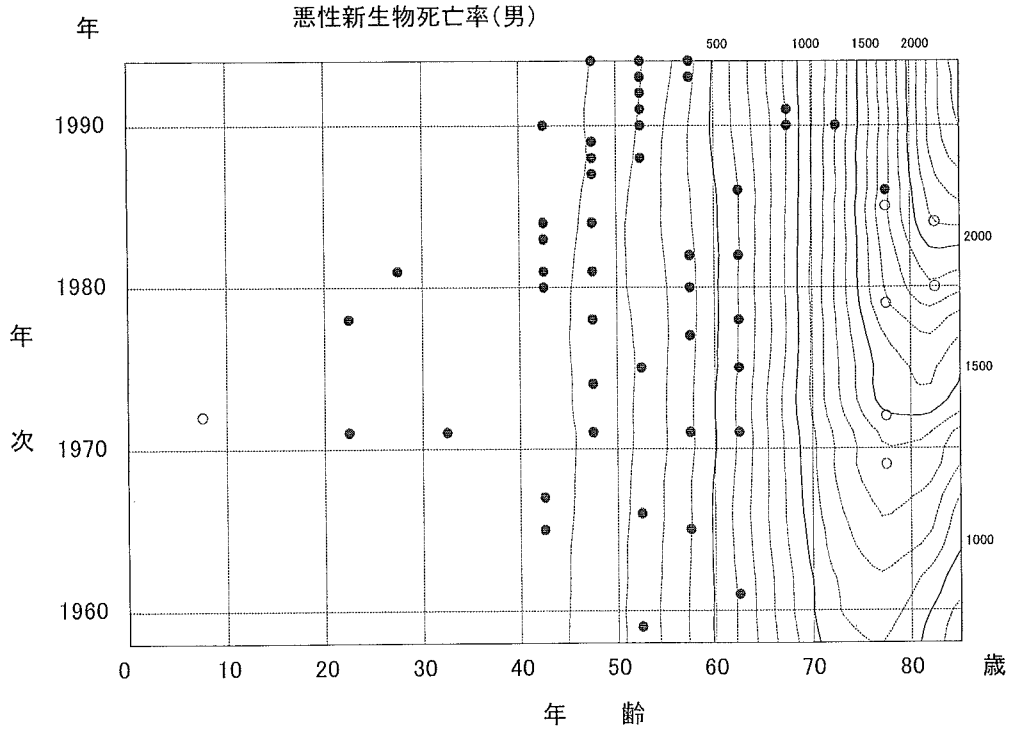


図7 年齢×年次についての白血病死亡率 (10万人対：男)

