

## 介護状態生命表の作成

-高齢期のライフコース分析のために-

田中 周二\* 長谷川敏彦†

2017年11月26日

### 概要

日本の高齢化問題はますます深刻化しており、死に至るまでの収入の確保（年金）も重要であるが日常生活の健康（QOL）の確保（医療・介護）もそれに劣らず重要である。高齢期のこのような問題の解決のための基本的なツールとして、健康状態の変化を組み込んだ多状態生命表の作成が必須である。本稿では、S 県 H 市の介護保険の要介護度の履歴データを含む縦断データ（2006-2016）により、マルコフ過程モデルを使って、性別・年齢別の推移確率を推定した。これにもとづき、H 市の多状態生命表を作成した。前半では、この作成過程について説明する。

後半では、この確率推移行列と、全国ベースの介護保険事業年報の要介護度別年齢階級別人口を組み合わせることによって、日本全国の多状態生命表を作成する。このための方法論の開発が本稿のもう一つの目的である。

最後に応用として、得られた推移確率を用いて、2040 年後の状態別の人口予測を行い、超高齢化社会の姿について議論する。今後の社会保障費用の見通しの中で、医療保険と介護保険の伸びが著しく、2000 年以降社会保障と税の負担、それに伴う国家財政の健全化は内政上の最重要課題であり、中長期的にもそうあり続けると考えられる。その観点からも高齢者の健康状態について正確な見通しを推計することは政策判断にとって、今後ますます重要性を増してくると思われる。

キーワード 介護保険, 要介護度, 認知度, 寝たきり度, マルコフ過程モデル, 推移確率行列, 多状態生命表

## 1 はじめに

2000 年に発足した介護保険は、今年で 17 年を経て年金・医療と並んで高齢期を支える社会保障の 3 本の柱としての 1 本として定着している。

---

\* 発表者, 日本大学文理学部

† 未来医療研究機構 代表理事

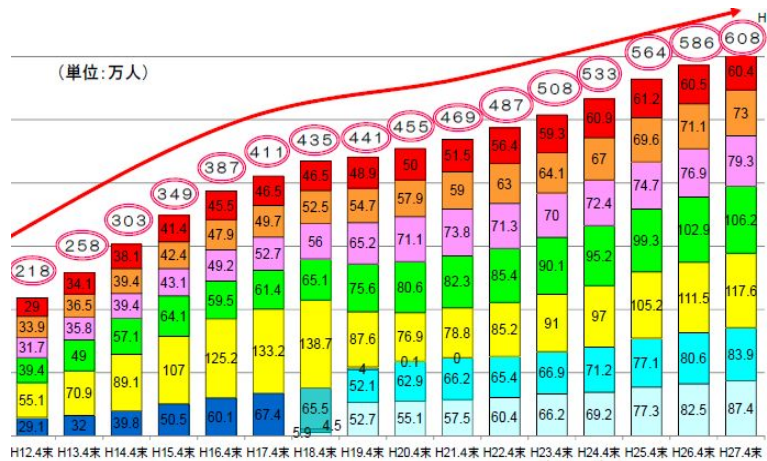


図1 要介護度別認定者数の推移

介護保険は、高齢者の介護を社会全体で支え合う仕組みであり、以下の3つの機能を有すると説明されている。

- 自立支援:単に介護を要する高齢者の身の回りの世話をするというを超えて、高齢者の自立を支援することを理念とする。
- 利用者本位:利用者の選択により、多様な主体から保健医療サービス、福祉サービスを総合的に受けられる制度。
- 社会保険方式:給付と負担の関係が明確な社会保険方式を採用

多くの国で、介護保険機能は民間保険会社が担っているが、日本はドイツと並んで広範な給付ときめ細かいサービスを提供する制度となっている。被保険者は、主に65歳以上の高齢者であるが、一定の介護状態になると認定を受け、必要であれば介護状態に応じたサービス給付<sup>\*1</sup>を受けることができる。制度の運営は、その性格から地域自治体に多くの部分が委任されている。

その財源は、40歳以上の被保険者から徴収する介護保険料と税金(国、都道府県、市町村)が折半となっている。

要介護認定は、申請があると認定調査員による第1次審査があり、コンピューターによる第1次判定のあと、要介護認定審査会による第2次判定によって確定する。

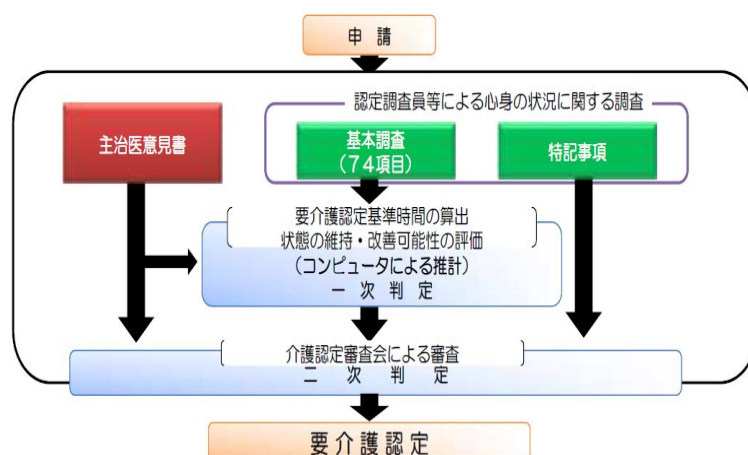


図2 要介護認定の流れ

<sup>\*1</sup> 月額の上限が約5万円(要支援1)から36万円(要介護5)まで

要介護認定により、各被保険者の要介護状態の推移が把握できると同時に、寝たきり度 (ADL)、認知症度も分かる。また、認定審査会には主治医による意見書も提出され介護判定に至った疾患名も分かる。

