

博士論文

繰り返しイベントデータに対する動的予測

– 動的擬似値を用いたランドマークアプローチ –

横 田 勲

目次

要旨	7
1 序文	8
1.1 経時データの考慮	12
1.2 ランドマークモデル	15
1.3 繰り返しイベントデータへの拡張	17
1.4 本研究の目的	22
2 方法	23
2.1 準備	23
2.1.1 記法	23
2.1.2 動的予測とランドマークモデル	24
2.2 動的擬似値	26
2.2.1 終端イベントを検討しない場合	29
2.2.2 終端イベントを検討する場合	33
2.3 動的擬似値を用いた一般化線形モデルによる動的予測確率の推定	34
2.3.1 ランドマーク時点ごとの予測確率の推定	34

2.3.2	ランドマークスーパーモデルによる、時点に対するスムージング	37
2.3.3	解析プログラムの作成	39
3	シミュレーション実験	41
3.1	終端イベントを検討しない場合	41
3.2	終端イベントを検討する場合	46
4	実データへの適用：大腸がん肝転移切除例の追跡データ	50
4.1	背景	51
4.2	切除不能再発発生による死亡に関する動的予測	54
4.3	切除不能再発が起こっていない下での再発に関する動的予測	63
5	考察	80
6	結論	91
7	謝辞	92
	文献	93

表目次

1	終端イベントを考慮しない場合の各シナリオにおけるパラメータ値	42
2	終端イベントを考慮しない場合における各シナリオ発生データの要約	44
3	終端イベントを考慮しない場合におけるバイアスと RMSE に関するシミュレーション結果	45
4	終端イベントを考慮した場合の各シナリオにおけるパラメータ値	46
5	終端イベントを考慮する場合における各シナリオ発生データの要約	48
6	終端イベントを考慮しない場合におけるバイアスと RMSE に関するシミュレーション結果	49
7	大腸がん肝転移切除例の追跡データ：ベースライン時における背景情報	52
8	大腸がん肝転移切除例の追跡データ：再発回数集計結果	53
9	死亡を動的予測するための提案したスーパーモデルの推定結果	59
10	前回切除時腫瘍個数によって再発を動的予測するためのスーパーモデルの推定結果	68

11	再発回数によって再発を動的予測するためのスーパーモデルの 推定結果	71
12	腫瘍個数と腫瘍径でサブグループ化した再発回数によって再発 を動的予測するためのスーパーモデルの推定結果	77

図目次

1	多状態モデルで表す競合リスクの例	18
2	終端イベントを検討しない場合の遷移図	30
3	illness-death モデルの遷移図	33
4	終端イベントを検討する場合の遷移図	33
5	OS,TSF,RFS についてイベント未発生確率に関する Kaplan-Meier 推定値	56
6	1,2,3 年ランドマーク時における OS,TSF,RFS についてイベント 未発生確率に関する Kaplan-Meier 推定値	57
7	RFS,TSF イベント有無組み合わせ別の 3 年死亡確率に関する動 的予測	58
8	RFS 有無別の 3 年死亡確率に関する動的予測	61
9	TSF 有無別の 3 年死亡確率に関する動的予測	61
10	RFS,TSF 有無別の c-index を用いた OS 予測性能評価	62
11	大腸がん肝転移切除例の追跡データにおける切除不能再発と死 亡が起こっていない条件下での 3 年イベント確率に関する積み 上げ面グラフ	64

12	前回切除時腫瘍個数別の3年イベント確率に関する積み上げグラフ	67
13	再発回数別の3年イベント確率に関する積み上げ面グラフ . . .	70
14	腫瘍個数と腫瘍径でサブグループ化した再発回数別の3年イベント確率に関する積み上げ面グラフ(死亡を打ち切りとして扱った場合)	73
15	腫瘍個数と腫瘍径でサブグループ化した再発回数別の3年イベント確率に関する積み上げ面グラフ(死亡を終端イベントとして扱った場合)	75

要旨

生存時間をアウトカムとした臨床研究では、追跡開始時のみならず追跡途中に得られる臨床経時データを利用した予後予測にしばしば興味がある。そのアプローチのひとつに w 年生存割合の考え方を拡張したランドマーク解析による動的予測があり、本研究では繰り返しイベントデータに適用可能な解析法を提案した。提案法がバイアスの少ない推定を行えることをシミュレーション実験にて確認した。さらに、大腸がん肝転移切除例の追跡データに提案法を適用し、切除不能再発が死亡に対する強い予測因子であること、前回切除時の腫瘍個数やそれまでの再発回数がその後の再発や死亡の予測に有用であることを示した。

1 序文

臨床研究では興味のあるイベントが発生するまでの時間、すなわち time-to-event をアウトカムとして、どのような対象者がイベントを起こしやすいかを予測することに関心は高く、その予測性能を高めるために追跡中に観察された中間イベントや対象者の健康状態を表すバイオマーカー等の臨床情報の利用が検討される。これらの臨床情報は追跡開始時のみならず経時的に繰り返し観察されるため、「経時データ」と呼ばれる。経時データのひとつである中間イベントの例として、大腸がん肝転移切除患者における全生存時間に興味がある研究において、追跡途中で発生する再発が挙げられる。このうち、初回再発までの時間である RFS(Recurrence-Free Survival) よりも、TSF(Time to Surgical Failure) のほうが代替エンドポイントとしてふさわしいと主張する研究 [1] がある。同じ個人で再発と再切除を複数回繰り返す大腸がん肝転移切除例では、再切除が行えないような再発が起こるまでの時間である TSF のほうが生存時間の予測性能がよい可能性が示唆されている。経時データの他の種類であるバイオマーカーの例として、前立腺がん再発に興味がある研究における、PSA(Prostate Specific Antigen) 値は、再発を予測する腫瘍マーカーとして繰り返し観察される [2]。

イベント発生の予測では、ほとんどの対象者がイベントを起こすような極端な場合をのぞけば、イベント発生時点そのものを予測することは難しい。むしろ、それぞれの時点におけるイベント発生のしやすさを検討するほうが現実的である。

イベント発生のしやすさは、臨床研究における time-to-event データの場合、「ハザード」を用いた議論が通常行われる [3-5]。これは、追跡途中の脱落や追跡終了時における生存例のような右側打ち切りに対処するためである。ハザードは、それまでイベントを起こしていない対象者が直後の微小時間にてイベントを起こす条件付き確率を表す。ハザード関数 $\lambda(\cdot)$ が定まれば、それまでにイベントを起こしていない確率を表す生存関数 $S(\cdot)$ も自動的に定まる。Nelson-Aalen 推定量 [6, 7] は累積ハザード関数のノンパラメトリック最尤推定量 [8] として知られ、Cox 比例ハザードモデル (以下、「Cox モデル」とよぶ) [9] はハザード関数に対する回帰モデルの代表例であり、推定されるハザード比は治療効果や予後因子、予測因子の影響の程度を評価するために頻用される。Nelson-Aalen 推定量の応用である Breslow 推定量 [9] によって Cox モデルを基にベースラインハザード関数を推定することができるため、個人ごとにハザードの予測値を得ることができる。

しかしながら、ハザード比のような相対指標であれば、対象者間でのイベントの起こりやすさを表現できるものの、ハザードの絶対値は解釈上の分かりにくさ、および生存関数との直感的理解における乖離が伴う。また、ハザードはその定義から 0 から無限大までの値をとり、時点のスケールにも依存する指標である。一方で、ある期間を設定することで、ハザードからイベント発生確率を得ることができる。確率は 0 から 1 をとり、絶対値の解釈が容易な指標であり、現実の臨床研究においてもハザード比だけでなく、“w 年生存割合” が利用されてい