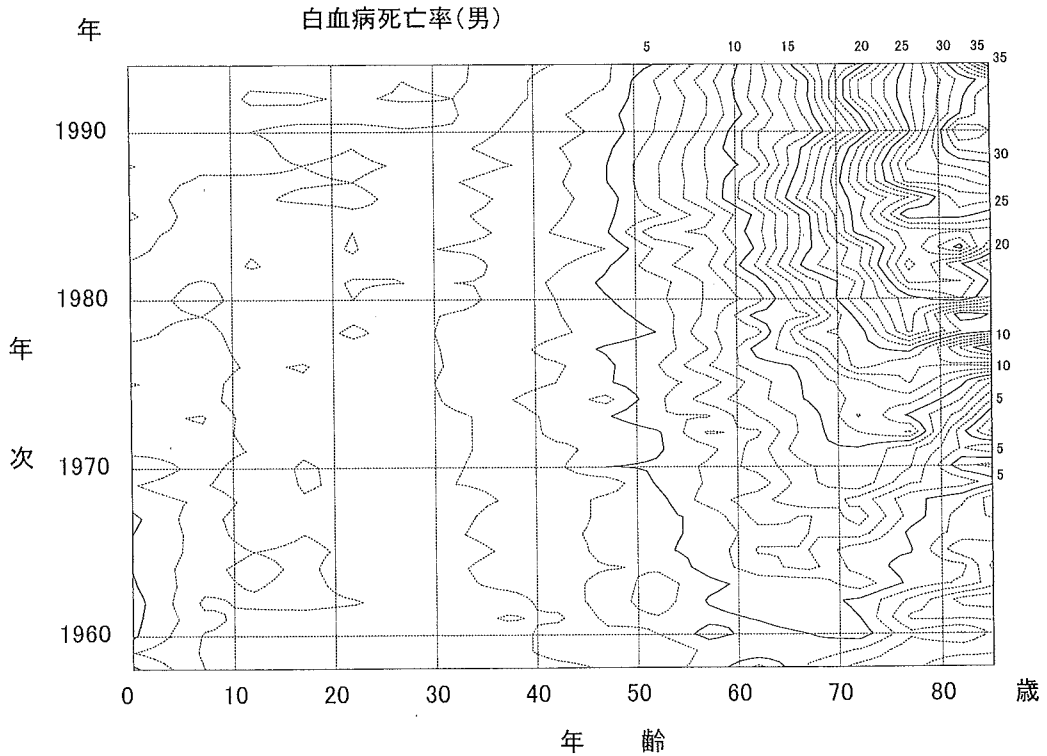
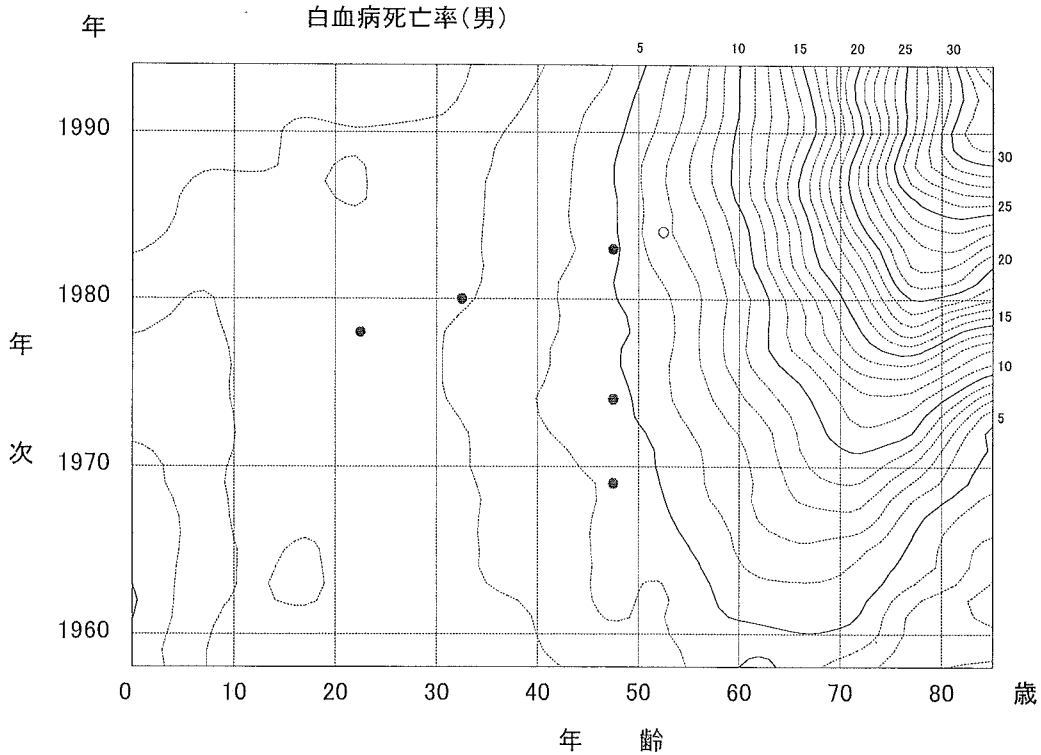


図8 年齢×年次についての白血病死亡率の等高線図 (10万人対：男)  
 —重み付き平均の繰り返し回数5回—  
 (●○は95%信頼限界以上・以下を表す)



