

喫煙状態と人口動態を考慮した肺がんの数値モデル

滋賀大学・データサイエンス教育研究センター 山口 崇幸

Takayuki Yamaguchi

The Center for Data Science Education and Research

Shiga University

1 はじめに

喫煙状態を考慮した肺がんの罹患数や死亡数に関する数値モデルを検討したい。喫煙は肺がんの危険因子の一つであることがよく知られており、喫煙者、非喫煙者、元喫煙者であるかによってリスクが異なる [1, 2, 3, 4, 5, 6]。したがって、数値モデルは喫煙状態の違いを考慮しなければならない。また、がんは高齢者のリスクが高く、罹患数や死亡数は人口動態に強く依存する。モデルを実際のデータにフィットさせるためには、時刻だけでなく年齢を考慮し、人口動態を扱う必要がある。

いくつかの状態がありその状態間の遷移を記述するためにコンパートメントモデルが用いられる。たとえば、ある感染症について、感受性者が感染者との接触により感染し、その

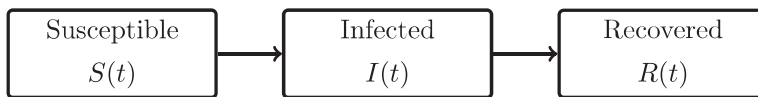


図1 SIRモデルのコンパートメント

後、回復するという図 1 のダイナミクスは次の SIR モデルで表される [7, 8, 9].

$$\begin{aligned}\frac{dS}{dt}(t) &= -\beta S(t)I(t), \\ \frac{dI}{dt}(t) &= \beta S(t)I(t) - \gamma I(t), \\ \frac{dR}{dt}(t) &= \gamma I(t).\end{aligned}$$

β は感染者との接触の頻度に関するパラメータ (伝達係数), γ はコンパートメント間の遷移の速度を表すパラメータ (回復率または隔離率) である. また, βI は感染力と呼ばれ, 単位時間あたり単位時間あたりの感染率である. このようなモデルは実際の感染症のアウトブレイクに対して用いられており, そのときには実際のデータからパラメータ β, γ を推定する問題になる.

時刻と年齢を考慮し, 人口動態を表す偏微分方程式としては, 次のマッケンドリック方程式が知られている [10, 8, 9].

$$\frac{\partial p}{\partial a}(t, a) + \frac{\partial p}{\partial t}(t, a) = -\mu(t, a)p(t, a).$$

ここで $p(t, a)$ は時刻 t のときに a 歳である人口である. $\mu(t, a)$ は死力と呼ばれ, 時刻 t , 年齢 a での (瞬間の) 死亡率である. この偏微分方程式は 1 階準線形なので

$$\tilde{p}(s) = p(t + s, a + s)$$

とおくと

$$\begin{aligned}\frac{d\tilde{p}}{ds}(s) &= \frac{\partial p}{\partial a}(t + s, a + s) + \frac{\partial p}{\partial t}(t + s, a + s) \\ \frac{d\tilde{p}}{ds}(s) &= -\mu(t + s, a + s)p(t + s, a + s) \\ \frac{d\tilde{p}}{ds}(s) &= -\mu(t + s, a + s)\tilde{p}(s)\end{aligned}\tag{1}$$

という常微分方程式となる. (1) は時刻 t で年齢 a であるコホートが死亡で減少していく過程を表現する. 元の偏微分方程式の解を数値的に得るためには, 境界条件と初期条件に合わせて t と a を変えて (1) を何度も解くことになる.

本稿では, 人口動態を扱うために時刻と年齢にマッケンドリック方程式を基にし, 喫煙状態によるコンパートメントを考え, 肺がんの罹患数や死亡数に関する数理モデルを検討する.