要介護認定を受けた被保険者は、それぞれのニーズに応じた介護サービスを受けることができる。

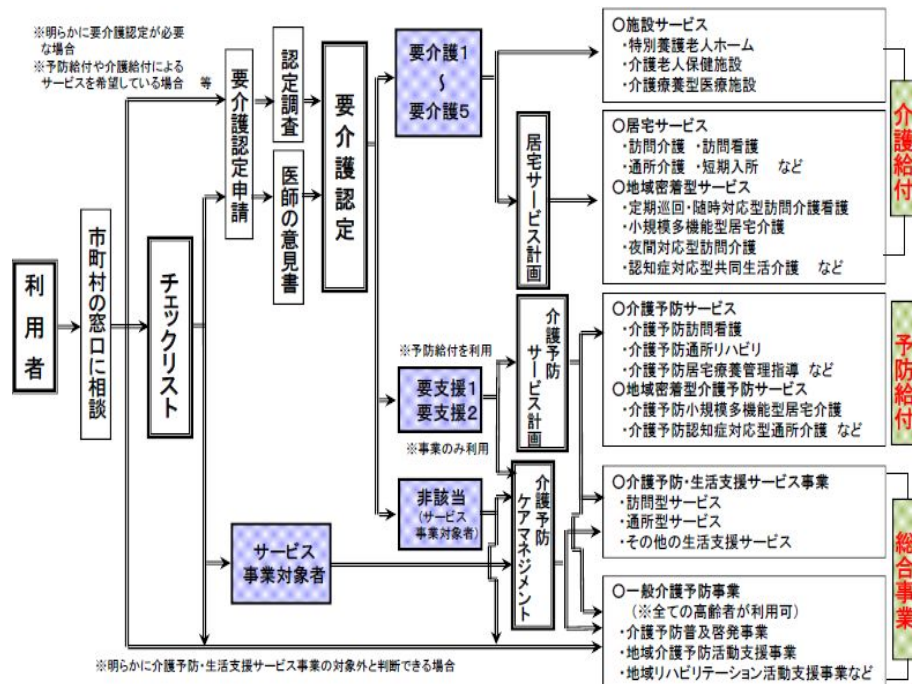


図3 介護保険のサービス利用

## 2 H市の介護保険データによる介護状態推移行列の推定

### 2.1 介護保険データについて

■H市のプロフィール H市はS県にある人口約11万4千人(2015年度国勢調査)の市であり、経済水準や人口構成など日本の平均的な特徴を有する。



図4 本州の中のS県

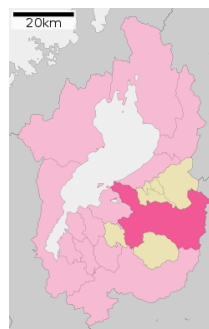


図5 S県の中のH市

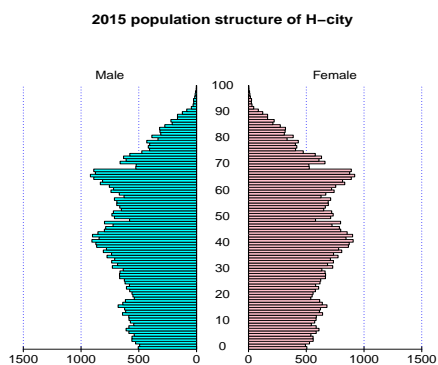


図6 総人口ピラミッド (2010年,2015年)

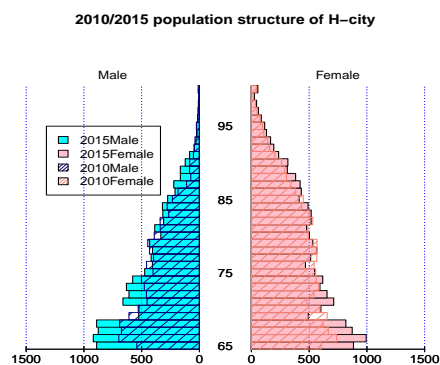


図7 65歳以上人口ピラミッド (2010年,2015年)

■H市の65歳以上人口および認定者の推移(2000-2016年) 65歳以上人口は2万8千人(男性1万3千人,女性1万5千人)で介護保険の第1号被保険者となっている。65歳以上の介護保険被保険者のうちの認定者数は以下のとおり推移している。

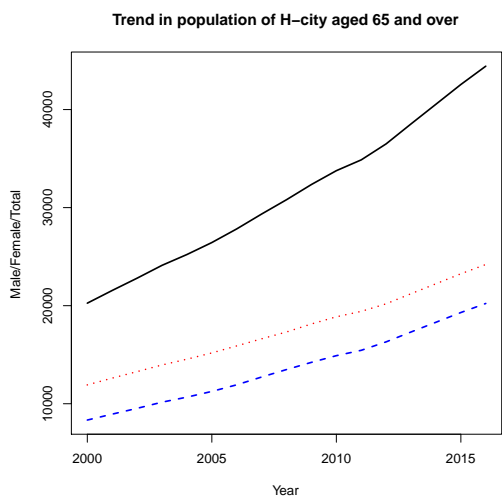


図8 H市の65歳以上人口の推移(男女別および総数)

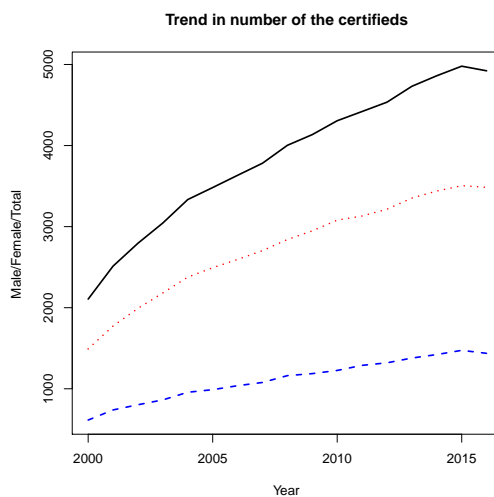


図9 介護保険認定者数の推移(男女別および総数)

## 2.2 介護保険の仕組みと収集データの概要

■収集データ H市の介護保険のデータファイルの種類と概要は以下のとおり。

名前	摘要	期間(データ長)	主要項目
資格台帳	被保険者の属性	2000/4-2016/9 (42145)	ID, 生年月日, 性別, 資格取得日, 死亡日
認定履歴記録	認定者の変更履歴	2006/4-2016/9 (45338)	ID, 認定日, 要介護度, 認知度, 寝たきり度
第1回認定記録	第1回認定記録	2000/4-2016/9 (15301)	ID, 認定日, 要介護度
認定履歴予備記録	認定者の変更履歴*2	2000/4-2016/9 (48708)	ID, 認定日, 要介護度, 認知度, 寝たきり度

その他に、2010/1-2015/1 までに限り、主治医意見書の3病名、介護保険費用(毎月)、医療費(毎月)のデータファイルが存在する。

■3つの認定履歴記録のチェック 第1回認定記録、認定履歴記録、認定履歴予備記録の3つの認定年度別比較を行うと、認定履歴予備記録の2006年以前には欠損値が多いと推測される。

年度	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
第1回認定記録		2392	801	722	708	750	713	724	754	853
認定履歴予備記録	13	806	1156	1252	1398	1790	2178	2689	3362	3068
認定履歴記録							1	2338	4351	3004
年度	2009	2010	2011	2012	2013	2014	2015	2016	2017	
第1回認定記録	787	884	809	914	951	912	934	693		
認定履歴予備記録	3885	4069	4154	3967	4087	4072	4188	2546		
認定履歴記録	3441	4192	4301	4061	4089	4124	4186	3954	84	