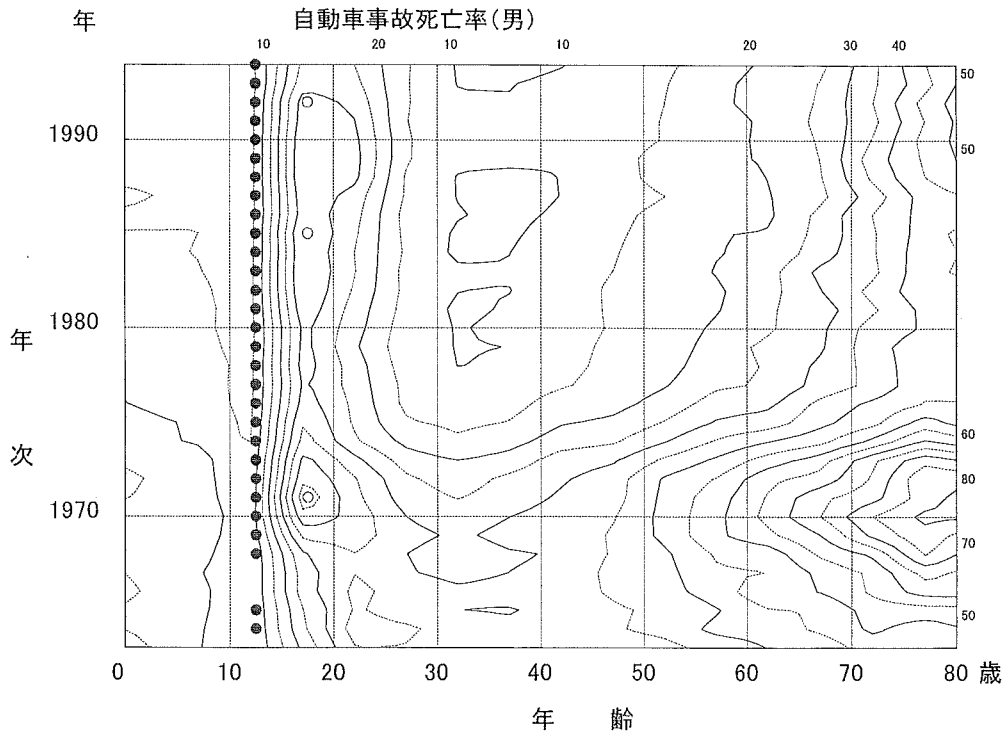
図であり、これは元データの等高線地図とほとんど同じである。全年次（1963年～1994年）を通して15～19歳での死亡率が、その前後の年齢に比して高く突出している。これは自動車交通事故死亡率（男）に特有な状況であり、実測値の年次についての平均値は、10～14歳は3.4であるのに対し、15～19歳年齢では33.6、20～24歳では29.4である。この特徴のため、平滑化を行うと図9からも明らかなように、それらの年齢階級での死亡率が大きく変化する。すなわち、15～19歳階級の死亡率は低められ、10～14歳階級の死亡率と20～24歳階級は引き上げられる。そのため重み付き平均を1回行うだけで、10～14歳階級に相当するセルのほとんどで信頼区間の上限を越えてしまう。結果として信頼区間を外れる割合は $\hat{P}_{out}(1) = 6.3\%$ にもなってしまう（図9）、平滑化を行うことはあたたかも不適当なケースであるかのようである。

しかし自動車交通事故については、「死亡率が滑らかである」と言う平滑化の前提条件は10歳～19歳の間では次の理由により疑わしい。すなわ

ち、自動二輪車の免許取得可能年齢は16歳、普通乗用車が18歳であり、未熟・無謀運転に起因する自動車死亡事故が20歳前後に非常に多いことである。したがって、免許取得可能年齢の前後では死亡率が滑らかであるとの仮定にはやや無理がありそうである。そこで、当該年齢（10～14歳、15～19歳）は、信頼区間を外れる割合の計算からは除くことにした。しかし、平滑化の対象には含むことにした。なぜなら、これらの年齢階級では平滑化によって信頼区間からは外れるものの、死亡率曲面の傾向や特徴を観察する上では支障となるほどの変化ではなかったからである。

その結果、信頼区間を外れる割合  $\hat{P}_{out}(n)$  は平滑化4回で  $\hat{P}_{out}(4) = 5.8\%$ 、5回で  $\hat{P}_{out}(5) = 8.3\%$ 、6回で  $\hat{P}_{out}(6) = 11.0\%$ であった。なお対象セル数は、上記の若年層を除いたので  $K = 14 \times 32 = 448$ セルであるから、(2)式の信頼区間を外れるセルの割合の95%信頼区間は  $P_{out(-)} = 3.0\% \sim P_{out(+)} = 7.0\%$ である。この基準から許容されるのは4回までであるが、等高線図を観察すると4回では

