

コピュラを用いた生存時間解析

— 関連のあるエンドポイントとメタ分析の活用 —

江村 剛志¹・道前 洋史²

(受付 2019 年 4 月 24 日；改訂 2020 年 1 月 18 日；採択 1 月 20 日)

要 旨

ここ 10 年間で医学・生物学関連データベースの公開が急速に発展し、個々の患者の詳細かつ正確な情報が誰にでも利用出来るようになった。これらデータベースでは、患者の生存期間、無増悪期間、腫瘍径、遺伝子発現量など患者の極めて詳細な情報が記録されている。加えて、複数のデータソースを統合して解析する「メタ分析」の重要性がここ 10 年間で高まり、複雑かつ膨大な生存期間データを古典的な生存時間解析法だけで解析することは不十分になりつつある。本稿では、コピュラを用いて 2 つの生存期間変数をモデリングする手法について総説する。モデルの実践的な有用性を示すため、より具体的に患者の生存期間と無増悪期間の相関をコピュラでモデル化する手法を考える。また複数のデータソースを統合して、患者の生存期間と術後無増悪期間を同時解析する「メタ分析」法のためのモデルを紹介する。モデルのパラメータを推定する際に、生存期間と無増悪期間が半競合リスク関係にあることを考慮した上で最尤法を用いる必要性を解説する。最後に、個別医療の問題に関連した応用例の一つとして、遺伝子発現量と増悪情報を用いて死亡率の動的予測を行う手法を紹介する。

キーワード：遺伝子発現量，個別医療，競合リスク，動的予測，臨床試験，Cox 比例ハザードモデル。

1. はじめに

生存時間解析 (Survival analysis) は、最も歴史のある学術分野の 1 つでその起源は生命表が考案された 1600 年代にまで遡る。生存時間解析で使用されている統計手法は極めて多岐に及び、いまなお多くの研究者が学術研究を行っていることは生存時間解析のハンドブック (Klein et al., 2014) から理解できる。生存時間解析は現実問題にも幅広く応用されており、医学研究における癌患者の生存率の計算法、工場で生産される部品の信頼性の計算法、政府統計または保険数理における生命表の計算法など多くの統計手法の数理的基礎を成す学術分野といえる。

医学研究における生存時間解析では、年齢や性別などの共変量が生存期間にどのように影響するのかを Cox 回帰法 (Cox, 1972) で調べることが頻繁に行われている。Cox 回帰法は生存時間解析で最も広く使用されてきた統計手法の一つであり、これからも医学研究に必要不可欠な手法であることは疑う余地が無い。

しかしながら、ここ 10 年間で医学・生物学関連の公開データベース (癌ゲノムのデータベー

¹ 長庚大学 情報管理学系：〒333 桃園市龟山区文化一路 259 号 (台湾)；takeshiemura@gmail.com

² 北里大学 薬学部：〒108-8641 東京都港区白金 5-9-1

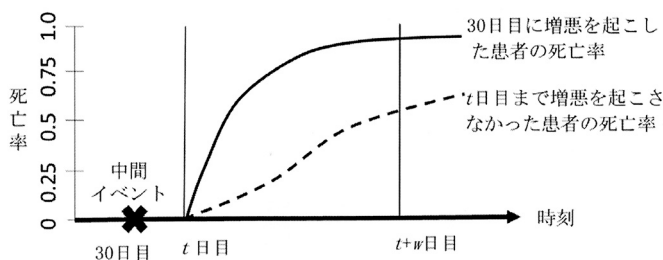


図 1. 術後 t 日目から開始する患者の死亡率の動的予測。

ス GEO や TCGA など) や統計解析フリーソフト (R など) が急速に発展し、個々の患者の詳細かつ正確な情報が誰にでも利用出来るようになった。これらデータベースでは、患者の生存期間、無増悪期間、腫瘍径、遺伝子発現量など患者の極めて詳細な情報が記録されている。加えて、複数のデータソースを統合して解析する「メタ分析」の重要性がここ 10 年間で高まり、複雑かつ膨大な生存期間データを古典的な生存時間解析法だけで解析することは不十分になりつつある。このような患者の個別データを有効利用し、個別医療の進展に貢献できる統計手法を開発することは生物・医学統計学者の重要な使命でもある。

本稿では、コンピュータを用いて 2 つの生存期間変数をモデリングする手法について総説する。モデルの実践的な有用性を示すため、より具体的に患者の生存期間と無増悪期間の相関をコンピュータでモデル化する手法を考える。また複数のデータソースを統合して、患者の生存期間と無増悪期間を同時解析する「メタ分析法」のためのモデルを紹介する。モデルのパラメータを推定する際、生存期間と無増悪期間が半競合リスク関係 (Fine et al., 2001) にあることを考慮した上で最尤法を用いる必要性を解説する。最後に、個別医療の問題に関連した応用例の一つとして、遺伝子発現量と増悪情報を用いて死亡率の動的予測を行う手法を紹介する (Emura et al., 2018)。すなわち、患者個々の術後経過と遺伝子情報を考慮した死亡率の個別予測式をメタ分析のモデルから導出し、患者レベルの死亡予測を行う手法 (図 1) を紹介する。このような動的予測の手法は多く提案されている (Sène et al., 2016; Proust-Lima et al., 2014; Rondeau et al., 2017 など) が、本稿で紹介する手法はメタ分析を使用する点と、コンピュータを利用する点で既出のものとは大きく異なる。

本稿は以下のように構成されている。2 節では生存時間解析における重要な用語「エンドポイント」を解説し、なぜコンピュータを利用するのかを癌の臨床研究の側面から説明する。3 節では、生存時間解析におけるコンピュータのモデルを定式化し、提案されているいくつかのモデルと統計手法を紹介する。4 節では Frailty モデルを用いたメタ分析の考え方を紹介する。5 節では Frailty モデルにコンピュータを組み合わせた統計モデルを紹介し、エンドポイント間の半競合リスク関係を解説する。6 節では個別医療の問題に関連した動的予測の手法を紹介する。5 節、6 節では卵巣癌患者のデータを用いた事例を用いて、統計手法の利用法を解説する。

2. 生存期間のエンドポイント

医学研究における生存時間解析の基本的な目的の 1 つは、特定の治療や予後因子が、患者の生存期間にどれだけ影響するか調べることである。医学研究において、生存期間はエンドポイントと呼ばれ、具体的には次のような複数の定義がある。

- 全生存期間 (OS: Overall survival)

- 無増悪期間 (TTP: Time to tumor progression)
- 無増悪生存期間 (PFS: Progression-free survival)
- 無病生存期間 (DFS: Disease-free survival)

増悪とは癌の再発、腫瘍増大、遠隔転移などを意味する。無増悪期間とは増悪がなかった期間である。例えば、手術後に残った直径 1cm の腫瘍がそのまま増大せず、かつ遠隔転移が見られなければ、無増悪とみなされる。

エンドポイントの定義は上記以外にもあるが、いずれにせよ定義が明確であり、かつ臨床的に重要な転帰を表すものでなければならない (Pazdur, 2008; Piedbois and Crowell, 2008; Le Tourneau et al., 2009)。医学研究者や臨床医は、これらエンドポイントから重要なものを選び、個々の患者に対して追跡・観察する。

治療効果や予後因子がこれらエンドポイントにどのように影響するのかを調べるためには通常 Cox 比例ハザードモデル (Cox, 1972) を用いる。例えば、生存期間を応答変数、治療群と予後因子を共変量とし、部分尤度法により共変量がハザード(瞬間死亡率)に与える影響を調べることが出来る。このような共変量の効果を調べる解析は、医学研究において極めて頻繁に行われている。

全生存期間は癌の研究において最も重要かつ明確に定義できるエンドポイントであり、真のエンドポイント (true endpoint) またはゴールドスタンダードエンドポイント (Pazdur, 2008; Shi and Sargent, 2009; Michiels et al., 2009; Oba et al., 2013) と呼ばれる。例えば、頭頸部癌の化学療法や放射線療法の効果を評価する際の最も重要なエンドポイントは全生存期間である (Michiels et al., 2009; Le Tourneau et al., 2009)。

無増悪期間、無増悪生存期間、無病生存期間などのエンドポイントはしばしば代替エンドポイント (surrogate endpoint) とよばれ、全生存期間に次いで重要なエンドポイントである。代替エンドポイントを使用する場合、“増悪”の判定は臨床医の恣意性が伴うため、RECIST ガイドラン (Eisenhauer et al., 2009) などに従ってすすめることが推奨されている。

多くの臨床研究では、患者から 2 つのエンドポイント (例えば、全生存期間と無増悪生存期間) を観察するためのフォローアップを行う。その後、得られたデータをもとに、Cox 回帰法などで治療効果や予後因子が各エンドポイントに与える影響を解析する。しかしながら、単純な Cox 回帰では 1 つのエンドポイントを応答変数とした周辺解析しか行うことができず、2 つ以上のエンドポイントを同時解析する場合には不十分である。

例えば、ある治療が生存期間と無増悪期間の両方に与える影響に興味があるとする。この場合、生存期間を応答変数とした Cox 回帰と、無増悪期間を応答変数とした周辺 Cox 回帰を別々に実行することが考えられるが、これらの 2 つの解析結果からは、全生存期間と無増悪期間の相関の推定値は得られない。無増悪期間が全生存期間とどのように関連しているかを調べるには相関の解析が必要であり、このような相関の情報は抗癌剤開発の観点からも重要である (Sherrill et al., 2008; Michiels et al., 2009; Rondeau et al., 2015)。加えて、無増悪期間が全生存期間の代替エンドポイントであることを実証したい場合や、無増悪期間を全生存期間の予後予測に利用する場合に、相関の推定値が必要であることを考慮すると、周辺 Cox 回帰法はデータを有効に活用しているとはいえない。

3. コピュラの生存時間モデル

コピュラは 2 つの確率変数間の相関構造をあらわす関数である。コピュラ概念は数学者 Abe Sklar により導入された (Sklar, 1959)。コピュラの数学的な定義と性質は Nelsen (2006) の著書に良くまとめられている。これは数学書であるが、大学初年度の微積分の知識で読める良

書である。以降では、コピュラを生存時間モデルに応用するための基礎知識や用語を概説する。

3.1 2次元生存時間モデル

コピュラは2つの生存期間変数間の相関構造をモデリングするために用いることが出来る。いま X と Y を生存期間変数とし、 $\mathbf{Z}_1 = (Z_{11}, \dots, Z_{1p_1})$ と $\mathbf{Z}_2 = (Z_{21}, \dots, Z_{2p_2})$ を X と Y それぞれに関連する共変量とする。ここで \mathbf{Z}_1 と \mathbf{Z}_2 は $\mathbf{Z}_1 = \mathbf{Z}_2$ であっても良い。また $S_X(x|\mathbf{Z}_1) = \Pr(X > x|\mathbf{Z}_1)$ と $S_Y(y|\mathbf{Z}_2) = \Pr(Y > y|\mathbf{Z}_2)$ を X と Y の生存関数とする。このとき、 X と Y の2次元生存関数を次のように定義する。

$$(3.1) \quad \Pr(X > x, Y > y|\mathbf{Z}_1, \mathbf{Z}_2) = C_\theta\{S_X(x|\mathbf{Z}_1), S_Y(y|\mathbf{Z}_2)\}.$$

ここで C_θ はコピュラ (Sklar, 1959; Nelsen, 2006) で、 θ は X と Y の相関の強さを定めるパラメータである。コピュラの種類は無数にあるが、本稿で主に扱うコピュラは次の3つである。

- 独立型: $C(u, v) = uv$, $0 \leq u, v \leq 1$,
- クレイトン型 (Clayton, 1978):

$$C_\theta(u, v) = (u^{-\theta} + v^{-\theta} - 1)^{-1/\theta}, \quad \theta > 0, \quad 0 \leq u, v \leq 1,$$

- ゲンベル型 (Gumbel, 1960):

$$C_\theta(u, v) = \exp[-\{(-\log u)^{\theta+1} + (-\log v)^{\theta+1}\}^{\frac{1}{\theta+1}}], \quad \theta \geq 0, \quad 0 \leq u, v \leq 1.$$

独立型のコピュラを選ぶと、 $\Pr(X > x, Y > y|\mathbf{Z}_1, \mathbf{Z}_2) = S_X(x|\mathbf{Z}_1)S_Y(y|\mathbf{Z}_2)$ となり、 X と Y は共変量を所与したときの条件付独立なモデルとなる。コピュラの定義により、上記3つの C_θ はどれもコピュラであるから、 C_θ は2次元分布関数であり、その周辺分布は $[0, 1]$ 上の一様分布である。実際、周辺分布は $C_\theta(u, 1) = u$ と $C_\theta(1, v) = v$ であり、これらは $[0, 1]$ 上の一様分布の分布関数である。

C_θ が2次元分布関数であるから、 $\Pr(U \leq u, V \leq v) = C_\theta(u, v)$ となる一様分布の確率変数のペア (U, V) を考えることが出来る。このとき、確率変数のペア (X, Y) を変換 $X = S_X^{-1}(U|\mathbf{Z}_1)$ と $Y = S_Y^{-1}(V|\mathbf{Z}_2)$ で定義すると、 (X, Y) の分布は式 (3.1) を満足することがわかる。これは、確率変数のペア (U, V) が相関構造のみを定め、 $S_X(\cdot|\mathbf{Z}_1)$ と $S_Y(\cdot|\mathbf{Z}_2)$ が周辺構造のみを定めていることを意味する。この周辺構造と相関構造が分離できることは統計モデリングの観点から極めて有用である。

上記の3つのコピュラの例は、どれも下記で定義するアルキメデス型コピュラ (Archimedean copula) に属することがわかる。いま、関数 $\phi_\theta: [0, 1] \mapsto [0, \infty]$ を連続な減少関数で $\phi_\theta(1) = 0$ を満足するとする。このとき、アルキメデス型コピュラは

$$C_\theta(u, v) = \phi_\theta^{-1}\{\phi_\theta(u) + \phi_\theta(v)\},$$

で定義される。ここで ϕ_θ は生成素 (generator) と呼ばれ、その逆関数を ϕ_θ^{-1} で表す。クレイトン型は $\phi_\theta(t) = (t^{-\theta} - 1)/\theta$, $\theta > 0$ である。その極限は $\lim_{\theta \rightarrow 0} \phi_\theta(t) = -\log(t)$ であり、これは独立型の生成素になっていることがわかる。ゲンベル型は $\phi_\theta(t) = \{-\log(t)\}^{\theta+1}$ で生成され、 $\theta = 0$ が独立型の生成素になる。すなわち、クレイトン型とゲンベル型では極限が $\lim_{\theta \rightarrow 0} C_\theta(u, v) = uv$ となっているので、独立型に帰着する。表 1 にコピュラとその生成素をまとめた。

3.2 ケンドール順位相関係数

ケンドール順位相関係数(以下、ケンドール相関と記す)は X と Y の相関を表す指標である。

表 1. コピュラの例.