## 2.3 介護度別人口数・割合の推移

■介護度別総人口数・割合の推移 男女計の要介護状態ごとの人口とその割合の推移は以下のとおり。

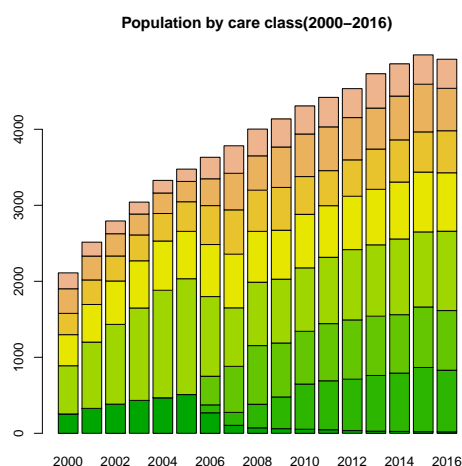


図10 介護度別総人口数の推移

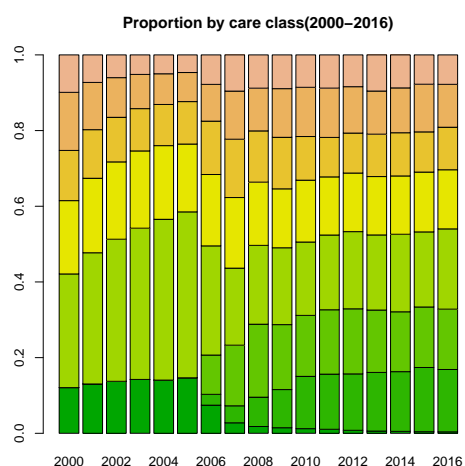


図11 介護度別総人口割合の推移

## 2.4 認定者の死亡率

■2006-2016年データより推定された年齢別粗死亡率 2006-2016年の非該当者の生存人口と非該当者の死亡人口、その比率を表にすると以下のとおり。

年齢区分	(40,45]	(45,50]	(50,55]	(55,60]	(60,65]	(65,70]	(70,75]
生存人口	37	76	157	464	5591	6524	7700
死亡人口	0	1	2	4	16	309	777
比率	0.0000	0.0132	0.0127	0.0086	0.0029	0.0474	0.1009
年齢区分	(75,80]	(80,85]	(85,90]	(90,95]	(95,100]	(100,105]	(105,110]
生存人口	8809	7219	4683	2861	972	172	28
死亡人口	754	664	322	110	32	5	0
比率	0.0856	0.0920	0.0688	0.0384	0.0329	0.0291	0.0000

### 3 介護状態間の推移確率行列

#### 3.1 マルコフ連鎖モデル

介護状態間の推移確率行列の推定については、標準的なマルコフ推移行列の推定方法を採用した。多状態の推移モデルにおいては、推移確率行列  $\mathbf{P}(x, y)$  と状態滞在確率  $\mathbf{S}(x, y)$  という2つの確率概念が登場する。

異なる2つの年齢  $x < y$  に対し、 $\mathbf{P}(x, y)$  は  $x, y$  間の推移確率の行列を表し、 $\mathbf{p}(x)$  は  $x$  歳の状態確率ベクトルを表す。 $x, y$  間に継続して同じ状態に滞在する確率を  $\mathbf{S}(x, y)$  で表し、特に年齢  $x$  の生存確率をその年齢で生存している確率と定義する。

$x$  から  $y$  までの間隔は、 $x, x+1, x+2, \dots, y$  と分割でき、これを推移確率行列で表すと、行列の掛け算になる。

$$\mathbf{P}(x, y) = \mathbf{P}(x, x+1)\mathbf{P}(x+1, x+2) \cdots \mathbf{P}(y-1, y) \quad (3.1)$$

$x$  歳の状態確率ベクトル  $\mathbf{p}(x)$  は、 $(y-x)$  年後に  $\mathbf{p}(y) = \mathbf{P}(x, y)\mathbf{p}(x)$  に移動する。

介護保険の場合には、 $t$  年度始めの  $x$  歳の状態別人口ベクトル  $\mathbf{l}(x) = (l_1(x), l_2(x), \dots, l_8(x) = 0)$

$=$ (非該当  $(x)$ , 要支援 1 $(x)$ , 要支援 2 $(x)$ , 要介護 1 $(x)$ ,  $\dots$ , 要介護 5 $(x)$ , 死亡 (5)) となり、死亡は吸収状態となる。

$x$  歳から  $(x+1)$  歳の状態人口ベクトルの変化は  $\mathbf{l}(x+1) = \mathbf{P}(x, x+1)\mathbf{l}(x)$  と表される。

一般に、推移行列  $\mathbf{P}(x, y)$  は、

$$\mathbf{P}(x, y) = \begin{bmatrix} p_{11}(x, y) & p_{12}(x, y) & \cdots & p_{17}(x, y) & q^{(1)}(x, y) \\ p_{21}(x, y) & & \ddots & & q^{(2)}(x, y) \\ & & \ddots & & q^{(3)}(x, y) \\ \vdots & & \cdots & p_{44}(x, y) & \cdots & \vdots & q^{(4)}(x, y) \\ \vdots & & & & \ddots & & q^{(5)}(x, y) \\ & & & & & & q^{(6)}(x, y) \\ p_{71}(x, y) & & \cdots & & & p_{77}(x, y) & q^{(7)}(x, y) \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

と表される。

#### 3.2 推定方法

推移確率の推定は、分母を経過件数、分子をイベント発生件数とした経験率より求める。ノンパラメトリックな方法のほか、分布を仮定するパラメトリックな方法 (以下は区分的指数分布を仮定して最尤推定法による Nelson-Aalen 法を紹介) がある。

- 推移確率の計算の基礎は、個人  $k$  が状態  $\{1, 2, \dots, I\}$  の特定の状態  $i$  から  $j$  に  $t_1$  から  $t_2$  に推移する数

${}_k N_{ij}(t)$  である.

- ${}_k Y_i(t)$  を  $k$  が時点  $t$  で状態  $i$  か否かを表す指示関数とし,  $Y_i(t) = \sum_{k=1}^K {}_k Y_i(t)$  は集団の中での状態  $i$  の人数を表し, リスク集合と呼ぶ.
- 計数過程  ${}_k N_{ij}(t)$  を強度が瞬間推移率と暴露の状態  ${}_k \lambda_{ij} = {}_k \mu_{ij}(t) {}_k Y_i(t)$  に依存するポアソン過程とする.
- $K$  人の集団で考えると, 計数過程は  $N_{ij}(t) = \sum_{k=1}^K {}_k N_{ij}(t)$  となる.
- $dM_{ij}(t) = dN_{ij}(t) - \lambda_{ij}(t)dt$ . (Aalen) と定義される  $M_{ij}$  をマルチンゲールと仮定する.
- 強度が瞬間推移率と暴露の状態に依存しているため, 尤度関数は, 計数過程  ${}_k N_{ij}(t)$  と強度過程  ${}_k \lambda_{ij}(t)$  によって表現される. 尤度関数は,

$$\left[ \prod_{n=1}^{{}_k N_{ij}(\omega)} {}_k \lambda_{ij}^n ({}_k T_{ij}^n) \exp \left[ - \int_0^\omega {}_k \lambda_{ij}^n(\tau) d\tau \right] \right] \exp \left[ \int_0^\omega - {}_k \lambda_{ij}^c(\tau) d\tau \right]$$

となる. ここで,  $\omega$  は最終年齢,  ${}_k T_{ij}^n$  は  $i \rightarrow j$  の  $n$  回目の推移が生じた時刻を表す.

- この尤度関数を最大化するパラメーターが推移率の最尤推定値である.
- 累積ハザード率  $A_{ij}(t) = \int_0^t dA_{ij}(\tau)$  であり,  $dA_{ij}(\tau)$  は微小区間の累積ハザード率の増分に等しい. これが連続過程の場合には,  $dA_{ij}(\tau) = \mu_{ij}(\tau) d\tau$  となる.
- 年齢  $t$  の自然な累積推移率の推定値は,  $\hat{A}_{ij} = \int_0^t \frac{dN_{ij}(\tau)}{Y_i(\tau)}$  である.

まず縦断データの場合の推移確率の推定のためには, 上のロジックで分かるように, リスク集合と状態間推移数を集計する必要がある. また縦断データには打ち切り, 切断など不完全データがあるため, データ処理が厄介になる.

