

科学研究費助成事業 研究成果報告書

令和 3 年 6 月 1 日現在

機関番号：15401

研究種目：若手研究

研究期間：2018～2020

課題番号：18K17381

研究課題名（和文）マルコフモデルによる病態推移予測の精度評価と費用効果分析への応用

研究課題名（英文）Accuracy evaluation of predicted disease state transition by Markov model and its application to cost-effectiveness analysis

研究代表者

秋田 智之（Akita, Tomoyuki）

広島大学・医系科学研究科（医）・講師

研究者番号：80609925

交付決定額（研究期間全体）：（直接経費） 3,100,000円

研究成果の概要（和文）：疾病の進行や費用効果分析においてマルコフモデルが用いられている。このモデルは実際の疫学・臨床研究のデータから算出された「推移確率」をもとに、将来の病態進行を予測することが可能であるが、これまで推移確率の元データ数が予測精度に与える影響を検討した研究ほとんどない。本研究では、マルコフモデルによる予測精度を信頼区間で評価する公式を開発し、数値シミュレーションにより、データ数と信頼区間の精度（被覆確率）について検討した。さらに、本公式をもとに具体的なマルコフモデルを用いた費用効果分析における、費用と効果の推定精度について検討した。

研究成果の学術的意義や社会的意義

検診・治療等の疾病対策導入を検討するとき、検診を導入した/導入しなかった場合の、生涯にかかる費用とQOLをそれぞれ見積り、導入に要した費用に似合うだけのQOL改善が見込まれるのかが評価されている。方法の一つであるマルコフモデルは、実際の疫学・臨床研究のデータから、1年間の疾患の発症率や進行率を出して、それをもとに仮想的に病態進行をシミュレーションを行う。元のデータの対象者数が少ない場合、予測の精度がよくないと考えられるが、これまでの研究では、ほとんど考慮されていない。そこで本研究では、予測の精度を「信頼区間」として表現するための公式を開発し、その方法の妥当性を理論と実用の両面から検討した。

研究成果の概要（英文）：Markov model is used to estimate the disease progression and cost-effectiveness analysis. This model predicts future pathological progression based on the "transition probability" calculated by data from epidemiological/clinical research data, but so far, few studies examined that the number of data from the original research has affected the prediction accuracy.

In this study, we have developed a formula to evaluate the prediction accuracy by Markov model in the "confidence interval". Next, a numerical simulation was performed to examine the number of data and the accuracy of the confidence interval (covering probability). It was also compared with the existing formula confidence intervals in a special case of the Markov model (two states irreversible). Furthermore, based on this formula, the estimation accuracy of cost and effect in concrete cost-effectiveness analysis was examined.

研究分野：数理疫学

キーワード：マルコフモデル 推定精度 費用効果分析 漸近論

1. 研究開始当初の背景

検診・治療導入等の疾病対策の導入を検討する際に、費用対効果を見積もることは重要である。例えば、肝癌の主要原因であるC型肝炎ウイルスは、経口薬(DAA)によりウイルス排除が可能となっているが、薬価が高価であり、導入に要した費用に似合う効果(余命延長、QOL改善)が見込まれるのかの評価が行われている。費用効果分析のためには病態推移や治療効果に関する臨床疫学情報と薬価等医療費に関する情報が必要である。

マルコフモデルは現在の状態と確率推移行列から将来の状態分布が決まる応用範囲の広いモデルであり、病態推移のモデルとして医療経済的評価に多用されている。しかし、これまでマルコフモデルの基となる「確率推移行列」の精度、すなわち推移確率算出のために用いたデータ数(人年)についてはあまり検討されていない。マルコフモデルによる推定患者数や推定死亡者数の精度は、確率推移行列の精度に大きく依存し、元データ数(人年)が少ないと費用効果分析の精度は悪くなると考えられる。