2 肺がんの数理モデル

モデルを構築するために、図 2 のコンパートメントを考える。喫煙状態は喫煙者 (smokers)、非喫煙者 (never-smokers)、元喫煙者 (ex-smokers) の 3 つとし、肺がんの罹患数と死亡数を扱うために肺がん罹患者も加える。時刻を t 、年齢を a とする。 $M(t, a)$, $S(t, a)$, $E(t, a)$ はそれぞれ喫煙者、非喫煙者、元喫煙者の数であり、 $L(t, a)$ は肺がん罹患者の数である。

出生した人口は非喫煙者になり、時間発展とともに非喫煙者の一部が喫煙者となり、元喫煙者となっていく。喫煙率のデータとしては、JT 全国喫煙者率調査 [11] があるが、喫煙者とそれ以外 (非喫煙者と元喫煙者) の割合しかわからないため、喫煙率のデータに適合させることを目的とすると、喫煙の過程については簡略化せざるを得ない。非喫煙者と元喫煙者の割合がわからないため、禁煙すると再喫煙しないという仮定した。非喫煙者から喫煙者、喫煙者から元喫煙者への遷移率は、それぞれ $\sigma(t, a)$, $\delta(t, a)$ とする。

非喫煙者、喫煙者、元喫煙者は肺がんのリスクが異なるとする。 $\lambda(t, a)$ は非喫煙者から肺がん人口への遷移率であり、喫煙者と元喫煙者に対しては、それぞれ相対リスク k_S , k_E を $\lambda(t, a)$ に係数として掛ける。死亡のリスクも、喫煙者と元喫煙者は非喫煙者よりも高くなる。 $\mu(t, a)$ は非喫煙者の死力であるが、喫煙者と元喫煙者の死力としては $\mu(t, a)$ に相対リスクを表す係数 q_S , q_E を掛けた死力を使う。また、肺がん罹患者の死力は、 $\mu(t, a)$ と肺がんによる死力 $\nu(t, a)$ の和とする。

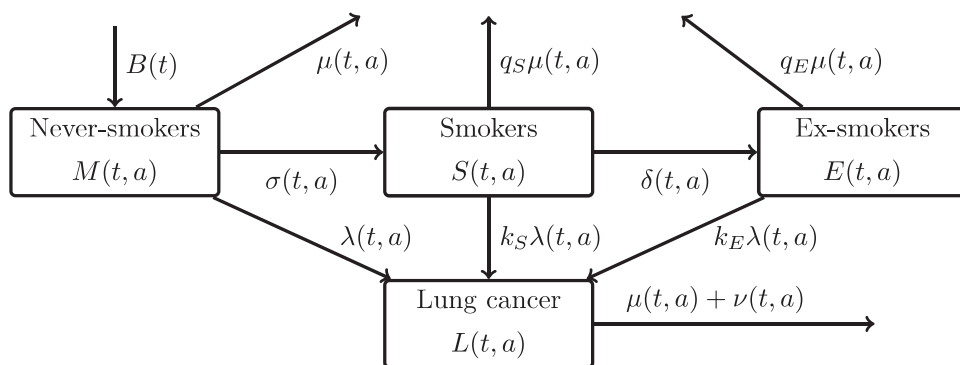


図 2 肺がんのモデルのコンパートメント

したがって、図 2 のコンパートメントは次の偏微分方程式で表される.

$$\frac{\partial M}{\partial t}(t, a) + \frac{\partial M}{\partial a}(t, a) = -(\mu(t, a) + \sigma(t, a) + \lambda(t, a))M(t, a), \quad (2)$$

$$\frac{\partial S}{\partial t}(t, a) + \frac{\partial S}{\partial a}(t, a) = \sigma(t, a)M(t, a) - (q_S\mu(t, a) + \delta(t, a) + k_S\lambda(t, a))S(t, a), \quad (3)$$

$$\frac{\partial E}{\partial t}(t, a) + \frac{\partial E}{\partial a}(t, a) = \delta(t, a)S(t, a) - (q_E\mu(t, a) + k_E\lambda(t, a))E(t, a), \quad (4)$$

$$\begin{aligned} \frac{\partial L}{\partial t}(t, a) + \frac{\partial L}{\partial a}(t, a) = & \lambda(t, a)(M(t, a) + k_S S(t, a) + k_E E(t, a)) \\ & - (\mu(t, a) + \nu(t, a))E(t, a). \end{aligned} \quad (5)$$