R のパッケージ Biograph では縦断データを効率的に集計・処理して, 生存時間解析や推移確率を推定する各種パッケージのデータ入力形式に合わせて変換する機能も揃っている. ただ, 前処理として Biograph object と呼ばれる特殊なデータ形式に変換しておかなくてはならない. 以下は, 介護保険の Biograph object の一部である.

#### Biograph object の例

```
ID born start end sex zip home ns path    dates.Tr1 dates.Tr2
1 42 -3 1275 1401 2 7 0 4 YCFX 1276 ...
2 240 46 1275 1401 2 6 0 3 YFX 1276 ...
3 422 61 1275 1401 1 1 0 3 YCX 1276
```

第1欄は個人コード, 第2欄は生年月, 第3,4欄はデータ開始日と終了日, 第5欄から第7欄は個人の属性, 第6欄は状態間推移回数, 第7欄は状態推移列 (第1行は YCFX と変遷した), 第8欄以降は状態推移日を記載する. なお年月表示は1900年1月からの月数を数えた CMC (Century Month Code) である.

なお連携できるパッケージは, Survival(survival analysis), eha(transition rate), etm, mvna(Nelson-Aalen, Aalen-Johansen) などである.

### 3.3 状態推移パス総数, 平均年齢

介護保険履歴データより集計すると以下の結果を得た.\*3

\*3 以下, 介護状態をアルファベットで表示.(Y=非該当, A=要支援 1, B=要支援 2, C=要介護 1, D=要介護 2, E=要介護 3, F=要介護 4, G=要介護 5, X=死亡)

	Y	A	B	C	D	E	F	G	X
Y	0	741	653	798	528	373	370	231	1201
A	36	0	250	220	66	35	39	17	124
B	10	193	0	201	209	94	63	23	153
C	4	93	196	0	453	238	176	97	318
D	1	22	72	195	0	363	223	110	437
E	2	8	18	56	183	0	283	154	510
F	0	5	5	27	68	138	0	225	698
G	1	0	2	6	12	26	90	0	678
X	0	0	0	0	0	0	0	0	0

表1 状態間推移パス総数 (男性)

	Y	A	B	C	D	E	F	G	X
Y		80.18	78.45	79.91	78.83	78.08	78.62	78.02	85.08
A	79.25		81.20	81.79	82.30	81.10	83.96	84.79	81.50
B	79.80	78.87		82.29	79.83	82.35	81.48	81.67	81.80
C	88.25	78.95	78.41		81.05	82.44	83.25	84.52	82.47
D	72.50	79.00	77.31	79.76		81.30	82.39	81.88	82.36
E	80.50	79.62	80.56	79.34	80.01		81.11	81.94	83.04
F		80.30	76.90	78.65	80.93	81.27		80.82	83.30
G	90.50		83.50	76.83	76.75	80.77	81.33		82.41
X									

表2 状態推移時点の平均年齢 (男性)

	Y	A	B	C	D	E	F	G	X
Y	0	1416	1191	1338	681	433	359	205	1789
A	55	0	615	429	124	53	53	15	121
B	20	467	0	527	391	139	137	51	181
C	7	142	463	0	1088	471	333	185	414
D	1	25	152	414	0	788	500	223	540
E	2	7	32	93	337	0	701	306	572
F	0	4	13	42	114	285	0	595	967
G	0	1	4	14	22	38	199	0	1144
X	0	0	0	0	0	0	0	0	0

表3 状態間推移パス総数 (女性)

### 3.4 確率推移行列

推定された確率推移行列 (年齢合算) は以下のとおり.



	Y	A	B	C	D	E	F	G	X
Y		81.08	80.80	82.18	82.34	81.91	82.56	81.93	86.72
A	80.97		83.37	83.41	84.27	84.18	84.29	81.43	82.29
B	80.05	81.67		84.53	84.74	84.97	85.51	87.05	85.24
C	81.79	81.13	81.71		84.75	85.17	86.64	86.52	86.16
D	79.50	82.26	82.47	83.35		85.47	86.91	86.44	86.69
E	84.50	79.50	83.12	83.79	83.97		86.11	86.61	87.96
F		83.75	80.42	85.02	85.72	85.04		85.87	88.52
G		91.50	84.00	82.50	85.05	88.63	85.53		87.79
X									

表4 状態推移時点の平均年齢(女性)

	Y	A	B	C	D	E	F	G	X
Y	88.66	1.45	1.37	1.62	1.19	0.85	0.83	0.51	3.51
A	1.67	62.74	9.79	8.67	3.69	2.11	2.07	1.00	8.26
B	0.46	6.20	63.00	6.81	7.18	3.68	2.66	1.17	8.85
C	0.14	2.20	4.41	58.88	9.73	5.64	4.36	2.47	12.16
D	0.05	0.69	1.89	4.59	59.39	8.33	5.60	2.95	16.51
E	0.08	0.32	0.68	1.84	5.35	56.86	7.95	4.51	22.41
F	0.01	0.19	0.25	0.96	2.29	4.22	55.48	6.36	30.23
G	0.06	0.02	0.13	0.35	0.71	1.36	4.08	51.19	42.11
X	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00

表5 推移確率行列(男性)

	Y	A	B	C	D	E	F	G	X
Y	86.64	2.22	2.03	2.21	1.27	0.83	0.72	0.43	3.64
A	1.16	68.32	11.47	8.28	3.29	1.60	1.52	0.63	3.73
B	0.37	6.54	68.73	7.69	5.99	2.62	2.51	1.15	4.40
C	0.10	1.60	4.70	64.75	10.39	5.24	4.04	2.39	6.79
D	0.03	0.44	2.01	4.93	63.40	9.25	6.56	3.37	10.00
E	0.04	0.16	0.62	1.62	5.07	63.83	10.40	5.27	12.99
F	0.00	0.09	0.27	0.76	1.89	4.31	63.82	8.94	19.91
G	0.00	0.03	0.11	0.32	0.53	0.90	3.91	65.85	28.36
X	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	100.00

表6 推移確率行列(女性)

### 3.5 多状態生命表

以上より,介護度を状態とする多状態生命表の作成例を図示する.起点は65歳の非該当者,75歳の要介護1の男女とする.