2. 研究の目的

本研究ではマルコフモデルにより推定した状態分布の将来予測の精度を評価するために、予測状態分布の95%信頼区間の構築法について検討した。

3. 研究の方法

1) マルコフモデルの信頼区間の構築

- Tanaka (2003)は4肝病態を設定し、臨床疫学データから患者の病態の推移数を集計し、確率推移行列を算出し、30年間の肝病態の自然史をシミュレートした(図1)。
- 表1において病態間の推移数(データ数、人年)も小さい場合は確率推移行列の精度もよくないため、推定精度が悪くなると考えられる。
- 本研究では、T年後の病態の状態分布を推移確率行列の関数で表現することにより、図2のエラーバーに示すような「信頼区間」の形で、推定精度を表すための方法論について、数理統計学的面から検討した。

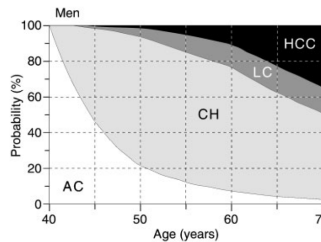


図1. C型肝炎ウイルスキャリアの自然史

表1. 肝病態間の推移数

前年\次年	AC	CH	LC	HCC
AC	35	5	0	0
CH	0	278	3	1
LC	0	0	15	0
HCC	0	0	0	1

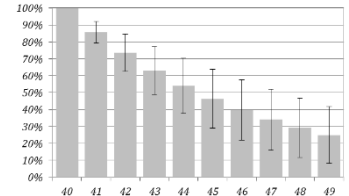


図2. 状態分布の推移と信頼区間(AC)

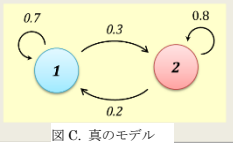
2) 「信頼区間」の公式の精度評価

① シミュレーションによる評価

- 上記1) で検討した予測値の信頼区間の公式について、推移数(人年)がどの程度であれば、「95%信頼区間」の被覆確率が95%に近いのかを確認するために、図3,4に示す数値シミュレーションを行った。

●数値実験の設定2 & 方法2

- 真のモデル 設定1と同様
 - シミュレーション回数 10000
 - 解析対象期間 7=5, 10, 15, 20, 25
- 方法1と同様に被覆確率を算出した。



図C. 真のモデル

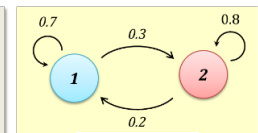
図4. 数値実験2の方法

●数値実験の設定1

- 真のモデル 図1
- 対象期間 10期
- 10期後の状態1 40.06% (図2)
- 行和 $v_1=v_2=15, 20, 30, 50, 100, 500, 1000$
- シミュレーション回数 10000回

●数値実験の方法1

- 二項乱数 $x_{12} \sim B(y_1, 0.3)$, $x_{21} \sim B(y_2, 0.2)$ を発生させて、表1を作成する。
- 表A. から確率推移行列を推定し、10期後の状態1の信頼区間を算出する。
- ii) i), ii) を10000回繰り返し、信頼区間に真の値40.06%を含む被覆確率を算出する。



図A. 真のモデル



図B. 真のモデルによる10期間の状態分布
表A. シミュレーションデータ

	状態1	状態2	行和
状態1	y_1-x_{12}	x_{12}	y_1
状態2	x_{21}	y_2-x_{21}	y_2

図3. 数値実験1の方法

② カプランマイヤー法との比較

- KM法は状態数2の不可逆確率過程の推定方法ともみなせ、マルコフ性を仮定せずに推定を行うため、マルコフモデルとKMの結果を比較することによりマルコフ性を仮定しない場合の信頼区間との比較が可能である。透析患者の長期コホートデータ (*Journal of Medical Virology*: 92, 3436-3447) を基に図5に示すようにN年生存率に指数曲線 $e^{-\lambda t}$ を近似させると

$$(S(t+1) \quad 1-S(t+1)) = (S(t) \quad 1-S(t)) \begin{pmatrix} 1-\lambda & \lambda \\ 0 & 1 \end{pmatrix}$$

となる。t期後の状態分布はマルコフモデルの確率推移行列を用いて

$$(S(t) \quad 1-S(t)) = (S(0) \quad 1-S(0)) \begin{pmatrix} 1-\lambda & \lambda \\ 1 & 0 \end{pmatrix}^t$$