コピュラの型	パラメータ	生成素 $\phi_\theta(t)$	ケンドール相関 τ_θ	オッズ比 $r_\theta(s) = -3\dot{\phi}_\theta(s)/\phi_\theta(s)$
独立型	なし	$-\log(t)$	0	1
クレイトン型	$\theta > 0$	$(t^{-\theta} - 1)/\theta$	$\theta/(\theta + 2)$	$1 + \theta$
グンベル型	$\theta \geq 0$	$\{-\log(t)\}^{\theta+1}$	$\theta/(\theta + 1)$	$1 - \theta/\log(s)$

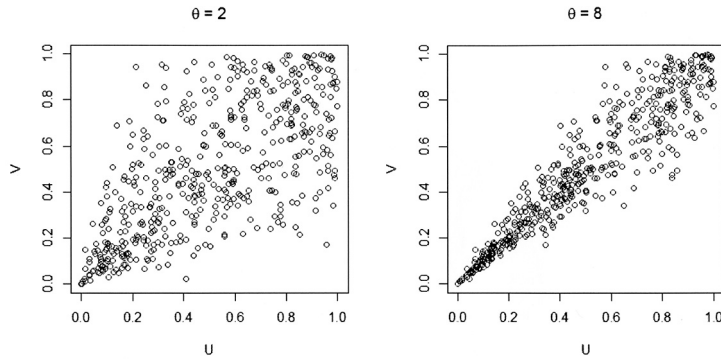


図 2. 500 組の (U, V) をクレイトン型コピュラから生成し、散布図にしたもの.

モデル(3.1)のもとで、 X と Y のケンドール相関はコピュラ C_θ のみに依存することがわかり、

$$\tau_\theta = 4 \int_0^1 \int_0^1 C_\theta(u, v) dC_\theta(u, v) - 1$$

と書ける. ケンドール相関は周辺分布の構造 $S_X(x|Z_1)$ と $S_Y(y|Z_2)$ に依存しない. よってコピュラ C_θ の型を定めれば、 τ_θ はパラメータ θ によって定まる. アルキメデス型コピュラでは、ケンドール相関は生成素を使って次のように計算できる.

$$\tau_\theta = 1 + 4 \int_0^1 \frac{\phi_\theta(t)}{\dot{\phi}_\theta(t)} dt, \quad \dot{\phi}_\theta(t) = d\phi_\theta(t)/dt.$$

この式を用いると、クレイトン型では $\tau_\theta = \theta/(\theta + 2)$ 、グンベル型では $\tau_\theta = \theta/(\theta + 1)$ となり、ともに無相関 $\tau_0 = 0$ から正の完全相関 $\tau_\infty = 1$ までとりうる.

図 2 は 500 組の (U, V) をクレイトン型から生成し、散布図にしたものである. 図から U と V との間に正の相関があることがわかる. 相関の程度は $\theta = 2$ ($\tau_\theta = 0.5$) よりも $\theta = 8$ ($\tau_\theta = 0.8$) が強いことがわかる. また、 $U \approx 0$ と $V \approx 0$ 付近でより強い相関を示し、これは下側裾従属性(lower tail dependence)と呼ばれるクレイトン型コピュラが有する特徴的な現象である. 一方、グンベル型では $U \approx 1$ と $V \approx 1$ 付近でより強い相関を示し、上側裾従属性(upper tail dependence)が現れる.

モデル(3.1)のもとで X と Y を生成するには、 $X = S_X^{-1}(U|Z_1)$ と $Y = S_Y^{-1}(V|Z_2)$ で変換すれば良い. 変換 S_X^{-1} と S_Y^{-1} は減少関数であるため、変換により裾従属の位置は反転する. すなわち、クレイトン型では X と Y 間に上側裾従属、グンベル型では X と Y 間に下側裾従属があらわれる.

3.3 オッズ比による相関

生存時間解析においては、 2×2 クロス集計表と関連した相関の指標が、推定や予測の局面で有用となることがある。例えば、死亡と増悪の関連をみるとき、死亡と増悪のあり・なしを 2×2 クロス集計表に集計し、そのオッズ比を計算して相関の度合いを調べることがある。このオッズ比の考え方は Clayton (1978) が提案した 2 次元生命表に遡り、Oakes (1989) がより明確に定式化した。

いま X を無増悪期間、 Y を全生存期間とし、モデル (3.1) を仮定する。オッズ比関数 (英語では Cross-ratio function) を次のように定義する。

$$\frac{\Pr(X = x, Y = y | \mathbf{Z}) \Pr(X > x, Y > y | \mathbf{Z})}{\Pr(X = x, Y > y | \mathbf{Z}) \Pr(X > x, Y = y | \mathbf{Z})} = R_\theta \{S_X(x | \mathbf{Z}_1), S_Y(y | \mathbf{Z}_2)\}$$

左辺はオッズ比の定義であり、表 2 の 2×2 クロス集計表に基づく。右辺は、モデル (3.1) から導出された結果であり、

$$R_\theta(u, v) = \frac{C_\theta^{[1,1]}(u, v) C_\theta(u, v)}{C_\theta^{[1,0]}(u, v) C_\theta^{[0,1]}(u, v)}$$

はコピュラのオッズ比関数である。ここで

$$C_\theta^{[1,0]}(u, v) = \frac{\partial}{\partial u} C_\theta(u, v), \quad C_\theta^{[0,1]}(u, v) = \frac{\partial}{\partial v} C_\theta(u, v),$$

はコピュラの条件付分布関数である。すなわち $U = u$ の条件下の V の分布関数は $\Pr(V \leq v | U = u) = C_\theta^{[1,0]}(u, v)$ であり、 $V = v$ の条件下の U の分布関数は $\Pr(U \leq u | V = v) = C_\theta^{[0,1]}(u, v)$ である。また、

$$C_\theta^{[1,1]}(u, v) = \frac{\partial^2}{\partial u \partial v} C_\theta(u, v)$$

はコピュラの密度関数を表す。

関数 $R_\theta(u, v)$ は点 (u, v) における局所的な相関を表し、次のように解釈できる。

- $R_\theta(u, v) > 1$; 局所正相関,
- $0 < R_\theta(u, v) < 1$; 局所負相関,
- $R_\theta(u, v) = 1$; 局所独立.

独立型コピュラのもとで $R_\theta(u, v) = 1$ が全ての $0 \leq u \leq 1$ と $0 \leq v \leq 1$ で成立する。クレイトン型コピュラではオッズ比関数が定数 $R_\theta(u, v) = 1 + \theta$ となり、これは原論文 Clayton (1978) に由来する。アルキメデス型コピュラのオッズ比関数は次のように書ける

$$R_\theta(u, v) = r_\theta \{C_\theta(u, v)\}, \quad r_\theta(s) = -s \ddot{\phi}_\theta(s) / \dot{\phi}_\theta(s), \quad \ddot{\phi}_\theta(t) = d^2 \phi_\theta(t) / dt^2.$$

オッズ比関数 R_θ が相対リスク (Relative risk) として解釈できることは生存時間解析において極めて有用である。例えば、 X を無増悪期間、 Y を全生存期間とし、増悪が生存期間に与える影響に興味があるとする。ある時刻 x において増悪したか否かの情報を共変量と考え、次の 2 つの条件付ハザード関数を定義する。

表 2. 2×2 クロス集計表。A, B, C, D はイベントが起こる頻度、オッズ比は AD/(BC)。

	死亡; $Y = y$	生存; $Y > y$
増悪あり; $X = x$	A	B
増悪なし; $X > x$	C	D

- $\lambda_Y(y|X = x, \mathbf{Z}) = \Pr(y \leq Y < y + dy | Y \geq y, X = x, \mathbf{Z})/dy$
時刻 x で増悪したときの条件付ハザード,
- $\lambda_Y(y|X > x, \mathbf{Z}) = \Pr(y \leq Y < y + dy | Y \geq y, X > x, \mathbf{Z})/dy$
時刻 x 以前に増悪しなかったときの条件付ハザード.

相対リスクは2つの条件付ハザード関数の比で定義される;

$$\frac{\lambda_Y(y|X = x, \mathbf{Z})}{\lambda_Y(y|X > x, \mathbf{Z})} = R_\theta \{S_X(x|\mathbf{Z}_1), S_Y(y|\mathbf{Z}_2)\}.$$

右辺を導出するには、モデル(3.1)の仮定が必要である。 $R_\theta > 1$ であれば、増悪したことが死亡リスクを上昇させることを意味する。クレイトン型コピュラでは、相対リスクが定数 $1 + \theta$ なので、これは比例ハザード性を意味する。また、 $\theta > 0$ なので、増悪と死亡が任意の時刻 (x, y) において局所正相関をもつモデルを表す。

Clayton (1978) は、データを表2の 2×2 クロス集計表にまとめ、そのオッズ比を計算することにより θ を推定することを提唱している。この手法は X と Y が共に右側打ち切りを受けている場合にも応用でき、さらに周辺分布の構造 $S_X(x|\mathbf{Z}_1)$ と $S_Y(y|\mathbf{Z}_2)$ に依存しないセミ・パラメトリックな推定法である。この手法はアルキメデス型コピュラに一般化可能であり (Emura et al., 2010), さらに半競合リスクデータ (Wang, 2003) や従属切断データ (Emura and Wang, 2010; Emura et al., 2011) にも適用可能である。

3.4 応用例

コピュラは生存時間のモデルに広く用いられてきたが、その中でもとりわけ臨床医学研究において広く利用されているのが Burzykowski et al. (2001) による2段階推定法 (two-step method) である。代替エンドポイントを用いる臨床試験の策定にあたり、現時点ではコピュラを用いた2段階推定法 (Burzykowski et al., 2005) が広く用いられている。最近、この2段階推定法を実行する R パッケージ (Rotolo et al., 2018) が開発され、利用が容易になった。この R パッケージではクレイトン型、グンベル型、プラケット型コピュラが利用できる。

いま、 i 番目の研究単位から得られた j 番目の患者に対して、 X_{ij} を無増悪期間、 Y_{ij} を全生存期間とする。Burzykowski et al. (2001) は X_{ij} と Y_{ij} の周辺ハザードに

$$\begin{cases} \lambda_{X_{ij}}(t|Z_{ij}) = \lambda_{X_i}(t) \exp(\beta_{X_i} Z_{ij}) \\ \lambda_{Y_{ij}}(t|Z_{ij}) = \lambda_{Y_i}(t) \exp(\beta_{Y_i} Z_{ij}) \end{cases}$$

のような Cox モデルを仮定した。ここで Z_{ij} は治療群を表す変数、 β はその回帰係数、 $\lambda_i(\cdot)$ は i 番目の研究単位の基準ハザード関数である。さらに、 X_{ij} と Y_{ij} の同時分布を、

$$\Pr(X_{ij} > x, Y_{ij} > y) = C_\theta \left[\exp \left\{ - \int_0^x \lambda_{X_{ij}}(t|Z_{ij}) dt \right\}, \exp \left\{ - \int_0^y \lambda_{Y_{ij}}(t|Z_{ij}) dt \right\} \right]$$

で定義し、 C_θ はコピュラとした。パラメータ推定の方法はいくつかある。例えば、まず通常の Cox 回帰法 (部分尤度法) やワイブル回帰法で $(\beta, \lambda_i(\cdot))$ を推定してから、次に θ を推定する手法がある。他にも、 $(\theta, \beta_{X_i}, \beta_{Y_i}, \lambda_{X_i}(\cdot), \lambda_{Y_i}(\cdot))$ を同時推定する全尤度法がある。さらに基準ハザードの共通性 $\lambda_{X_i}(\cdot) = \lambda_{X_0}(\cdot)$ と $\lambda_{Y_i}(\cdot) = \lambda_{Y_0}(\cdot)$ を仮定して推定する場合もある。いずれにせよ、独立な右側打ち切りのある生存期間データに適用可能である。詳細は Burzykowski et al. (2001, 2005), Rotolo et al. (2018) を参照されたい。

上記手法において、エンドポイント間の相関を考慮しているにもかかわらず推定の段階において、独立右側打ち切りしか想定していない点に注意すべきである。例えば、無増悪期間を

X_{ij} として β_{X_i} を Cox 回帰法(部分尤度法)で推定する場合, 死亡を独立右側打ち切りとみなして解析していることになる. しかし死亡による打ち切りは, 無増悪期間と明らかに独立でないため, 多くの患者が早期に(増悪する前に)死亡している場合には回帰推定値に大きなバイアスが入る. この問題は従属打ち切りの問題として知られ, 打ち切り数が多いほど, また相関が強いほどバイアスが大きくなることが知られている (Emura and Chen, 2016, 2018).

