
 特 集 IV

長寿革命に係る人口学的観点からの総合的研究 (その3)

 都道府県別の満年齢別・死亡年別死亡数を
 出生年別に分割するモデル開発に関する研究

堀口侑*・石井太**

国立社会保障・人口問題研究所「日本版死亡データベース」(JMD)は、国際的な死亡データベースである Human Mortality Database (HMD) と整合性をもち、わが国の生命表を死亡研究に最適化して総合的に再編成した死亡データベースである。HMD では、満年齢別・死亡年別・出生年別である「レキシストライアングル (Lexis triangle)」別死亡数を基礎データとして生命表を作成しており、JMD でも同様であるが、都道府県別のレキシストライアングル別死亡数は1947~50年を除き、人口動態調査の個票が利用可能である1972年以降しか得られないことから、現在、1975年以降の生命表のみの作成を行っているところである。

一方、HMD の方法論の中では、満年齢別・死亡年別の死亡数しか得られない場合、これをさらに出生年別のレキシストライアングル別死亡数に分割するための方法が述べられており、基礎データとしてレキシストライアングル別死亡数が得られない国や時代の生命表の推計を行うために用いられている。これは、乳児死亡率や隣接するコーホートサイズの比などを説明変数とし、満年齢別・死亡年別の死亡数に含まれる隣接する2つの出生コーホートのうち年少コーホートの死亡数の割合である「下方レキシストライアングル割合」を被説明変数とする線形回帰モデルを構築し、これを用いて年次別・年齢別死亡数を出生年別に分割するものである。

本研究では、このHMDの方法を基礎とし、わが国の1971年以前の都道府県別・満年齢別・死亡年別死亡数を出生年別に分割し、レキシストライアングル別死亡数を推計するためのモデルを開発することを目的として研究を行った。研究の結果から、検討対象とした6種類のモデルのうち、1947~1975年のデータに基づいて、 $\log \text{IMR}(t)$ と全ての年齢階級のダミー変数との交差項を考え、ここから交差項以外を保持したまま、AICを基準とする変数減少法によって変数を選択した Model VI が最も有効であることが明らかとなった。本研究で提案したモデルは、これまでJMDで提供が行われていなかった1975年以前の都道府県生命表の作成に応用可能なものであり、JMDのさらなる拡張に資するものと考えられる。

キーワード: 都道府県別死亡, レキシストライアングル, 日本版死亡データベース

* 慶應義塾大学・院

** 慶應義塾大学

I. はじめに

国立社会保障・人口問題研究所「日本版死亡データベース」(Japanese Mortality Database, 以下, JMD という)は, 国際的な死亡データベースである Human Mortality Database (以下, HMD という)と整合性をもち, わが国の生命表を死亡研究に最適化して総合的に再編成した死亡データベースである. JMD ではその作成にあたり, HMD の方法論 (Wilmoth et al. 2021) を基本としつつ, 比較可能性を阻害しない範囲で, 日本の死亡状況により適合させるための改善を行っている (石井 2015).

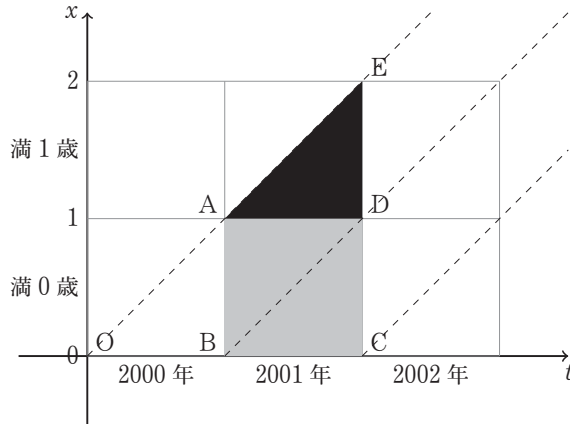
HMD では, 満年齢別・死亡年別・出生年別である「レキシストライアングル (Lexis triangle)」別死亡数を基礎データとして生命表を作成する. JMD でも同様に, レキシストライアングル別死亡数を用いて生命表を作成しているが, 都道府県別のレキシストライアングル別死亡数は1947~50年を除き, 人口動態調査の個票が利用可能である1972年以降しか得られないことから, 現在, 1975年以降の生命表のみの作成を行っているところである.

一方, HMD の方法論の中では, 満年齢別・死亡年別の死亡数しか得られない場合, これをさらに出生年別のレキシストライアングル別死亡数に分割するための方法が述べられており, 基礎データとしてレキシストライアングル別死亡数が得られない国や時代の生命表の推計を行うために用いられている. これは, 乳児死亡率や隣接するコーホートサイズの比などを説明変数とし, 満年齢別・死亡年別の死亡数に含まれる隣接する2つの出生コーホートのうち年少コーホートの死亡数の割合である「下方レキシストライアングル割合」を被説明変数とする線形回帰モデルを構築し, これを用いて年次別・年齢別死亡数を出生年別に分割するものである. そこで, 本研究は, この方法を基礎とし, わが国の1971年以前の都道府県別・満年齢別・死亡年別死亡数を出生年別に分割し, レキシストライアングル別死亡数を推計するためのモデルを提案する.

II. 先行研究と本研究の位置付け

HMD では, 統一的な方法で生命表を作成する観点から, 作成手順書である Method Protocol を策定している (Wilmoth et al. 2021). それによれば, HMD の生命表は, おおまかに以下の6段階の手順に基づいて作成される (石井 2015).

- 出生数: 可能な限り長期間の性別年間出生数を収集する.
- 死亡数: 可能な限り詳細なレベルの死亡数を収集し, 生データがまとめられている場合, 統一的な手法によって, 満年齢別・死亡年別・出生年別の死亡数を推定する.
- 人口規模: 各年1月1日現在推計人口を, 統計データから得るか, またはセンサスと出生・死亡数に基づいて推計する.
- リスク対応生存延べ年数 (exposure): ある年齢×時間区間において, 死亡リスクに



資料: 筆者作成

図1 レキシス図による年齢別死亡数と出生コーホート別死亡数の模式図

さらされる生存延べ年数を推計する。

- 死亡率: 死亡率は、ある年齢×時間区間に属する死亡数の、対応する区間の生存延べ年数に対する比として計算する。
- 生命表: 生命表を作成するため、まず死亡確率が死亡率より計算され、これにより生命表が求められる。

このように、HMDでは各国の様々な形式から成る基礎統計に対して、最初に「満年齢別・死亡年別・出生年別死亡数」、すなわち、「レキシストライアングル別死亡数」を統一的手法によって推計することがその基礎となっている。そして、その後、各年1月1日現在の人口を推計し、さらにある年齢×時間区間において死亡リスクにさらされる生存延べ年数であるリスク対応生存延べ年数 (exposure) を推計して、ここから死亡率を推定して生命表を作成するという方法が採られている。

このようにHMDの手順書においては、どのようなケースにおいてもまずレキシストライアングル別死亡数を作成することが前提となっているが、生データの死亡数が満年齢別・死亡年別にしか得られず、出生年別に分かれていない場合もある。そこで、このような場合、Wilmoth et al. (2021)では、これを出生年別に分割し、レキシストライアングル別死亡数を推計するための線形回帰モデルを提案している。そこで、まずここではその方法論についてレビューすることとする。

最初に、レキシストライアングルの概念について整理しておく。図1は、レキシス図を用いて、満年齢別・死亡年別死亡数と出生年との関係を示したものである。灰色で示された正方形ABCDは、2001年における満0歳死亡数を表す。ここで、2000年生まれコーホートはOBを始点とする生命線、2001年生まれコーホートはBCを始点とする生命線で表されることから、正方形ABCDにおける死亡点は、三角形ABDに含まれる2000年生まれ

コーホートと三角形 BCD に含まれる 2001 年生まれコーホートという隣接する二つの出生コーホートの死亡点から成る。このように、レキシス図における満年齢別・死亡年別の正方形領域は、隣接する二つの出生コーホートの生命線が通過する二つの直角三角形領域に分けることができる。そのそれぞれの三角形が満年齢別・死亡年別・出生年別の領域、すなわちレキシストライアングルを表しているが、このうち、年長のコーホートに対応する左上の直角三角形領域を上方レキシストライアングル (upper Lexis triangle)、年少のコーホートに対応する右下の直角三角形領域を下方レキシストライアングル (lower Lexis triangle) と呼ぶ。2001 年における満 0 歳を表す正方形 ABCD では、三角形 ABD が上方レキシストライアングル、三角形 BCD が下方レキシストライアングルに相当する。

なお、2001 年における 2000 年生まれコーホートの死亡点は満 0 歳だけではなく、満 1 歳にも存在する。図 1 の三角形 ADE は 2001 年における満 1 歳の 2000 年生まれコーホートの生命線が通過する領域を示したものである。従って、年次別・出生年別死亡数は図の平行四辺形 ABDE という領域内の死亡点に相当することから、年次別に出生コーホート別の死亡数を得たい場合にも、レキシストライアングル別死亡数が必要となる。