と表せるため、 λ の95%信頼区間を元に状態分布の信頼区間推定し、KM法のGreenwoodの公式による95%信頼区間と比較し、マルコフ性

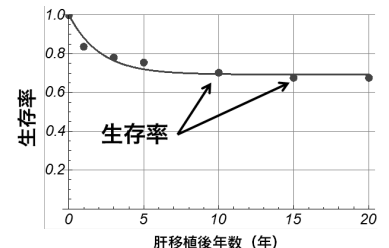


図5. 生存率の指数曲線による近似

の検討を行った。

3) 費用効果分析における検証

- マルコフモデルを用いた費用効果分析を行う際に、病態推移確率の元データである「病態間の推移数 (人年)」が、費用と効果の推定精度に与える影響について、以下の例をもとに試算した。
- 試算には、うつ病を発症した糖尿病 かつ/または 冠動脈性心疾患 (CHD) 患者に対する治療の費用対効果を検討した研究 (*BMJ Open*, 2016;6:e012514) のデータを参考に以下のモデルを設定した：病態推移モデルは3状態 (A非うつ、Bうつ、C死亡) のマルコフモデル、推移確率はA→B 37%、A→C 0%、B→A 27%、B→C 2%、C→C 100%、費用はA £1516、B £1730、HRQoL scoreはA 0.196、B 0.158である (図6)。
- 割引率を2%とし、全員がAの状態からスタートし、追跡開始から5年間の病態推移およびその間の費用とQALYについて、マルコフモデルによりシミュレーションした

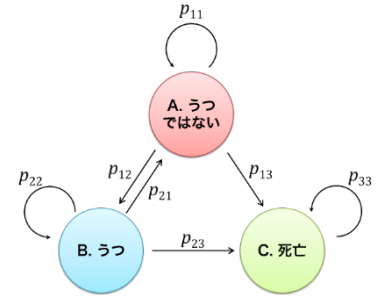


図6. うつ病の推移モデル

4. 研究成果

1) マルコフモデルの信頼区間の構築

有限状態マルコフ連鎖モデルの定義：離散型確率過程 $\{X_n\}$ の取り得る値 (状態) を $\{1, 2, \dots, S\}$ とする。このとき、任意の n 、任意の $x_0, x_1, \dots, x_n, x_{n+1}$ に対して $P(X_{n+1}=x_{n+1}|X_0=x_0, X_1=x_1, \dots, X_n=x_n)=P(X_{n+1}=x_{n+1}|x_n=x_n)$ が成立するとき、この確率過程を有限状態マルコフ連鎖モデル、 $p_{jk}=P(X_{n+1}=k|x_n=j)$ を (j, k) 成分とした行列 P を確率推移行列と定義する。

確率推移行列の最尤推定量および n 期後の状態分布の推定値：議論の簡略化のため、データが一つのマルコフ連鎖の履歴 $x_0x_1\dots x_n$ から作られているとする。このとき、状態 i から状態 j への推移の回数を y_{ij} とおくと履歴が $x_0x_1\dots x_n$ となる確率は $P(X_0=x_0, X_1=x_1, \dots, X_n=x_n)=\prod_i \prod_j p_{ij}^{y_{ij}}$ で与えられる。これを制約条件 $p_{ij} \geq 0, \sum_j p_{ij}=1$ の下で最大化することによる最尤推定量 $\hat{p}_{ij}=y_{ij}/y_i$ を得る。 \hat{p}_{ij} を (i, j) 成分とした推定確率推移行列 \hat{P} により、初期分布を π_0 としたとき n 期後の状態確率分布が $\pi_n=\pi_0 \hat{P}^n$ で推定され、特に状態 i にいる確率の推定値はあるベクトル値行列関数と \hat{p}_0 の内積 $\hat{\pi}_{ni}=\pi_{ni}(\hat{P})=f_{ni}(\hat{P})\pi_0$ で表現することができる。