図9 年齢×年次についての自動車交通事故死亡率の等高線図 (10万人対:男)  
 全年齢を対象—重み付き平均の繰り返し回数1回—  
 (●○は95%信頼限界以上・以下を表す)



やや不十分であった。そこで、悪性新生物と同様に、図4(b)に示すように  $P_{out(+)}$  と  $\hat{P}_{out}(n)$  との関係を見ると、 $P_{out(+)} - \hat{P}_{out}(4) = 1.2\%$  は  $\hat{P}_{out}(5) - P_{out(+)} = 1.3\%$  と同程度であった。そこで、 $n=5$  回も許容可能な範囲であるとした。

結果を図10に示す。これより、最初にあげた特徴の他に、25歳～60歳ぐらいまでは大きな変化はないが60歳以降の高年齢層では年齢が上がるにつれて死亡率が増加している様子がはっきりしてきた。

#### IV 考 察

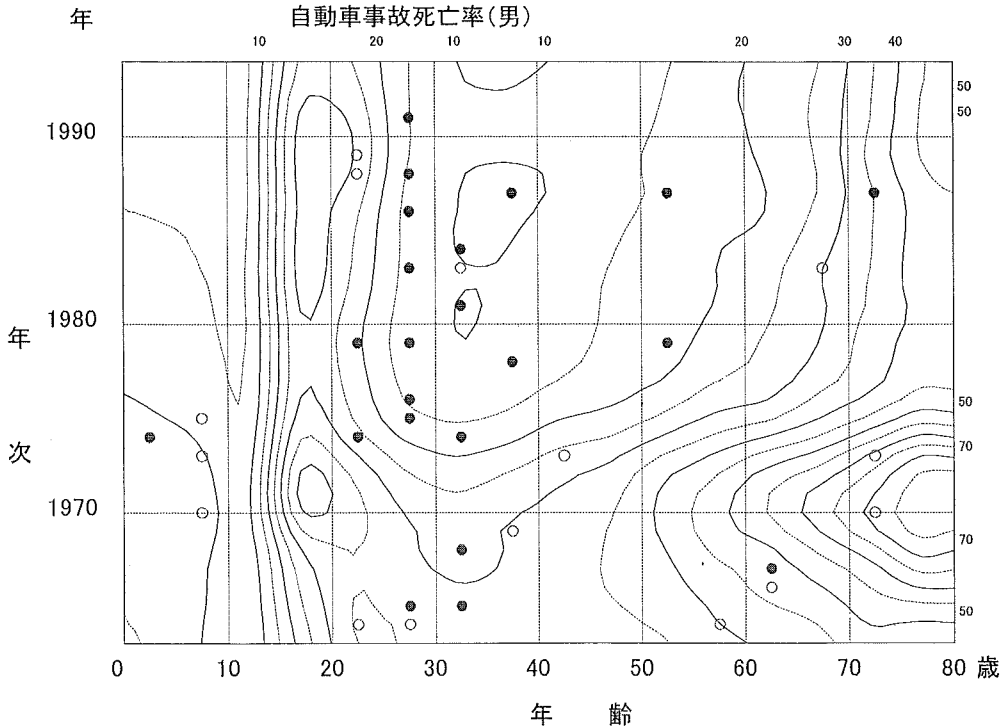
本稿で扱った平滑化法は移動平均法に類似した方法である。しかし、通常の移動平均法では滑らかな成分(真値)は低次の多項式で近似できると想定して、重み係数を工夫することに主眼を置き、1回だけの平滑化を行っているようである<sup>12)</sup>。それに対し、ここでの方法は重み係数を予め適当に決めておいて、操作回数を加減することによって平滑化を行っている。1回だけの平滑化で済ます

ためには、適切な重み係数を選択する必要があるが、死因毎にそれを定めることは大変な煩わしさを伴う。そこで本報では適切な操作回数によって平滑化の程度を定める方が実用的であると考えた。