正方形領域の死亡数をレキシストライアングル別死亡数に分割する方法として最も単純なものは、上方レキシストライアングルと下方レキシストライアングルに半数ずつの死亡数を割り当てるものである。しかしながら、Wilmoth et al. (2021) によれば、全年齢の死亡数を期間・コーホート両方の形式で整備するためには、少なくとも以下の二つの理由からより精密な方法が必要であるとされている。一点目は、0 歳では死亡数が下方レキシストライアングルに集中していて、半数ずつを割り当てるべきではないことである。また二点目は、2 回の大戦による出生数の急激な上下などの歴史的イベントによる顕著な変動による隣接するコーホートの相対的サイズの違いによって、上方・下方トライアングルの死亡数分布が影響を受けることである。

そこで、以上の二点を考慮し、Wilmoth et al. (2021) では、下方レキシストライアングル死亡数の正方形領域の死亡数に対する割合である $\pi_d(x, t)$: 下方レキシストライアングル割合を被説明変数とする線形回帰モデルを推定し、これを用いて、満年齢別・死亡年別死亡数からレキシストライアングル別死亡数を推計する方法を提案している。ここで、

$$\pi_d(x, t) = \frac{D_L(x, t)}{D(x, t)} = \frac{D_L(x, t)}{D_L(x, t) + D_U(x, t)}$$

である。ただし、 x は満年齢、 t は年次、 $D(x, t)$ は、 t 年 x 歳の死亡数であり、 $D_L(x, t)$ 、 $D_U(x, t)$ は、それぞれ、年次 t 、満年齢 x 歳の下方、上方レキシストライアングルの死亡数である。

そして、線形回帰モデルによって $\pi_d(x, t)$ の推計値 ($\hat{\pi}_d(x, t)$) が得られたとすれば、

$$\begin{aligned} \hat{D}_L(x, t) &= \hat{\pi}_d(x, t) \cdot D(x, t) \\ \hat{D}_U(x, t) &= [1 - \hat{\pi}_d(x, t)] \cdot D(x, t) \end{aligned}$$

によって、満年齢別・死亡年別死亡数からレキシストライアングル別死亡数を推計することができる。

Wilmoth et al. (2021) による回帰モデルは以下の式(1)で表されるものである。ただし、記法については本研究のモデルに合わせている。

$$\begin{aligned}
 \pi_d(x, t) = & \text{constant} + \sum_{i=0}^{22} \beta_i A_i + \beta_{23} \cdot [\pi_b(x, t) - 0.5] \\
 & + \beta_{24} \cdot I(t=1918) + \beta_{25} \cdot I(t=1919) \\
 & + \beta_{26} \cdot \log \text{IMR}(t) \\
 & + \beta_{27} \cdot \log \text{IMR}(t) \cdot A_0 \\
 & + \beta_{28} \cdot \log \text{IMR}(t) \cdot A_1 \\
 & + \beta_{29} \cdot [\log \text{IMR}(t) - \log(0.01)] \cdot A_0 \cdot I(\text{IMR}(t) < 0.01)
 \end{aligned} \tag{1}$$

ここで、 x : 年齢、 t : 年次であり、 A_i は、 $A_i = I(x \in C_i)$ で定義される年齢（階級）のダミー変数である。ただし、

$$(C_0, C_1, C_2, C_3, C_4, \dots, C_{22}) = (\{0\}, \{1\}, \{2,3,4\}, \{5, \dots, 9\}, \dots, \{100, \dots, 104\})$$

であり、 A_i の回帰係数 β_i , ($i = 0, \dots, 22$) の総和は 0 となる ($\sum_{i=0}^{22} \beta_i = 0$) ように正規化されているため、レファレンスカテゴリーは存在しない。また、 $I(\cdot)$: 指示関数は、括弧内の命題が真であるとき 1 を、偽であるとき 0 を取る関数を表す。

ここで注意が必要なのは、この回帰モデルは、生命表などの詳細な指標の推計前の段階において、その作成のための基礎データを得るために使用されるモデルであることである。したがって、ここで説明変数として利用できるデータは出生数と死亡数という限られたデータだけであり、これらのみから得られる単純な変数であることが求められる。

まず、説明変数の一つとして投入されているのが $\pi_b(x, t)$: 年少コーホートサイズ割合、すなわち、 $B(t)$ を t 年の出生数として、

$$\pi_b(x, t) = \frac{B(t-x)}{B(t-x) + B(t-x-1)} \tag{2}$$

で定義される変数である。これは、分割対象となる満年齢 x 歳、年次 t 年の正方形領域について、下方レキシストライアングルの対象となる年少コーホートのサイズが、上方・下方両者の対象となるコーホートサイズに占める割合を、出生時のコーホートサイズを用いて求めたものである。これは、先述した精密な方法が必要となる二点目の理由である、顕著な変動による隣接するコーホートの相対的サイズの違いを考慮するための変数である。

一方、 $\pi_d(x, t)$ の年次変化に関してもいくつかの単純な説明変数が用いられている。一つは、 $I(t=1918)$, $I(t=1919)$ という、それぞれ1918年と1919年に 1、それ以外の年次で

0を取るダミー変数であり、これらはスペイン・インフルエンザの影響を反映させるための変数である。

また、多くの年齢で、 $\pi_d(x, t)$ が年次を追って増加する傾向が観察されていることから、年次 t における乳児死亡率 (Infant Mortality Rate, 以下、IMR という) $IMR(t)$ の自然対数値である $\log IMR(t)$ が、このような年次変化に関する死亡水準やパターンの変化を代表する変数として投入されている。また、 $\log IMR(t)$ と全ての年齢階級のダミー変数との交差項を検討したところ、0歳と1歳のみ統計的な有意性が認められたことから、0歳と1歳の交差項を説明変数に含めたとされている。さらに、0歳の $\pi_d(0, t)$ については、 $IMR(t)$ が0.01を下回ると逆に低下する傾向を示すことが観察されたことから、 $IMR(t)$ が0.01を下回る場合のみ $\log IMR(t) - \log(0.01)$ を説明変数とするため、これと $I(IMR(t) < 0.01)$, $\log IMR(t) - \log(0.01)$, A_0 との交差項を説明変数として投入したとされている。

ここで、 $IMR(t)$ については、限られたデータから推定を行う観点から、Pressat (1972) により提案された単純な推定式である式(3)が用いられている。

$$IMR(t) = \frac{D(0, t)}{\frac{1}{3}B(t-1) + \frac{2}{3}B(t)} \quad (3)$$

ただし、 $D(0, t)$ は t 年における0歳死亡数を、 $B(t)$ は t 年の出生数を表す。

回帰係数の推定にあたっては、スウェーデン (1901~1999年)、日本 (1950~1998年)、フランス (1907~1997年) の0~104歳の死亡データを用い、以下のウェイトを用いた加重最小二乗法によって行っている。

$$\omega(x, t) = \frac{D(x, t)}{\sum_x D(x, t)} \quad (4)$$

ここで $D(x, t)$ は、 t 年 x 歳の死亡数である。以上のような手法に基づき、Wilmoth et al. (2021) では、限られた単純な変数のみを用いてレキシストライアングル別死亡数を推計するためのモデルを提案している。

一方、わが国において、下方レキシストライアングル割合を用いた先行研究として林他 (2021) が挙げられる。これは人口動態統計などで得られる各年の出生数、各年・各歳の死亡数に基づいて生年コーホート別の人口の再構成を行った研究であり、この中で日本全体に関する下方レキシストライアングル割合を用いた出生コーホート別死亡数推計が行われている。ただし、この研究で推計が行われているのは出生コーホート別死亡数の公表値が存在しない1919~1936年の日本全体のみについてであり、また対象期間が限定されていることから、1917年と1937年の $\pi_d(x, t)$ の平均値を用いて推計が行われている。

しかしながら、JMD で必要となる、都道府県別のレキシストライアングル別死亡数が存在しない第二次大戦後の1951~1971年の推計にあたっては、林他 (2021) の研究と異なる

り対象期間が短くないことや、都道府県のレキシストライアングルの推計が目的であることから、HMD で用いているようなモデルを構築することが効率的と考えられる。しかしながら、わが国の都道府県別死亡に関して、このような研究はこれまで行われていない。

そこで、本研究は、JMD の都道府県別生命表の構築に必要な基礎データを作成する観点から、Wilmoth et al. (2021) で行われているような、満年齢別・死亡年別死亡数を出生年別に分割してレキシストライアングル別死亡数を推計するためのモデルを、わが国の1971年以前の都道府県別死亡に適用可能な形として提案することを目的とするものである。

III. データと方法

1. データ

本研究では、死亡データとして、厚生労働省「人口動態調査」による、全国（1947～2021年）および都道府県別（1972～2021年）の男女別レキシストライアングル別死亡数（0～109歳）を用いた¹⁾。また、出生データとしては全国（1873～2021年）および都道府県別（1947～2021年）の男女別出生数を用いた（1899年以降は厚生労働省「人口動態調査」、1898年以前は内閣府統計局「帝国統計年鑑」による）。

2. 方法

本研究で提案する線形回帰モデルは、以下の式(5)によって表される。

$$\begin{aligned} \pi_d(x, t) = & \text{constant} + \sum_{i=0}^{22} \beta_i A_i \\ & + \beta_{23} \cdot [\pi_b(x, t) - 0.5] + \beta_{24} \cdot \log \text{IMR}(t) \\ & + \sum_{i=25}^{47} \beta_i A_{i-25} \cdot \log \text{IMR}(t) \\ & + \beta_{48} \cdot [\log \text{IMR}(t) - \log(0.01)] \cdot A_0 \cdot I(\text{IMR}(t) < 0.01) \end{aligned} \quad (5)$$