る。このような確率による表現が、イベント発生予測の結果を活用することを考えたときに特に有用と思われる。例えば、がん患者の中でステージの良好な対象者がそうでない対象者に比べ、死亡ハザード比が0.6倍であったと想定する。生存時間分布に指数分布を仮定すれば、ステージの良好な対象者の3年生存割合が40%であったとすれば、ステージの悪い対象者のそれは21%であることを意味し、ステージの悪い対象者には副作用リスクは大きくともよりよい効果が望まれる治療を選択すべきという判断が下されるかもしれない。しかし、上記と同じハザード比であったとしても、ステージの良好な対象者の3年生存割合が98%であった場合、ステージの悪い対象者のそれは97%となり、その差に大きな違いはなく、ステージ間の予後に大きな違いはないという結論に至るかもしれない。

w 年生存割合は通常、追跡開始から w 年経過後の生存確率を表す。ここで個人予測を考えると、例えばがんと診断後2年間生存した対象者が、その後さらに3年後まで生存する確率にも関心がもたれる。このようなある時点まで生存した下での条件付きの生存確率、もしくはイベント発生確率を推定することが「動的予測 (dynamic prediction)」と呼ばれ [10, 11]、近年、生存時間解析分野で多くの研究がなされている。また、臨床論文での適用も“条件付き生存 (conditional survival)”という呼び名で近年応用例がいくつかみられ [12–18]、ハザード比よりも結果の解釈が容易であることが指摘されている [19–21]。

本研究では動的予測を、ある時点 s にてイベントを起こしていない条件の下で、

時点 $s + w$ までにイベントを起こす確率を推定することと定義する。すなわち、

$$F(s + w | T > s) = 1 - S(s + w | T > s) = \Pr\{T \in (s, s + w] | T > s\} \quad (1)$$

を推定する。ただし、 T はイベント時点を表す確率変数であり、 $F(\cdot | T > s)$ 、 $S(\cdot | T > s)$ はそれぞれ条件付きのイベント発生・未発生確率である。以降 s を動的予測時点、 w を動的予測幅と呼ぶ。なお、通常用いられる w 年生存割合は、上式で $s = 0$ とした場合である。

上記のような動的予測を行うことは、今後ますます進むといわれる個別化医療において治療方針の意思決定を行う際には有用と思われる。例えば、Taylor ら [2] によって放射線治療のみ行った前立腺がん患者に対して、経時測定される PSA 値を Web フォームで入力することで、予測される PSA の推移と前立腺がん再発リスクの動的予測結果を瞬時に示す Web ツールが既に開発されている。さらに、このツールは、副作用のおそれはあるものの前立腺がん再発に有効とされているホルモン療法を現在日において行った場合の今後の再発リスクも表示可能であり、個人に対して治療を行うかどうかの意思決定に役立つ。

また、動的予測はバイオマーカーや中間イベントなどの最終予後に対する代替性能の評価にも活用できると考えられる。動的予測によって予測確率が得られることから、解釈が容易な確率の関数である c-index(受信者動作特性曲線;ROC の曲線下面積)などの予測性能を表す指標を利用できると思われ、本研究においても代替性能評価のひとつのアプローチとして動的予測の検討を行う。

1.1 経時データの考慮

有用な動的予測のモデルを作成するには、研究開始時点で観察されるベースライン情報だけでなく、経時データ情報を柔軟に取り入れることが重要である。Coxモデルの原著 [9] にもみられるように、経時データを時間依存性共変量として説明変数に組み込む回帰モデルは古くから考えられてきたが、Prentice や Kalbfleisch により、時間依存性共変量を用いるためには、その変数が外生変数の仮定を満たしている必要があることが指摘された [4, 22]。時点 t までに得られる共変量履歴 (経時データ) $Z(t) = \{Z(u); 0 \leq u < t\}$ が外生変数であるとは、 $0 < u \leq t$ を満たす任意の u, t について、

$$\Pr\{u \leq T < u + \Delta u \mid Z(u), T \geq u\} = \Pr\{u \leq T < u + \Delta u \mid Z(t), T \geq u\}$$

すなわち、

$$\Pr\{Z(t) \mid Z(u), T \geq u\} = \Pr\{Z(t) \mid Z(u), T = u\}$$

が成立することをさし、時点 u でのハザードはそれまでの共変量履歴には影響を受けないが、 $[u, u + \Delta u)$ におけるイベント発生と将来の共変量パスが独立であることを意味する。外生変数の例としては、年齢のように時間経過すれば自動的に値が決定するもの、PM2.5 のような大気汚染に関する物質水準のように対象者のイベント情報とは関係なく定まるものなどがある。

一方で上記の外生性を満たさない、共変量パスがイベント発生状態に影響を受

けてしまう変数を内生変数と呼ぶ。臨床研究で測定される対象者の健康状態を表す経時データの多くは、それ自体が測定された時点で対象者がイベント未発生であることが確定してしまうため、内生変数に分類される。一般に、ハザード関数 $\lambda(t)$ とイベント未発生確率 $\Pr(T > t)$ の間には、1対1の関係式が成立するが、経時データが内生変数の場合、その関係が成立しない。すなわち、内生経時データ $Z(t)$ を条件づけた下で、

$$\Pr\{T > t | Z(t)\} \neq \exp\left[-\int_0^t \lambda\{u | Z(u)\} du\right]$$

であるため、イベント未発生確率を推定するにはハザード関数のほか、内生変数の確率過程に関する分布のモデル化が必要となる。これまで提案されてきた内生変数へ対処する動的予測モデルは、多状態モデル、同時モデル、ランドマークモデルの3通りに大別される。

多状態モデル [23–25] は、複数タイプのイベントに同時に興味がある場合に用いられるイベント履歴解析 [26–28] の代表的なモデルであり、状態間の医学的遷移関係を仮定し、それらの遷移ハザードに対するモデル化を行う手法である。その歴史は古く、競合リスクモデル [25, 29–31] や illness-death モデル [32, 33] を中心に臨床論文での適用も多い。多状態モデルを利用した動的予測では、中間イベント発生をひとつの遷移状態として扱い、それぞれの状態から予測したいイベント発生へ遷移するハザードをモデル化することで条件付きの状態確率を推定する。状態確率の推定は Kaplan-Meier 推定量 [34] を多状態モデルに一般化した

Aalen-Johansen 推定量 [35] を用いることが一般的である。

動的予測に多状態モデルを用いる上での欠点はいくつか挙げられている [36]。まず、多状態モデルから得られた結果は、想定した遷移関係の誤特定に対し敏感といわれている。また、多状態モデルでは予測に用いたい経時データを状態によってモデル化する必要があるため、様々な経時データを柔軟に取り込むことはできない。さらに、遷移ハザードに回帰モデルをあてはめ、状態確率を Aalen-Johansen 推定量により推定する場合、その妥当性を保証するには状態遷移に関してマルコフ性の仮定が必要である。これらの理由もあり、後述する同時モデルやランドマークモデルに基づく動的予測方法の開発のほうが盛んと考えられる。

一方、内生変数による動的予測に対するもう一つのアプローチである同時モデル [37,38] は、生存時間データと経時データをリンクする共通のパラメータを想定した shared parameter アプローチを中心に、1980 年代後半から情報のある打ち切りの補正方法 [39,40] として開発が進んだ。1990 年代に入ると、離散的に測定される経時データに対して連続的な測定過程を仮定し、Cox モデルのあてはめを行う手法の提案 [41]、個人ごとの変量効果で条件付けることで経時データと time-to-event アウトカムが独立とする考えや EM アルゴリズムを利用した推定法の提案 [42] といった発展を受けて、2000 年代以降様々な適用場面での応用例が散見されるようになった [43]。最近では、経時データの測定過程に何らかのモデルを仮定し、そのモデルの下での予測生存時間分布を利用して、動的予測へ

の応用が試みられている [2, 11, 44–46]。前述の Taylor ら [2] による前立腺がん再発リスクの動的予測はその一例である。

このような同時モデルは、経時データの測定過程に対する統計モデルが妥当であれば、そこから得られる結論も妥当である一方で、モデルにおいた仮定の破たんに対する脆弱性という欠点がある。一般的に互いに複雑な関連性をもつ経時データに対する統計モデルの構築は難しく、そのための情報量は十分でないのが通常であり、そのモデルの誤特定が生じやすいと考えられる。また、shared parameter を含む同時尤度からパラメータ推定を行うには、モンテカルロ・シミュレーション法のような数値計算が必要となるため、解析プログラム上の困難さや負荷が大きく、その適用が簡便とはいえない。