この従属打ち切り問題を統計的に正しく扱うには, データ構造を半競合リスク (Fine et al., 2001) の枠組みで捉えた統計手法を用いると良い. すなわち, 死亡は増悪の観測を打ち切るため競合リスクであるが, 増悪は死亡の観測を打ち切ることがないため競合リスクでないとする考えである. 半競合リスクデータの枠組みで正しく回帰推定値を得る場合には, 部分尤度法 (Cox 回帰法) ではなく, 全尤度を利用した推定法 (Chen, 2012; Emura et al., 2017) を利用することになる. 半競合リスクデータを扱う統計手法は 5 節で再考する.

上記以外にも, コピュラは極めて多くの生存時間解析法で利用されてきた. 例えば, クラスタ生存時間解析 (Rotolo et al., 2013; Emura et al., 2017; Peng et al., 2018), 競合リスク解析 (Lo and Wilke, 2010; de Uña-Álvarez and Veraverbeke, 2013, 2017; Shih and Emura, 2018; Shih et al., 2019; Zhou et al., 2019; Emura et al., 2019c; Wang et al., 2020), 半競合リスク解析 (Wang, 2003; Chen, 2012; Peng, 2019; Peng and Xiang, 2019), 従属打ち切り下での生存時間解析 (Emura and Michimae, 2017; Moradian et al., 2019; Emura and Chen, 2016, 2018), 従属切断下での生存時間解析 (Chaieb et al., 2006; Emura and Wang, 2012; Emura and Murotani, 2015; Emura and Pan, 2020), ヴァインコピュラを用いた多変量生存時間データ (Barthel et al., 2018), 繰り返しイベント時間解析などである (Ling et al., 2016; Li et al., 2019, 2020).

4. メタ分析の活用

メタ分析とは, 複数の研究単位から集められたデータを統合し, より一般的な母集団に対する推論を行う手法である. メタ分析は出版された複数の研究論文から要約統計量(相対リスクなど)を抽出して統合解析を行う手法を指すこともあるが, 本稿でいうメタ分析とは患者個人のデータを用いる, いわゆる IPD メタ分析 (Individual-patient data meta-analysis) を指す. 複数の研究単位から集めた患者の個別データに基づく統合解析は, 全生存期間と代替エンドポイントの相関解析に不可欠とされている (Burzykowski et al., 2001, 2005; Shi and Sargent, 2009; Rotolo et al., 2018; Sofeu et al., 2019, 2020).

IPD メタ分析を通じて全生存期間とその代替エンドポイントの相関を推定する研究が数多く行われている. 大腸癌 (Buyse et al., 2008), 頭頸部癌 (Michiels et al., 2009), 胃癌 (Oba et al., 2013), 結腸癌 (Alonso and Molenberghs, 2008) の患者において全生存期間と無増悪生存期間の間に強い正相関があることが報告されている一方, 乳癌の患者においては正相関がそれほど強くないことが報告されている (Michiels et al., 2016). このように全生存期間と無増悪生存期間の間に意味のある相関があるかどうかは癌種や進行度などにより異なる. これらのメタ分析において, 相関の推定にはコピュラが用いられてきた(3.4 節).

推定効率の点からも, 2 つ以上のエンドポイントをコピュラでモデル化した同時解析には利点がある. 実際, Riley (2009) は, 連続変量の多変量メタ分析において, 同時モデルで推定した回帰係数が, 周辺モデルで推定したものより推定精度が良く(推定量の平均二乗誤差が小さく)なるとしている.

ほとんどのメタ分析においては, 研究単位間の異質性を考慮した解析が行われる. 一般的に, 異なる時期, 異なる地域・国で行われた研究を統合する場合, 試験間で, 参加者の背景, エンドポイントや予後因子の定義など, さまざまな要因が厳密には異なることが一般的であ

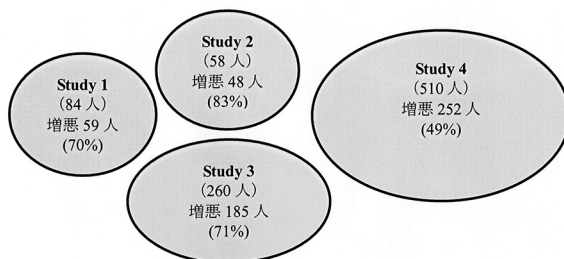


図 3. 卵巣癌患者を 4 つのデータソースから集めたメタ分析のデータ (Ganzfried et al., 2013; Emura et al., 2018). データは R パッケージ *joint.Cox* (Emura, 2019) で入手可能.

る。また、設備が整い経験豊富なスタッフが揃った国は、そうでない国より、患者の死亡や増悪リスクが低いと考えられる。ほとんどのメタ分析の例において、このような研究単位間の異質性は、患者から観測されている共変量で説明しきれないことが多く、これを観測不能なランダム効果として解析モデル内に取り入れる。このような解析モデルをランダム効果モデル (Random-effects model) と呼ぶ。

図 3 は 4 つの研究単位から集められた患者をもとにしたメタ分析のためのデータの例を示している。データは合計 912 人の外科手術後の卵巣癌患者を追跡調査し、増悪があったかどうかを調べたものである。図 3 から増悪率が研究単位間で大きく異なることがわかる。例えば、最も高い研究単位で 83% (Study 2)、最も低いもので 49% である (Study 4)。この増悪率の差異は追跡期間にもよるため、Study 2 が Study 1 より増悪率が高いとは断言できず、増悪率の異質性はデータから容易に観測できるものではない。

生存時間解析において、ランダム効果モデルに相当するものが、Frailty モデルである。とりわけ Shared frailty といわれるモデルにおいては、研究単位内の全ての患者で共通した (Share された) 観測されない因子が存在することを仮定し、この因子の分散パラメータで異質性の度合いをモデル化する (Duchateau et al., 2002; Rondeau et al., 2015)。

図 3 のデータを解析する統計手法を簡単に紹介する。G 個の研究単位を考え、i 番目の研究単位が N_i 人の患者を含むとする ($i = 1, 2, \dots, G$)。i 番目の研究単位の j 番目の患者 ($j = 1, 2, \dots, N_i$) に対し、 X_{ij} をイベント時間、 \mathbf{Z}_{ij} を共変量ベクトルと定義する。図 3 の例では、 $G = 4$ で $N_1 = 84, N_2 = 58, N_3 = 260, N_4 = 510$ であり、合計 912 人となる。

Shared frailty モデルでは、i 番目の研究単位の j 番目の患者 ($j = 1, 2, \dots, N_i$) のハザード関数を次のようにモデル化する

$$\lambda_{ij}(t|u_i, \mathbf{Z}_{ij}) = u_i \lambda_0(t) \exp(\beta' \mathbf{Z}_{ij}).$$

ここで $\beta = (\beta_1, \dots, \beta_p)'$ は未知の回帰係数、 $\lambda_0(\cdot)$ は未知の基準ハザード関数、 $u_i > 0$ ($i = 1, 2, \dots, G$) は観測不能な Frailty 項である。上記モデルにおいて、Frailty 項 u_i が全ての患者で共有されていることが Shared frailty と呼ばれる由来である。

Frailty 項 $u_i > 0$ は確率変数の実現値とみなす。以下では、最も広く使用される平均、分散のガンマ分布を用いる。すなわち、確率密度関数が

$$(4.1) \quad f_\eta(u) = \frac{1}{\Gamma(1/\eta)\eta^{1/\eta}} u^{1/\eta-1} \exp\left(-\frac{u}{\eta}\right), \quad \eta > 0, \quad u > 0$$

であると仮定する。他の研究単位と比較して高リスクな研究単位は $u_i > 1$ 、低リスクな研究単

位は $0 < u_i < 1$ で与えられ、その分散 η は研究単位間の異質性の度合いを決めるパラメータである。極限は $\eta \rightarrow 0$ 研究単位間に異質性が無いことを意味する。実現値 u_i は観測することが出来ないが、その分散 η はデータから推定することが出来る。

パラメータ $(\eta, \beta, \lambda_0(\cdot))$ を最尤法で推定する様々な方法が提案されており、そのうち Frailty 分布としてガンマ分布や対数正規分布を使用する場合は、R のパッケージ等を利用して最尤推定量の計算が可能である (Hirsch and Wienke, 2012; Rondeau and Gonzalez, 2005)。

実データではイベント時間 X_{ij} は右側打ち切りにより観測出来ないケースが多い。そのため、右側打ち切り時間 C_{ij} を定義し、観測される時間を $T_{ij} = \min(X_{ij}, C_{ij})$ とする。そのため i 番目の研究単位の j 番目の患者に関するデータは $(T_{ij}, \delta_{ij}, \mathbf{Z}_{ij})$ となり、 $\delta_{ij} = I(T_{ij} = X_{ij})$ はイベントが起こったか否かを示す指標である。

パラメータ $(\eta, \beta, \lambda_0(\cdot))$ の推定はデータ $\{(T_{ij}, \delta_{ij}, \mathbf{Z}_{ij}); i = 1, 2, \dots, G; j = 1, 2, \dots, N_i\}$ が与えられたときの尤度関数を構成し、対数尤度関数 $(\eta, \beta, \lambda_0(\cdot))$ に関して最大化することによって行う (詳しくは Emura et al., 2019b の 2.5 節を参照)。

5. メタ分析による 2 つのエンドポイントの同時解析

本節では 2 つのエンドポイントを同時に解析するための方法を紹介する。とりわけ、無増悪期間と全生存期間を同時解析し、その関連を調べるために患者の個別データ (Individual-Patient-Data, IPD) を用いたメタ分析を考える。このような分析法は「IPD メタ分析」と呼ばれており、要約データに基づくメタ分析 (個別データを用いない分析) と本質的に異なる。3 節で示したようにエンドポイント間の相関をコンピュータでモデル化し、4 節で示したように研究単位間の異質性を Frailty 項でモデル化するために Joint frailty-copula モデル (Emura et al., 2017) を導入する。

5.1 エンドポイントの測定と半競合リスク

複数のエンドポイントを扱う統計手法では、エンドポイントがどのような打ち切りを受けているか明確にする必要がある。これは、データ構造を明らかにするだけでなく、尤度関数を正しく構成するために必要なプロセスである。

G 個の研究単位を考え、 i 番目の研究単位が N_i 人の患者を含むとする ($i = 1, 2, \dots, G$)。 i 番目の研究単位の j 番目の患者 ($j = 1, 2, \dots, N_i$) に対し、

- X_{ij} = 無増悪期間,
- D_{ij} = 全生存期間,
- C_{ij} = 右側打ち切り期間 (独立かつ無情報な打ち切りを仮定),

と定義する。患者は 3 つの変数 X_{ij} , D_{ij} , C_{ij} を有しているが、どれが観測されるかは、患者の状態や打ち切り期間によって決まる。例えば、患者の状態が安定していれば、 X_{ij} と D_{ij} は観測されず、打ち切り期間 C_{ij} のみが観測される。この場合、打ち切り期間 C_{ij} までに増悪と死亡が起こらなかった情報が記録される。

図 4 で患者がとりうる 4 つのケース (A~D) を示す。ケース A では、患者が増悪し、その後死亡しているので、 X_{ij} と D_{ij} が観測される。ケース B では、患者は増悪したが、死亡する前に打ち切りとなっているので、 X_{ij} と C_{ij} が観測される。すなわち、 D_{ij} が C_{ij} によって右側打ち切りされている。ケース C では、患者が増悪せずに死亡しているので、 D_{ij} のみが観測される。この場合、 X_{ij} が D_{ij} によって右側打ち切りされている。 X_{ij} と D_{ij} に相関があるので、この打ち切りは通常の「独立打ち切り」でないことに注意する。ケース D では、 X_{ij} も D_{ij} も観

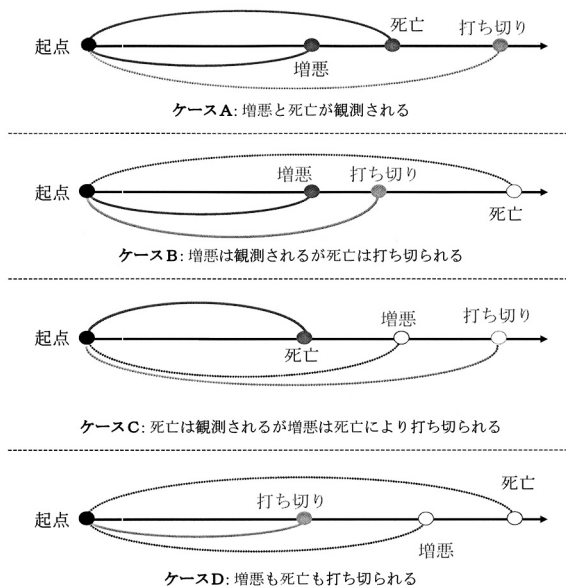


図 4. 患者がとりうる 4 つのケース (A~D). 観測されるイベントは●, 観測されないイベントは○で表す. 観測される期間は実線, 観測されない期間は点線で表す.