マルコフモデルによる推定値の信頼区間：状態 i から各状態への推移回数 $y_i=(y_{i1}, y_{i2}, \dots, y_{iS})'$ は S 項分布に従うので最尤推定量の漸近正規性から S 変量正規分布 $N_S(y_i \pi_i, H)$, $y_i=1_S' y_i$ に近似的に従う。ここで行列 $H=(h_{ij})$ は $h_{jk}=-n_{ij}n_{ik}/n_i$ ($i \neq j$), $h_{jj}=n_{ij}(n_i-n_{ij})/n_i$ で定義される。従って、 $\text{Var}(\hat{p}_{ij}) \sim -n_{ij}n_{ik}/n_i^3$ ($i \neq j$), $\text{Var}(\hat{p}_{ii}) \sim n_{ij}(n_i-n_{ij})/n_i^3$ で近似できるので、 π_{ni} の信頼区間は $\hat{\pi}_{ni} \pm 1.96 \text{SE}(\hat{\pi}_{ni})$ で得られる。ここに標準誤差は $\text{SE}(\hat{\pi}_{ni})=\text{grad}(\pi_{ni}(\hat{P}))\Sigma(\hat{P})\text{grad}(\pi_{ni}(\hat{P}))$, $\Sigma(\hat{P})$ は $\text{Var}(\hat{p}_{ij})$ を (i, j) 成分に持つ行列である。

状態数が 2 の場合に状態分布の信頼区間を出力する R のプログラム例について図 7 に、その実行結果を図 8 にそれぞれ示した。

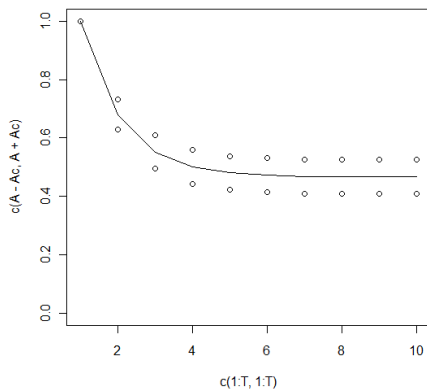


図8. 図7のRのプログラムの実行結果

```

●Rのプログラム例
# 対象期間
T <- 10

# データの入力
X <-
matrix(c(68,32,28,72),nc=2,byrow=T)

# 推移確率の算出
n1 <- X[1]+X[3]; p <- X[3]/n1
n2 <- X[2]+X[4]; q <- X[4]/n2

# 確率推移行列とその微分
P <- matrix(c(1-p,p,q,1-q),nc=2,byrow=T)
Pp <- matrix(c(-1,1,0,0),nc=2,byrow=T)
Pq <- matrix(c(0,0,1,-1),nc=2,byrow=T)

# 初期状態
S <- c(1,0)

# 推定値を代入するベクトル
A <- (NULL); Ac <- (NULL);
Ap <- (NULL); Aq <- (NULL)
Rp <- matrix(0,nc=2,nr=2)
Rq <- matrix(0,nc=2,nr=2)

# 行列のべき
pow <- function(A,k){
  W <- diag(2)
  if (k==0) {return(W)}
  else { for(i in 1:k) W <- W %*% A
  return(W)}
}

# 各状態の点推定値・信頼区間を算出
for(i in 1:T){
  A <- c(A,S[1]); # 状態iの値を追加
  S <- S %*% P # 1期後の状態分布を算出
  Ap <- c(A0,B0) %*% Rp
  Aq <- c(A0,B0) %*% Rq
  Ac <-
  c(Ac, 1.96*sqrt(Ap[1]^2*p*(1-p)/n1+Aq[1]^2*q*(1-q)/n2))
  # 分散を計算し、信頼区間の幅を追加
  Rp <- P %*% Rp + Pp %*% pow(P,i-1)
  Rq <- P %*% Rq + Pq %*% pow(P,i-1)
  # この2行に先の微分の公式を利用
}

# 状態Aの推移と信頼区間のプロット
plot(c(1:T,1:T),c(A-Ac,A+Ac),ylim=c(0,1))
lines(A)

```

図7.Rのプログラム

2) 「信頼区間」の公式の精度評価

① シミュレーションによる評価

2状態のマルコフモデルにおいて、10期先の予測値の95%信頼区間の被覆確率のシミュレーションを行ったところ、行和が30程度以上で被覆確率がほぼ95%に近い値となった(表2)。また、推定期間を25期先まで伸ばした推定値の95%信頼区間においても、行和が50程度以上で被覆確率がほぼ95%に近い値となった(表3)。