Wilmoth et al. (2021) による式(1)との違いとして、 C_i について、

$$\begin{aligned} & (C_0, C_1, C_2, C_3, C_4, C_5, C_6, \dots, C_{22}) \\ & = (\{0\}, \{1\}, \{2, 3, 4\}, \{5, \dots, 9\}, \{10, \dots, 14\}, \\ & \quad \{15, \dots, 19\}, \{25, \dots, 29\}, \dots, \{105, \dots, 109\}) \end{aligned}$$

とし、20-24歳階級はレファレンスカテゴリとして C_i に含まず、 A_i の回帰係数 β_i ,

1) 本研究で使用した「人口動態調査」に関する分析結果には、統計法第32条の規定に基づき、調査票情報を二次利用したものが含まれている。また、死亡票の独自集計であるため、分析結果が公表数値とは一致しない場合がある。

($i = 0, \dots, 22$) の総和が 0 となる ($\sum_{i=0}^{22} \beta_i = 0$) という制約を外した点、また、0 歳、1 歳以外の年齢階級に関する交差項も説明変数としている点がある。

一方で、 $\pi_b(x, t)$: 年少コーホートサイズ割合、 $\text{IMR}(t)$ については、Wilmoth et al. (2021) によるものと同じ定義によるものであり、式(2)、(3)により計算される。なお、本研究においては、 $\pi_b(x, t)$ については安定性の観点から都道府県別の値ではなく全国値を用い、出生数データが得られない $t - x - 1 < 1873$ については $\pi_b(x, t) = 0.5$ とした。一方、 $\text{IMR}(t)$ については都道府県別出生数を用いて計算を行ったが、1947年については1946年の出生数が得られないため、 $\text{IMR}(1947) = \frac{D(0, 1947)}{B(1947)}$ とした。

推計は、Wilmoth et al. (2021) と同様、式(4)によるウェイトを用いた加重最小二乗法によった。

本研究では、式(5)によるモデルとして、以下の Model I~VI の 6 通りのモデルを全国のデータに基づいて推定して検討した。

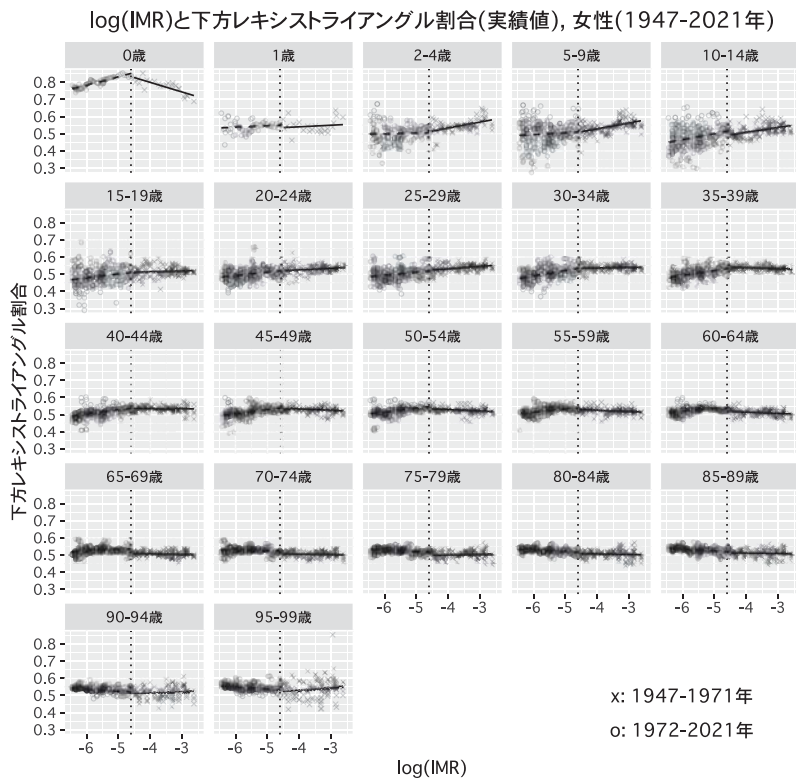
- Model I: $\log \text{IMR}(t)$ との交差項として 0 歳と 1 歳の項のみと、 $\log \text{IMR}(t) - \log(0.01)$, A_0 , $I(\text{IMR}(t) < 0.01)$ の交差項を説明変数とした HMD モデルと同じ説明変数を用い、1947~2021年の全年次のデータを用いて推定するモデル
- Model II: Model I で1947~1975年のデータを用いて推定するモデル
- Model III: $\log \text{IMR}(t)$ と全ての年齢階級のダミー変数との交差項を考え、1947~2021年の全年次のデータを用いて推定するモデル
- Model IV: Model III で1947~1975年のデータを用いて推定するモデル
- Model V: Model III から交差項以外を保持したまま、AIC を基準とする変数減少法によって変数を選択したモデル
- Model VI: Model IV から交差項以外を保持したまま、AIC を基準とする変数減少法によって変数を選択したモデル

これらのモデルを検討対象とした理由について述べる。Model I, III, V と II, IV, VI は対象とする年次をそれぞれ1947~2021年、1947~1975年としたものであり、それ以外の内容は同じであることから、まず、Model I, III, V について述べる。Model I は Wilmoth et al. (2021) と同じ説明変数を日本のデータに適用した場合の評価を示すものである。一方、Wilmoth et al. (2021) では $\log \text{IMR}(t)$ との交差項として、 $\log \text{IMR}(t)$ と全ての年齢階級のダミー変数との交差項を検討したところ、0 歳と 1 歳のみ統計的な有意性が認められたことから、0 歳と 1 歳の交差項を説明変数に含めたとしており、同様の検討を行うものが Model III となる。ここで $\log \text{IMR}(t)$ は、本来、乳児の死亡水準であり、直接的に高齢の下方レキシストライアングル割合に影響を与えるものではないが、先述の通り、このモデルは生命表などの詳細な指標の推計前の段階において、その作成のための基礎データを得るために使用されるモデルであり、利用可能な説明変数は出生数と死亡数しかなく、両者のみから得られる $\log \text{IMR}(t)$ を、その時点における全体的な死亡水

準を代表する変数として用いる考え方に立っている。このため、全体的な死亡水準と下方レキシストライアングル割合の間について、高齢部分も含めた全ての年齢階級との関係を想定したものである。ただし、Wilmoth et al. (2021) が最終的に0歳と1歳の交差項のみを用いたのと同様、わが国のデータに適用した場合にも、全体的な死亡水準と年齢階級ごとの下方レキシストライアングル割合の関係の強さは異なることが考えられることから、AICを基準とする変数減少法によって変数を選択したのが Model V ということになる。

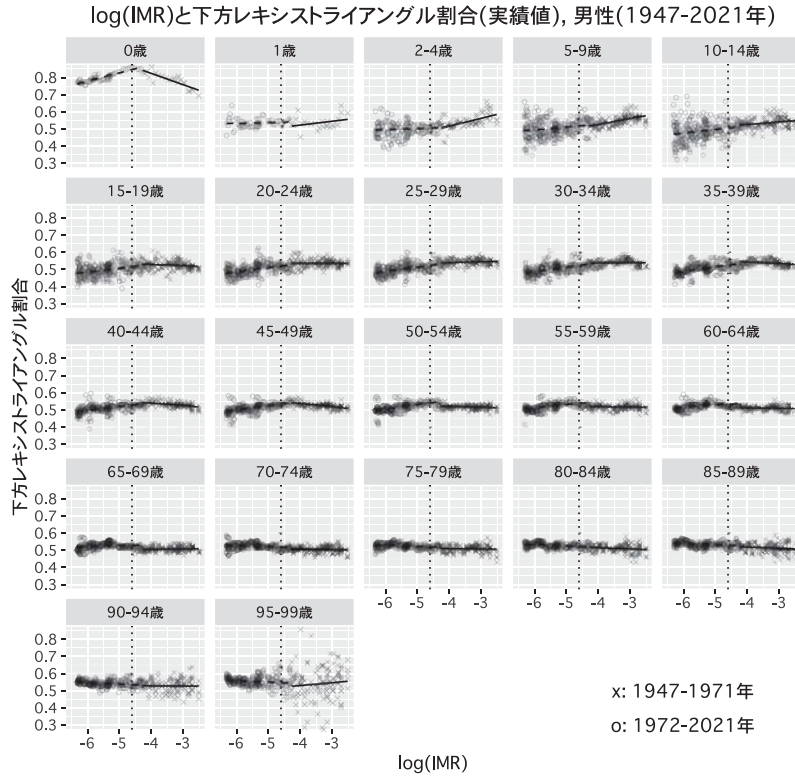
そこで、実際に図2、3から、年齢階級別に $\log\text{IMR}(t)$ と下方レキシストライアングル割合 $\pi_d(x, t)$ との関係を観察してみることにする。図の横軸は $\log\text{IMR}(t)$ を示し、年次推移によって低下することから、概ね右から左に向かって古い年次から新しい年次へと推移することとなる。下方レキシストライアングル割合は1971年以前は×のマーカ、1972年以降は○のマーカで示されており、縦の点線は $\log\text{IMR}(t) = \log 0.01$ を示している。