測されず, 打ち切り期間 C_{ij} のみが観測される. すなわち, X_{ij} も D_{ij} も右側打ち切りされている.

まとめると i 番目の研究単位の j 番目の患者 ($j = 1, 2, \dots, N_i$) に対して観測されるデータは $(T_{ij}, T_{ij}^*, \delta_{ij}, \delta_{ij}^*, Z_{1,ij}, Z_{2,ij})$ と書ける. ここで,

- $T_{ij} = \min(X_{ij}, D_{ij}, C_{ij})$; 最初に起こったイベントの時間,
- $\delta_{ij} = I(T_{ij} = X_{ij})$; 増悪の指標 (無増悪 = 0; 増悪 = 1),
- $T_{ij}^* = \min(D_{ij}, C_{ij})$; 最後に起こったイベントの時間,
- $\delta_{ij}^* = I(T_{ij}^* = D_{ij})$; 死亡の指標 (生存 = 0; 死亡 = 1),
- $Z_{1,ij}$; 増悪に関連する p_1 次元共変量,
- $Z_{2,ij}$; 死亡に関連する p_2 次元共変量.

ペア $(\delta_{ij}, \delta_{ij}^*)$ は上記 A, B, C, D の 4 つのケースを区別するための情報であり, $(\delta_{ij}, \delta_{ij}^*) = (1, 1), (1, 0), (0, 1), (0, 0)$ のいずれかを取りえる (表 3). 例えば, ケース A は $(\delta_{ij}, \delta_{ij}^*) = (1, 1)$ に対応し, この場合 $T_{ij} = X_{ij}$ と $T_{ij}^* = D_{ij}$ が観測される.

上記のような構造をもつ生存期間データは「半競合リスクデータ」と呼ばれている (Fine et al., 2001). これは, 死亡のような終着イベント (terminal event) が増悪のような非終着イベント (non-terminal event) の競合リスクになっていることに由来する. ここで「終着イベント」とは, それが起こった後の一切の情報が得られなくなるようなイベントを指す. 死亡は明らかに終着イベントである. 増悪後には死亡時刻などの情報が得られるため, 増悪は終着イベントでない. 半競合リスクデータでは増悪だけでなく, より一般の中間イベント (非終着イベント) を考えることが出来る. 半競合リスクデータの統計解析法の最新の動向は Peng (2019) を参照されたい.

表 3. 患者がとりうる 4 つのケース (A~D).

	最初のイベント	最後のイベント	T_{ij}	T_{ij}^*	δ_{ij}	δ_{ij}^*
ケース A	増悪	死亡	X_{ij}	D_{ij}	1	1
ケース B	増悪	打ち切り	X_{ij}	C_{ij}	1	0
ケース C	死亡	死亡	D_{ij}	D_{ij}	0	1
ケース D	打ち切り	打ち切り	C_{ij}	C_{ij}	0	0

5.2 Joint frailty-copula モデル

Rondeau et al. (2015) は研究単位間の異質性を考慮した半競合リスクデータに基づくメタ分析の手法を提案している. 彼女らが提案している Joint frailty モデルにおいて, Frailty 項 $u_i > 0$ は $\text{Var}_\eta(u_i) = \eta$ のガンマ分布のモデル (4.1) に従うと仮定し, 分散 η は研究単位間の異質性の度合いを決めるパラメータである. X_{ij} と D_{ij} の条件付 ($u_i, \mathbf{Z}_{1,ij}, \mathbf{Z}_{2,ij}$ を与えたとき) ハザード関数が次のような Joint frailty モデルに従うと仮定する.

$$\begin{cases} r_{ij}(t|u_i) = u_i r_0(t) \exp(\beta'_1 \mathbf{Z}_{1,ij}) \\ \lambda_{ij}(t|u_i) = u_i^\alpha \lambda_0(t) \exp(\beta'_2 \mathbf{Z}_{2,ij}). \end{cases}$$

ここで β は $\mathbf{Z}_{\cdot,ij}$ の効果を決める未知パラメータで, $r_0(\cdot)$ と $\lambda_0(\cdot)$ は未知の基準ハザード関数である. 他の研究単位と比較して高リスクな研究単位は $u_i > 1$, 低リスクな研究単位は $0 < u_i < 1$ で与えられ, パラメータ α は増悪率と死亡率の異質性の差を表す. 例えば, $\alpha = 0$ は増悪率に異質性があるが死亡率には異質性がないモデル, $\alpha = 1$ は増悪率と死亡率に同等の異質性があるモデルに対応する. Rondeau et al. (2015) のモデルにおいて重要な仮定は, ($u_i, \mathbf{Z}_{1,ij}, \mathbf{Z}_{2,ij}$) が与えられたもとで, X_{ij} と D_{ij} が条件付独立となることである.

Emura et al. (2017) は条件付独立の仮定を除いたコピュラのモデルである Joint frailty-copula モデルを提案した. C_θ をコピュラとしたとき, Joint frailty-copula モデルは次の式で与えられる;

$$(5.1) \quad \begin{cases} r_{ij}(t|u_i) = u_i r_0(t) \exp(\beta'_1 \mathbf{Z}_{1,ij}) \\ \lambda_{ij}(t|u_i) = u_i^\alpha \lambda_0(t) \exp(\beta'_2 \mathbf{Z}_{2,ij}) \\ \Pr(X_{ij} > x, D_{ij} > y|u_i) = C_\theta[S_{X_{ij}}(x|u_i), S_{D_{ij}}(y|u_i)]. \end{cases}$$

ここで θ は未知パラメータである. このモデルにおいて, X_{ij} と D_{ij} の生存関数は次のように書ける

$$\begin{cases} S_{X_{ij}}(x|u_i) = \exp\{-u_i R_0(x) \exp(\beta'_1 \mathbf{Z}_{1,ij})\} \\ S_{D_{ij}}(y|u_i) = \exp\{-u_i^\alpha \Lambda_0(y) \exp(\beta'_2 \mathbf{Z}_{2,ij})\}. \end{cases}$$

ここで $R_0(x) = \int_0^x r_0(t) dt$ と $\Lambda_0(y) = \int_0^y \lambda_0(t) dt$ は累積基準ハザード関数である.

コピュラ C_θ は X_{ij} と D_{ij} の相関構造を定める. 独立型 $C_\theta(v, w) = vw$ は Rondeau et al. (2015) のモデルに帰着する. クレイトン型ではケンドール相関は $\tau_\theta = \theta/(\theta + 2)$ であり, グンベル型ではケンドール相関は $\tau_\theta = \theta/(\theta + 1)$ である. Burzykowski et al. (2001) のコピュラモデルと同じ原理で, ケンドール相関は, 無増悪期間が全生存期間の代替エンドポイントになりえるかどうかの指標の 1 つである.

Joint frailty-copula モデル内の未知パラメータは観測されたデータ $\{(T_{ij}, T_{ij}^*, \delta_{ij}, \delta_{ij}^*, \mathbf{Z}_{1,ij}, \mathbf{Z}_{2,ij}), i = 1, 2, \dots, G, j = 1, 2, \dots, N_i\}$ を用いて最尤法で推定することが出来る. まず次のよう

な記号を導入する；

$$R_{ij}(t) = R_0(t) \exp(\beta'_1 \mathbf{Z}_{1,ij}), \quad r_{ij}(t) = dR_{ij}(t)/dt = r_0(t) \exp(\beta'_1 \mathbf{Z}_{1,ij}),$$

$$\Lambda_{ij}(t) = \Lambda_0(t) \exp(\beta'_2 \mathbf{Z}_{2,ij}), \quad \lambda_{ij}(t) = d\Lambda_{ij}(t)/dt = \lambda_0(t) \exp(\beta'_2 \mathbf{Z}_{2,ij}).$$

モデル(4.1)と(5.1)のもとで、対数尤度関数は次のように構成される；

$$\begin{aligned} \ell(\alpha, \eta, \theta, \beta_1, \beta_2, r_0, \lambda_0) &= \sum_{i=1}^G \left[\sum_{j=1}^{N_i} \{ \delta_{ij} \log r_{ij}(T_{ij}) + \delta_{ij}^* \log \lambda_{ij}(T_{ij}^*) \} \right. \\ &\quad \left. + \log \int_0^\infty \left\{ u^{m_i + \alpha m_i^*} \prod_{j=1}^{N_i} \psi_\theta [uR_{ij}(T_{ij}), u^\alpha \Lambda_{ij}(T_{ij}^*)]^{\delta_{ij}} \psi_\theta^* [uR_{ij}(T_{ij}), u^\alpha \Lambda_{ij}(T_{ij}^*)]^{\delta_{ij}^*} \right. \right. \\ &\quad \left. \left. \times \Theta_\theta [uR_{ij}(T_{ij}), u^\alpha \Lambda_{ij}(T_{ij}^*)]^{\delta_{ij} \delta_{ij}^*} D_\theta [uR_{ij}(T_{ij}), u^\alpha \Lambda_{ij}(T_{ij}^*)] \right\} f_\eta(u) du \right], \end{aligned}$$

ここで $m_i = \sum_{j=1}^{N_i} \delta_{ij}$ は i 番目の研究単位で増悪した患者数, $m_i^* = \sum_{j=1}^{N_i} \delta_{ij}^*$ は i 番目の研究単位で死亡した患者数であり, $D_\theta[s, t] = C_\theta[\exp(-s), \exp(-t)]$,

$$\psi_\theta[s, t] = \frac{D_\theta^{[1,0]}[s, t]}{D_\theta[s, t]}, \quad \psi_\theta^*[s, t] = \frac{D_\theta^{[0,1]}[s, t]}{D_\theta[s, t]}, \quad \Theta_\theta[s, t] = \frac{D_\theta^{[1,1]}[s, t] D_\theta[s, t]}{D_\theta^{[1,0]}[s, t] D_\theta^{[0,1]}[s, t]},$$

$D_\theta^{[1,0]}[s, t] = -\partial D_\theta[s, t]/\partial s$, $D_\theta^{[0,1]}[s, t] = -\partial D_\theta[s, t]/\partial t$, $D_\theta^{[1,1]} = \partial^2 D_\theta[s, t]/\partial s \partial t$ はコンピュータを微分して得られる. 対数尤度の導出は Emura et al. (2017, 2019b) を参照されたい.

例えばクレイトン型では, $D_\theta[s, t] = A_\theta(s, t)^{-1/\theta}$, $\psi_\theta[s, t] = \exp(\theta s)/A_\theta(s, t)$, $\psi_\theta^*[s, t] = \exp(\theta t)/A_\theta(s, t)$, $\Theta_\theta[s, t] = 1 + \theta$ とおき, ここで $A_\theta(s, t) = \exp(\theta s) + \exp(\theta t) - 1$ である. これらを用いて対数尤度を計算する.

基準ハザード関数 $r_0(\cdot)$ と $\lambda_0(\cdot)$ のモデル化はいくつか提案されている. Emura et al. (2017) はスプライン関数でモデル化し, 罰則付き最尤法で推定することを提案している. スプライン関数は単調増加(減少)型, 凸型, 凹型などのハザードを柔軟にモデル化できる (Shih and Emura, 2020). Peng et al. (2018) と Peng (2019) は完全にノンパラメトリックなモデルを提案している. Li et al. (2019) はパラメトリックなモデルを考え, パラメータに事前分布を仮定したベイズ法を用いている. 篠原 (2019) と Wu et al. (2020) はワイブル分布を用いている. Schneider et al. (2020) でもワイブル分布と区分的指数分布のモデルを考えているが, Rondeau et al. (2015) と同様に, 条件付独立性を仮定している. スプライン関数を用いる場合, 推定値の標準誤差(SE)と信頼区間(CI)は罰則付き対数尤度のヘッセ行列から計算でき (Emura et al., 2017, 2019b), R パッケージ *joint.Cox* (Emura, 2019) で実行可能である. ノンパラメトリック最尤法を用いる場合, ヘッセ行列を利用すると SE を過小評価する傾向があるため, 層化ブートストラップ法が推奨されている (Peng, 2019).

5.3 卵巣癌データへの適用例

Ganzfried et al. (2013) の卵巣癌データを用いて, 2つのエンドポイントを同時に解析する方法を説明する. Ganzfried et al. (2013) は IPD メタ分析を用いて, 遺伝子 *CXCL12* の発現量の全生存期間に与える影響が有意であることを示した. 本稿では, Joint frailty-copula モデルの適用例として, *CXCL12* 遺伝子が2つのエンドポイント(無増悪期間と全生存期間)に与える影

表 4. 卵巣癌データ (Ganzfried et al., 2013; Emura et al., 2017). 研究単位は GEO (Gene Expression Omnibus) アクセス番号をあらわす. イベント発生率 (%) は研究単位ごとに特定のイベント (増悪, 死亡, 打ち切り) の発生率を計算した.