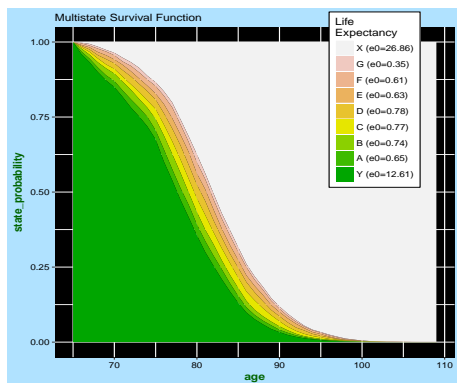


図 12 非該当者生存関数 (65 歳, 男性)

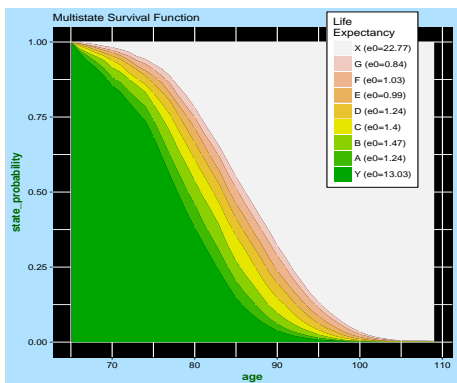


図 13 非該当者生存関数 (65 歳, 女性)

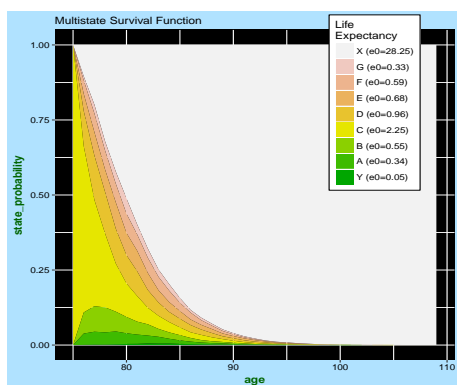


図 14 要介護 1 生存関数 (75 歳, 男性)

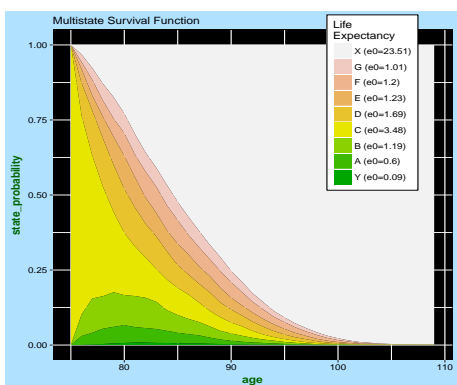


図 15 要介護 1 生存関数 (75 歳, 女性)

## 4 日本全国ベースの多重状態生命表の試作

H 市の確率推移行列は、データのみに基づく粗推移確率であり、生命表の作成でしばしば行われる補整や外挿などの処理を行っていない。今回は、全国レベルで使用できる「多状態生命表」を H 市の経験推移率と第 21 回生命表および過去の要介護状態別年齢別人口統計を使用して「校正」することにより試作する。この過程で、推定を行う場合の諸課題を発見することも目的とする。

第 21 回生命表は、2015 年の国勢調査にもとづく 5 年に 1 度公表される完全生命表である。介護保険データ期間が、2006-2016 年ということもあり、その間に公表された直近の完全生命表ということにより採用した。男女別・年齢別・要介護状態別人口統計は、厚生労働省が毎年公表している介護保険事業状況報告（年報）の中の要介護（要支援）認定者数により把握することができる。ただし、年齢別は 5 歳年齢階級別となっている。

以下、手順を箇条書きにして説明する。

1. まず H 市の介護保険データを男女別に分け、以下の手順を実行する。
2. 年齢別要介護状態別の経験推移率の分母の経過契約数と分子の推移件数を 5 歳階級別のデータに集約する。結果として、65-69 歳、70-74 歳、...、95 歳以上の 7 つの年齢階級の各状態間の推移データが作成できる。
3. 各状態からの死亡推移率（死亡率）については、一般化線形モデルによりまとめてポアソン回帰を行ったところ、非該当から要支援 1,2 までと要介護とでは明らかに異なる傾向であった。前者では年齢別死亡率

が低いレベルから急速に高くなるのに対し、後者は最初からかなり高く、上がり方も緩やかであった。このため、2グループで別々にポアソン回帰を実行した。説明変数は年齢の2次式とカテゴリー変数として介護状態とした。

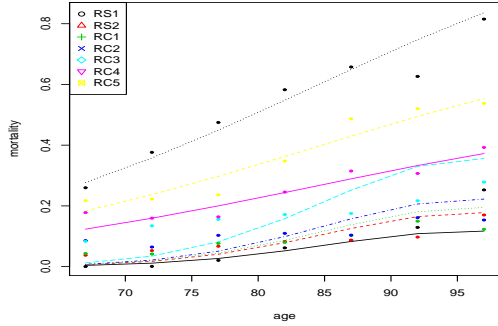


図 16 要介護度別死亡率 (男性)

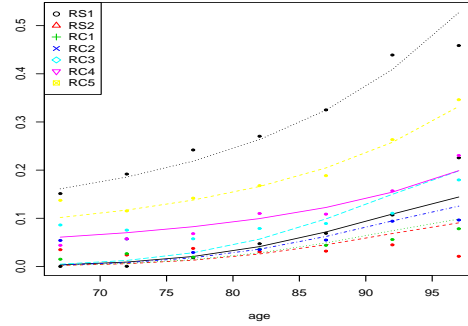


図 17 要介護度別死亡率 (女性)

死亡を除く各状態間の推移率は、それぞれが相当異なる動きをしていたため、単独で年齢の2次式としてポアソン回帰を行った。このポアソン回帰後の推移行列を  $M$  とする。以下は 65-69 歳の非該当状態から死亡を除く各状態への推移率を粗推移確率と平滑化後の推移確率を示している。

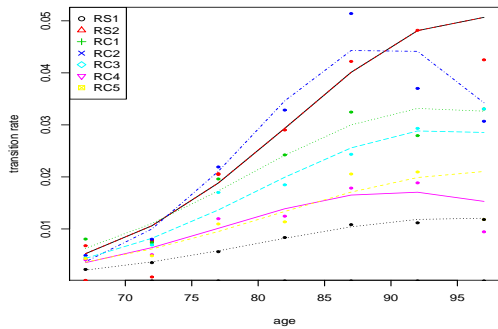


図 18 非該当から多状態への推移確率 (男性, 65-69 歳)

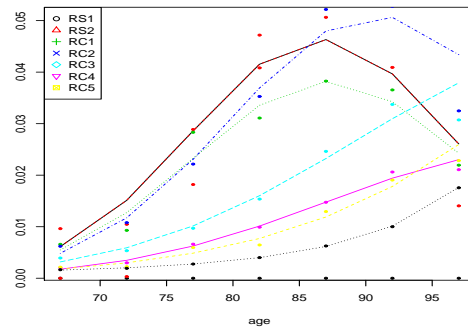


図 19 非該当から多状態への推移確率 (女性, 65-69 歳)

1. 各状態別死亡率を各状態別人口で加重平均をとったものがその年齢の第 21 回生命表の死亡率に一致しなければならない。このため、介護保険事業状況報告 (年報) と総人口から各状態別人口を計算し、これを制約条件とした
2. 各状態別人口は、過去の改正により 2009 年から 2016 年までの 8 年間しかデータの連続性がない。また 5 歳階級別の人口をとったため、5 年ごとの組み合わせをとると、2009 年と 2014 年、2010 年と 2015 年、2011 年と 2016 年の 3 組しかない。
3. 従って、例えば 65-69 歳の人口について、推移行列を  $M$  とすると  $M$  をかけると 1 年後の 66-70 歳の人口となるはずである。しかしながら、H 市と推移確率が異なるため、一部の要素を修正する。
4. 人口は 5 歳階級別になっているので松山・尾山 (2013) の積分補間の方法で 1 年後の人口を求める。
5. 死亡数については第 22 回生命表の死亡率にもとづき決定する。

6. これらの制約条件のもとで次節以降で説明する Elastic Net の方法を用いて行列の各要素を推定する.