この図を観察すると、今回の対象となる年次である1971年以前は $\log\text{IMR}(t) > \log 0.01$ に多くの点が存在しており、縦の点線の右側の関係をモデリングすることが重要であることがわかる。また、Wilmoth et al. (2021) では多くの年齢で、 $\pi_d(x, t)$ が年次を追って



資料: 筆者推計

図2 IMR($\log\text{IMR}(t)$) と下方レキシストライアングル割合 ($\pi_d(x, t)$) (女性, 1947~2021年)



資料: 筆者推計

図3 IMR($\log\text{IMR}(t)$)と下方レキシストライアングル割合($\pi_d(x, t)$)(男性, 1947~2021年)

増加する傾向が観察されていることから $\log\text{IMR}(t)$ をこのような年次変化に関する死亡水準やパターンの変化を代表する変数として投入したとあり、年次を追って $\log\text{IMR}(t)$ は概ね低下することに注意すると、このことはグラフを回帰した直線の傾きが負であることに対応する。図には1971年以前と1972年以降のデータ対する回帰直線をそれぞれ実線・破線で示したが、これを見ると、わが国でも50歳代後半から80歳代までは全年次で概ねそのような傾向が観察されるものの、年齢階級によっては正の傾きを持つものや、縦の点線の左右で傾向が違うものも存在していることがわかる。

Wilmoth et al. (2021) のモデルでは、0歳と1歳には交差項が投入されており、他の年齢階級とは異なる動きをすることが假定されていることと、また、0歳については、 $\log\text{IMR}(t) < \log 0.01$ となる場合に傾きが変わることがモデルに含められており、わが国のデータをみても1歳では全年次で正の傾きとなっていること、また、0歳については縦の点線の左右で傾きが正と負で逆転していることから、0歳と1歳についてはHMDのモデルと共通する動きが確認されるものの、他の年齢階級の動きは1971年の前後で変わるものもあること、さらに、本研究の対象となる1971年以降の動きのみを見ても年齢階級ごとにその傾きは異なっており、0歳と1歳以外の年齢階級についても交差項の必要性が

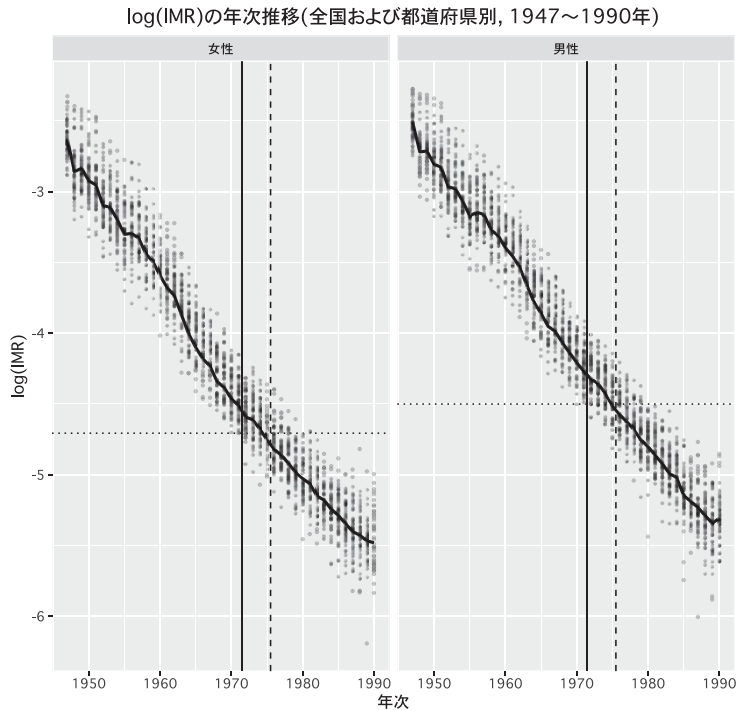
示唆されているといえ、Model I, III, V を検討対象とする必要性が理解できる。

次に、Model I, III, V に対して、Model II, IV, VI として、対象とする年次を1947～1975年に限定したモデルを検討対象とした理由について述べる。

先述の通り、都道府県別のレキシストライアングル別死亡数は1947～50年を除き、人口動態調査の個票が利用可能である1972年以降しか得られないため、本研究では1971年以前の都道府県別出生年別死亡数の推計に資するモデル構築が最大の目的となる。一般には、より多くの年次に対して適用可能なモデルがよいと考えられるが、実際には1972年以降の都道府県別・レキシストライアングル別死亡数は個票から直接得られるため、モデルを用いて推定する必要はなく、本研究の目的に照らせば、1971年以前をより効率的に推定できるモデルが望ましいことになる。

もちろん、利用可能な説明変数と下方レキシストライアングル割合の関係性が時系列的に変化しないのであれば、年次を限定したモデルを検討する必要性は高くないと考えられるが、図2、3の観察からは、1971年以前と1972年以降では $\log\text{IMR}(t)$ と $\pi_d(x, t)$ との関係が異なっていることが示唆される。これが対象とする年次を1975年以前に限定した Model II, IV, VI を検討対象に加えた理由である。

なお、Model II, IV, VI において、限定の最終年次を1975年としたのは以下の理由に



資料: 筆者推計

図4 $\log\text{IMR}(t)$ の年次推移 (全国および都道府県別, 1947～1990年)

よる。図4は男女別の $\log\text{IMR}(t)$ の年次推移 (1947~1990年) を示したものである。図で○のマーカで示されているのは各都道府県の値を示したもので、実線は全国値を示したものである。また、垂直な実線は1971年と1972年の間を、垂直な破線は1975年と1976年の間を示しており、水平な点線は1947~1971年における都道府県別 $\log\text{IMR}(t)$ の最小値を示している。これを見ると、男女とも全国値の $\log\text{IMR}(t)$ の1975年における値が水平な点線を下回っており、1947~1971年における都道府県別 $\log\text{IMR}(t)$ の下限が1975年以前の全国値の範囲内に入っていることがわかる。これより、全国値の年次の限定について、1947~1975年とすることとしたものである。

なお、以上の全国データによる回帰モデル以外に、都道府県別データによる回帰モデルも検討したが、後述する通り、全国データによる回帰モデルでは、全年次によるモデル (Model I, III, V) よりも1947~1975年によるモデル (Model II, IV, VI) の方が当てはまりがよく、連続的にデータが得られる1972年以降のみの都道府県別データを用いた回帰モデルでは、特に0歳の $\pi_d(0, t)$ の推定にバイアスを生じるなど有効な結果が得られなかった。このため、本研究では、以降、全国データによる回帰モデルのみを検討の対象とする。

IV. 結果と考察

1. 記述統計量

表1は、線形回帰モデル(5)の各変数に関する記述統計量をまとめたものである。

表1 記述統計量

	女性				男性			
	全期間		1975年以前		全期間		1975年以前	
	平均	標準偏差	平均	標準偏差	平均	標準偏差	平均	標準偏差
π_d	0.5262	0.0740	0.5316	0.0861	0.5248	0.0782	0.5294	0.0887
A_0	0.0092	0.0955	0.0093	0.0962	0.0093	0.0959	0.0095	0.0969
A_1	0.0092	0.0955	0.0093	0.0962	0.0093	0.0959	0.0095	0.0969
A_2	0.0276	0.1639	0.0280	0.1651	0.0279	0.1646	0.0284	0.1662
A_3	0.0460	0.2095	0.0467	0.2111	0.0465	0.2105	0.0474	0.2125
A_4	0.0460	0.2095	0.0467	0.2111	0.0465	0.2105	0.0474	0.2125
A_5	0.0460	0.2095	0.0467	0.2111	0.0465	0.2105	0.0474	0.2125
A_6	0.0460	0.2095	0.0467	0.2111	0.0465	0.2105	0.0474	0.2125
A_7	0.0460	0.2095	0.0467	0.2111	0.0465	0.2105	0.0474	0.2125
A_8	0.0460	0.2095	0.0467	0.2111	0.0465	0.2105	0.0474	0.2125
A_9	0.0460	0.2095	0.0467	0.2111	0.0465	0.2105	0.0474	0.2125
A_{10}	0.0460	0.2095	0.0467	0.2111	0.0465	0.2105	0.0474	0.2125
A_{11}	0.0460	0.2095	0.0467	0.2111	0.0465	0.2105	0.0474	0.2125
A_{12}	0.0460	0.2095	0.0467	0.2111	0.0465	0.2105	0.0474	0.2125
A_{13}	0.0460	0.2095	0.0467	0.2111	0.0465	0.2105	0.0474	0.2125
A_{14}	0.0460	0.2095	0.0467	0.2111	0.0465	0.2105	0.0474	0.2125
A_{15}	0.0460	0.2095	0.0467	0.2111	0.0465	0.2105	0.0474	0.2125
A_{16}	0.0460	0.2095	0.0467	0.2111	0.0465	0.2105	0.0474	0.2125
A_{17}	0.0460	0.2095	0.0467	0.2111	0.0465	0.2105	0.0474	0.2125
A_{18}	0.0460	0.2095	0.0467	0.2111	0.0465	0.2105	0.0474	0.2125
A_{19}	0.0460	0.2095	0.0467	0.2111	0.0465	0.2105	0.0474	0.2125
A_{20}	0.0460	0.2095	0.0467	0.2111	0.0465	0.2105	0.0474	0.2125
A_{21}	0.0458	0.2090	0.0461	0.2097	0.0440	0.2051	0.0408	0.1979
A_{22}	0.0339	0.1809	0.0197	0.1388	0.0269	0.1618	0.0118	0.1078
π_b	0.5004	0.0148	0.5016	0.0128	0.5004	0.0139	0.5016	0.0120
IMR	0.0134	0.0165	0.0291	0.0176	0.0157	0.0189	0.0343	0.0192
N	8149		3104		8072		3061	