1.2 ランドマークモデル

動的予測に対するもう一つ別なアプローチであるランドマークモデル [10, 47, 48] は、よりモデルに対する頑健性をもち [49]、バイアスの少ない動的予測を行うために提案された方法であり、がんの臨床試験において腫瘍に対する抗がん剤の奏功程度による生存時間の比較を行う場面で用いられるランドマーク解析 [50, 51] の考えを基にしている。ランドマーク解析の原著は 1983 年の Anderson ら [50] まで遡るが、そこでは多発性骨髄腫に対するエンドキサンの効果を検討する臨床試験において、治療が奏功することは予後によい影響を与えることを示すため

に、治療開始後 56 日目 (ランドマーク時点) において生存している患者に限定して奏功したか否かで 2 群に分け、ランドマーク時点以降の生存時間を比較するという解析方法であった。これまで散見された解析では、追跡期間中に奏功したか否かで群分けし、追跡開始からイベント発生までの全期間を比較していた。しかし、奏功が確認されるまでには当然ある程度の時間を要するため、仮に奏功が後に全く影響を与えない場合であったとしても、奏功が確認された対象者はそうでない対象者に比べ、イベント発生までの時間は長くなる。この *immortality bias* を考慮するためにランドマーク解析は提案された [52]。

vanHouwelingen [10] はこの考えを基に、あるランドマーク時点 s において at risk である条件の下で、時間依存性共変量もランドマーク時点にて得られる情報 $Z(s)$ のみ利用し、ベースライン共変量と同様の扱いにすることで、内生変数を対処する方法を提案した。また、ランドマーク時点ごとに回帰モデルを想定するため、回帰係数が時点によって変化することを許す *time-varying coefficient* モデル [53, 54] のひとつとみなせ、生存時間解析で頻用される Cox モデルの比例ハザード性の仮定を緩めることも可能である。さらに、ランドマークモデルはデータセットの加工のみ行えば、通常の Cox モデルのあてはめが行える解析パッケージを利用して簡便に適用できることも特長である。R では *dynpred* パッケージ [55] にてデータセット加工が行える。また、乳がん術後補助療法のアロマターゼ阻害薬に関する治療効果が経時変化することを表現するために、時間依存性ハ

ザード比をランドマークモデルを用いて解析した臨床論文のような適用例 [56] もみられる。本研究では、ランドマークモデルに基づく動的予測を考える。

1.3 繰り返しイベントデータへの拡張

通常の生存時間解析では、それぞれの対象者について観察されるアウトカムは、単一で繰り返しが無いイベントを想定している。一方、近年方法論の応用が進んでいる分野のひとつに、同じ対象者について発生するいくつかのイベントをまとめて扱うイベント履歴解析があり、代表的なものとして競合リスクイベントと繰り返しイベントが挙げられる。前者は複数種類のイベントのうちいずれか一つのイベントだけが各個人に観察される状況であり、後者は同一のイベントが繰り返し複数回発生する状況である。

競合リスクは、例えばがん患者が、原病による死亡を起こすか、他病による死亡を起こすかのような、個人についていずれか片方のイベントしか観察されない状況をいう。あるイベントの累積発生割合を、他の競合リスクを打ち切りとして扱うことの問題点は古くから指摘されており [57]、現在では図 1 に示すような多状態モデルの特殊な場合として競合リスクは扱われる。回帰分析型アプローチには、状態間の遷移ごとに各イベントに対するハザードを想定し、Cox モデルのような古典的な生存時間データに対する回帰モデルをあてはめる、原因別ハザードモデルが Prentice ら [58] により提案されている。しかし、原因別ハザードモデ

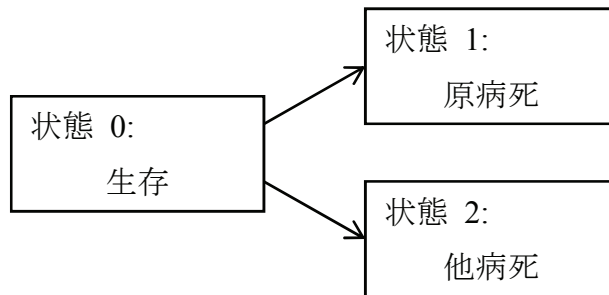


図 1: 多状態モデルで表す競合リスクの例

ルでは、他の競合リスクイベントに対する共変量の効果を見捨てるため、累積発生割合とハザード関数の間に直接的な関係性が必ずしも成立しない [59]。そのため、例えば、Breslow 推定量にて累積発生割合を推定するには、競合リスクは互いに独立という、データから検証不能 [60] かつ極めて非現実的な仮定が必要となる [61]。そこで、別な回帰モデルとして提案された Fine-Gray モデル [62] は、累積発生割合、すなわち、多状態モデルの文脈での状態確率をモデル化するアプローチである。原因別ハザードモデルと Fine-Gray モデルの使い分けは大まかに言えば、ハザード関数と状態確率のどちらの解釈を優先するかがポイントと考えられ、これらの議論は現在も盛んである [61, 63]。それゆえ、解釈上分かりやすい動的予測は魅力的な適用場面であると思われる。

前述の動的予測を目的としたランドマークモデルの競合リスクへの拡張は 2 通りの方法でなされた。一つが Fine-Gray モデルを拡張したもの [64] であり、もう一つが観察値自体ではなくデータにある関数変換を施した擬似値 (pseudo-value) の考え方による動的擬似値 [65] を利用した方法である。後者の擬似値は、統計学

においては Quenouille [66] により提案され、Tukey [67] によってより一般的な状況へ拡張されたジャックナイフ法 [68,69] の枠組みで提案された。ジャックナイフ法による推定量ではバイアスのオーダーを $O(n^{-2})$ まで減らし、頑健な分散推定が可能となることが当初の目的であった。これが生存時間分野では Andersen ら [70] によって、右側打ち切り観察値を不完全データとみなし、全ての観察データを擬似値によって置き換えることで、多状態モデルにおける状態確率の推定に用いる方法として提案された。特に、競合リスクモデルにおいて有用であるとの報告がなされ [71–73]、現在でもいくつかの拡張法が提案されている [74–77]。この擬似値をランドマークモデルのために再定義したものが動的擬似値であり、もともとは競合リスクイベントに対して定義されたものの、この考え方を応用することで繰り返しイベントデータ [78–80] に対してランドマークモデルを構築することが可能と考えられる。

本研究のモチベーションとなった大腸がん肝転移切除後の再発をはじめ、肝細胞がん切除後の異時多中心性発がんによる再発 [81]、がんの骨転移患者における病的骨折や放射線治療のような骨関連イベント [82]、骨髄移植後の感染イベント [83] のように、同一対象者についてイベント発生が繰り返される状況がある。原発性肝がん、転移性肝がんともに適応可能であれば肝切除手術が有効であり、再発しても適応可能である限り再度の切除手術が施され、対象者によっては再発と手術を複数回繰り返す。特に大腸がんの血行性転移をきたす臓器の中で最も高