表2. 数値実験1の結果

$p \setminus y_1$	15	20	30	50	100	500	1000
0.1	0.775	0.845	0.897	0.909	0.926	0.946	0.943
0.3	0.888	0.921	0.932	0.944	0.942	0.953	0.949
0.5	0.897	0.909	0.925	0.936	0.941	0.954	0.953
0.7	0.902	0.905	0.922	0.932	0.942	0.948	0.949
0.9	0.911	0.903	0.928	0.933	0.943	0.947	0.951

表3. 数値実験2の結果

$T \setminus y_1$	15	20	30	50	100	500	1000
5	0.933	0.926	0.926	0.945	0.941	0.896	0.827
10	0.888	0.921	0.932	0.944	0.942	0.953	0.949
15	0.884	0.912	0.930	0.937	0.947	0.951	0.945
20	0.859	0.929	0.914	0.950	0.944	0.951	0.950
25	0.881	0.909	0.925	0.937	0.942	0.952	0.949

② カプランマイヤー法との比較

Kaplanマイヤー法による累積死亡率(1-生存率)とマルコフモデルによる累積死亡率はほぼ一致していた。一方、95%信頼区間については、本提案手法による信頼区間は、グリーンウッドの公式による信頼区間よりも非常に狭くなった。これは1)マルコフ過程が非可逆であること、2)マルコフ性の仮定が強いことから、分散が小さくなることが理由として考えられた(図9)。

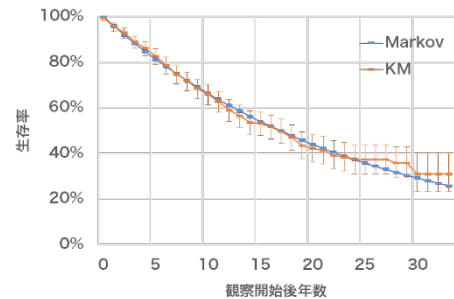


図9. KM法とMarkovモデルの信頼区間の比較

3) 費用効果分析における検証

マルコフモデルに基づく5年間の病態推移を基に算出すると、一人当たりの延べ費用は£9,687、QALYは1.028となる。しかし、推移確率算出のもととなる各状態からの推移数を100とした場合、一人当たりの延べ費用の95%信頼区間は£8,107~11,269、同QALYの95%CIは0.856~1.200となり、相対精度が±20%程度となっていた。実際の臨床データでは、各状態からの推移数が100を超えないケースもあると考えられるため、相対精度が±20%を超えるケースがあると考えられる。

4) 結語

本研究では、推移確率算出の基のデータをもとに根拠に基づき費用効果分析の結果を信頼区間という形で評価することが可能であり、医学分野で有用な方法であると考えられた。

5. 主な発表論文等

〔雑誌論文〕 計2件（うち査読付論文 2件 / うち国際共著 0件 / うちオープンアクセス 0件）

1. 著者名 Tada T, Kumada T, Toyoda H, Ohisa M, Akita T, Tanaka J	4. 巻 53
2. 論文標題 Long-term natural history of liver disease in patients with chronic hepatitis B virus infection: an analysis using the Markov chain model	5. 発行年 2018年
3. 雑誌名 Journal of Gastroenterology	6. 最初と最後の頁 1196-205
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

1. 著者名 Yamasaki K, Tanaka J, Kurisu A, Akita T, Ohisa M, Sakamune K, Ko K, Sugiyama A, Yasaka T, Shirahama S	4. 巻 90
2. 論文標題 Natural course of persistent hepatitis B virus infection in HBe antigen-positive and -negative cohorts in Japan based on the Markov model	5. 発行年 2018年
3. 雑誌名 Journal of Medical Virology	6. 最初と最後の頁 1800-13
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計1件（うち招待講演 0件 / うち国際学会 0件）

1. 発表者名 秋田智之、杉山文、今田寛人、栗栖あけみ、田中純子
2. 発表標題 MarkovモデルとAge-Cohortモデルによる献血の需要と供給の将来推計の試み
3. 学会等名 第31回 日本疫学会学術総会
4. 発表年 2021年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

6. 研究組織

氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8 . 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------