従来の移動平均法における平滑化の程度を決める指針としては、平滑化の程度と実測値からの乖離とのバランスを計るものがある。すなわち、平滑化されたデータ系列  $m^{(1)}$  の滑らかさの程度 ( $m^{(1)}$  の差分) と、 $m^{(1)}$  と実測値  $m^{(0)}$  との差はトレードオフの関係にあるので、滑らかさの程度と実測値との差の1次結合である評価関数を作り、それが最小となる場合が求めるべきものとする方法である<sup>12,13)</sup>。この指針に基づくものの例として、平滑化スプラインなどがある<sup>13)</sup>。しかし、実際上は評価関数の最小値を与える1次結合をユニークに定めるのは困難であり、例えばSAS/GRAPHのG3GRIDプロシージャでは、平滑化の程度を決めるのは使用者が恣意的に行うようになっている<sup>14)</sup>。

本報の方法は、実測値とのズレを信頼区間に基

図10 年齢×年次についての自動車交通事故死亡率の等高線図（10万人対：男）  
 除10～14歳階級と15～19歳階級—重み付き平均の平滑化回数5回—  
 (●○は95%信頼限界以上・以下を表す)



づいた許容限度内に抑えることを第1の条件として平滑化を実現する。このような方法は筆者らの知る限りでは例が無いようであるが、信頼区間は推測統計学の基本であり受け入れやすい方法と思われる。また、本報では死亡率を扱ったが、信頼区間を求め得る他の量、例えば、学童の平均身長などについて平滑化を行う際にも適用できよう。

本報では  $\hat{P}_{out}(n)$  の信頼係数は95%としたが、その取り方には任意性がある。それで信頼係数を80%とした場合についても試みてみた。悪性新生物では繰り返し数2回で14.1%、3回で24.8%であった。白血病では20%を越えたのは25回であった(21.5%)。自動車交通事故では、特異的な10歳～14歳と15歳～19歳階級を除いた場合、信頼区間を外れる割合  $\hat{P}_{out}(n)$  は3回で15.2%、4回で22.5%となった。以上のごとく、信頼係数80%での  $\hat{P}_{out}(n)$  に基づく適切な回数は、信頼係数95%を用いた時とほぼ同様であった。

しかし、自動車交通事故で特異的な10～14歳、15～19歳年齢階級(2×32年次)を除外せずに全

年齢階級について行った場合は、2回で16.9%、3回で24.6%であった。すなわち95%信頼区間を用いて全年齢階級を対象としたときには1回で標準割合を超過したので、このようなケースでは信頼係数を低く設定する方が良いようである。

本報で用いた平滑化は、死亡率曲面を平面化する方向に働くものであり、繰り返し回数を増やしていくときに信頼区間から外れるセルは、死亡率の変化が大きいところで起こる。特に、自動車交通事故の場合、死亡率は10～14歳で最小で、15～19歳ではその10倍近くにも急増していた為に、10～14歳の死亡率は1回の平滑化でも信頼区間の上限を越えるほど高められ、図9でみたように●記号が特異的に多くなっていた。このように、死亡率が急激に(非線形的に)変化するところでは、平滑化後の値と実測値とのズレが大きくなってしまふ。これに対処するためには、重み係数の与え方に何らかの工夫を行う必要があるが、それは今後に残された課題である。

なお、以上の点は自動車交通事故に特有な事柄



である。慢性疾患などを中心とした主要死因の場合、死亡率はほぼ連続的に変化しているので問題がない。例えば、全悪性新生物は、死亡率が急増する50歳付近で、死亡率が年齢軸に沿ってやや下に凸の傾向にあるとは言うものの、図1の平滑化による等高線地図を得るに際して、特に不都合な点はみいだせなかった。

次に、死亡率が小さい白血病では、平滑化を30回近く行っても信頼区間の条件を満たしていた点について考察を加えてみる。(1)式の信頼限界で、 $\hat{m} \equiv m_{ij}^{(0)}$ と略記すると、セル位置の分母人口をNとして、

$$\begin{aligned} \hat{m} \pm \Delta m &= \hat{m} \pm 1.96 \sqrt{\frac{\hat{m}(1-\hat{m})}{N}} \\ &\approx \hat{m} \pm 2 \sqrt{\frac{\hat{m}}{N}} \end{aligned} \quad (1')$$