資料:筆者推計

男女とも年齢は0～109歳で110通りあるため、全期間75年分の標本の大きさは8250、1975年以前29年分の標本の大きさは3190となるべきところであるが、全期間では女性の100-104歳に2、105-109歳に99、男性の100-104歳に20、105-109歳に158の π_d の欠測値があり、1975年以前では女性の100-104歳に2、105-109歳に84、男性の100-104歳に20、105-109歳に109の欠測値があることから、全期間の女性は8149、男性は8072、1975年以前の女性は3104、男性は3061となっている。

また、IMRの平均値を見ると、全期間では女性で0.0134、男性で0.0157であるのに対して、1975年以前については女性で0.0291、男性で0.0343となっており、図4で観察した通り、全期間と1975年以前ではIMRの水準に大きな違いがあることが理解できる。

2. モデル推定結果

表2、3は、全国のデータに基づき、回帰式(5)によって回帰分析を行った結果である。ここでは、Model I, II, IV, VIの結果のみを示している。

まず女性の結果である表2を見ると、どのモデルにおいても $\pi_b - 0.5$ の係数は有意となっていて、年少コーホートサイズ割合が下方レキシストライアングル割合の推定において有効であることが理解できる。また、Model I, IIの A_i ($i \geq 2$)の係数は、レファレンスカテゴリである20-24歳の下方レキシストライアングル割合との相対的格差を示している。例えば、35-39歳、40-44歳の A_8, A_9 の符号はModel Iではマイナス、Model IIではプラスとなっているが、これは図2の35-39歳、40-44歳の下方レキシストライアングル割合が○のマーカーでは低いことに対応していると見ることができる。

一方、LIやこれとの交差項の係数から、 $\log \text{IMR}(t)$ に対する下方レキシストライアングル割合の反応についてもモデルごとの違いを見ることができる。Model IではLI ($\log \text{IMR}(t)$)の係数が -0.0073 と負であり、これはWilmoth et al. (2021)で得られた観察である、多くの年齢で $\pi_d(x, t)$ が年次を追って増加、すなわち、 $\log \text{IMR}(t)$ の低下に対して $\pi_d(x, t)$ が増加することに対応するものと見ることができる。また、 $\text{LI} \cdot A_0$ の係数が -0.0543 と負であり、 $\text{LIC}([\log \text{IMR}(t) - \log(0.01)] \cdot A_0 \cdot I(\text{IMR}(t) < 0.01))$ の係数が 0.0966 と正であることは、図2の0歳の回帰直線で観察した通り、 $\text{IMR}(t) < 0.01$ の範囲では傾きが正、 $\text{IMR}(t) < 0.01$ の範囲では負になることと整合している。

ただし、図2で観察した通り、 $\log \text{IMR}(t)$ と下方レキシストライアングル割合との関係は年齢階級別に見ると一様ではなく、また、1971年以前に限定した場合の動きは全年次とは異なっていた。Model Iにおいて基礎データを1975年以前に限定したModel IIの結果を見ると、LIの係数は -0.0031 とその絶対値はModel Iと比較して小さいものとなっている。また、LICについては1975年以前では $\text{IMR}(t) < 0.01$ となる年次が少ないため、有意な変数ではなくなっていることがわかる。

次に、自由度調整済決定係数を見ると、Model Iが0.8798であったのに対してModel IIでは0.9274となっており、当てはまりが改善していることが理解できる。なお、表には示していないが、Model III, Vの自由度調整済決定係数はどちらも0.8981であり、基礎デー

表2 回帰モデル推定結果 (女性, Model I, II, IV, VI)

	Model I			Model II			Model IV			Model VI		
	coef	se	p 値	coef	se	p 値	coef	se	p 値	coef	se	p 値
const	0.4999	0.0021	0.000	0.5211	0.0031	0.000	0.5605	0.0132	0.000	0.5617	0.0085	0.000
A ₀	0.0611	0.0068	0.000	0.0482	0.0084	0.000	0.0089	0.0153	0.560	0.0074	0.0112	0.505
A ₁	0.1323	0.0106	0.000	0.1637	0.0174	0.000	0.1244	0.0213	0.000	0.1231	0.0188	0.000
A ₂	0.0346	0.0028	0.000	0.0330	0.0033	0.000	0.1195	0.0199	0.000	0.1183	0.0171	0.000
A ₃	0.0236	0.0033	0.000	0.0268	0.0039	0.000	0.0832	0.0233	0.000	0.0820	0.0210	0.000
A ₄	-0.0063	0.0039	0.106	-0.0033	0.0047	0.481	0.0167	0.0274	0.541	-0.0032	0.0046	0.495
A ₅	-0.0127	0.0032	0.000	-0.0121	0.0038	0.002	-0.0475	0.0216	0.028	-0.0487	0.0191	0.011
A ₆	0.0037	0.0027	0.175	0.0068	0.0033	0.042	0.0086	0.0186	0.643	0.0074	0.0032	0.023
A ₇	0.0002	0.0027	0.928	0.0043	0.0033	0.196	-0.0185	0.0187	0.320	-0.0198	0.0157	0.208
A ₈	-0.0032	0.0026	0.215	0.0019	0.0033	0.552	-0.0437	0.0182	0.017	-0.0450	0.0152	0.003
A ₉	-0.0051	0.0025	0.037	0.0016	0.0032	0.618	-0.0451	0.0178	0.011	-0.0463	0.0146	0.002
A ₁₀	-0.0073	0.0023	0.002	-0.0049	0.0030	0.108	-0.0623	0.0171	0.000	-0.0635	0.0139	0.000
A ₁₁	-0.0081	0.0022	0.000	-0.0083	0.0029	0.004	-0.0785	0.0165	0.000	-0.0797	0.0130	0.000
A ₁₂	-0.0090	0.0022	0.000	-0.0123	0.0028	0.000	-0.0701	0.0159	0.000	-0.0713	0.0123	0.000
A ₁₃	-0.0144	0.0021	0.000	-0.0229	0.0027	0.000	-0.0889	0.0153	0.000	-0.0901	0.0115	0.000
A ₁₄	-0.0170	0.0021	0.000	-0.0264	0.0026	0.000	-0.0818	0.0148	0.000	-0.0830	0.0108	0.000
A ₁₅	-0.0201	0.0021	0.000	-0.0301	0.0026	0.000	-0.0760	0.0145	0.000	-0.0773	0.0104	0.000
A ₁₆	-0.0209	0.0020	0.000	-0.0299	0.0026	0.000	-0.0609	0.0144	0.000	-0.0621	0.0103	0.000
A ₁₇	-0.0192	0.0020	0.000	-0.0285	0.0026	0.000	-0.0621	0.0148	0.000	-0.0633	0.0108	0.000
A ₁₈	-0.0148	0.0021	0.000	-0.0232	0.0027	0.000	-0.0631	0.0161	0.000	-0.0644	0.0125	0.000
A ₁₉	-0.0057	0.0021	0.007	-0.0140	0.0033	0.000	-0.0473	0.0209	0.024	-0.0486	0.0184	0.008
A ₂₀	0.0070	0.0022	0.002	-0.0013	0.0061	0.830	-0.0227	0.0428	0.595	0.0076	0.0062	0.219
A ₂₁	0.0216	0.0030	0.000	0.0178	0.0208	0.391	-0.0271	0.1494	0.856	0.0270	0.0204	0.185
A ₂₂	0.0349	0.0076	0.000	0.0560	0.1044	0.592	0.2656	0.6442	0.680	0.0636	0.1020	0.533
$\pi_b - 0.5$	0.9295	0.0159	0.000	0.8236	0.0279	0.000	0.8263	0.0274	0.000	0.8266	0.0273	0.000
LI	-0.0073	0.0002	0.000	-0.0031	0.0006	0.000	0.0086	0.0038	0.026	0.0089	0.0024	0.000
LI · A ₀	-0.0543	0.0019	0.000	-0.0556	0.0024	0.000	-0.0672	0.0045	0.000	-0.0677	0.0032	0.000
LI · A ₁	0.0271	0.0030	0.000	0.0403	0.0053	0.000	0.0286	0.0064	0.000	0.0283	0.0057	0.000
LI · A ₂							0.0273	0.0059	0.000	0.0270	0.0051	0.000
LI · A ₃							0.0170	0.0068	0.013	0.0166	0.0061	0.007
LI · A ₄							0.0059	0.0080	0.462			
LI · A ₅							-0.0104	0.0063	0.098	-0.0108	0.0056	0.052
LI · A ₆							0.0004	0.0054	0.945			
LI · A ₇							-0.0069	0.0053	0.194	-0.0073	0.0044	0.099
LI · A ₈							-0.0134	0.0052	0.010	-0.0138	0.0042	0.001
LI · A ₉							-0.0136	0.0050	0.006	-0.0140	0.0040	0.001
LI · A ₁₀							-0.0165	0.0048	0.001	-0.0169	0.0038	0.000
LI · A ₁₁							-0.0199	0.0046	0.000	-0.0203	0.0036	0.000
LI · A ₁₂							-0.0165	0.0045	0.000	-0.0169	0.0034	0.000
LI · A ₁₃							-0.0187	0.0043	0.000	-0.0190	0.0032	0.000
LI · A ₁₄							-0.0159	0.0042	0.000	-0.0162	0.0030	0.000
LI · A ₁₅							-0.0134	0.0041	0.001	-0.0137	0.0029	0.000
LI · A ₁₆							-0.0094	0.0041	0.022	-0.0098	0.0029	0.001
LI · A ₁₇							-0.0102	0.0042	0.015	-0.0105	0.0029	0.000
LI · A ₁₈							-0.0118	0.0045	0.008	-0.0121	0.0033	0.000
LI · A ₁₉							-0.0101	0.0055	0.066	-0.0105	0.0046	0.024
LI · A ₂₀							-0.0073	0.0105	0.488			
LI · A ₂₁							-0.0130	0.0358	0.717			
LI · A ₂₂							0.0504	0.1584	0.751			
LIC	0.0966	0.0058	0.000	-0.0067	0.0752	0.928	-0.0072	0.0734	0.921			
N	8149			3104			3104			3104		
調整 R ²	0.8798			0.9274			0.9307			0.9308		