以上の手続きにより,5 歳年齢階級の年齢と介護状態の要素間が滑らかに接続する 9 行 9 列の確率推移行列が得られた. 年齢階級は 7 つあるので, 年齢階級を含めると 3 次元配列となる. 生命表は, 性別と年齢階級  $i$ , 介護状態  $j$  を与えると, 初期値ベクトルと  $\mathbf{e}_j = (0, \dots, 0, 100000(j \text{ 番目}), 0 \dots 0)'$  と行列  $M_i$  を選択し, 5 年後の介護状態別人口分布は  $\mathbf{e}_j M_i^5$  となる. 10 年後は,  $\mathbf{e}_j M_i^5 M_{i+1}^5$  となり, 以下, 5 年ごとに延長して, 生命表が完成する. 全部をアウトプットすると膨大な量となるため, PC 上にファイルを置いてその都度, 作成できるようにすればよいだろう. イメージは 3.5 で示したような生存曲線と表が出力されることになる.

## 5 LASSO, Ridge, Elastic Net

### 5.1 L1 正則化, L2 正則化とその混合

多くの説明変数を用いる重回帰分析の変数選択に効力を発揮する.

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \quad (5.1)$$

係数における罰金項をどのような関数を用いるかにより LASSO, Ridge, Elastic Net になる.

LASSO (L1 正則化) 罰金項を  $\|\boldsymbol{\beta}\|_1 = \sum_k |\beta_k|$  として

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} (\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_1^2 - \lambda \|\boldsymbol{\beta}\|_1) \quad (5.2)$$

$\lambda = 0$  のときには OLS になる.

Ridge (L2 正則化) 罰金項を  $\|\boldsymbol{\beta}\|_2 = \sqrt{\sum_k \beta_k^2}$  として

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} (\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2 - \lambda \|\boldsymbol{\beta}\|_2^2) \quad (5.3)$$

Elastic Net (L1/L2 正則化) LASSO と Ridge の割合  $\alpha$  による混合.

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} \alpha (\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_1^2 - \lambda \|\boldsymbol{\beta}\|_1) + (1 - \alpha) (\|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|_2^2 - \lambda \|\boldsymbol{\beta}\|_2^2) \quad (5.4)$$

### 5.2 確率推移行列の要素の推定

H 市の性別  $k$ , 年齢階級  $l$  の確率推移行列  $M_{kl}$  を校正して, 同じ性別・年齢階級の確率推移行列  $\tilde{M}_{kl}$  を推定する.  $\tilde{M}_{kl}$  に求められる要件は, 同じ性別・年齢階級の要介護状態別人口が 1 年後に 1 歳上の年齢階級 (例えば 65-69 歳ならば 66 歳-70 歳) の要介護状態別人口にできるだけ「近くなる」ように推定することである. ところが, 人口は 5 歳階級ごとにしか公表されていない. このため, 5 年ごとに合わせれば正確であるが, サンプル数が減ってしまう.

このため, 松山・尾山 (2017) によるアルゴリズムを利用して 1 歳ずらした人口を推定した. この方法は 5 歳年齢階級しか公表されていない統計数値において各歳別の数値を求める場合に適した方法である. 各年齢での数値を求める上で,

1. 各年齢における数値の年齢区分における合計は元データと一致すること.
2. 年齢区分による境界点において連続であること.

という二つの制約を満たす手法として以下の計算ステップを導入し, 以下では積分補間と呼ぶことにする.

### 5.3 積分補間

前提として公的統計から区分  $i = 1, 2, \dots, n$  に対するデータ  $\{(x_i, y_i)\}$  が得られているとする。  $x_i$  は年齢区分を表す。

- (ステップ 1) 観測値から累積度数  $Y_i = \sum_{j=1}^i y_j$  を求め、データ  $\{(x_i, Y_i)\}$  を得る。
- (ステップ 2)  $Y_i$  の傾きを見ながら連続した年齢区分を選択し滑らかな累積度数関数  $F(x)$  を 3 次元スプライン関数により推定する。区分を  $m$  個選択した場合、 $F(x)$  は選択した  $(x_i, Y_i)$  の点を全て通過する。
- (ステップ 3) 年齢区分の選択により複数個の  $F(x)$  が求まるが重複部分に関しては和半等により調整を行い、0 歳とデータにおける最終年齢に関しては定数補外による調整を行う。
- (ステップ 4) 各歳の値  $f(x)$  は

$$f(x) = F(x+1) - F(x) \quad (x = 0, 1, \dots, \omega) \quad (5.5)$$

により求める。

ただし  $f(x)$  が負値とならないようにステップ 3 において調整を行う。

### 5.4 Elastic Net による推定

以下の手続きは  $k, l$  によらず同じなので、 $k, l$  は省略する。

- まず、 $\tilde{M} - M = E$  として、修正行列  $E = (e_{ij})$  の係数を推定する問題に置き換える。
- $t$  年度の要介護状態ベクトル  $l_t = (l_t^j)_{j \in C}$   $C$ : 要介護状態の集合は 1 歳年齢が上昇した  $l_{t+1}$  となるが、この  $l_{t+1}$  は積分補間の方法で求めておく。
- こうすると、

$$l_{t+1} = \tilde{M}l_t = (M + E)l_t = Ml_t + El_t \quad (5.6)$$

- また、死亡率を与えられた死亡率  $q$  とすると、 $q \sum_j l_t^j = \sum_j (\tilde{M}l_t)^j$  が成立する。

これを、行列  $E$  の制約条件として Elastic Net による推定を行う。

## 5.5 行列要素の線形回帰モデル

$y = E\beta + \epsilon$ ,  $y$ : 119(17 \* 7) 要素のベクトル,  $E$ : 119 \* 65,  $\beta$ : 係数ベクトル

ただし, 切片 ( $\beta_0$ )=0, 列和  $\sum_{j=1}^n E_{ij} = 1$  であつ  $\tilde{M} \geq 0$  すなわち  $E \geq -M$  という制約付きの回帰計算を行う.

式番号	y	E1	E2	E3	E9	E10	E11	E65
1	15.65	224.24	3.18	3.56	0.00	0.00	0.00	0.00
2	-3.13	0.00	0.00	0.00	224.24	3.18	3.56	0.00
3	-3.36	0.00	0.00	0.00	0.00	0.00	0.00	0.00
4	-2.99	0.00	0.00	0.00	0.00	0.00	0.00	0.00
5	-2.04	0.00	0.00	0.00	0.00	0.00	0.00	0.00
6	-2.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00
7	-0.76	0.00	0.00	0.00	0.00	0.00	0.00	0.00
8	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
9	0.00	1000.00	1000.00	1000.00	0.00	0.00	0.00	1000.00
10	0.00	0.00	0.00	0.00	1000.00	1000.00	1000.00	1000.00
11	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1000.00
12	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1000.00
13	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1000.00
14	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1000.00
15	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1000.00
16	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1000.00
17	-2.14	0.00	0.00	0.00	0.00	0.00	0.00	252.70

表7 回帰モデルの応答変数と説明変数 (一部)