研究単位	サンプル数	観測されたイベント数 (イベント発生率 %)		
		増悪 ($\delta_y = 1$)	死亡 ($\delta_y^* = 1$)	打ち切り ($\delta_y^* = 0$)
GSE17260	$N_1 = 110$	76 (69%)	46 (42%)	64 (58%)
GSE30161	$N_2 = 58$	48 (83%)	36 (62%)	22 (38%)
GSE9891	$N_3 = 278$	185 (67%)	113 (41%)	165 (59%)
TCGA	$N_4 = 557$	266 (48%)	290 (52%)	267 (48%)
合計	$\sum_{i=1}^4 N_i = 1003$	575 (57%)	485 (48%)	518 (52%)

表 5. 卵巣癌データにおける無増悪期間と全生存期間を同時解析した結果. *CXCL12* 遺伝子発現量の相対リスクを考えている. 相対リスクが 1 以上とは, *CXCL12* 値が高いと増悪 (死亡) リスクが上昇することを意味する.

コピュラ	クレイトン型	独立型
	推定値 (95% CI)	推定値 (95% CI)
増悪の相対リスク ; $\exp(\beta_1)$	1.22 (1.13-1.32)	1.24 (1.14-1.35)
死亡の相対リスク ; $\exp(\beta_2)$	1.18 (1.08-1.29)	1.17 (1.07-1.29)
異質性の分散 ; η	0.033 (0.006-0.187)	0.028 (0.004-0.180)
相関パラメータ ; θ	2.35 (1.90-2.90)	0.00 (固定)
ケンドール相関 ; $\tau = \theta / (\theta + 2)$	0.54 (0.49-0.59)	0.00 (固定)

響の解析事例を紹介する.

Ganzfried et al. (2013) のデータは多くの研究単位を含む IPD メタ分析用のデータであるが, 多くの研究単位では 1 つのエンドポイントしか測定されていない. そのため, Emura et al. (2017) で行われたように, 無増悪期間と全生存期間の両方が測定されている 4 つの研究単位から 1,003 人の卵巣癌患者を取り出した (表 4). この取り出したデータは R パッケージ *joint.Cox* に収められている. Ganzfried et al. (2013) に従い, *CXCL12* 遺伝子の発現量のみを共変量とした.

卵巣癌データを Joint frailty-copula のクレイトン型コピュラモデルに当てはめた結果を表 5 に示す. *CXCL12* 遺伝子が全生存期間に与える影響は有意 (P 値 < 0.05) であり, その相対リスク (Relative Risk) は $RR = 1.18$ (95% CI: 1.08-1.29) である. また, *CXCL12* 遺伝子が無増悪期間に与える影響も有意 (P 値 < 0.05) であり, その影響は全生存期間のそれよりも大きい ($RR = 1.22$, 95% CI: 1.13-1.32). このことは, *CXCL12* の発現量は卵巣癌患者の術後増悪を予測するバイオマーカーとなりうることを示している. これら相対リスクの値は, *CXCL12* の発現量が SD 1 つ分増加したときのリスク増加を表す. コピュラのパラメータ推定値 ($\hat{\theta} = 2.35$, 95% CI: 1.90-2.90) は, 増悪と死亡の間に中程度の相関があることを示す ($\tau = 0.54$, 95% CI: 0.38-0.70). このことは, 増悪で死亡をある程度予測できることを意味する.

独立型コピュラを適用した結果を表 5 に示す. この場合, 増悪と死亡間の相関を考慮できないため, このモデルの妥当性は弱い. しかし, 相対リスクの推定値は, クレイトン型のものと比べてほとんど変化がない. これは, 相対リスクの推定値はコピュラの誤特定に関してある程

度の頑健性を持つことを示している。このような頑健性は Emura et al. (2017) のシミュレーション解析でも報告されている。

6. 動的予測と個別医療

6.1 生存期間の予測

生存時間解析における重要なテーマの1つが、生存期間の予測法の開発である。過去の患者の生存期間データをもとにして、新たな患者の生存期間を予測するためのモデル構築に関する手法は数多く提案されており、これは統計学の手法に加えて機械学習の手法も多い。例えば、乳癌 (Gómez et al., 2016; Shukla et al., 2018)、卵巣癌 (Yoshihara et al., 2010; Emura et al., 2018)、膀胱癌 (Guinney et al., 2017)、その他の癌種 (Kim et al., 2018; Choi et al., 2018) について、患者の生存期間の予測式が提案されている。

正確な予測式を構築することは、患者本人やその主治医に対してより正確な人生設計や治療計画を考える礎となるため、個別医療 (personalized medicine) の中核となる。個別医療の定義によると、それは遺伝子 (分子) 情報を基に健康増進を図る医療である (Schleiden et al., 2013; Hayes et al., 2014)。これは遺伝子 (分子) 情報の活用が患者の予後予測に重要な役割を果たすためである。

Waldron et al. (2014) は卵巣癌患者の大規模な IPD メタ分析用データを用い、遺伝子を利用した 14 種類の予測法の予測精度を比較した。彼らは、12 種類の予測法が全生存期間の予測に有意 (P 値 < 0.05) に働くことを示したが、同時にその予測能力はそれほど高くない (C 統計量 $= 0.56 - 0.60$) ことも指摘した。彼らはまた、これら生存期間予測法が臨床現場で実際に有効利用されるためには、さらなる改良や検証が必要であることも指摘している。

生存期間予測法を改良するためのアイデアは数多くあるが、以下では動的予測法とコンピュータを IPD メタ分析用データに適用した予測手法を考える。

6.2 動的予測

動的予測 (dynamic prediction) では、通常の共変量 (年齢、腫瘍径など) に加えて、中間イベント (増悪など) の情報を予測因子に取り入れる (van Houwelingen and Putter, 2011)。治療開始時点 (例えば手術日や投与開始日) を起点とする通常の予測の場合、手術日に測定される年齢や腫瘍径を共変量として扱い、手術日から時刻 w 後 (例えば $w = 5$ 年) に死亡する確率を予測する。動的予測では、手術後に経過観察した後にしか得られない増悪イベントの情報などを取り入れるため、起点を術後 (例えば術後 t 日目) に設定し、時刻 t まで生存した患者が、時刻 $t + w$ までに死亡する (または生存している) 確率を予測する (Thompson et al., 2007)。

いま D を全生存期間、 X を無増悪期間、 Z を遺伝子情報を含む共変量とする。共変量の値 Z は起点 $t = 0$ (例えば手術日) に測定され、その後変化しないと仮定する (時間依存性共変量は考えない)。

図 5 はで 4 人の患者 (患者 1 ~ 患者 4) に対して動的予測を考えている。患者 1, 2 は時刻 t 以前に死亡しているので、予測の対象とはならない。患者 3, 4 は時刻 t において生存しているので、予測を行うことが出来る。患者 3 は時刻 t 以前に増悪しているので、増悪時刻 $X = x$ で条件付けを行った死亡率を予測する。患者 4 は、時刻 t 以前には増悪していないので $X > t$ で条件付けを行った死亡率を予測する。

動的予測を実践するにあたり、予測時刻 t は前もって決めておく必要がある。例えば、時刻 t は術後 3 ヶ月ごとにと決めておく。この予測スケジュールは患者の状態によって変更してはならず、また患者の死亡や増悪に関連のある事象に依存してはならない。例えば、患

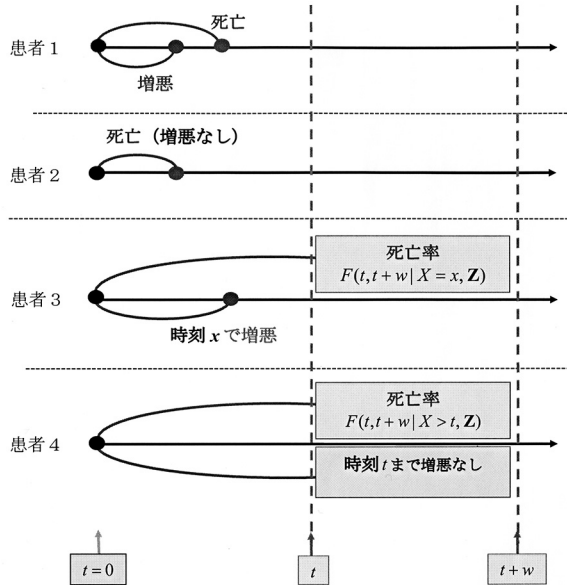


図 5. 増悪情報に基づく死亡率の動的予測. 患者 1, 2 は時刻 t 以前に死亡しているため, 動的予測の対象とならない. 患者 3, 4 に対する動的予測は, 時刻 t 以前に観測されている増悪情報に基づいて行われる.

者が何らかの不調を訴えて来院した時刻や, 増悪した時刻を t に設定すると, 妥当な予測は行えなくなる.

動的予測を定式化するために次のような 2 つの関数を利用する.

(i) 条件付死亡率関数 (Conditional Failure Function)

動的予測の基礎となるのが, 次の 2 つの条件付死亡率関数

$$F(t, t+w | X > t, Z) = \Pr(D \leq t+w | D > t, X > t, Z),$$

$$F(t, t+w | X = x, Z) = \Pr(D \leq t+w | D > t, X = x, Z), \quad x \leq t$$

である. 2 つの式にある条件 $\{D \geq t\}$ は, 予測の対象である患者が時刻 t において生存していることを前提としている. 1 つ目の式にある条件 $\{X > t\}$ は, 時刻 t においてまだ増悪していないことを意味する. 2 つ目の式にある条件 $\{X = x\}$ は, 時刻 t 以前に増悪したことを意味する. この場合, 増悪した時刻 $X = x$ は既知なので予測式に組み入れることができる. 条件付死亡率関数は, これら条件下で, 患者が時刻 t から $t+w$ の間に死亡する確率である. 通常の (動的でない) 予測では予測時刻を $t=0$ とし, $w=5$ 年などに設定して予測を行う. 予測時刻を $t=0$ と設定すると, 増悪の情報は不明なので, 予測式は $F(t, t+w | Z) = \Pr(D \leq t+w | D > t, Z)$ のみである.

増悪した患者は, そうでない患者に比べて, その死亡率が上昇する. したがって, 直感的に不等式 $F(t, t+w | X = x, Z) > F(t, t+w | X > t, Z)$ が全ての $w > 0$ に対して成立すると考えられる (図 1). X と D が正相関を持つ場合, この不等式は通常成立する (数学的に成立するかどうかはモデルに依存). X と D が (Z を与えたときの) 条件付独立であれば, 等式 $F(t, t+w | X = x, Z) = F(t, t+w | X > t, Z)$ が成立することは明らかである.

(ii) 条件付ハザード関数 (Conditional Hazard Function)

条件付ハザード関数は、条件付死亡率関数を用いて次のように表現できる。

$$\lambda(t|X = x, \mathbf{Z}) = F(t, t + dw|X = x, \mathbf{Z})/dw,$$

$$\lambda(t|X > x, \mathbf{Z}) = F(t, t + dw|X > x, \mathbf{Z})/dw.$$

これらの式により、条件付ハザード関数は、上記の条件付死亡率関数において w が微小な値 dw をとる場合と解釈できる。2つの条件付ハザード関数の比をとると、 X と D の関連を表す指標である次のようなオッズ比関数(3.3節)となる。

$$\frac{\lambda(t|X = x, \mathbf{Z})}{\lambda(t|X > x, \mathbf{Z})} = \frac{\Pr(X = x, D = t|\mathbf{Z})\Pr(X > x, D > t|\mathbf{Z})}{\Pr(X = x, D > t|\mathbf{Z})\Pr(X > x, D = t|\mathbf{Z})}, \quad t > 0, \quad x > 0.$$

オッズ比が1より大きいときは正相関を表し、オッズ比が1より小さいときは負相関を表す。オッズ比が1のときは X と D は (\mathbf{Z} を与えたときの)条件付独立である。

Clayton (1978)が提案した次のセミパラメトリックモデル

$$\Pr(X > x, D > t|\mathbf{Z}) = [S_X(x|\mathbf{Z})^{-\theta} + S_D(t|\mathbf{Z})^{-\theta} - 1]^{-\frac{1}{\theta}}, \quad t > 0, \quad x > 0, \quad \theta > 0$$

を考えると、このモデルの下では、2つの条件付ハザード関数の関係をオッズ比定数 ($\theta + 1$) を用いて

$$\lambda(t|X = x, \mathbf{Z}) = (\theta + 1)\lambda(t|X > x, \mathbf{Z}), \quad t > 0, \quad x > 0.$$

と表すことができる。これは、患者の増悪イベント $\{X = x\}$ があったときと、そうでないとき $\{X > x\}$ の相対リスクが $(\theta + 1)$ となるクレイトン型 (Clayton, 1978)コピュラモデルである(3.3節を参照)。Day et al. (1997)はクレイトン型モデルにランドマーク解析を使った動的予測を提案している。

動的予測において、条件付死亡率関数、条件付ハザード関数のどちらも有用な予測指標である。時刻 t を予測時刻と呼び、これは臨床医が患者に対して予測を始める時刻である。これら予測指標をデータから推定する場合、通常何らかのモデルを利用する。van Houwelingen and Putter (2011)はランドマーク解析法を採用している。これは予測時刻 t 以前に収集されたデータを基に条件付 Cox 回帰法を当てはめて予測モデルを構築するものである。ランドマーク解析法では、 t を複数選び、複数の条件付 Cox モデルを構築したときにモデル間の一致性が成立しないなどのいくつかの諸問題がある。詳細については van Houwelingen and Putter (2011)と Suresh (2018)に詳しく議論されている。

ランドマーク解析法の諸問題の解決策として、結合モデル (Joint model) を用いた動的予測式が提案されている (Proust-Lima et al., 2014; Rondeau et al., 2017; Sène et al., 2016 など)。これらのモデルにおいては、Frailty 項を用いて応答変数間の相関を考慮している。Frailty 項の替わりに、コピュラで相関を考慮するモデルは Emura et al. (2018), Suresh et al. (2019)などを参照されたい。

6.3 Joint frailty-copula モデルを利用した動的予測

以下では、Frailty 項とコピュラで構成した Joint frailty-copula モデルによる動的予測法を紹介する (Emura et al., 2018)。この手法では、全生存期間と無増悪期間の相関を解析する際に用いる IPD メタ分析データを利用する。このため、Frailty 項は相関を考慮するというよりも、研究単位間の異質性をモデル化するために導入されている。

無増悪期間の生存関数を $S_X(x|u) = \Pr(X > x|u, \mathbf{Z})$ 、全生存期間の生存関数を $S_D(y|u) =$

$\Pr(D > y|u, \mathbf{Z})$ で定義し, ここで u は共変量 \mathbf{Z} で説明しきれない Frailty 項である. u の分布は, 式(4.1)で与えられるように, 平均 1, 分散 η のガンマ分布であるとする. Joint frailty-copula モデルの仮定をおくと,

$$\Pr(X > x, D > y|u, \mathbf{Z}) = C_\theta[S_X(x|u), S_D(y|u)],$$

となり, $C_\theta(v, w)$ はコピュラ, θ は X と D の相関を表すパラメータである. このモデルの下で, 条件付死亡率関数は次のように表せる.