注: LI は $\log\text{IMR}(t)$, LIC は $[\log\text{IMR}(t) - \log(0.01)] \cdot A_0 \cdot I(\text{IMR}(t) < 0.01)$ を表す。
 資料: 筆者推計

タを1975年以前に限定したモデルより低いものとなっていることから、ここでは基礎データを1975年以前に限定した Model IV, VI についてのみ考察を行っている。

次に交差項について検討するため $\log\text{IMR}(t)$ と全ての年齢階級のダミー変数との交差項を投入した Model IV を見ると、 $\text{LI} \cdot A_0$ の係数は年齢階級ごとに異なる係数となっており、特に、2-4 歳、5-9 歳に対応する $i=2,3$ で正であるのに対し、30 歳以上に対応する $i \geq 7$ の多くで負になっていることはここでの LI の係数は、レファレンスカテゴリーである 20-24 歳について、図 2 の 1971 年以前の回帰直線の傾きが図 2 で 2-4 歳、5-9 歳の実線の傾きが 20-24 歳のそれより大きく、30 歳以上の多くの年齢階級では小さくなっていることと整合的である。

ただし、Model IV では交差項のいくつかに p 値が大きいものが含まれており、特に年齢階級が 100 歳以上のところでは推定結果の安定性に影響を及ぼすことが考えられる。

表3 回帰モデル推定結果 (男性, Model I, II, IV, VI)

	Model I			Model II			Model IV			Model VI		
	coef	se	p 値	coef	se	p 値	coef	se	p 値	coef	se	p 値
const	0.5086	0.0019	0.000	0.5247	0.0029	0.000	0.5363	0.0118	0.000	0.5294	0.0033	0.000
A ₀	0.0527	0.0063	0.000	0.0309	0.0076	0.000	0.0196	0.0136	0.149	0.0265	0.0075	0.000
A ₁	0.1187	0.0108	0.000	0.1691	0.0174	0.000	0.1577	0.0203	0.000	0.1646	0.0169	0.000
A ₂	0.0307	0.0027	0.000	0.0276	0.0031	0.000	0.1698	0.0184	0.000	0.1767	0.0145	0.000
A ₃	0.0229	0.0031	0.000	0.0231	0.0035	0.000	0.1111	0.0209	0.000	0.1179	0.0176	0.000
A ₄	0.0005	0.0037	0.899	0.0024	0.0043	0.573	0.0366	0.0255	0.152	0.0434	0.0229	0.058
A ₅	-0.0099	0.0027	0.000	-0.0103	0.0033	0.002	-0.0579	0.0188	0.002	-0.0511	0.0150	0.001
A ₆	0.0037	0.0025	0.137	0.0063	0.0030	0.035	0.0124	0.0171	0.469	0.0063	0.0029	0.028
A ₇	-0.0048	0.0024	0.051	-0.0020	0.0030	0.510	0.0085	0.0174	0.624	-0.0019	0.0029	0.526
A ₈	-0.0075	0.0023	0.001	-0.0048	0.0029	0.105	-0.0131	0.0167	0.431	-0.0045	0.0028	0.111
A ₉	-0.0113	0.0022	0.000	-0.0091	0.0028	0.001	-0.0492	0.0159	0.002	-0.0424	0.0112	0.000
A ₁₀	-0.0124	0.0021	0.000	-0.0120	0.0027	0.000	-0.0762	0.0153	0.000	-0.0694	0.0103	0.000
A ₁₁	-0.0132	0.0020	0.000	-0.0206	0.0026	0.000	-0.0589	0.0147	0.000	-0.0521	0.0095	0.000
A ₁₂	-0.0127	0.0019	0.000	-0.0221	0.0024	0.000	-0.0349	0.0140	0.013	-0.0218	0.0024	0.000
A ₁₃	-0.0159	0.0019	0.000	-0.0295	0.0024	0.000	-0.0470	0.0135	0.001	-0.0402	0.0075	0.000
A ₁₄	-0.0170	0.0018	0.000	-0.0316	0.0023	0.000	-0.0333	0.0132	0.012	-0.0313	0.0022	0.000
A ₁₅	-0.0172	0.0018	0.000	-0.0322	0.0023	0.000	-0.0435	0.0131	0.001	-0.0318	0.0022	0.000
A ₁₆	-0.0164	0.0018	0.000	-0.0294	0.0023	0.000	-0.0389	0.0134	0.004	-0.0290	0.0023	0.000
A ₁₇	-0.0127	0.0018	0.000	-0.0263	0.0024	0.000	-0.0487	0.0144	0.001	-0.0419	0.0089	0.000
A ₁₈	-0.0061	0.0019	0.001	-0.0211	0.0028	0.000	-0.0533	0.0175	0.002	-0.0465	0.0134	0.001
A ₁₉	0.0030	0.0020	0.134	-0.0090	0.0043	0.037	-0.0311	0.0288	0.281	-0.0083	0.0042	0.047
A ₂₀	0.0138	0.0026	0.000	-0.0050	0.0106	0.636	0.0252	0.0740	0.734	-0.0043	0.0103	0.679
A ₂₁	0.0273	0.0056	0.000	0.0211	0.0401	0.600	0.0606	0.2657	0.820	0.0218	0.0388	0.574
A ₂₂	0.0467	0.0224	0.037	0.0278	0.1636	0.865	-0.3160	0.8548	0.712	0.0283	0.1581	0.858
$\pi_b - 0.5$	0.8261	0.0171	0.000	0.7529	0.0290	0.000	0.7615	0.0283	0.000	0.7636	0.0282	0.000
LI	-0.0050	0.0002	0.000	-0.0030	0.0006	0.000	0.0004	0.0035	0.899	-0.0016	0.0008	0.039
LI · A ₀	-0.0616	0.0018	0.000	-0.0653	0.0022	0.000	-0.0687	0.0041	0.000	-0.0666	0.0022	0.000
LI · A ₁	0.0260	0.0031	0.000	0.0468	0.0056	0.000	0.0433	0.0064	0.000	0.0454	0.0054	0.000
LI · A ₂							0.0459	0.0057	0.000	0.0479	0.0045	0.000
LI · A ₃							0.0273	0.0063	0.000	0.0293	0.0053	0.000
LI · A ₄							0.0103	0.0076	0.174	0.0124	0.0068	0.069
LI · A ₅							-0.0142	0.0055	0.010	-0.0121	0.0043	0.005
LI · A ₆							0.0018	0.0050	0.725			
LI · A ₇							0.0029	0.0050	0.561			
LI · A ₈							-0.0026	0.0048	0.595			
LI · A ₉							-0.0115	0.0046	0.012	-0.0095	0.0031	0.002
LI · A ₁₀							-0.0184	0.0044	0.000	-0.0164	0.0028	0.000
LI · A ₁₁							-0.0111	0.0043	0.010	-0.0090	0.0026	0.001
LI · A ₁₂							-0.0038	0.0041	0.348			
LI · A ₁₃							-0.0051	0.0039	0.195	-0.0031	0.0020	0.129
LI · A ₁₄							-0.0007	0.0039	0.849			
LI · A ₁₅							-0.0034	0.0038	0.377			
LI · A ₁₆							-0.0029	0.0039	0.456			
LI · A ₁₇							-0.0064	0.0041	0.121	-0.0043	0.0023	0.062
LI · A ₁₈							-0.0089	0.0049	0.067	-0.0069	0.0035	0.049
LI · A ₁₉							-0.0062	0.0077	0.415			
LI · A ₂₀							0.0073	0.0190	0.701			
LI · A ₂₁							0.0097	0.0678	0.886			
LI · A ₂₂							-0.0942	0.2293	0.681			
LIC	0.1360	0.0063	0.000									
N	8072			3061			3061			3061		
調整 R ²	0.8871			0.9406			0.9443			0.9444		