## 5.6 較正前後の確率推移行列 (男性,75-79 歳)

	非該当	要支援 1	要支援 2	要介護 1	要介護 2	要介護 3	要介護 4	要介護 5	死亡
非該当	0.93	0.01	0.01	0.01	0.01	0.00	0.00	0.00	0.03
要支援 1	0.01	0.62	0.12	0.13	0.04	0.02	0.02	0.01	0.03
要支援 2	0.00	0.09	0.59	0.09	0.11	0.04	0.03	0.01	0.04
要介護 1	0.00	0.04	0.07	0.53	0.14	0.07	0.05	0.04	0.04
要介護 2	0.00	0.01	0.03	0.08	0.57	0.14	0.08	0.04	0.06
要介護 3	0.00	0.00	0.00	0.00	0.09	0.69	0.00	0.00	0.20
要介護 4	0.00	0.01	0.01	0.04	0.07	0.00	0.46	0.11	0.30
要介護 5	0.00	0.00	0.00	0.01	0.01	0.02	0.07	0.44	0.45
死亡	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00

表8 較正後の  $\tilde{M}$

	非該当	要支援 1	要支援 2	要介護 1	要介護 2	要介護 3	要介護 4	要介護 5	死亡
非該当	0.88	0.02	0.02	0.02	0.01	0.01	0.01	0.01	0.03
要支援 1	0.02	0.62	0.12	0.13	0.03	0.02	0.02	0.00	0.04
要支援 2	0.01	0.09	0.59	0.08	0.11	0.04	0.03	0.01	0.04
要介護 1	0.00	0.04	0.07	0.53	0.14	0.07	0.05	0.04	0.04
要介護 2	0.00	0.01	0.03	0.08	0.57	0.14	0.07	0.04	0.07
要介護 3	0.00	0.00	0.00	0.00	0.09	0.69	0.00	0.00	0.20
要介護 4	0.00	0.00	0.01	0.04	0.07	0.00	0.46	0.11	0.30
要介護 5	0.00	0.00	0.00	0.01	0.01	0.02	0.07	0.44	0.45
死亡	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00

表9 較正前の  $M$

	RMSE	$R^2$
較正前	190917.1	0.9721
LASSO	92665.7	0.9945
Ridge	121183.4	0.9901
Elastic Net	89821.1	0.9949

表 10 LASSO,Ridge,ElasticNet( $\alpha=0.95$ ) の精度比較

## 5.7 較正方法の概要と精度比較

較正には R パッケージの `glmnet` を利用した。 $\alpha = 1$  のとき LASSO,  $\alpha = 0$  のとき Ridge, それ以外は Elastic Net となる. Elastic Net の場合にはまず最適な  $\alpha$  を決め, 対応する  $\lambda$  を求める.

$$\text{glmnet}(E, Y, \text{alpha} \in [0, 1], \text{intercept} = \text{FALSE}, \text{lower.limits} = \text{Lower}) \quad (5.7)$$

\*切片ゼロ,  $\tilde{M}$  が正となるよう下限を設定 LASSO,Ridge,ElasticNet( $\alpha=0.95$ ) のそれぞれの方法による精度比較を行う. 比較尺度は,  $RMSE$  と  $R^2$  である.

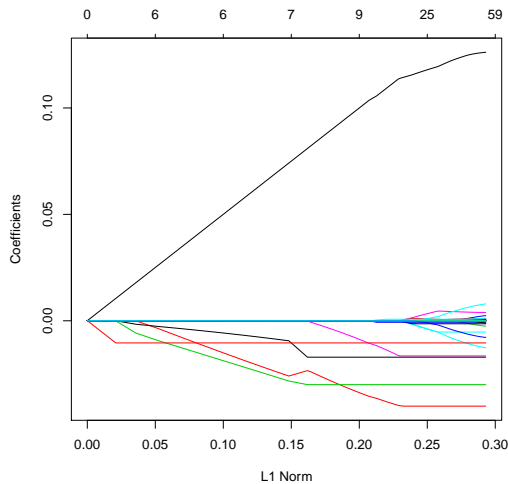


図 20 Elastic Net の係数樹状図

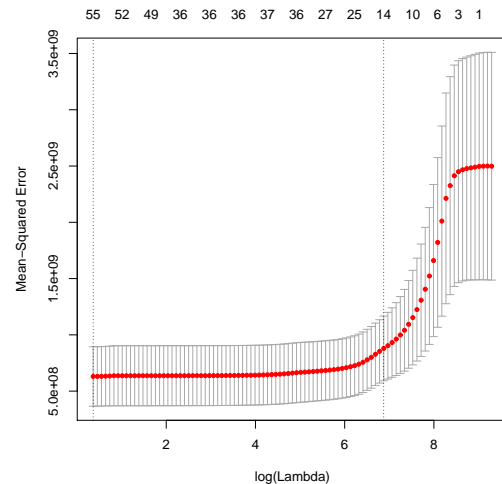


図 21 クロスバリデーションによる変数選択

## 6 要介護状態別の将来人口推計

社会保障人口問題研究所では,5 年に一度,日本の人口の死亡率予測(上位,中位,下位)を実施しており,その予測に基づいて将来人口の推計を行っている.この将来人口推計は,公的年金の財政予測に利用されるなど多くの応用がある.ところが,将来人口推計の結果を見せることにより人口ピラミッドが壺型に変化する様子を見て高齢化の深刻さについてそれなりに理解できるが,さらに高齢者人口の健康状態の詳細が分かると具体的な施策につながる有用な情報が得られると考えられる.ここに多状態生命表の応用分野がある.

将来人口推計の性・年齢別の人口に整合的な確率推移行列の較正とそれを用いた介護状態別の分布を求める手順は以下のとおりである.

1. 例えば 2015 年の 65-69 歳の介護状態別人口ベクトル  $\mathbf{l}_{67}$  は,5 年後に 2020 年 70-74 歳の人口ベクトルに移る.
2. 2015 年生命表にもとづく確率推移行列  $M_{67}$  があるので, $M_{67}^5 \mathbf{l}_{67}$  の死亡以外の要素の和が 2020 年 70-74 歳の将来人口推計の値に一致すればよい.
3. 確率推移行列  $M_{67}$  の死亡率に一定率を乗じて,2020 年 70-74 歳の将来人口推計との最小二乗誤差を最小化するように一定率を決定する.
4. 2020 年の 65-69 歳の介護状態別人口ベクトルを得るには,2015 年の 60-64 歳の人口 (非該当) から求める. 以下,2025 年,2030 年もより若い年齢層の人口を利用する.
5. 2025 年 75-79 歳人口ベクトルは,2020 年 70-74 歳の人口ベクトルから求める. 以下, 同様にして 2045 年までの介護状態別人口分布が得られる.

以下がこの計算結果である. 最初の図は,2020,2030,2040 年度の男女の介護状態別人口分布である. 男女とも非該当の人口はほとんど変化しないが, 要介護状態の人口が増えてゆく. 第 2 図以下は各年度の年齢別・介護状態別分布である. 男女で大きく異なる. 女性の寿命が長いので, 後期高齢者の人口は女性の伸びが大きい. 後期高齢者では要介護者しかも重度の要介護者が増大する.2030 年までは上昇するが, その後は緩慢な増加という傾向が見られる.