頻度でみられる肝転移では、切除手術適応である場合が多く、切除可能であればそうでない場合に比べ生存予後はよいとされる [84]。

繰り返しイベントデータに対する Cox モデルの拡張は 1980 年代に盛んに行われた中では、Andersen-Gill モデル [85](以下、AG モデル)、Prentice-Williams-Peterson モデル [86](以下、PWP モデル)、Wei-Lin-Weissfeld モデル [87](以下、WLW モデル) の 3 つのモデルが有名である。AG モデルは、過去の繰り返しイベント数によらずに、共通のベースラインハザード関数を仮定した比例ハザードモデルであり、基本的に非斉時 Poisson 過程から繰り返しイベントが生起されると仮定する。PWP モデルは、それぞれのイベント回数ごとに異なるベースラインハザード関数を仮定し、 k 回目イベントの at risk は $k-1$ 回目イベント時点から開始するとした条件付き尤度を構成する。時点の扱いも、追跡開始時からの経過時間を用いる total times モデルと、前回イベント発生時からの経過時間を用いる gap times モデルの 2 通りが提案されており、それぞれイベント発生に対するマルコフ性とセミマルコフ性に対応させている。前回イベント発生時点で at risk となることから、 k 回目のイベントに関する total times モデルでは、対象者のイベント情報は $k-1$ 回目イベント時点にて左側切断 (遅延登録) されたとみなす扱いをする。一方、WLW モデルは、繰り返しイベントにも適用可能な、より一般的な多変量生存時間解析のモデルとして提案された。それぞれの繰り返しイベントは別々の過程から生起するとみなし、各イベント時点の周辺分布に対して Cox モデルを仮

定する。PWP total times モデルと比べ、イベント回数ごとに異なるベースラインハザードを仮定し、追跡開始時からの経過時間を時点に用いる点は共通であるが、 k 回目イベントの at risk である限り、 $k+1$ 回目以降のイベントも必ず at risk として扱う点が異なる。また、GEEs(Generalized Estimating Equations; 一般化推定方程式) [88] 同様の推定方法を利用し、同一対象者について複数回利用されるイベント間の相関を考慮したサンドウィッチ分散を利用することで、回帰係数に関して妥当な推測を行っている [89]。このように WLW モデルは AG モデルや PWP モデルのような、繰り返しイベント間の構造を仮定していないという点で、頑健なモデルといわれる。一方で、回帰係数にて表される治療効果は、後に続くイベントに対するモデルの回帰係数へ影響を与えることの批判 [90] もあり、それぞれのモデルに対する議論は現在も続いている [91, 92]。

繰り返しイベントを追跡する上での問題点のひとつに終端イベント (terminal events) の観察がある [93, 94]。終端イベントとは、例えば死亡のように、その発生によって以降の繰り返しイベントが発生しなくなるようなものをさす。終端イベントの発生が繰り返しイベントの発生と独立であるという仮定は現実的ではなく、それゆえ終端イベントを単純な打ち切りとして扱う解析手法には問題が生じる。そこで終端イベントを情報のある打ち切りとみなして、その補正を行うアプローチ [95–97] と、終端イベントも興味のあるイベントとしてモデル化するアプローチの2つが考えられる。後者のような終端イベントと非終端イベントの2つ

が存在するセミ競合リスク [98] の状況では、illness-death 過程を仮定すると見通しがよくなると報告されている [99]。本研究では、終端イベント (大腸がん肝転移切除症例では死亡) も一つの重要なイベントであり、illness-death モデルを拡張することで終端イベントが存在する下での繰り返しイベントを捉えることとする。

1.4 本研究の目的

本研究では、これまでに提案されている経時データを利用したランドマーク解析による動的予測法を繰り返しイベントが観察される状況に拡張する。さらに、終端イベントにも関心がある状況下での繰り返しイベントデータへの適用も念頭におく。

提案手法を大腸がん肝転移切除例追跡データに適用し、切除不能再発が死亡に対するよりよい代替指標となりうるかの検討と、切除不能再発が起こる前であれば、どの程度の再発・死亡リスクがあるかを前回切除時の腫瘍情報・切除履歴等を用いて予測する。

動的予測という評価方法によって、ある対象者がどの程度イベントを起こしやすいかを「確率」という直感的に理解しやすい数値で表現することが可能になり、本研究によって、繰り返しイベント・競合リスクを含む実際に観察されるほぼすべての種類の time-to-event データに対する、ランドマークモデルに基づいた動的予測が可能となる。

2 方法

2.1 準備

2.1.1 記法

対象者 $i (i = 1, \dots, n)$ の $j (j = 1, \dots, J)$ 回目の潜在的な繰り返しイベント時点 T_{ij} に関して、

$$T_{i1} < \dots < T_{iJ} \quad (2)$$

のような順序性がある。ただし、最大の繰り返しイベント回数 J_i が $J_i < J$ であるような対象者については、便宜的に $T_{i,J_i+1} = \dots = T_{i,J} = \infty$ とする。対象者 i の終端イベント時点を T_i^D と表す。以降、終端イベントを打ち切りと考慮しない場合であっても $T_i^D = \infty$ とみなすことで、終端イベントと打ち切りを区別する場合と統一して表記を行う。対象者 i の打ち切り時点を C_i 、時点 t における共変量ベクトルを $\mathbf{Z}_i(t) = (Z_{i1}(t), \dots, Z_{ip}(t))^T$ とし、以下のような仮定をおく。

(A1) 異なる対象者から得られる潜在データ $\{T_{i1}, \dots, T_{iJ}, T_i^D, C_i, \mathbf{Z}_i(t)\}$ は互いに独立である。

以上より、観察される繰り返しイベント時点は $\tilde{T}_{ij} = \min(T_{ij}, T_i^D, C_i)$ 、観察される終端イベント時点は $\tilde{T}_i^D = \min(T_i^D, C_i)$ として得られる。それぞれのイベント観察に関する指示変数は指示関数 $I(\cdot)$ を用いて、 $\delta_{ij} = I(T_{ij} < T_i^D, T_{ij} < C_i)$ 、 $\delta_i^D = I(T_i^D < C_i)$ と表す。対象者 i に関して観察されるデータセットは、

$$\{\tilde{T}_{i1}, \dots, \tilde{T}_{iJ}, \tilde{T}_i^D, C_i, \delta_{i1}, \dots, \delta_{iJ}, \delta_i^D, \mathbf{Z}_i(t)\}$$

で表される。

2.1.2 動的予測とランドマークモデル

1節の(1)式で定義したように、ある時点にてイベントを起こしていない条件の下で、その後のイベントを起こす確率を推定することを動的予測と定義した。

この定義を繰り返しイベントに対応するために再定義する。

動的予測を行う時点を $s_m (m = 1, \dots, M)$ 、動的予測幅を定数 w と表せば、 s_m にて at risk にある下で、 $s_m + w$ までに終端イベントを起こさずに $k (k = 0, \dots, K \leq J)$ 回繰り返しイベントを起こす条件付き確率は、

$$\begin{aligned} & F_k(s_m + w | s_m) \\ &= \sum_j^{J-k} \Pr(T_{j+k} \leq s_m + w, T_{j+k+1} > s_m + w | T_j \leq s_m, T_{j+1} > s_m, T^D > s_m + w) \end{aligned} \quad (3)$$

と表される。 $k \geq 1$ にて(2)式に示した順序性より $\Pr[\{T_{j+1}, \dots, T_{j+k}\} \in (s, s + w)] = \Pr(T_{j+k} \leq s + w)$ が成立するゆえ、 k 回のイベント時点が含まれることが(3)式右辺のように表現される。同様に s_m にて at risk にある下で、 $s_m + w$ までに終端イベントを起こす条件付き確率は、

$$F^D(s_m + w | s_m) = \Pr(T^D \leq s_m + w | T^D > s_m) \quad (4)$$

で表される。なお、(3)式、(4)式の定義より、

$$\sum_{k=0} F_k(s_m + w | s_m) + F^D(s_m + w | s_m) = 1 \quad (5)$$

が成立する。(5)式は s_m にて at risk である対象者が $s_m + w$ にて終端イベントを起こした状態もしくは終端イベントを起こさず、任意の回数の繰り返しイベントを起こした状態のいずれかであることに対応した制約である。

ここで、対象者が観察下にある状態である at risk とは、通常の単一イベントや競合リスクイベントをアウトカムとしていればイベントと打ち切りのいずれも起こしていない状態を意味するが、本研究では繰り返しイベントをアウトカムとするため、打ち切りや終端イベントのいずれも起こしていない状態とする。すなわち、繰り返しイベントないし終端イベントを起こさずの状態のことを意味する。