初期条件と境界条件は

$$\begin{aligned} M(t_0, a) &= M_0(a), \quad S(t_0, a) = S_0(a), \quad E(t_0, a) = E_0(a), \quad L(t_0, a) = L_0(a), \\ M(t, a_0) &= B(t), \quad S(t, a_0) = 0, \quad E(t, a_0) = 0, \quad L(t, a_0) = 0 \end{aligned} \quad (6)$$

である. t_0 はある初期時刻, a_0 は考慮している年齢の最小値, $B(t)$ は時刻 t で a_0 歳になる人数である. $M_0(a)$, $S_0(a)$, $E_0(a)$, $L_0(a)$ は対応するコンパートメントの年齢に関する分布である.

このモデルは非常にシンプルなので, 他の要素を考慮したい場合はなんらかの拡張する必要がある. たとえば, 喫煙期間を考慮する場合を考える. 喫煙の総量が肺がんのリスクを高めることが知られており [1, 2, 3, 5], 喫煙開始年齢が早いほど喫煙のリスクが高いことも知られている [4]. 喫煙習慣が変わらなく, 毎日の喫煙量が一定だとすれば, 喫煙期間から喫煙の総量に変換できる. 喫煙者の喫煙開始年齢は喫煙期間と年齢から簡単に計算できる.

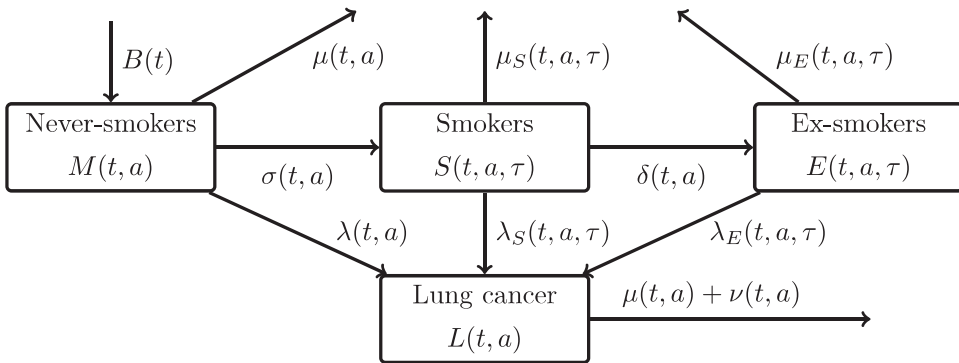


図 3 喫煙期間を考えたモデルのコンパートメント

喫煙期間を τ とする. 図 3 のモデルは次の偏微分方程式で表される.

$$\begin{aligned}
 \frac{\partial M}{\partial t}(t, a) + \frac{\partial M}{\partial a}(t, a) &= -(\mu(t, a) + \sigma(t, a) + \lambda(t, a))M(t, a), \\
 \frac{\partial S}{\partial t}(t, a, \tau) + \frac{\partial S}{\partial a}(t, a, \tau) + \frac{\partial S}{\partial \tau}(t, a, \tau) &= -(\mu_S(t, a, \tau) + \delta(t, a) + \lambda_S(t, a, \tau))S(t, a, \tau), \\
 \frac{\partial E}{\partial t}(t, a, \tau) + \frac{\partial E}{\partial a}(t, a, \tau) &= \delta(t, a)S(t, a, \tau) - (\mu_E(t, a, \tau) + \lambda_E(t, a, \tau))E(t, a, \tau), \\
 \frac{\partial L}{\partial t}(t, a) + \frac{\partial L}{\partial a}(t, a) &= \int_0^a (\lambda_S(t, a, \tau)S(t, a, \tau) + \lambda_E(t, a, \tau)E(t, a, \tau)) d\tau \\
 &\quad + \lambda(t, a)M(t, a) - (\mu(t, a) + \nu(t, a))E(t, a).
 \end{aligned}$$