である ( $\hat{m} \ll 1$ )。したがって、死亡率に対する信頼区間の相対的な幅は

$$\frac{\Delta m}{\hat{m}} \approx 2 \sqrt{\frac{1}{N\hat{m}}} \quad (4)$$

である。ここで、全悪性新生物と白血病の死亡率を  $\hat{m}_1, \hat{m}_2$  とすれば、多くのセルで100倍以上異なるので、 $\hat{m}_1 \approx 100\hat{m}_2$  と置くことができる。すると、両死因について(4)式で与えられる量の比は

$$\begin{aligned} \sqrt{\frac{1}{N\hat{m}_1}} : \sqrt{\frac{1}{N\hat{m}_2}} &\approx \sqrt{\frac{1}{100\hat{m}_2}} : \sqrt{\frac{1}{\hat{m}_2}} \\ &= \frac{1}{10} : 1 \end{aligned} \quad (5)$$

程度である。すなわち、死亡率が小さい白血病の信頼区間幅は、全悪性新生物よりも相対的に10倍程度広いことになる。平滑化に伴う  $\hat{m}$  の変化は  $\hat{m}$  にほぼ比例するので、同じ回数での平滑化であっても、死亡率が非常に小さい傷病では信頼区間から外れる割合  $\hat{P}_{out}$  が少ない結果になる。

白血病の例では平滑化が10回以下でも、偶然変動に伴う凹凸が十分除去されたときみなせたにもかかわらず、30回近くまで信頼区間を外れる割合が5%未満に留まっていたのは以上に述べた理由によるものと思われる。

また、信頼区間の幅に対する死亡率の変化の相対値である  $S_n$  の様子を表した図3のグラフの様子も、上で論じたことを反映した結果であるとして理解することができよう。すなわち、人口10万人対の年齢調整死亡率の平方根(1993年)は、悪

性新生物 =  $\sqrt{216.4} = 14.7$ 、自動車交通事故 =  $\sqrt{19.1} = 4.4$ 、白血病 =  $\sqrt{5.4} = 2.3$  であり、傷病間の  $S_n$  の大小順序と一致している。

さらに、対象を全国から都道府県などの小集団に取って行う場合を考えると、 $\hat{m}$  が全国と同程度の傷病であってもNが小さくなるために  $\frac{\Delta m}{\hat{m}}$  は全国の場合よりも大きくなる。したがって、そのような場合にも全痛に対する白血病のケースと類似した状況が当てはまるので、平滑化を行うことが有効であると言えよう。

ここで、操作回数を多くした場合に、遠くのセルの影響がどのように効いてくるかについて考えてみる。図1のモデルでは、年齢(年次)が3歳(3年)以上離れたセルから受ける影響は、 $n=10$  では1/5程度であるが、 $n=30$  では2/3ほどにも達する。このことから、操作回数はできるだけ少ない方が良いと言える。すなわち、白血病の例のようなケースで、必要以上に平滑化を行うことを避けるために、第2の副条件すなわち、平滑化の収束を判定する指標  $S_n$  を導入したのは以上の事情による。

悪性新生物と自動車交通事故では、適切と判断した回数での  $S_n$  は0.01よりも大きかった(悪性新生物では0.036、自動車交通事故では0.017)。これらの傷病では死亡率が比較的大きいため、(4)式の信頼区間との比が小さくなるために、平滑化の少ない段階でも信頼区間から外れる割合  $\hat{P}_{out}$  が多くなったと考えられる。これらの例では、95%信頼区間に基づく第1の基準と平滑化の収束度  $S_n$  に基づく第2の基準を、同時に十分満足させることは難しいようである。

本報での平滑化法は簡単な重み付き平均を繰り返すもので、平易かつ直観的であり、実際の死亡率データに適用し易い。そこで、適切な平滑化の繰り返し回数を、推測統計学に基づいて客観的に決定する方法を明らかにした。また、平滑化の収束の程度をみる指標を設定して、適切な回数を決するための第2の規準とした。典型的な3死因に適用した結果、本方法は死亡構造を視覚的に観察し把握する上で有用であることをみた。

(受付 2000. 5.29)  
(採用 2001. 3.23)