次に、ランドマークモデルに基づく動的予測を述べる。動的予測は時点 s_m にて at risk である条件付きのイベント発生確率を推定することから、時点 s_m までに打ち切りや終端イベントを起こした対象者のデータは除外し、時点 s_m にて at risk であるサブグループに限定した解析を行うことは自然である。さらに、臨床経時データ $\mathbf{Z}(t)$ について、時点 s_m での観察値 $\mathbf{Z}(s_m)$ のみ利用し、 $F_k(s_m + w | s_m), F^D(s_m + w | s_m)$ のモデル化を行う。このモデル化により、臨床経時データがほとんどの場合に該当する内生変数へ対処することができる。

ランドマークモデルをあてはめる上で、イベント時点や経時データの加工を行うことに伴い、表記を定める。まず、イベント時点のうちランドマーク時点以前のものはアウトカムとして利用しないため、ランドマーク時点以降のイベントを

順に数え上げる。ランドマーク時点 s_m における対象者 i の k 回目の潜在的な繰り返しイベント時点を T_{mik}^* と表す。この対象者がランドマーク時点までに J_{mi} 回の潜在繰り返しイベント時点を持っていたならば、 T_{mik}^* は、 $T_{mik}^* = T_{i,(k+J_{mi})}$ となる。このうち、実際に観察される (打ち切りや終端イベント時点よりも先にある) イベント時点を、 $\tilde{T}_{mik}^* = \min(T_{mik}^*, T_i^D, C_i)$ と表す。各 T_{mij}^* に対応するイベントに関する指示変数を δ_{mik}^* と表せば、 $\delta_{mik}^* = \delta_{i,(k+J_{mi})}$ となる。最後に、共変量ベクトルは、ランドマーク時点までに得られる情報 $\mathbf{Z}_i(s_m)$ のみとする。以上より、ランドマーク時点 s_m において、対象者 i に関して観察されるデータセットは、

$$\{\tilde{T}_{mi1}^*, \dots, \tilde{T}_{miJ}^*, \tilde{T}_i^D, C_i, \delta_{mi1}^*, \dots, \delta_{miJ}^*, \delta_i^D, \mathbf{Z}_i(s_m)\}$$

で表される。以降、表記の簡略化のため、解析目的によって任意に指定されるランドマーク時点の違いを表す添字 m は省略する。

2.2 動的擬似値

互いに独立かつ同一分布に従う確率変数 T について、ある関数 f で変換した後の期待値 θ に興味があるとする。これは打ち切りデータが存在しない場合、対象者 i ($i = 1, \dots, n$) の観察生存時間 T_i を用いて、

$$\theta = E(f(T)) = \frac{1}{n} \sum_i f(T_i) \quad (6)$$

と単純に得られる。もし右側打ち切りの不完全データが存在する、すなわち $f(T_i)$ がすべて観察されないならば、打ち切りがランダムという前提で妥当な推定量を用いて $\hat{\theta}$ を求めればよい。例えば、時点 t での生存確率 $S(t)$ を推定したい場合、Kaplan-Meier 推定量 $S(t) = E(I(T > t))$ はその一つである。さて、対象者 i のみ除いたデータセットに対し $\hat{\theta}$ と同じ推定を行った、leave-one-out 推定量を $\hat{\theta}^{(-i)}$ と表記し、その推定量を用いて、

$$\hat{\theta}_i = n \cdot \hat{\theta} - (n - 1) \cdot \hat{\theta}^{(-i)} \quad (7)$$

を定義する。この $\hat{\theta}_i$ はジャックナイフ法における擬似値として知られる。擬似値は n 個の独立な推定値のように扱うことができ [67]、その平均であるジャックナイフ統計量は一致性をもつ [70]。擬似値は通常 $\hat{\theta}$ のバイアスや精度についての議論を行うために用いられるが、ここでは不完全データが含まれる $f(T_i)$ の代わりに、全ての対象者に対し観察されるデータとして用いる。

(3),(4) 式の条件付き確率をそれぞれ母数 $\theta_k(s), \theta^D(s)$ とおけば、これらは以下の指示関数の期待値として与えられる。

$$\theta_k(s) = F_k(s + w | s) = E \left\{ I \left(T_k^* \leq s + w, T_{k+1}^* > s + w \mid T_k^* > s \right) \right\} \quad (8)$$

$$\theta^D(s) = F^D(s + w | s) = E \left\{ I \left(T^D \leq s + w \mid T^D > s \right) \right\} \quad (9)$$

$\theta_k(s), \theta^D(s)$ は (6) 式の形で表現でき、かつ (A1) の仮定より、(7) 式の定義になり、以下のような擬似値を定義する。

$$\hat{\theta}_{ik}(s) = n_s \cdot \hat{F}_k(s+w|s) - (n_s - 1) \cdot \hat{F}_k^{(-i)}(s+w|s) \quad (10)$$

$$\hat{\theta}_i^D(s) = n_s \cdot \hat{F}^D(s+w|s) - (n_s - 1) \cdot \hat{F}^{D,(-i)}(s+w|s) \quad (11)$$

ただし、 n_s は時点 s における at risk 数である。このようなランドマーク時点ごとに at risk である対象者 i について定義する擬似値は、Nicolaie ら [65] によって動的擬似値と名付けられた。なお以降、(8)-(11) 式の添字 k, D は省略し、統一して表現する場合がある。

動的擬似値の漸近的性質については、擬似値についての漸近的性質を深く考察した Graw ら [100] によって与えられている。そのために以下のような仮定を置く。

(A2) 打ち切り時点 C_i は潜在イベント時点 $T_{i1}, \dots, T_{iJ}, T_i^D$ や共変量 $\mathbf{Z}_i(t)$ と確率的に独立である。

(A3) ランドマーク時点 s は、 $s+w < \tau$ であるようなある τ について $G(\tau) > 0$ を満たすようなもののみを考える。ただし、 $G(\cdot)$ は打ち切りに関する生存関数である。

(10),(11) 式の右辺の推定値に Aalen-Johansen 推定量を用いることを前提とした場合、以下のような命題 (P1)-(P3) が証明される。なお、この数式的証明は Graw ら [100]、Nicolaie ら [65] を参照されたい。

(P1) $\hat{\theta}_i(s)$, $i = 1, \dots, n_s$ は $n \rightarrow \infty$ につれ漸近的に *i.i.d.* である。

(P2) $\hat{\theta}_i(s)$ は $i \neq i'$ かつ $s \neq s'$ であるような $\hat{\theta}_{i'}(s')$ と $n \rightarrow \infty$ につれ漸近的に独立である。

(P3) $n \rightarrow \infty$ につれ漸近的に以下のような共変量での条件付き期待値が等しい。

$$E\{\hat{\theta}_{ik}(s) \mid \mathbf{Z}_i(s), T_i > s\} = E\{I(T_k^* \leq s + w, T_{k+1}^* > s + w \mid \mathbf{Z}_i(s), T_k^* > s)\}$$

$$E\{\hat{\theta}_i^D(s) \mid \mathbf{Z}_i(s), T_i > s\} = E\{I(T^D \leq s + w \mid \mathbf{Z}_i(s), T^D > s)\}$$

(P1),(P2) は異なる対象者同士であれば、ランドマーク時点が同じか異なるかによらず、動的擬似値が独立であることを意味する。(P3) は、通常の観察値と同様に動的擬似値をアウトカムとした回帰モデルにて推論を行うことが可能なことを意味する。

2.2.1 終端イベントを検討しない場合

終端イベントを検討しない、すなわち (10) 式のような擬似値のみ生成する状況をまず考える。ここでは動的擬似値の生成方法として Aalen-Johansen 推定量に基づく方法と生存関数の差にて定義する方法の 2 通りを提案する。

終端イベントを検討しない場合、繰り返しイベントは図 2 に示すような遷移関係をたどる。この場合、最大の繰り返しイベント数を J_s として、 $J_s + 1$ 個の状態をもつ多状態モデルを仮定できる。これら遷移に対するマルコフモデルの仮定