初期条件と境界条件は

$$\begin{aligned}
 S(t, a, 0) &= \sigma(t, a)M(t, a) \\
 M(t_0, a) &= M_0(a), \quad S(t_0, a, \tau) = S_0(a, \tau), \quad E(t_0, a, \tau) = E_0(a, \tau), \quad L(t_0, a) = L_0(a), \\
 M(t, a_0) &= B(t), \quad S(t, a_0, \tau) = 0, \quad E(t, a_0, \tau) = 0, \quad L(t, a_0) = 0
 \end{aligned}$$

である. τ に関係する部分で方程式が変更されている. S と E については, 喫煙期間である τ にも依存する. 死亡や肺がんのリスクは $\mu_S(t, a, \tau)$, $\mu_E(t, a, \tau)$, $\lambda_S(t, a, \tau)$, $\lambda_E(t, a, \tau)$ を用いて τ への依存を表現する. $E(t, a, \tau)$ の τ は喫煙していた期間を表しているが, この定式では禁煙してからの期間がわからなくなるので, 元喫煙者の喫煙開始年齢は考えられていないことに注意が必要である.

3 モデルを用いた肺がん罹患数・死亡数の予測

最後に, モデルの活用例として, (2)-(6) を実データに適合させ, 肺がん罹患数と死亡数の予測を行った [12] の概要を述べる. 日本では, 男性, 女性ともに喫煙率は減少してきた. 男性の喫煙率は 1965 年では 82.3% であるが 2016 年では 29.7% であり, 女性の喫煙率はそれぞれ, 15.7%, 9.7% である. このような喫煙率の減少にも関わらず, 男性の肺がん罹患数と肺がん死亡数はそれぞれ 1975 年では 12,701, 10,711, 2014 年では 77,617, 52,505 であり, 増加している. 女性も同様の増加傾向であり, それぞれ 1975 年では 4,971, 4,048, 2014 年では 36,933, 20,891 である. 肺がんの増加は, 高齢化によるものと考えられ, 危険因子である喫煙の減少と高齢化の両方が関係している肺がんのダイナミクスを定量化するためには, 前セクションで述べたような数理モデルを用いる必要がある.

男性と女性で別々に (2)-(6) のモデルを考え, $\mu(t, a)$, $\lambda(t, a)$, $\sigma(t, a)$, $\delta(t, a)$ に階段関数を仮定し, 階段関数の値を最尤推定により求めた. 相対リスクについては, [13] の値を