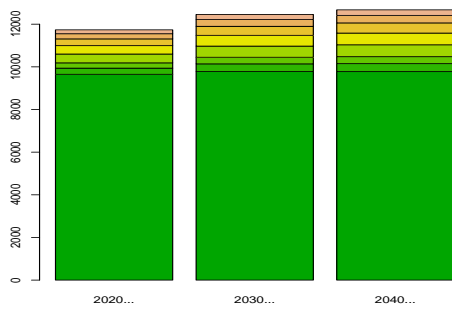


図 22 2020 年,2030 年,2030 年介護状態別人口分布 (男性)

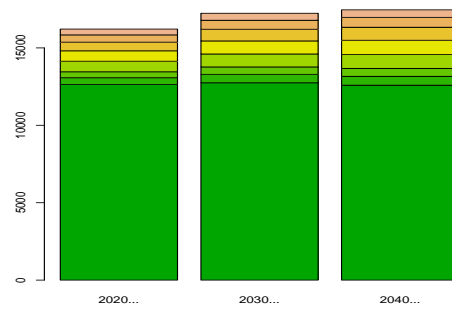


図 23 2020 年,2030 年,2030 年介護状態別人口分布 (女性)

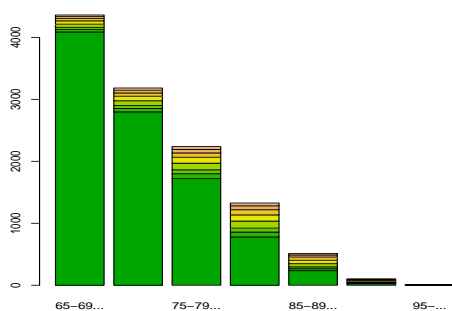


図 24 2020 年年齢階級別介護状態別人口分布 (男性)

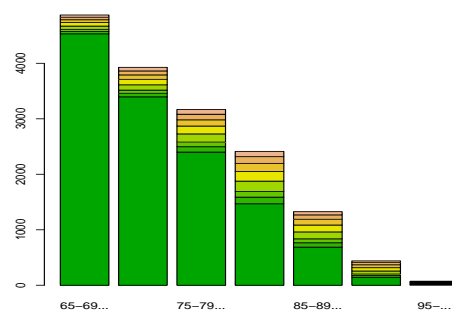


図 25 2020 年年齢階級別介護状態別人口分布 (女性)



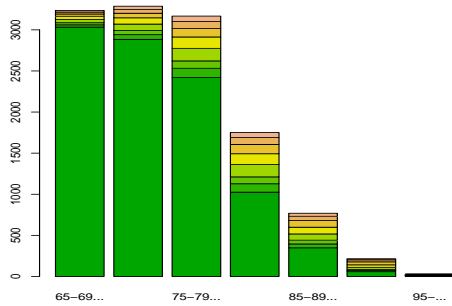


図 26 2030 年年齢階級別介護状態別人口分布 (男性)

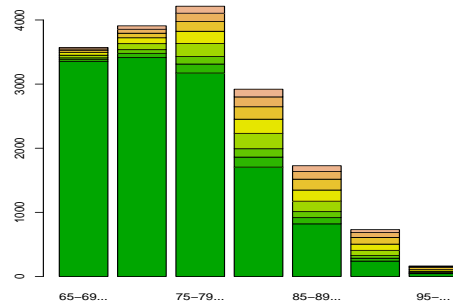


図 27 2030 年年齢階級別介護状態別人口分布 (男性)

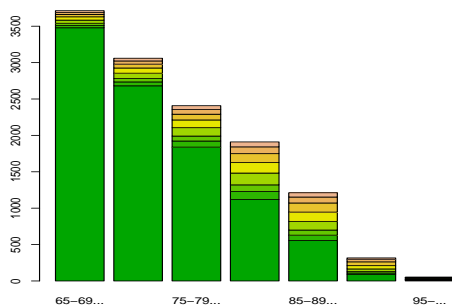


図 28 2040 年年齢階級別介護状態別人口分布 (男性)

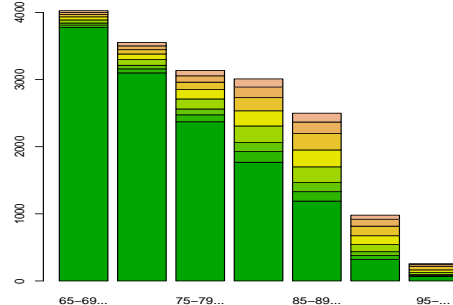


図 29 2040 年年齢階級別介護状態別人口分布 (男性)

## 7 結論

第 4 章の冒頭でも述べたように、多状態生命表作成のためには確率推移行列の要素のロバストな推定手法を確立しなければならない。今回は、ややアドホックな方法をとっているが、本来的には、H 市のデータにもとづき、全国の人口データを較正する体系的な方法論を確立しなければならない。まだ十分な分析ができていないが、ある介護状態での死亡率とその状態からの死亡以外の状態への推移確率には何らかの関係があるように見えるため、その関係を推定に利用することにより推定精度が改善する可能性がある。また、推定方法もベイズ手法や機械学習のような手法も有効かもしれない。また、他の自治体のデータが入手できれば、地域差などの情報を取り入れることにより状態推移に関する理解が深まるかもしれない。

多状態生命表には様々な利用分野がある。まず、多状態生命表によって新たに健康寿命(余命)を定義できる。ある年齢から死亡するまで非該当の状態にいる平均余命を健康余命と定義すると、悪くても非該当から要支援 2 までの状態にいる平均余命を準健康余命と定義して、健康状態の段階を反映した健康寿命概念に拡張できる。あるいは要介護 3 から 5 まで継続して滞在する期間を重度介護期間と定義し、これをできるだけ短くするという施策を考えることもできる。

介護保険費用の将来シミュレーションも容易にできるようになる。介護状態に応じて毎月の費用がほぼ決まっているので、介護状態人口ベクトルが推定できれば、介護費用も推定できる。介護保険の保険料も性・年

齢・介護状態の区分で計算可能なので、公的介護保険だけでなく民間の介護保険商品の設計にも利用できるであろう。

将来人口推計は全国ベースで公表されるほか、都道府県別、市町村別にも公表されている。今後の課題としては、介護保険事業は、地域レベルでの活動が重要であるので、ミクロの介護状態別人口推計は大きな利用価値があると考えられる。

## 参考文献

- [1] 「公的介護保険制度の現状と今後の役割」, 厚生労働省老健局総務課, 2015年 [http://www.mhlw.go.jp/stf/seisakunitsuite/bunya/hukushi\\_kaigo/kaigo\\_koureisha/gaiyo/index.html](http://www.mhlw.go.jp/stf/seisakunitsuite/bunya/hukushi_kaigo/kaigo_koureisha/gaiyo/index.html)
- [2] 「介護保険事業状況報告（年報）」, 厚生労働省老健局介護保険計画課, 平成12年度～平成27年度 <http://www.mhlw.go.jp/toukei/list/84-1.html>
- [3] Willekens, Frans, "Multistate Analysis of Life Histories with R", Springer, 2014
- [4] Shuji Tanaka, "A Proposal for Redesigning Social Security: Long-Term Care Pension", Geneva Papers on Risk and Insurance, Issues and Practice, SPECIAL ISSUE ON LONGEVITY Volume 41, Issue 1, pp.98-117, 2015.11