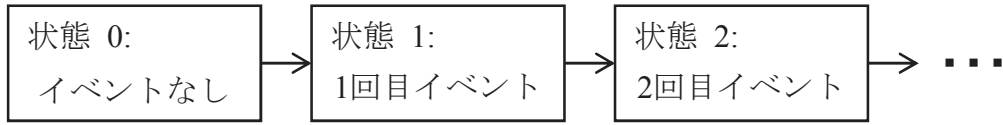


図 2: 終端イベントを検討しない場合の遷移図

と (A1)、(A2) の仮定より、状態確率は Aalen-Johansen 推定量を用いて、 $n^{1/2}$ のオーダーで一致推定可能である [8]。また、以降で述べる共変量での条件付けを行わない場合であれば、マルコフ性の仮定なしに Aalen-Johansen 推定量は一致性を持つことが示されている [101]。

時点 t における状態 q から状態 r へ遷移するハザード関数を $\lambda_{qr}(t)$ と表す。この累積ハザード関数 $\Lambda_{qr}(t_0, t) = \int_{t_0}^t \lambda_{qr}(u) du$ のノンパラメトリック最尤推定量の代表的なものが Nelson-Aalen 推定量であり、イベント数を at risk 数で除したハザードの t までにおける総和として、以下のように得られる。

$$\hat{\Lambda}_{qr}(t_0, t) = \sum_{u: u \in (t_0, t]} \hat{\lambda}_{qr}(u) = \sum_{u: u \in (t_0, t]} \frac{N_{qr}(u)}{Y_q(u)}, \quad (q \neq r) \quad (12)$$

ただし、 $N_{qr}(\cdot)$ が状態 q にある対象者が状態 r のイベントを起こした数、 $Y_q(\cdot)$ が状態 q にいる at risk 数を表す。(12) 式のようなハザードを基に、積積分 (product-integral) を用いて遷移確率 $P_{qr}(t_0, t)$ を求めた推定量が Aalen-Johansen 推定量である。各遷移間のハザードを成分とした遷移ハザード行列 $\hat{\lambda}(t) = [\hat{\lambda}_{qr}(t)]$ (ただし、 $\hat{\lambda}_{qq}(t) = - \sum_{r: r \neq q} \hat{\lambda}_{qr}(t)$ と定義する) とおけば、遷移確率行列 $\mathbf{P}(t_0, t) = [P_{qr}(t_0, t)]$ は、

$$\hat{P}(t_0, t) = \prod_{u: u \in (t_0, t]} \{\mathbf{1} + \hat{\lambda}(u)\} \quad (13)$$

のように推定される。ただし、 $\mathbf{1}$ は単位行列である。なお、この段落での時点は離散的なものを考える。そのため積積分は (13) 式のような総積で表すことが可能である。

2 回目イベントまで考慮する場合に、(13) 式を用いて遷移確率を求める例を示す。時点 t_0 では全員状態 0 にあるため、時点 t において対象者が状態 k にある確率は $P_{0k}(t)$ で表される。

$$\begin{aligned} \hat{P}_{00}(t_0, t) &= \exp \left\{ - \int_{t_0}^t \hat{\lambda}_{01}(u) du \right\} \\ \hat{P}_{01}(t_0, t) &= \int_{t_0}^t \hat{P}_{00}(t_0, u) \hat{\lambda}_{01}(u) \hat{P}_{11}(u, t) du \\ \hat{P}_{02}(t_0, t) &= \int_{t_0}^t \hat{P}_{00}(t_0, u) \hat{\lambda}_{01}(u) \hat{P}_{11}(u, u') \hat{\lambda}_{12}(u') du du' \\ \hat{P}_{11}(t, t') &= \exp \left\{ - \int_t^{t'} \hat{\lambda}_{12}(u) du \right\} \end{aligned}$$

以上より、(10) 式の動的擬似値は $\hat{F}_k(s + w | s) = \hat{P}_{0k}(s, s + w)$ から、

$$\hat{\theta}_{ik}(s) = n_s \cdot \hat{P}_{0k}(s, s + w) - (n_s - 1) \cdot \hat{P}_{0k}^{(-i)}(s, s + w) \quad (14)$$

と提案される。

(14) 式で提案した Aalen-Johansen 推定量は行列計算が必要であり、その計算が煩雑となるため、(2) 式のイベント時点に関する順序性を利用することで、より簡便に計算可能な Kaplan-Meier 推定量に基づく動的擬似値を本研究では提案する。(8) 式右辺は、(2) 式を用いて以下のように書き換えられる。

$$\begin{aligned}
& E \left\{ I \left(T_k^* \leq s + w, T_{k+1}^* > s + w \mid T_k^* > s \right) \right\} \\
&= E \left\{ I \left(T_k^* \leq s + w \mid T_k^* > s \right) \cdot I \left(T_{k+1}^* > s + w \mid T_k^* > s \right) \right\} \\
&= E \left[I \left(T_k^* \leq s + w \mid T_k^* > s \right) \cdot \left\{ 1 - I \left(T_{k+1}^* \leq s + w \mid T_k^* > s \right) \right\} \right] \\
&= E \left\{ I \left(T_k^* \leq s + w \mid T_k^* > s \right) - I \left(T_{k+1}^* \leq s + w \mid T_k^* > s \right) \right\} \\
&= E \left\{ I \left(T_k^* \leq s + w \mid T_k^* > s \right) \right\} - E \left\{ I \left(T_{k+1}^* \leq s + w \mid T_k^* > s \right) \right\} \\
&= E \left\{ I \left(T_k^* \leq s + w \mid T_k^* > s \right) \right\} - E \left\{ I \left(T_{k+1}^* \leq s + w \mid T_{k+1}^* > s \right) \right\} \\
&= \left\{ 1 - S_k^*(s + w \mid s) \right\} - \left\{ 1 - S_{k+1}^*(s + w \mid s) \right\} \\
&= S_{k+1}^*(s + w \mid s) - S_k^*(s + w \mid s) \tag{15}
\end{aligned}$$

ただし、 $S_k^*(\cdot)$ は、 T_k^* に関する生存関数である。すなわち、 $k + 1$ 回目のイベントに関する生存確率と k 回目のイベントに関する生存確率の差が k 回イベントを起こす確率であるということである。 $S^*(\cdot)$ は、Kaplan-Meier 推定量、ないし Nelson-Aalen 推定量を生存関数に変換することで、どちらも $n^{1/2}$ のオーダーで一致推定可能である [8]。収束速度も同じゆえ、以下に示す動的擬似値に関して命題 (P1)-(P3) は Aalen-Johansen 推定量を用いた場合と同様に証明される。

$$\begin{aligned}
\hat{\theta}_{ik}(s) &= n_s \cdot \left\{ \hat{S}_{k+1}^*(s + w \mid s) - \hat{S}_k^*(s + w \mid s) \right\} \\
&\quad - (n_s - 1) \cdot \left\{ \hat{S}_{k+1}^{*(-i)}(s + w \mid s) - \hat{S}_k^{*(-i)}(s + w \mid s) \right\} \tag{16}
\end{aligned}$$

このイベント回数別に考える発想は WLW モデル [87] と同じであり、切片のみの WLW モデルにおいて、Breslow 推定量を用いてハザードを推定することは、Nelson-Aalen 推定量を用いて $S^*(\cdot)$ を推定することに等しい。