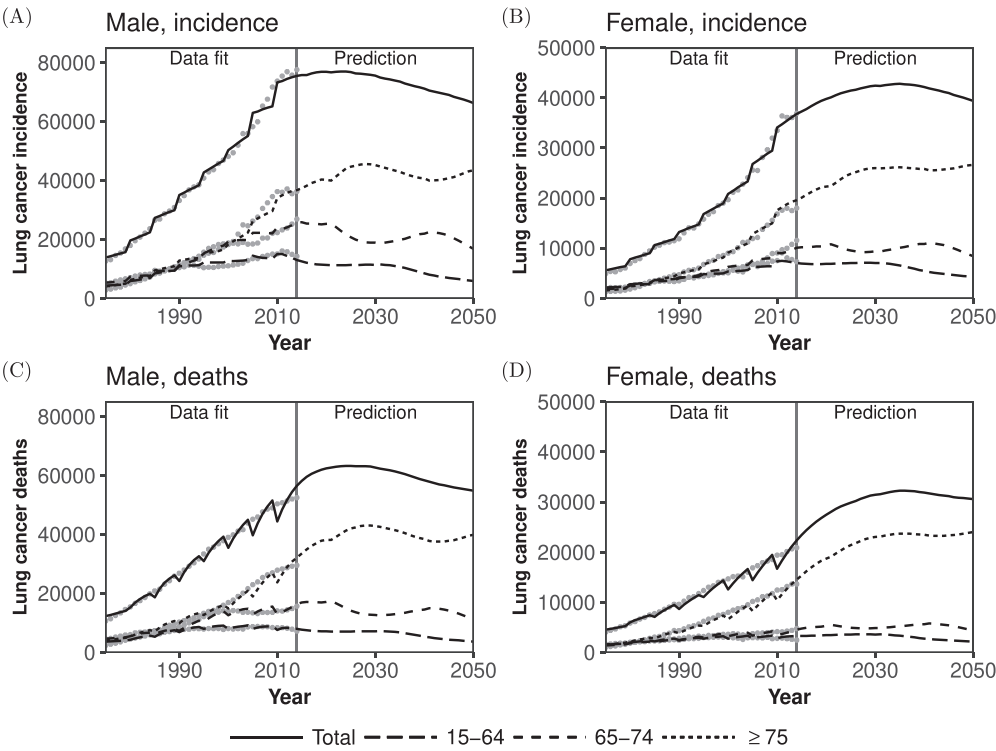


図4 [12] の Figure A2. (A) 男性の肺がん新規罹患数, (B) 女性の肺がん新規罹患数, (C) 男性の肺がん死亡数, (D) 女性の肺がん死亡数. 灰色の点は実データである. 実線は全年齢の予測値, 3つの破線はそれぞれ15歳から64歳, 65歳から74歳, 75歳以上の予測値を表す. 2014年まではモデルをデータに適合させた結果であり, 2015年以降は将来推計人口を使い, 2014年のパラメータにより計算した予測である. 95%信頼区間は非常に狭かったため, グラフには表示していない.

使用し, 男性のモデルでは $k_S = 4.94$, $k_E = 2.20$, $q_S = 0.486185$, $q_E = 0.201306$, 女性のモデルでは $k_S = 4.25$, $k_E = 2.19$, $q_S = 0.509629$, $q_E = 0.455788$ とした. 最尤推定に使用したデータは, JT 全国喫煙者率調査 [11], 人口推計 [14], 人口動態統計時系列データ [15, 16, 17], 国立がん研究センターがん情報サービス「がん登録・統計」[18] である. また, 将来の予測を計算するために日本の将来推計人口 (平成 29 年推計) [19] も使用した.

モデルは 1 階準線形の偏微分方程式なので, 特性曲線法で出生年毎の人口の時間発展を表す常微分方程式に変換し, 4 次の Runge-Kutta 法で解いた. そのときに, 死亡や肺がん罹患の累積のコンパートメントを追加し, データに対応する 1 年間での死亡数や罹患数を

計算し、それを期待値とし、尤度関数を計算した。喫煙者の数は二項分布、人口、死亡数、肺がん罹患数、肺がん死亡数はポアソン分布に従うと仮定して次のような形の尤度関数を用いた。

$$L(\theta) = \left(\prod_i \frac{\lambda_i e^{-\lambda_i}}{k_i!} \right) \left(\prod_j \binom{n_j}{k_j} (p_j)^{k_j} (1 - p_j)^{n_j - k_j} \right).$$

$L(\theta)$ の前半がポアソン分布の部分で後半が二項分布の部分である。 θ は未知のパラメータからなるベクトルであり、 k_i は実データ、 n_i はサンプルサイズである。 λ_i はポアソン分布の期待値、 p_i は二項分布の確率であり、どちらもモデルから計算される。 $L(\theta)$ を最大にする θ を求めるのが最尤推定であるが、具体的な数値アルゴリズムとしては BFGS 法や DFP 法などの準ニュートン法を用いる。