(I) 時刻 t 以前に患者が増悪していない場合

$X > t$ で条件付けを行い, 条件付死亡率関数は

$$\begin{aligned} F(t, t+w|X > t, \mathbf{Z}) &= \Pr(D \leq t+w|D > t, X > t, \mathbf{Z}) \\ &= \frac{\int_0^\infty (C_\theta[S_X(t|u), S_D(t|u)] - C_\theta[S_X(t|u), S_D(t+w|u)])f_\eta(u)du}{\int_0^\infty C_\theta[S_X(t|u), S_D(t|u)]f_\eta(u)du} \end{aligned}$$

となる. ただし $f_\eta(u)$ は式(4.1)で与えられるガンマ分布の確率密度関数である. また, 条件付ハザード関数は

$$\lambda(t|X > t, \mathbf{Z}, u) = \lambda_D(t|u) \frac{S_D(t|u)C_\theta^{[0,1]}[S_X(t|u), S_D(t|u)]}{C_\theta[S_X(t|u), S_D(t|u)]}$$

となる. ここで $\lambda_D(t|u) = -\partial \log S_D(t|u)/\partial t$, $C_\theta^{[0,1]}(v, w) = \partial C_\theta(v, w)/\partial w$ である.

(II) 時刻 t 以前に患者が増悪している場合

増悪時刻 $X = x$ で条件付けを行い, 条件付死亡率関数は

$$\begin{aligned} F(t, t+w|X = x, \mathbf{Z}) &= \Pr(D \leq t+w|D > t, X = x, \mathbf{Z}) \\ &= \frac{\int_0^\infty (C_\theta^{[1,0]}[S_X(x|u), S_D(t|u)] - C_\theta^{[1,0]}[S_X(x|u), S_D(t+w|u)])\lambda_X(x|u)S_X(x|u)f_\eta(u)du}{\int_0^\infty C_\theta^{[1,0]}[S_X(x|u), S_D(t|u)]\lambda_X(x|u)S_X(x|u)f_\eta(u)du} \end{aligned}$$

となる. ここで $\lambda_X(x|u) = -\partial \log S_X(x|u)/\partial x$ である. また, 条件付ハザード関数は

$$\lambda(t|X = x, \mathbf{Z}, u) = \lambda_D(t|u) \frac{S_D(t|u)C_\theta^{[1,1]}[S_X(x|u), S_D(t|u)]}{C_\theta^{[1,0]}[S_X(x|u), S_D(t|u)]}$$

となる. ここで $C_\theta^{[1,0]}(v, w) = \partial C_\theta(v, w)/\partial v$, $C_\theta^{[1,1]}(v, w) = \partial^2 C_\theta(v, w)/\partial v \partial w$ である.

上式の導出は Emura et al. (2018, 2019b)を参照されたい. IPD メタ分析用データなどを用いて条件付死亡率関数や条件付ハザード関数内のパラメータは推定することができる(詳しくは Emura et al., 2018, 2019b を参照).

条件付死亡率関数は Frailty 項 u で積分している. これは予測すべき患者の Frailty 項 u が観測不能なため, u の分布で平均化していることを意味する. 積分は数値的に計算する. 簡便な方法であるが, 積分する代わりに平均値 $u = 1$ で置き換える方法もある.

条件付ハザード関数は Frailty 項 u で条件付けを行っている. u は未知なので積分する方が望ましいが, 条件付けを行うことにより, 次のような数学的性質が利用できる.

$$\frac{\lambda(t|X = t, \mathbf{Z}, u)}{\lambda(t|X > t, \mathbf{Z}, u)} = R_\theta[S_X(t|u), S_D(t|u)].$$

ここで R_θ は 3.3 節で既出のオッズ比関数である. クレイトン型コピュラでは $R_\theta(v, w) = 1 + \theta$ が $\{X = t\}$ の $\{X > t\}$ に対する相対リスクとなる. 先に $\lambda(t|X > t, \mathbf{Z}, u)$ を計算すれば, それに $1 + \theta$ をかけることにより $\lambda(t|X = t, \mathbf{Z}, u)$ が得られる. 予測に使用する Frailty 項は $u = 1$

とする。

動的予測を臨床現場で実践する場合、まだ死亡していない新たな患者を対象にする。ここで「新たな」とは、すでに死亡・生存の状態が確定しているデータ内の患者と区別するための語である。現実にもそのような患者を有していなくても、仮想的な患者を想定し、どのように予測のプロセスが実行されるのかを考察することは重要である。

まず患者が有している共変量情報を $\mathbf{Z} = (\mathbf{Z}_1, \mathbf{Z}_2, \text{CC}_1, \text{CC}_2)$ とし、これらは：

- \mathbf{Z}_1 ; 増悪に関連する p_1 次元の臨床的共変量,
- \mathbf{Z}_2 ; 全生存に関連する p_2 次元の臨床的共変量,
- $\text{CC}_1 = w_1 V_1 + \dots + w_{q_1} V_{q_1}$; 増悪に関連する遺伝子発現量の複合共変量,
- $\text{CC}_2 = \varpi_1 W_1 + \dots + \varpi_{q_2} W_{q_2}$; 全生存に関連する遺伝子発現量の複合共変量,

と定義する。ここで (V_1, \dots, V_{q_1}) は q_1 個の増悪に関連する遺伝子発現量, (W_1, \dots, W_{q_2}) は q_2 個の全生存に関連する遺伝子発現量である。遺伝子の個数 q_1 (または q_2) と重み w_k (または ϖ_k) はデータから計算される (Emura et al., 2018, 2019b)。全ての遺伝子発現量は、遺伝子ごとに平均 0, 標準偏差 1 で標準化されていると仮定する。複合共変量 (compound covariate) の基本的な考え方については Matsui (2006), Emura et al. (2012, 2019a, 2019b) を参照されたい。

データを Joint frailty-copula モデルに当てはめることにより、無増悪期間と全生存期間の生存関数が次のように推定される。

$$\begin{aligned} \hat{S}_X(t|u) &= \exp \left\{ -u \hat{R}_0(t) \exp \left(\hat{\beta}'_1 \mathbf{Z}_1 + \hat{\gamma}_1 \frac{\text{CC}_1 - \hat{\mu}_1}{\hat{\sigma}_1} \right) \right\}, \\ \hat{S}_D(t|u) &= \exp \left\{ -u^\alpha \hat{\Lambda}_0(t) \exp \left(\hat{\beta}'_2 \mathbf{Z}_2 + \hat{\gamma}_2 \frac{\text{CC}_2 - \hat{\mu}_2}{\hat{\sigma}_2} \right) \right\}. \end{aligned}$$

ここで $\hat{R}_0(t) = \int_0^t \hat{r}_0(x) dx$ と $\hat{\Lambda}_0(t) = \int_0^t \hat{\lambda}_0(x) dx$ は推定された累積基準ハザード関数, $(\hat{\theta}, \hat{\eta}, \hat{\beta}_1, \hat{\beta}_2, \hat{\gamma}_1, \hat{\gamma}_2, \hat{r}_0, \hat{\lambda}_0)$ は推定値されたパラメータの値である。 $\hat{\mu}$ は CC の平均, $\hat{\sigma}$ は CC の標準偏差である。

これら推定値を条件付死亡率関数や条件付ハザード関数の未知パラメータに代入することにより、新たな患者への動的予測が可能となる。例えば、条件付死亡率関数は次のように計算される。

$$\hat{F}(t, t+w | X > t, \mathbf{Z}) = \frac{\int_0^\infty (C_\theta^\alpha[\hat{S}_X(t|u), \hat{S}_D(t|u)] - C_\theta[\hat{S}_X(t|u), \hat{S}_D(t+w|u)]) f_{\hat{\eta}}(u) du}{\int_0^\infty C_\theta[\hat{S}_X(t|u), \hat{S}_D(t|u)] f_{\hat{\eta}}(u) du},$$

$$\begin{aligned} \hat{F}(t, t+w | X = x, \mathbf{Z}) &= \frac{\int_0^\infty (C_\theta^{[1,0]}[\hat{S}_X(x|u), \hat{S}_D(t|u)] - C_\theta^{[1,0]}[\hat{S}_X(x|u), \hat{S}_D(t+w|u)]) \hat{\lambda}_X(x|u) \hat{S}_X(x|u) f_{\hat{\eta}}(u) du}{\int_0^\infty C_\theta^{[1,0]}[\hat{S}_X(x|u), \hat{S}_D(t|u)] \hat{\lambda}_X(x|u) \hat{S}_X(x|u) f_{\hat{\eta}}(u) du}. \end{aligned}$$

6.4 卵巣癌患者への適用例

Emura et al. (2018) は Ganzfried et al. (2013) の 912 人の卵巣癌患者をもとに、次のような複合共変量を得た。

- 増悪に関連する 158 個の遺伝子 (P 値 < 0.001) ;
 $\text{CC}_1 = (0.249 \times \text{CXCL12}) + (0.235 \times \text{TIMP2}) + (0.222 \times \text{PDPN}) + \dots + (-0.152 \times \text{MMP12}).$
- 死亡に関連する 128 個の遺伝子 (P 値 < 0.001) ;

$$CC_2 = (0.237 \times NCOA3) + (0.223 \times TEAD1) + (0.263 \times YWHAB) + \dots + (-0.157 \times KCNH4).$$

ここで遺伝子発現量は遺伝子名 (*CXCL12* など) で表し, 全患者で平均 0, 標準偏差 1 となるよう標準化してある. CC_1 の平均は $\hat{\mu}_1 = 0.338$, 標準偏差は $\hat{\sigma}_1 = 10.468$ であり, CC_2 の平均は $\hat{\mu}_2 = 0.222$, 標準偏差は $\hat{\sigma}_2 = 7.894$ であった. P 値が 0.001 より小さい遺伝子を選択する基準はマイクロアレイ解析でしばしば利用される (Matsui et al., 2012).

データを Joint frailty-copula モデルに当てはめた結果,

$$\Pr(X > x, D > y | u) = C_{\hat{\theta}}[\hat{S}_X(x|u), \hat{S}_D(y|u)] = [\hat{S}_X(x|u)^{-\hat{\theta}} + \hat{S}_D(y|u)^{-\hat{\theta}} - 1]^{-1/\hat{\theta}},$$

となり, ここで $\hat{\theta} = 1.9$ (ケンドール相関 $\hat{\tau} = 0.49$),

$$\hat{S}_X(x|u) = \exp \left\{ -u \hat{R}_0(t) \exp \left(\hat{\gamma}_1 \frac{CC_1 - \hat{\mu}_1}{\hat{\sigma}_1} \right) \right\},$$

$$\hat{S}_D(y|u) = \exp \left\{ -\hat{\Lambda}_0(t) \exp \left(\hat{\beta}_2 Z_2 + \hat{\gamma}_2 \frac{CC_2 - \hat{\mu}_2}{\hat{\sigma}_2} \right) \right\},$$

となった. ここで Z_2 は (0 または 1) をとる術後腫瘍径 (≤ 1 cm or > 1 cm) である. 腫瘍径が無増悪期間に与える影響は有意でなかった (P 値 > 0.05) ので $\hat{S}_X(x|u)$ のモデルから除いてある. 回帰推定値は $\hat{\gamma}_1 = 0.39$, $\hat{\beta}_2 = 0.16$, $\hat{\gamma}_2 = 0.44$ であった. 基準ハザード関数の推定値は, $M \cdot(t)$ を M スプライン基底関数 (Emura et al., 2017, 2019b) とすると,

$$\hat{r}_0(t) = d\hat{R}_0(t)/dt = 0.85 \times M_1(t) + 2.14 \times M_2(t) + 0 \times M_3(t) + 0.07 \times M_4(t) + 0 \times M_5(t),$$

$$\hat{\lambda}_0(t) = d\hat{\Lambda}_0(t)/dt = 0.17 \times M_1(t) + 1.05 \times M_2(t) + 1.24 \times M_3(t) + 0.27 \times M_4(t) + 0 \times M_5(t),$$

$t \in [0, 6420]$ であり, 6420 (日) は全患者の最長追跡日数である. 異質性分散の推定値は $\text{Var}(u) = \hat{\eta} = 0.04$ であった. ステージなどの他の共変量もいくつか測定されているが, これらを追加してもモデルの当てはまりは改善しなかった.