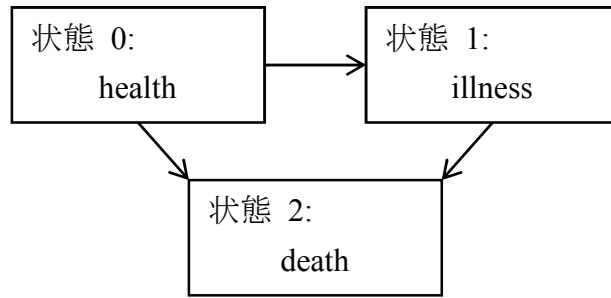


図 3: illness-death モデルの遷移図

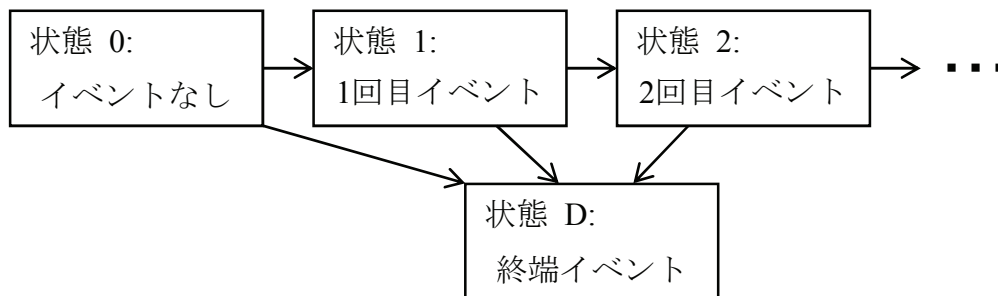


図 4: 終端イベントを検討する場合の遷移図

2.2.2 終端イベントを検討する場合

終端イベントと非終端イベントが1つずつ存在する場合は、図3に示すような illness-death モデルとして多状態モデルで扱うことができる。すなわち、終端イベントをいわゆる“death”、非終端イベントを“illness”がそれぞれ発生したとみなす考えである。

この illness-death 過程を拡張することで、図4にて示すような終端イベント存在下での繰り返しイベントに対するモデルを考える。この状況における状態確率は、(13)式に示した Aalen-Johansen 推定量を用いれば状態確率を推定可能であり、(10)式の動的擬似値は(14)式と同じ形で、(11)式の動的擬似値は、

$$\hat{\theta}_i^D(s) = n_s \cdot \hat{P}_{0D}(s, s+w) - (n_s - 1) \cdot \hat{P}_{0D}^{(-i)}(s, s+w) \quad (17)$$

のように表現することができる。

2.3 動的擬似値を用いた一般化線形モデルによる動的予測確率の推定

2.3.1 ランドマーク時点ごとの予測確率の推定

経時データをランドマークモデルにあてはめ、動的予測を行う。時点 s で at risk である対象者は時点 $s+w$ にて潜在的に k 回目のイベントもしくは終端イベントのいずれかの状態にあり、統計学的には多値アウトカム変数と捉えることができる。各アウトカムの条件付き確率 $\theta(s)$ は (8),(9) 式にて定義した。終端イベントを考慮する場合は終端イベントを検討しない場合に比べ、アウトカムの水準が一つ増えるだけなので、表記上簡略な、終端イベントを考慮しない場合について記述する。

対象者 i の時点 $s+w$ におけるイベント回数を表す多値アウトカムはベクトルを用いて、 $\mathbf{Y}_i(s) = [Y_{i1}(s), \dots, Y_{iJ_s}(s)]^T$ のように表す。ただし、各成分 $Y_{ik}(s)$ はイベント回数が k である時のみ 1、そうでなければ 0 となる指示変数である。冗長とならないよう、一度もイベントを起こしていない対象者の多値アウトカムの成分はすべて 0 となる。各成分は 2 値で表される多値アウトカムに対して、多変量 2 項モデル [102, 103] のあてはめを行い、対応する条件付き確率 $\theta_{ik}(s)$ を推定する。そのモデルは、2 水準のアウトカムである場合に頻用される一般化線形モ

デルを多水準の場合に素直に拡張したモデルであり、対象者間の相関を考慮するための混合効果モデルや GEEs への拡張を念頭に提案されている。特にロジットリンクとする一般化線形モデルでは共変量の構造によって、名義アウトカムである場合に標準的に利用される一般化ロジット (generalized logit) モデル [104]/ベースラインカテゴリロジット [105]をはじめ、累積ロジットモデルや隣接水準ロジット (adjacent-categories logit) モデル [106] など様々なロジスティック回帰モデルを表現可能な広いモデルである。

多値アウトカムベクトルに対応するよう、対象者 i のイベント確率ベクトル $\theta_i(s) = [\theta_{i1}(s), \dots, \theta_{iJ_s}(s)]^T$ を定める。 $\theta_i(s)$ と切片項を含んだ共変量 $\mathbf{Z}_i^*(s)$ の関係は、回帰係数ベクトル $\boldsymbol{\beta}$ を用いて、以下に示す一般化線形モデルにて与えられるものとする [107]。

$$g\{\theta_i(s)\} = \boldsymbol{\beta}(s)^T \mathbf{Z}_i^*(s)$$

ただし、 $g(\cdot)$ はリンク関数であり、リンク関数の逆関数、すなわち平均関数 $\mu(\cdot)$ は、

$$\theta_i(s) = g^{-1}\{\boldsymbol{\beta}(s)^T \mathbf{Z}_i^*(s)\} = \mu\{\boldsymbol{\beta}(s)^T \mathbf{Z}_i^*(s)\} \quad (18)$$

のように定義される。例えば一般化ロジットモデルを表現するには、リンクをロジット関数とし、共変量を以下のような構造にする。

$$\mathbf{Z}_i^*(s) = \begin{bmatrix} 1 & \mathbf{Z}_i(s)^\top & & & \\ & & 1 & \mathbf{Z}_i(s)^\top & \\ & & & \ddots & \\ & & & & 1 & \mathbf{Z}_i(s)^\top \end{bmatrix}$$

以下のような一般化線形モデルの推定方程式を解くことで $\boldsymbol{\beta}$ は推定される。

$$U\{\boldsymbol{\beta}(s)\} = \sum_i \mathbf{D}_i(s)^\top \mathbf{V}_i^{-1}(s) \{Y_i(s) - \boldsymbol{\theta}_i(s)\} = \mathbf{0}, \quad \mathbf{D}_i(s) = \frac{\partial \boldsymbol{\theta}_i(s)}{\partial \boldsymbol{\beta}(s)} \quad (19)$$

ただし、分散 $\mathbf{V}_i(s)$ は多項分布に基づき、

$$\mathbf{V}_i(s) = \begin{bmatrix} \theta_{i1}(s)\{1 - \theta_{i1}(s)\} & -\theta_{i1}(s)\theta_{i2}(s) & \cdots & -\theta_{i1}(s)\theta_{iJ_s}(s) \\ -\theta_{i2}(s)\theta_{i1}(s) & \theta_{i2}(s)\{1 - \theta_{i2}(s)\} & & \vdots \\ \vdots & & \ddots & -\theta_{i,(J_s-1)}(s)\theta_{iJ_s}(s) \\ -\theta_{iJ_s}(s)\theta_{i1}(s) & \cdots & \theta_{iJ_s}(s)\theta_{i,(J_s-1)}(s) & \theta_{iJ_s}(s)\{1 - \theta_{iJ_s}(s)\} \end{bmatrix}$$

と表される。ここまで表記してきた多値アウトカム $Y_{ik}(s)$ は右側打ち切りにより観察されない場合があるため、代わりに擬似値 $\hat{\theta}_{ik}(s)$ を用いて (19) 式を解くことで、 $\boldsymbol{\beta}(s)$ の推定を行う。この場合、 $\boldsymbol{\beta}(s)$ の一致性は (P3) によって保証される。

最後に、 $\boldsymbol{\beta}(s)$ の漸近分散は以下に示す Fisher 情報量の逆数であるモデル分散より得られる。

$$\text{Var}\{\hat{\boldsymbol{\beta}}(s)\} = \left[\sum_i \hat{\mathbf{D}}_i(s)^\top \hat{\mathbf{V}}_i^{-1}(s) \hat{\mathbf{D}}_i(s) \right]^{-1} \quad (20)$$

ランダムマーク時点 s における動的予測確率 $\boldsymbol{\theta}_i(s)$ は、得られた $\hat{\boldsymbol{\beta}}(s)$ を用いて (18) 式に示した平均関数によって変換することで推定する。また、(20) 式から得られる $\hat{\boldsymbol{\beta}}(s)$ の分散共分散行列を基に、デルタ法を用いて動的予測確率の分散を得る。