注: LI は $\log\text{IMR}(t)$, LIC は $[\log\text{IMR}(t) - \log(0.01)] \cdot A_0 \cdot I(\text{IMR}(t) < 0.01)$ を表す。
 資料: 筆者推計

Model VI は Model IV から AIC によって変数選択を行ったものであり, LI · A_i (i=4, 6, 20, 21, 22) が除かれるとともに, LIC も除かれている。また自由度調整済決定係数も Model IV, VI は 0.9307, 0.9308 と Model II よりも大きく, さらに Model VI が Model IV を上回ったことから, 本研究では Model VI を最終的なモデルとして採用することとした。

男性の結果である表 3 でも概ね女性と同様の結果が得られている。ただし, 男性では, Model II について, 1975 年以前に $\text{IMR}(t)$ が 0.01 を下回る年次がないことから, LIC の係数は推定ができなかった。また, Model VI では, 最終的に残された交差項は女性よりも少なく, LI · A_i (i=6, 7, 8, 12, 14, 15, 16, 19, 20, 21, 22) が除かれた。

自由度調整済決定係数を見ると, Model I は 0.8871 であるのに対して, 1975 年以前による Model II は 0.9406 と女性同様大きいものとなっている。表には示していないが, Model

III, V の自由度調整済決定係数はどちらも0.9042であり、女性と同様、基礎データを1975年以前に限定したモデルより低いものとなっている。また、全てのモデルの中で自由度調整済決定係数が最も高いのは Model VI の0.9444であり、男性についても Model VI を最終的なモデルとして採用することとした。

次に、これらのモデルから推定される下方レキシストライアングル割合を評価するため、全てのモデルについて、年齢階級別に実績値と推計値の差の平均値を示したものが図5, 6, 表4である。

これを見ると、男女とも Model I では10~49歳で推計値が過小推定となっているのに対して、60~89歳で過大推定となっていることがわかる。Model II では Model I に比べて特に若年層での過小推定はなくなるが、2-4歳, 5-9歳で過大推定となっていることが観察できる。Model III, V では男女とも50歳代付近で過大推定があるとともに、80歳以上では大きく過小推定がある。Model IV, VI は全体的にバイアスが小さく、概ね同程度の差となっているが、90歳以上でも男性の95-99歳以外はバイアスが小さいことが観察できる。

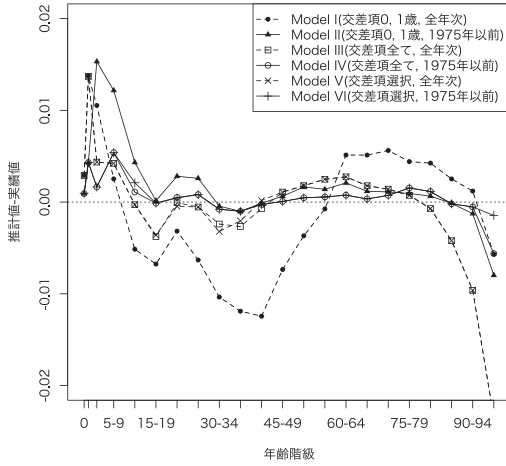
次に、モデルによって推定された都道府県別下方レキシストライアングル割合を評価する。ただし、1971年以前（1947~50年を除く）については都道府県別下方レキシストライアングル割合の実績統計は存在しないため、各モデルによる都道府県別下方レキシストライアングル割合の推計値を実績値と直接比較することはできない。しかしながら、下方レキシストライアングル割合の都道府県別の値は、全国の実績値の周りに分布すると考えられることから、都道府県別下方レキシストライアングル割合の推計値と全国の実績値を比較することによって評価が可能であると考えられる。ただし、全国値でも下方レキシストライアングル割合の変動は激しいため、実績値との直接比較からは必ずしも有効な評価が得られない可能性があると考えられる。

そこで、男女別・年齢階級別に、下方レキシストライアングル割合の全国実績値を $\log\text{IMR}(t)$ を説明変数とする単回帰を行い、その回帰直線に基づく「全国実績単回帰推計値」に対して、各モデルによる都道府県別推計値がどの程度乖離があるかによって評価を行うこととした。ただし、回帰を行う年次は1947~1975年とし、年少コーホート割合が他のコーホートと大きく異なる1966, 1967年生まれは除外した。

表5は、Model II, V, VI による下方レキシストライアングル割合の都道府県推計値と全国実績単回帰推計値の乖離について、男女別・年齢階級別に RMSE（平均二乗誤差の平方根）で示したものである。

Model II と VI は交差項の違いを示し、Model V と VI は対象期間（全期間、1975年以前）の違いを示すことから、表には Model II と VI の RMSE の差 (II - VI)、Model V と VI の RMSE の差 (V - VI) を示している。これを見ると、Model VI の RMSE が II, V より大きい箇所は、女性の1歳の V - VI で-0.0007である以外は-0.0002以下であり、ほぼ全ての年齢階級で Model VI の RMSE が Model II, V よりも小さいものとなっていることがわかる。

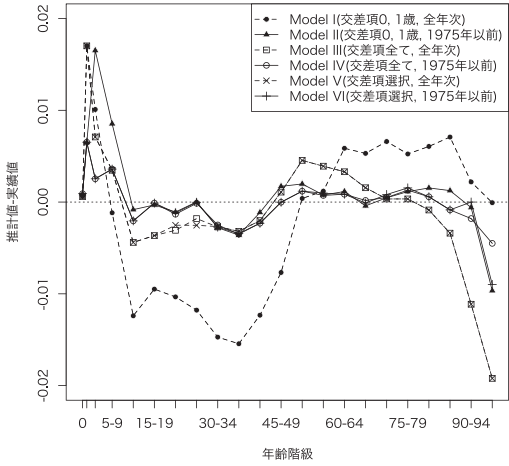
下方レキシストライアングル割合(全国)の推計値と実績値の差の平均
女性(1947-71年)



資料:筆者推計

図5 下方レキシストライアングル割合(全国)の推計値と実績値の差の平均(女性, 1947-71年)

下方レキシストライアングル割合(全国)の推計値と実績値の差の平均
男性(1947-71年)



資料:筆者推計

図6 下方レキシストライアングル割合(全国)の推計値と実績値の差の平均(男性, 1947-71年)

表4 下方レキシストライアングル割合(全国)の推計値と実績値の差の平均(1947-71年)

年齢階級	女性						男性					
	Model I	Model II	Model III	Model IV	Model V	Model VI	Model I	Model II	Model III	Model IV	Model V	Model VI
0	0.0029	0.0009	0.0029	0.0009	0.0029	0.0010	0.0006	0.0009	0.0006	0.0009	0.0006	0.0009
1	0.0137	0.0043	0.0137	0.0043	0.0137	0.0043	0.0170	0.0065	0.0170	0.0065	0.0170	0.0065
2-4	0.0105	0.0153	0.0044	0.0016	0.0044	0.0016	0.0101	0.0165	0.0071	0.0025	0.0071	0.0025
5-9	0.0025	0.0122	0.0042	0.0054	0.0042	0.0054	-0.0012	0.0085	0.0035	0.0036	0.0035	0.0036
10-14	-0.0051	0.0043	-0.0003	0.0011	-0.0003	0.0021	-0.0124	-0.0008	-0.0044	-0.0020	-0.0044	-0.0021
15-19	-0.0068	0.0001	-0.0038	-0.0001	-0.0036	-0.0001	-0.0095	-0.0003	-0.0037	-0.0001	-0.0037	-0.0001
20-24	-0.0032	0.0028	-0.0001	0.0005	-0.0005	0.0004	-0.0103	-0.0011	-0.0031	-0.0013	-0.0025	-0.0012
25-29	-0.0063	0.0026	-0.0005	0.0008	-0.0005	0.0008	-0.0118	0.0000	-0.0018	-0.0001	-0.0025	0.0000
30-34	-0.0104	-0.0004	-0.0024	-0.0008	-0.0032	-0.0008	-0.0147	-0.0028	-0.0027	-0.0025	-0.0027	-0.0027
35-39	-0.0119	-0.0010	-0.0027	-0.0010	-0.0020	-0.0010	-0.0154	-0.0036	-0.0032	-0.0035	-0.0032	-0.0034
40-44	-0.0124	-0.0001	-0.0007	-0.0003	0.0002	-0.0003	-0.0123	-0.0011	-0.0020	-0.0023	-0.0020	-0.0023
45-49	-0.0073	0.0006	0.0011	0.0001	0.0011	0.0001	-0.0077	0.0017	0.0011	0.0000	0.0011	0.0000
50-54	-0.0037	0.0017	0.0018	0.0005	0.0018	0.0005	0.0004	0.0020	0.0045	0.0012	0.0045	0.0012
55-59	-0.0007	0.0014	0.0025	0.0006	0.0025	0.0006	0.0012	0.0008	0.0039	0.0007	0.0039	0.0010
60-64	0.0051	0.0021	0.0027	0.0008	0.0027	0.0008	0.0059	0.0011	0.0033	0.0009	0.0033	0.0009
65-69	0.0051	0.0012	0.0018	0.0003	0.0018	0.0003	0.0053	-0.0004	0.0016	0.0002	0.0016	-0.0001
70-74	0.0056	0.0011	0.0014	0.0007	0.0014	0.0007	0.0066	0.0005	0.0003	0.0005	0.0003	0.0008
75-79	0.0044	0.0009	0.0007	0.0015	0.0007	0.0015	0.0052	0.0011	0.0003	0.0013	0.0003	0.0015
80-84	0.0042	0.0006	-0.0007	0.0012	-0.0007	0.0012	0.0061	0.0015	-0.0009	0.0006	-0.0009	0.0006
85-89	0.0025	-0.0001	-0.0042	-0.0002	-0.0042	-0.0002	0.0071	0.0012	-0.0034	-0.0009	-0.0034	-0.0009
90-94	0.0012	-0.0013	-0.0096	-0.0005	-0.0096	-0.0005	0.0022	-0.0006	-0.0111	-0.0018	-0.0111	0.0000
95-99	-0.0057	-0.0080	-0.0235	-0.0056	-0.0235	-0.0014	-0.0001	-0.0097	-0.0192	-0.0045	-0.0192	-0.0090