次に, 以下のような仮想的な二人の卵巣癌患者を考え, どのように予測のプロセスが実行されるのかを説明する.

- 患者 1; 高リスク遺伝子 ($CC_1 = 10$, $CC_2 = 10$), 腫瘍径 > 1 cm ($Z_2 = 1$).
- 患者 2; 低リスク遺伝子 ($CC_1 = -10$, $CC_2 = -10$), 腫瘍径 ≤ 1 cm ($Z_2 = 0$).

例えば患者 2 に対して条件付死亡率関数 $\hat{F}(t, t+w | X = x, \mathbf{Z})$ を $t = 1000$, $t+w < 6420$, $x = 600$ の設定で計算するときは, R パッケージ *joint.Cox* を利用し, 「付録」にあるコードを実行すれば良い. CC を計算するには, 患者の 158 個の遺伝子を調べ, 重み付け和を計算する必要がある. ここではそのプロセスを省略した.

図 6 は二人の患者の条件付死亡率の推移を表す. $t = 500$ 日目を起点に予測を始めると, 患者 1 は患者 2 よりも高い死亡率を示す (図 6 の左パネル). これは患者 1 が患者 2 より高リスク遺伝子を持ち, かつ腫瘍径が大きいことに由来する. 図 6 の左パネルにおいては予測時刻 $t = 500$ 日目において両患者ともに増悪がないと仮定した. 時刻を進め, $t = 1000$ 日目を起点とし, 患者 2 だけに増悪があった ($x = 600$ 日目) と仮定すると, 患者 2 の死亡率が急激に上昇し, 患者 1 の死亡率を上回る (図 6 の右パネル). このことは, 増悪の死亡に与える影響が, 遺伝子と腫瘍径の死亡に与える影響よりも強いことを示している.

図 7 は二人の卵巣癌患者の条件付ハザード関数 $\hat{\lambda}(t | X = t, \mathbf{Z}, u)$ と $\hat{\lambda}(t | X > t, \mathbf{Z}, u)$ を $u = 1$ で計算したものを示している. 患者 1 は危険因子 (高リスク遺伝子と腫瘍径 > 1 cm) を持つため, 患者 2 よりも高いハザード値を示していることがわかる. 他方, 患者 2 は時間とともに自然に減少するハザード値を示している. 同一の患者で, 増悪した場合としなかった場合を比較

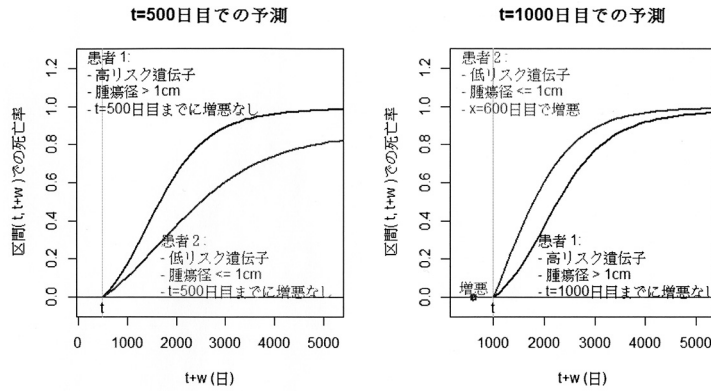


図 6. 二人の卵巣癌患者の条件付死亡率の推移.

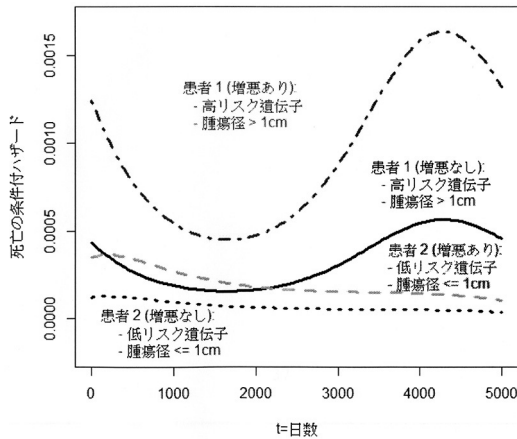


図 7. 二人の卵巣癌患者の条件付ハザードの推移.

するときに、これら 2 つのハザード間に比例関係が存在し、

$$\hat{\lambda}(t|X = t, \mathbf{Z}, u) = (1 + \hat{\theta})\hat{\lambda}(t|X > t, \mathbf{Z}, u) = 2.9\hat{\lambda}(t|X > t, \mathbf{Z}, u)$$

と表せることに注意する。これはクレイトン型コピュラの仮定から導かれるものである。したがって、増悪した場合は、増悪しなかった場合と比較してハザードが約 3 倍になることがわかる。逆に増悪しなければ、患者 2 の死亡リスクはかなり低下することがわかる。また、増悪しなかった場合でも、患者 1 のように危険因子(高リスク遺伝子と腫瘍径 > 1 cm)を持つ患者は、将来的に死亡リスクが上昇する可能性があることがわかる。

7. 今後の課題

コピュラの型は無数にあるため、研究者はその目的に応じて最も適したコピュラを選択することができる。しかし実際には、生存期間において変数間の相関をモデリングするためにはクレイトン型のコピュラが最も広く利用されてきた。本稿で示したようにその数学的に簡潔な関数形、ケンドール相関やオッズ比の特徴からその利便性・有用性は理解されるであろう。また

相関パラメータが相対リスクとして解釈できる医学上の重要性も強調した。このような理由から、本稿でも主にクレイトン型コピュラを用いて解説してきたが、これは母集団がクレイトン型コピュラに従うことを必ずしも意味していない。母集団の相関構造をよりよく近似するコピュラを用いたい場合は、クレイトン型、ゲンベル型、フランク型など特徴が異なるいくつかのコピュラ間でモデル選択を行うことが望ましい。しかしながら、尤度や予測モデルの計算が複雑になることから、現時点では 5 節、6 節の手法は R パッケージ *joint.Cox* においてクレイトン型のみ利用可能となっている。

6 節においては、患者の生存期間の予測にコピュラをどのように利用するかを、動的予測法の枠組みで解説した。患者の生存期間を正確に予測することは、個別医療の観点からも重要な課題であるため、現場の臨床医や医師が利用できる状態にすることが重要である。1 つの取り組みとして、臨床医向けに動的予測の利用法ウェブ・アプリケーションを作成する手法を解説した論文を執筆している (Emura et al., 2020)。

国内における上記動的予測の手法は、間質性肺炎合併肺癌患者へ応用された例がある (高見澤, 2018)。またこの動的予測の手法をワイブル分布で行うことにより、長期予測法を実現する方法も提案されている (篠原, 2019)。

謝 辞

本特集号の企画者である吉羽要直先生と加藤昇吾先生には、執筆の機会を頂いたことに心より感謝申し上げます。また改稿に当たり、有益なコメントをいただいた査読者と編集委員の方々に感謝申し上げます。最後に、図 1 をご提供頂いた篠原紗也華さんに感謝申し上げます。

参 考 文 献

- Alonso, A. and Molenberghs, G. (2008). Evaluating time to cancer recurrence as a surrogate marker for survival from an information theory perspective, *Statistical Methods in Medical Research*, **17**(5), 497-504.
- Barthel, N., Geerdens, C., Killiches, M., Janssen, P. and Czado, C. (2018). Vine copula based likelihood estimation of dependence patterns in multivariate event time data, *Computational Statistics & Data Analysis*, **117**, 109-127.
- Burzykowski, T., Molenberghs, G., Buyse, M., Geys, H. and Renard, D. (2001). Validation of surrogate end points in multiple randomized clinical trials with failure time end points, *Applied Statistics*, **50**(4), 405-422.
- Burzykowski, T., Molenberghs, G. and Buyse, M. (eds.) (2005). *The Evaluation of Surrogate End-points*, Springer, New York.
- Buyse, M., Burzykowski, T., Michiels, S. and Carroll, K. (2008). Individual-and trial-level surrogacy in colorectal cancer, *Statistical Methods in Medical Research*, **17**, 467-475.
- Chaieb, L., Rivest, L. P. and Abdous, B. (2006). Estimating survival under a dependent truncation, *Biometrika*, **93**(3), 655-669.
- Chen, Y. H. (2012). Maximum likelihood analysis of semicompeting risks data with semiparametric regression models, *Lifetime Data Analysis*, **18**, 36-57.
- Choi, J., Oh, I., Seo, S. and Ahn, J. (2018). G2Vec: Distributed gene representations for identification of cancer prognostic genes, *Scientific Reports*, **8**(1), 13729.
- Clayton, D. G. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence, *Biometrika*, **65**(1), 141-151.

- Cox, D. R. (1972). Regression models and life-tables (with discussion), *Journal of the Royal Statistical Society, Series B*, **34**, 187-220.
- Day, R., Bryant, J. and Lefkopoulou, M. (1997). Adaptation of bivariate frailty models for prediction, with application to biological markers as prognostic indicators, *Biometrika*, **84**(1), 45-56.
- de Uña-Álvarez, J. and Veraverbeke, N. (2013). Generalized copula-graphic estimator, *TEST*, **22**(2), 343-360.
- de Uña-Álvarez, J. and Veraverbeke, N. (2017). Copula-graphic estimation with left-truncated and right-censored data, *Statistics*, **51**(2), 387-403.
- Duchateau, L., Janssen, P., Lindsey, P., Legrand, C., Nguti, R., et al. (2002). The shared frailty model and the power for heterogeneity tests in multicenter trials, *Computational Statistics & Data Analysis*, **40**(3), 603-620.
- Eisenhauer, E. A., Therasse, P., Bogaerts, J., Schwartz, L. H., Sargent, D., et al. (2009). New response evaluation criteria in solid tumours: Revised RECIST guideline (version 1.1), *European Journal of Cancer*, **45**(2), 228-247.
- Emura, T. (2019). joint.Cox: Joint frailty-copula models for tumour progression and death in meta-analysis, *CRAN*.
- Emura, T. and Chen, Y. H. (2016). Gene selection for survival data under dependent censoring, a copula-based approach, *Statistical Methods in Medical Research*, **25**(6), 2840-2857.
- Emura, T. and Chen, Y. H. (2018). *Analysis of Survival Data with Dependent Censoring, Copula-based Approaches*, JSS Research Series in Statistics, Springer, Singapore.
- Emura, T. and Michimae, H. (2017). A copula-based inference to piecewise exponential models under dependent censoring, with application to time to metamorphosis of salamander larvae, *Environmental and Ecological Statistics*, **24**(1), 151-173.
- Emura, T. and Murotani, K. (2015). An algorithm for estimating survival under a copula-based dependent truncation model, *TEST*, **24**(4), 734-751.
- Emura, T. and Pan, C. H. (2020). Parametric likelihood inference and goodness-of-fit for dependently left-truncated data, a copula-based approach, *Statistical Papers*, **61**(1), 479-501.
- Emura, T. and Wang, W. (2010). Testing quasi-independence for truncation data, *Journal of Multivariate Analysis*, **101**, 223-239.
- Emura, T. and Wang, W. (2012). Nonparametric maximum likelihood estimation for dependent truncation data based on copulas, *Journal of Multivariate Analysis*, **110**, 171-188.
- Emura, T., Lin, C. W. and Wang, W. (2010). A goodness-of-fit test for Archimedean copula models in the presence of right censoring, *Computational Statistics & Data Analysis*, **54**, 3033-3043.
- Emura, T., Wang, W. and Hung, H. N. (2011). Semi-parametric inference for copula models for dependently truncated data, *Statistica Sinica*, **21**, 349-367.
- Emura T., Chen, Y. H. and Chen, H. Y. (2012). Survival prediction based on compound covariate under Cox proportional hazard models, *PLoS One*, **7**(10), e47627, DOI:10.1371/journal.pone.0047627.
- Emura, T., Nakatochi, M., Murotani, K. and Rondeau, V. (2017). A joint frailty-copula model between tumour progression and death for meta-analysis, *Statistical Methods in Medical Research*, **26**(6), 2649-2666.
- Emura, T., Nakatochi, M., Matsui, S. and Michimae, H. and Rondeau, V. (2018). Personalized dynamic prediction of death according to tumour progression and high-dimensional genetic factors: Meta-analysis with a joint model, *Statistical Methods in Medical Research*, **27**(9), 2842-2858.
- Emura, T., Matsui, S. and Chen, H. Y. (2019a). compound.Cox: Univariate feature selection and compound covariate for predicting survival, *Computer Methods and Programs in Biomedicine*, **168**, 21-37.
- Emura, T., Matsui, S. and Rondeau, V. (2019b). *Survival Analysis with Correlated Endpoints, Joint*