図 4 は [12] の結果である。最尤推定により得たパラメータを使って、モデルを計算したものである。2014 年まではモデルをデータに適合させた結果であり、2015 年以降は将来推計人口を使い、2014 年のパラメータにより計算した予測である。男性の肺がん罹患数と死亡数は 2024 年に最大となり、それぞれ 76,978 (95% 信頼区間: 76,630–77,253), 63,284 (95% 信頼区間: 62,991–63,507) であり、女性の肺がん罹患数は 2035 年に最大となり、42,838 (95% 信頼区間: 42,601–43,095) であり、女性の肺がん死亡数は 2036 年に最大となり、32,267 (95% 信頼区間: 32,063–32,460) であるという予測を得た。

謝辞

本研究は JSPS 科研費 JP19K19429 の助成を受けたものです。

参考文献

- [1] M. Ando, K. Wakai, et al. “Attributable and absolute risk of lung cancer death by smoking status: Findings from the Japan collaborative cohort study”. In: *International Journal of Cancer* 105.2 (2003), pp. 249–254.
- [2] T. Marugame, T. Sobue, et al. “Lung cancer death rates by smoking status: Comparison of the Three-Prefecture Cohort study in Japan to the Cancer Prevention Study II in the USA”. In: *Cancer Science* 96.2 (2005), pp. 120–126.
- [3] K. Wakai, M. Inoue, et al. “Tobacco smoking and lung cancer risk: an evaluation based on a systematic review of epidemiological evidence among the Japanese

- population". In: *Japanese Journal of Clinical Oncology* 36.5 (May 2006), pp. 309–324.
- [4] M. Hara, M. Inoue, et al. "The Association Between Cancer Risk and Age at Onset of Smoking in Japanese". In: *Journal of Epidemiology* 20.2 (Mar. 2010), pp. 128–135.
 - [5] T. Seki, Y. Nishino, et al. "Cigarette smoking and lung cancer risk according to histologic type in Japanese men and women". In: *Cancer Science* 104.11 (2013), pp. 1515–1522.
 - [6] K. Nakamura, S. Ukawa, et al. "Characteristics and prognosis of Japanese male and female lung cancer patients: The BioBank Japan Project". In: *Journal of Epidemiology* 27.3, Supplement (2017), S49–S57.
 - [7] W. O. Kermack and A. G. McKendrick. "A Contribution to the Mathematical Theory of Epidemics". In: *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 115.772 (1927), pp. 700–721.
 - [8] 稲葉 寿. 数理人口学. 東京大学出版会, 2002.
 - [9] 稲葉 寿. 感染症の数理モデル. 培風館, 2008.
 - [10] A. G. M'Kendrick. "Applications of Mathematics to Medical Problems". In: *Proceedings of the Edinburgh Mathematical Society* 44 (1925), pp. 98–130.
 - [11] 日本たばこ産業株式会社. *JT 全国喫煙者率調査*.
 - [12] T. Yamaguchi and H. Nishiura. "Predicting the Epidemiological Dynamics of Lung Cancer in Japan". In: *Journal of Clinical Medicine* 8.3 (2019).
 - [13] K. Ozasa. "Smoking and mortality in the Japan Collaborative Cohort Study for Evaluation of Cancer (JACC)". In: *Asian Pacific Journal of Cancer Prevention* 8 Suppl (2007), pp. 89–96.
 - [14] 総務省統計局. 人口推計.
 - [15] 一般財団法人 厚生統計協会. 人口動態統計時系列データ. DVD. 2014.
 - [16] 一般財団法人 厚生統計協会. 平成 25 年 人口動態統計時系列データ (単年分) . CD. 2014.
 - [17] 一般財団法人 厚生統計協会. 平成 26 年 人口動態統計時系列データ (単年分) . CD. 2015.
 - [18] 国立がん研究センター. がん情報サービス「がん登録・統計」.
 - [19] 国立社会保障・人口問題研究所. 日本の将来推計人口 (平成 29 年推計) . Apr. 2017.