資料: 筆者推計

表5 下方レキシストライアングル割合の都道府県推計値と全国実績単回帰推計値のRMSE (1947-71年)

年齢階級	女性					男性				
	Model II	Model V	Model VI	II - VI	V - VI	Model II	Model V	Model VI	II - VI	V - VI
0	0.0108	0.0130	0.0109	0.0000	0.0021	0.0094	0.0098	0.0095	-0.0001	0.0002
1	0.0174	0.0167	0.0174	0.0000	-0.0007	0.0164	0.0162	0.0164	0.0000	-0.0002
2-4	0.0243	0.0109	0.0093	0.0149	0.0016	0.0277	0.0118	0.0083	0.0194	0.0035
5-9	0.0218	0.0119	0.0111	0.0107	0.0008	0.0180	0.0103	0.0096	0.0084	0.0007
10-14	0.0155	0.0130	0.0120	0.0035	0.0011	0.0131	0.0121	0.0104	0.0027	0.0017
15-19	0.0109	0.0137	0.0109	0.0000	0.0028	0.0106	0.0141	0.0100	0.0006	0.0041
20-24	0.0133	0.0111	0.0101	0.0032	0.0010	0.0104	0.0104	0.0101	0.0002	0.0003
25-29	0.0129	0.0113	0.0101	0.0027	0.0012	0.0103	0.0109	0.0101	0.0002	0.0008
30-34	0.0097	0.0114	0.0094	0.0004	0.0021	0.0099	0.0102	0.0097	0.0002	0.0006
35-39	0.0068	0.0101	0.0069	-0.0001	0.0032	0.0068	0.0085	0.0067	0.0001	0.0018
40-44	0.0097	0.0132	0.0097	0.0000	0.0035	0.0084	0.0117	0.0073	0.0010	0.0044
45-49	0.0102	0.0123	0.0099	0.0003	0.0024	0.0108	0.0136	0.0072	0.0036	0.0063
50-54	0.0103	0.0116	0.0090	0.0014	0.0026	0.0076	0.0093	0.0062	0.0014	0.0031
55-59	0.0104	0.0115	0.0099	0.0004	0.0015	0.0073	0.0087	0.0075	-0.0002	0.0012
60-64	0.0116	0.0122	0.0108	0.0007	0.0014	0.0083	0.0094	0.0083	0.0000	0.0011
65-69	0.0095	0.0102	0.0091	0.0003	0.0011	0.0078	0.0091	0.0078	0.0000	0.0013
70-74	0.0089	0.0107	0.0089	0.0000	0.0018	0.0076	0.0091	0.0077	-0.0001	0.0014
75-79	0.0089	0.0110	0.0089	0.0000	0.0022	0.0076	0.0086	0.0078	-0.0002	0.0008
80-84	0.0070	0.0095	0.0070	0.0000	0.0025	0.0059	0.0066	0.0057	0.0002	0.0008
85-89	0.0054	0.0092	0.0054	0.0000	0.0038	0.0049	0.0064	0.0045	0.0004	0.0019
90-94	0.0050	0.0149	0.0041	0.0009	0.0107	0.0030	0.0135	0.0029	0.0002	0.0107
95-99	0.0134	0.0305	0.0039	0.0094	0.0266	0.0147	0.0259	0.0137	0.0010	0.0122

資料: 筆者推計

このように、Model VI を用いて推計した都道府県別下方レキシストライアングル割合推計値は、男女別・年齢階級別に見ても全国値との整合性が高く、本モデルの有効性を示しているといえよう。

V. おわりに

本研究では、Wilmoth et al. (2021) で提案された、下方レキシストライアングル割合 ($\pi_d(x, t)$) に関する線形回帰モデルをわが国の死亡データに応用し、都道府県別下方レキシストライアングル割合推計のためのモデル開発を行った。

冒頭でも述べた通り、JMD の都道府県生命表は、現在、1975年以降のみの提供となっている (石井 2015)。都道府県別の年次別・年齢別・生年別死亡数は、個票が利用できない1971年以前については直接実績値が得られないことから、これ以前の生命表を作成するにあたっては、満年齢別・死亡年別死亡数を出生年別に分割してレキシストライアングル別死亡数を推計する必要があるが、本研究で提案したモデルを用いることによって、1975年以前の都道府県生命表の作成・提供を行うことが可能となる。

一方、本研究で提案したモデルについては残された課題も存在している。本研究では、1947年以降のデータに対してモデル開発を行った。これは、第二次世界大戦後とそれ以前

では、死亡率水準が大きく異なることが考えられること、また、都道府県別生命表の提供についても、まずは全国と同様、1947年以降の優先度が高いと考えられたことによるものである。

しかしながら、今後、全国の生命表とともに、都道府県別生命表についても第二次大戦前の作成の検討が必要と考えられ、その際には第二次大戦前についても同様のモデルが必要と考えられる。このような点については今後の課題としたい。

(2023年10月4日査読終了)

謝辞

本論文の草稿をお読み頂き、貴重なご指摘を頂いた匿名の査読者に感謝申し上げます。なお、残された誤謬は筆者らの責任である。

本研究は、国立社会保障・人口問題研究所「超長寿社会における人口・経済・社会のモデリングと総合分析」の研究成果であり、本稿で使用した「人口動態調査」に関する分析結果には、統計法第32条の規定に基づき、調査票情報を二次利用したものが含まれている。また、死亡票の独自集計であるため、分析結果が公表数値とは一致しない場合がある。

参考文献

石井太 (2015) 「日本版死亡データベースの構築に関する研究」, 『人口問題研究』, 第71巻, 第1号, pp.3-27. 国立社会保障・人口問題研究所「日本版死亡データベース」.

<https://www.ipss.go.jp/p-toukei/JMD/index.asp> (2023年8月31日最終アクセス)

林玲子・別府志海・石井太・山内昌和 (2021) 「日本人生年コホートからみた出生・死亡指標の変曲点について」, 『厚生労働省行政推進調査事業費補助金 (政策科学総合研究事業(政策科学推進研究事業)) 「長期的人口減少と大国際人口移動時代における将来人口・世帯推計の方論的發展と応用に関する研究」 令和2年度総括研究報告書』, pp.131-148.

Human Mortality Database. Max Planck Institute for Demographic Research (Germany), University of California, Berkeley (USA), and French Institute for Demographic Studies (France). Available at www.mortality.org (2023年8月31日最終アクセス)

Pressat, R. (1972) *Demographic Analysis : Methods, Results, Applications*, London: Edward Arnold.

Wilmoth, J. R., K. Andreev, D. Jdanov, D. A. Gleijer, and T. Riffe (2021) “Methods Protocol for the Human Mortality Database (Version 6)”, available at

<https://www.mortality.org/Public/Docs/MethodsProtocol.pdf> (2023年8月31日最終アクセス)

Study for a Model Splitting Death Counts by Age and Time into Designating by Birth Cohort

HORIGUCHI Yu and ISHII Futoshi

The Japanese Mortality Database (JMD) is a database of life tables in Japan, which is compatible with the Human Mortality Database (HMD). The life tables in the HMD are constructed based on deaths by the *Lexis triangle*, namely, by age, time, and birth cohort, which is similar to the JMD. In Japan, however, the number of deaths by prefecture and by the Lexis triangle is available only after 1972 (excluding 1947–1950), which resulted in the availability of prefectural life tables in the JMD only after 1975.

In the HMD protocol, a method for splitting death data by age and time into the Lexis triangle is used for constructing life tables in countries or time periods in which death counts by the Lexis triangle are unavailable. This method proposes a linear model for "the proportion of the lower Lexis triangle" with infant mortality rate and ratio of adjacent cohorts, among others, as the explanatory variables and is used to estimate death counts by the Lexis triangle.

This study aims to develop a model for estimating death counts by the Lexis triangle from the data by age and time in Japan before 1972 based on the HMD method. On the basis of results, the study concluded that Model VI of the six models is the most efficient one, which is based on data from 1947 to 1975 with interaction terms for the logarithm of infant mortality rates and dummy variables for age classes. Moreover, it presents variables derived using the backward selection method with AIC. The study posits that the proposed model is applicable to the estimation of prefectural life tables prior to 1975, enabling further extension of the JMD.

Keywords: deaths by prefecture, Lexis triangle, the Japanese Mortality Database