- Frailty-Copula Models*, JSS Research Series in Statistics, Springer, Singapore.
- Emura, T., Shih, J. H., Ha, I. D. and Wilke, R. A. (2019c). Comparison of the marginal hazard model and the sub-distribution hazard model for competing risks under an assumed copula, *Statistical Methods in Medical Research*, DOI: 10.1177/0962280219892295.
- Emura, T., Michimae, H. and Matsui, S. (2020). Building web applications for personalized risk prediction using R packages joint.Cox and Shiny (submitted).
- Fine, J. P., Jiang, H. and Chappell, R. (2001). On semi-competing risks data, *Biometrika*, **88**, 907-920.
- Ganzfried, B. F., Riester, M., Haibe-Kains, B., Risch, T., Tyekucheva, S., et al. (2013). Curated ovarian data: Clinically annotated data for the ovarian cancer transcriptome, *Database*, Article ID bat013, DOI:10.1093/database/bat013.
- Gómez, I., Ribelles, N., Franco, L., Alba, E. and Jerez, J. M. (2016). Supervised discretization can discover risk groups in cancer survival analysis, *Computer Methods and Programs in Biomedicine*, **136**, 11-19.
- Guinney, J., Wang, T., Laajala, T. D., Winner, K. K., Bare, J., et al. (2017). Prediction of overall survival for patients with metastatic castration-resistant prostate cancer: Development of a prognostic model through a crowdsourced challenge with open clinical trial data, *The Lancet Oncology*, **18**(1), 132-142.
- Gumbel, E. J. (1960). Distributions de valeurs extremes en plusieurs dimensions, *Publications de l'Institut de statistique de l'Université de Paris*, **9**, 171-173.
- Hayes, D. F., Markus, H. S., Leslie, R. D. and Topol, E. J. (2014). Personalized medicine: Risk prediction, targeted therapies and mobile health technology, *BMC Medicine*, **12**(1), 37.
- Hirsch, K. and Wienke, A. (2012). Software for semiparametric shared gamma and log-normal frailty models: An overview, *Computer Methods and Programs in Biomedicine*, **107**(3), 582-597.
- Kim, M., Oh, I. and Ahn, J. (2018). An improved method for prediction of cancer prognosis by network learning, *Genes*, **9**, 478.
- Klein, J. P., Van Houwelingen, H. C., Ibrahim, J. G. and Scheike, T. H. (eds.) (2014). *Handbook of Survival Analysis*, CRC Press, New York.
- Le Tourneau, C., Michiels, S., Gan, H. K. and Siu, L. L. (2009). Reporting of time-to-event end points and tracking of failures in randomized trials of radiotherapy with or without any concomitant anticancer agent for locally advanced head and neck cancer, *Journal of Clinical Oncology*, **27**(35), 5965-5971.
- Li, Z., Chinchilli, V. M. and Wang, M. (2019). A Bayesian joint model of recurrent events and a terminal event, *Biometrical Journal*, **61**(1), 187-202.
- Li, Z., Chinchilli, V. M. and Wang, M. (2020). A time-varying Bayesian joint hierarchical copula model for analysing recurrent events and a terminal event: An application to the Cardiovascular Health Study, *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **69**(1), 151-166.
- Ling, M., Hu, T. and Sun, J. (2016). Cox regression analysis of dependent interval-censored failure time data, *Computational Statistics & Data Analysis*, **103**, 79-90.
- Lo, S. M. and Wilke, R. A. (2010). A copula model for dependent competing risks, *Journal of the Royal Statistical Society, Series C*, **59**(2), 359-376.
- Matsui, S. (2006). Predicting survival outcomes using subsets of significant genes in prognostic marker studies with microarrays, *BMC Bioinformatics*, **7**, 156.
- Matsui, S., Simon, R., Qu, P., Shaughnessy, J. D., Barlogie, B. and Crowley, J. (2012). Developing and validating continuous genomic signatures in randomized clinical trials for predictive medicine, *Clinical Cancer Research*, **18**(21), 6065-6073.
- Michiels, S., Le Maître, A., Buyse, M., Burzykowski, T., Maillard, E., et al. (2009). Surrogate end-

- points for overall survival in locally advanced head and neck cancer: Meta-analyses of individual patient data, *Lancet Oncology*, **10**(4), 341-350.
- Michiels, S., Pugliano, L., Marguet, S., Grun, D., Barinoff, J., et al. (2016). Progression-free survival as surrogate end point for overall survival in clinical trials of HER2-targeted agents in HER2-positive metastatic breast cancer, *Annals of Oncology*, **27**(6), 1029-1034.
- Moradian, H., Larocque, D. and Bellavance, F. (2019). Survival forests for data with dependent censoring, *Statistical Methods in Medical Research*, **28**(2), 445-461.
- Nelsen, R. B. (2006). *An Introduction to Copulas*, 2nd ed., Springer, New York.
- Oakes, D. (1989). Bivariate survival models induced by frailties, *Journal of the American Statistical Association*, **84**, 487-493.
- Oba, K., Paoletti, X., Alberts, S., Bang, Y. J., Benedetti, J., et al. (2013). Disease-free survival as a surrogate for overall survival in adjuvant trials of gastric cancer: A meta-analysis, *Journal of the National Cancer Institute*, **105**(21), 1600-1607.
- Pazdur, R. (2008). Endpoints for assessing drug activity in clinical trials, *The Oncologist*, **13**, 19-21.
- Peng, M. (2019). *Analysis of Complex Survival Data Subject to Semi-competing Risks*, Doctoral Thesis, Nanyang Technological University, Singapore.
- Peng, M. and Xiang, L. (2019). Joint regression analysis for survival data in the presence of two sets of semi-competing risks, *Biometrical Journal*, DOI: 10.1002/bimj.201800137.
- Peng, M., Xiang, L. and Wang, S. (2018). Semiparametric regression analysis of clustered survival data with semicompeting risks, *Computational Statistics & Data Analysis*, **124**, 53-70.
- Piedbois, P. and Crosswell, M. J. (2008). Surrogate endpoints for overall survival in advanced colorectal cancer: A clinician's perspective, *Statistical Methods in Medical Research*, **17**(5), 519-527.
- Proust-Lima, C., Séne, M., Taylor, J. M. and Jacqmin-Gadda, H. (2014). Joint latent class models for longitudinal and time-to-event data: A review, *Statistical Methods in Medical Research*, **23**(1), 74-90.
- Riley, R. D. (2009). Multivariate meta-analysis: The effect of ignoring within-study correlation, *Journal of the Royal Statistical Society, Series A*, **172**, 789-811.
- Rondeau, V. and Gonzalez, J. R. (2005). frailtypack: A computer program for the analysis of correlated failure time data using penalized likelihood estimation, *Computer Methods and Programs in Biomedicine*, **80**(2), 154-164.
- Rondeau, V., Pignon, J. P. and Michiels, S. (2015). A joint model for dependence between clustered times to tumour progression and deaths: A meta-analysis of chemotherapy in head and neck cancer, *Statistical Methods in Medical Research*, **24**(6), 711-729.
- Rondeau, V., Mauguen, A., Laurent, A., Berr, C. and Helmer, C. (2017). Dynamic prediction models for clustered and interval-censored outcomes: Investigating the intra-couple correlation in the risk of dementia, *Statistical Methods in Medical Research*, **26**(5), 2168-2183.
- Rotolo, F., Legrand, C. and Van Keilegom, I. (2013). A simulation procedure based on copulas to generate clustered multi-state survival data, *Computer Methods and Programs in Biomedicine*, **109**(3), 305-312.
- Rotolo, F., Paoletti, X. and Michiels, S. (2018). surroSurv: An R package for the evaluation of failure time surrogate endpoints in individual patient data meta-analyses of randomized clinical trials, *Computer Methods and Programs in Biomedicine*, **155**, 189-198.
- Schleiden, S., Klingler, C., Bertram, T., Rogowski, W. H. and Marckmann, G. (2013). What is personalized medicine: Sharpening a vague term based on a systematic literature review, *BMC Medical Ethics*, **14**(1), 55.
- Schneider, S., Demarqui, F. N., Colosimo, E. A. and Mayrink, V. D. (2020). An approach to model clustered survival data with dependent censoring, *Biometrical Journal*, **62**(1), 157-174.

- Sène, M., Taylor, J. M., Dignam, J. J., Jacqmin-Gadda, H. and Proust-Lima, C. (2016). Individualized dynamic prediction of prostate cancer recurrence with and without the initiation of a second treatment: Development and validation, *Statistical Methods in Medical Research*, **25**(6), 2972-2991.
- Sherrill, B., Amonkar, M., Wu, Y., Hirst, C., Stein, S, et al. (2008). Relationship between effects on time-to-disease progression and overall survival in studies of metastatic breast cancer, *British Journal of Cancer*, **99**, 1542-1548.
- Shi, Q. and Sargent, D. J. (2009). Meta-analysis for the evaluation of surrogate endpoints in cancer clinical trials, *International Journal of Clinical Oncology*, **14**(2), 102-111.
- Shih, J. H. and Emura, T. (2018). Likelihood-based inference for bivariate latent failure time models with competing risks under the generalized FGM copula, *Computational Statistics*, **33**(3), 1293-1323.
- Shih, J. H. and Emura, T. (2020). Penalized Cox regression with a five-parameter spline model (submitted).
- Shih, J. H., Lee, W., Sun, L. H. and Emura, T. (2019). Fitting competing risks data to bivariate Pareto models, *Communications in Statistics — Theory and Methods*, **48**(5), 1193-1220.
- 篠原紗也華 (2019). Dynamic prediction using joint frailty-copula model: Long-term prediction by Weibull distribution, 薬学学士論文, 北里大学薬学部薬学科臨床統計学教室.
- Shukla, N., Hagenbuchner, M., Win, K. T. and Yang, J. (2018). Breast cancer data analysis for survivability studies and prediction, *Computer Methods and Programs in Biomedicine*, **155**, 199-208.
- Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges, *Publications de l'Institut de Statistique de L'Université de Paris*, **8**, 229-231.
- Sofeu, C., Emura, T. and Rondeau, V. (2019). One-step validation method for surrogate endpoints in multiple randomized cancer clinical trials with failure-time endpoints, *Statistics in Medicine*, **38**(16), 2928-2942.
- Sofeu, C., Emura, T. and Rondeau, V. (2020). A joint frailty-copula model for meta-analytic validation of failure time surrogate endpoints in clinical trials (in revision).
- Suresh, K. (2018). *Dynamic Methods for the Prediction of Survival Outcomes using Longitudinal Biomarkers*, Ph.D Dissertation, University of Michigan Library.
- Suresh, K., Taylor, J. M., Tsodikov, A. (2019). A Gaussian copula approach for dynamic prediction of survival with a longitudinal biomarker, *Biostatistics*, DOI: 10.1093/biostatistics/kxz049.
- 高見澤勇 (2018). 化学療法が対象となる間質性肺炎合併肺がん患者の死亡率の動的予測: Joint frailty-copula モデルの臨床データへの適用, *RIMS Kôkyûroku*, No.2091, 31-37.
- Thompson, R. H., Leibovich, B. C., Lohse, C. M., Cheville, J. C., Zincke, H., et al. (2007). Dynamic outcome prediction in patients with clear cell renal cell carcinoma treated with radical nephrectomy: The D-SSIGN score, *The Journal of Urology*, **177**(2), 477-480.
- Yoshihara, K., Tajima, A., Yahata, T., Kodama, S., Fujiwara, H., et al. (2010). Gene expression profile for predicting survival in advanced-stage serous ovarian cancer across two independent datasets, *PLoS One*, **5**(3), e9615.
- van Houwelingen, H. C. and Putter, H. (2011). *Dynamic Prediction in Clinical Survival Analysis*, CRC Press, New York.
- Waldron, L., Haibe-Kains, B., Culhane, A. C., Riestler, M., Ding, J., et al. (2014). Comparative meta-analysis of prognostic gene signatures for late-stage ovarian cancer, *Journal of the National Cancer Institute*, **106**(5), dju049.
- Wang, W. (2003). Estimating the association parameter for copula models under dependent censoring, *Journal of the Royal Statistical Society, Series B*, **65**(1), 257-273.

- Wang, Y. C., Emura, T., Fan, T. H., Lo, S. and Wilke, R. (2020). Likelihood-based inference for a frailty-copula model based on competing risks failure time data, *Quality and Reliability Engineering International*, DOI: 10.1002/qre.2650.
- Wu, B. H., Emura, T. and Michimae, H. (2020). Meta-analysis of individual patient data with semi-competing risks under the Weibull joint frailty-copula model, *Computational Statistics*, DOI: 10.1007/s00180-020-00977-1.
- Zhou, Y., Lu, Z., Shi, Y. and Cheng, K. (2019). The copula-based method for statistical analysis of step-stress accelerated life test with dependent competing failure modes, *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, **233**(3), 401-418.

Survival Analysis Using Copulas

—Meta-analysis with Correlated Endpoints—

Takeshi Emura¹ and Hirofumi Michimae²

¹Department of Information Management, Chang Gung University

²School of Pharmacy, Kitasato University

With the rapid development this decade of open databases for biomedical researchers, we have gained access to complex and yet accurate patient-level information. For cancer patients in particular, these databases record information on individual patients, including overall survival time, time-to-tumor progression, tumor size, and gene expressions. In addition, there has been significant development of meta-analytical methodologies for analyzing data from different sources. To fully utilize such complex survival data, it is insufficient to apply the classical tools used in survival data analysis, such as Cox regression. In this paper, we review copula-based statistical methods for analyzing two survival time variables, namely overall survival time and time-to-tumor progression. We also review the joint frailty-copula model for individual patient data (IPD) meta-analysis methods, which account for the heterogeneity of patients from different sources. Regarding the construction of an appropriate likelihood function for a given dataset, we explain the importance of acknowledging the semi-competing risks relationship between overall survival time and time-to-tumor progression. Finally, we introduce a dynamic prediction method for overall survival time according to gene expressions and tumor progression, which may contribute to the development of personalized medicine.