## 文 献

- 1) Rothman KJ, Greenland S. *Modern Epidemiology* 2nd ed. Philadelphia: Lippincot—Raven 1998; 490.
- 2) James IR, Segal MR. On a Method of Mortality Analysis Incorporating Age-Year Interaction, with Application to Prostate Cancer Mortality. *Biometrics* 1982; 38: 433-443.
- 3) Ohtaki M, Kim D, Munaka M. A Nonparametric Method for Estimating Interaction Effect of Age and Period on Mortality. *Environmental Health Perspectives* 1990; 87: 115-121.
- 4) 丹後俊郎. コホート分析における年齢・時代・出生コホートの3効果の推定について. *応用統計学* 1984; 13: 11-26.
- 5) 久永富士朗, 吉永一彦, 重松峻夫, 他. 単年出生コホートによるわが国死亡率の解析—研究方法と全死因死亡の解析. *日本公衛誌* 1990; 37: 979-990.
- 6) Robertson C, Boyle P. Age-Period-Cohort Analysis of Chronic Disease Rates I: Modeling Approach. *Statistics in Medicine* 1998; 17: 1305-1323.
- 7) Robertson C, Boyle P. Age-Period-Cohort Analysis of Chronic Disease Rates II: Graphical Approaches. *Statistics in Medicine* 1998; 17: 1325-1340.
- 8) Devesa SS, Donaldson J, Fears T. Graphical Presentation of Trends in Rates. *Am J Epidemiol* 1995; 141: 300-4.
- 9) 根岸聖一, 横山英明. パソコンによる死亡率データ観察システムの試み. *日本公衛誌* 1999; 46: 848-857.
- 10) 人口動態統計. 厚生省大臣官房統計情報部, 編. 東京: 厚生統計協会, 1963-93.
- 11) 竹内 啓, 藤野和建. 2項分布とポアソン分布. 東京: 東京大学出版会, 1981; 24-25.
- 12) 竹内 啓, 他, 編. *統計学事典*. 東京: 東洋経済, 1989; 378-380.
- 13) 柏木宣久, 柳本武美. 平滑化とそのソフトウェア. *応用統計学*, 1990; 19: 79-92.
- 14) SAS/GRAPH Reference SAS Institution Cary, NC 27513.

## CRITERIA FOR DEGREES OF A SMOOTHING METHOD APPLIED TO MORTALITY RATES

Seiichi NEGISHI\*, Hideaki YOKOYAMA\*

**Key words:** Age-period-specific mortality rates, Weighted average, Smoothing method, Confidence intervals

**Purpose** Age-period-specific mortality rates are smoothed by methods to clarify their characteristics and trends. In the present study a procedure to determine appropriate iteration times for weighted averages was devised, and examined by application to mortality data.

**Materials and Methods** Mortality data for leukemia, solid neoplasms, and traffic accidents caused by cars in Japan were employed. The smoothing method was iterating weighted averages over age-period-specific death rates and their four nearest neighbors in age-period data tables. The primary condition, which the iteration times should satisfy, was that the differences between crude data and smoothed ones were not significantly large, i.e., smoothed mortality data should be within 95% confidence intervals of crude rates. Therefore, the following steps were adopted to determine appropriate times:

- (1) estimation of 95% confidence intervals for crude age-period-specific mortality rates.
- (2) determination of iteration times under the primary condition that the proportion of cells left out of the confidence intervals should be around 5%.

In cases where the proportion of cells outside of confidence intervals was still below 5% even with 10 or more times of iteration, an index of convergence of smoothing was introduced and used to determine appropriate iteration times.

**Results** For solid neoplasms, appropriate times of iteration was three. With leukemia, the proportion of cells left out of their confidence intervals remained within 5% with iteration 27 times. Therefore, an index of convergence for the smoothing process was used to avoid over-smoothing. As a result, a value of five was obtained. In the case of car traffic accidents, even only a single smoothing did not fulfill the primary condition, because mortality rates increased very rapidly from age classes of 10-14 to 15-19, the mortality rates for the latter being about ten times larger. To deal with this, the proportion of cells left out of their confidence intervals was calculated excluding these age classes. As a result, five times was found to be appropriate.

**Conclusion** The smoothing method in this study is simple and easy to apply to mortality data. And procedure to determine appropriate iteration times is based on an inferential statistical method. Applying the method to three kinds of mortality data, it was found to be practical and useful for clarification of their characteristics and trends.

---

\* Department of Medical Information Science, Jichi Medical School