2.3.2 ランドマークスーパーモデルによる、時点に対するスムージング

前節では、ランドマーク時点ごとに異なる一般化線形モデルを考えていた。そのため、動的予測結果は離散的に得られる。一般に、ランドマーク時点の推移とともに予測確率は、単調性や増加・減少のような、自然な関連性を示すことが多い。そこで、回帰係数を時点の関数としてスムージングすることで、隣り合うランドマーク時点間においても連続的に動的予測が行えることを考える。さらに、複数のランドマーク時点から得られる情報をすべて用いることで、回帰係数の精度向上、ひいては動的予測の正確性を高めることにもつながる。

具体的には時点 s の多項式 $f(s)$ を共変量に加え、複数時点間にわたる多項アウトカムをまとめて、GEEs を利用した推定方程式を考える。対象者 i のアウトカムはすべての時点をまとめた $\mathbf{Y}_i = [\mathbf{Y}_i(s_0)^\top, \dots, \mathbf{Y}_i(s_M)^\top]^\top$ を用い、共変量は、

$$\tilde{\mathbf{Z}}_i^* = \begin{bmatrix} \mathbf{Z}_i^*(s_0) & f_1(s_0)\mathbf{Z}_i^*(s_0) & \dots & f_h(s_0)\mathbf{Z}_i^*(s_0) \\ & & \vdots & \\ \mathbf{Z}_i^*(s_M) & f_1(s_M)\mathbf{Z}_i^*(s_M) & \dots & f_h(s_M)\mathbf{Z}_i^*(s_M) \end{bmatrix} \quad (21)$$

とする。例えば、多項式を2次式とするならば、 $f_1(s) = s$, $f_2(s) = s^2$ である。これらデータに対し、以下に表す擬似スコア方程式を解くことで $\tilde{\boldsymbol{\beta}}$ の推定値を得る。

$$U(\tilde{\boldsymbol{\beta}}) = \sum_i \tilde{\mathbf{D}}_i^\top \mathbf{V}_i^{-1} \{\mathbf{Y}_i - \boldsymbol{\theta}_i\} = \mathbf{0}, \quad (22)$$

$$\tilde{\mathbf{D}}_i = \frac{\partial \boldsymbol{\theta}_i}{\partial \tilde{\boldsymbol{\beta}}}, \quad \boldsymbol{\theta}_i = [\boldsymbol{\theta}_i(s_0)^\top, \dots, \boldsymbol{\theta}_i(s_M)^\top]^\top$$

ただし、分散 $V_i = \text{Cov}[Y_{ik}(s), Y_{ik'}(s')]$ の各成分は、

$$\text{Cov}[Y_{ik}(s), Y_{ik'}(s')] = \begin{cases} \theta_{ik}(s)\{1 - \theta_{ik}(s)\} & \text{if } k = k', s = s' \\ -\theta_{ik}(s)\theta_{ik'}(s) & \text{if } k \neq k', s = s' \\ \frac{\text{Corr}[Y_{ik}(s), Y_{ik'}(s')]}{[\theta_{ik}(s)\{1 - \theta_{ik}(s)\}\theta_{ik'}(s')\{1 - \theta_{ik'}(s')\}]^{1/2}} & \text{if } k \neq k', \text{ any } s, s' \end{cases} \quad (23)$$

のように表される。前節同様に、 $Y_{ik}(s)$ の代わりに $\hat{\theta}_{ik}(s)$ を用いて、(22) 式の解を反復計算する。一般的な GEEs では (23) 式の $\text{Corr}[Y_{ik}(s), Y_{ik'}(s')]$ の構造に依らずに $\tilde{\beta}$ の一致推定は可能であるが、ランドマーク時点で at risk である場合のみアウトカムが定義される現在の状況では、 $\text{Corr}[Y_{ik}(s), Y_{ik'}(s')] = 0$ 、すなわち、独立構造を作業相関行列に置かねばならない。もし独立構造以外を仮定した場合、分散 V_i が対象者の生存状態に依存してしまい、(22) 式に表される擬似スコア方程式の一致性が担保されず、したがって $\tilde{\beta}$ の一致性が保証されないと指摘されている [108]。

$\tilde{\beta}$ のモデル分散は相関構造が真でない限り一致性が保証されないため、相関構造の誤特定しても妥当な推定量を与える、以下に示すサンドウィッチ分散を用いる [88]。

$$\text{Var}(\hat{\beta}) = \left[\sum_i \hat{D}_i^\top \hat{V}_i^{-1} \hat{D}_i \right]^{-1} \left[\sum_i \hat{D}_i^\top \hat{V}_i^{-1} (Y_i - \hat{\theta}_i)(Y_i - \hat{\theta}_i)^\top \hat{V}_i^{-1} \hat{D}_i \right] \left[\sum_i \hat{D}_i^\top \hat{V}_i^{-1} \hat{D}_i \right]^{-1} \quad (24)$$

2.3.3 解析プログラムの作成

全てのシミュレーションでは SAS® Ver 9.3 (SAS Institute, Cary, NC)、実データ解析ではこれに加え、R 3.1.1 を用いた。4.2 節で述べる死亡予測では、2 値アウトカムゆえ、Cox モデルを扱う SAS·phreg プロシジャと一般化線形モデルを扱う SAS·genmod プロシジャを用いた。一方、多値アウトカムについて擬似値を用いた推定は genmod プロシジャで実装不可能であるため、シミュレーションと 4.3 節で述べる再発・死亡イベント予測では、SAS/IML にて独自に作成した行列計算プログラムを用いた。作成したプログラムは、適当に発生させた数値例を用いて、多値アウトカムの最も単純な例である 2 値アウトカムの場合や、擬似値ではなく通常の離散アウトカムをとる場合に genmod プロシジャ、logistic プロシジャを用いた結果と一致することで、その妥当性を確認した。また、IML にて推定方程式を解く際、非線形最小二乗問題を解く場合に広く用いられる Levenberg-Marquardt 法を実装した NLPLM サブルーチンを用いた。なお、反復計算の収束基準は、勾配の絶対値が 10^{-5} 以下とした。

実データ解析では、提案法で動的予測した確率について、2 値アウトカムの場合と多値アウトカムの場合で異なる 95%Wald 信頼区間を構成した。2 値アウトカムではリンク関数である logit スケールにて得た信頼区間をリンクの逆関数で変換したが、多値アウトカムでは確率のスケールにおいて、デルタ法によって推定した標準誤差を用いて信頼区間を得た。信頼区間の上限や下限が 0 や 1 を超え

る場合は0ないし1で切断を行った。これは、多値アウトカムの一般化ロジットモデルでは線形予測子の推定値をリンクの逆関数で変換してもそのままでは予測確率とならないためである。

3 シミュレーション実験

3.1 終端イベントを検討しない場合

提案する動的擬似値が、適用を想定するサンプルサイズの規模においてバイアスの少ない推定を行えることをシミュレーション実験を通して確認する。関連のある繰り返しイベントデータのデータ発生方法は、frailtyモデルとコピュラモデルに基づく方法の2通り [109] が考えられるが、繰り返しイベントデータの発生には対象者間のイベントの起こりやすさに関する異質性やイベント回数を重ねることにより生物学的に変化するイベントの起こりやすさを考慮しやすい frailtyモデルに基づく方法が適している [110]。具体的には、繰り返しイベントは最大2回発生したとし、対象者 i の1回目のイベント時点 t_{i1} と、2回目のイベントについて1回目のイベントからの経過時間 $t_{i2} - t_{i1}$ はそれぞれ以下のハザード $\lambda_{i1}, \lambda_{i2}$ をパラメータとする指数分布に従い発生させた [111, 112]。

$$\lambda_{i1} = u_i \lambda_{01}, \quad \lambda_{i2} = u_i \lambda_{02}$$

u_i がガンマ分布に従う frailty 項、 $\lambda_{01}, \lambda_{02}$ が各繰り返しイベントのベースラインハザードである。それぞれ対象者間のイベントの起こりやすさに関する異質性の有無、マルコフ性/セミマルコフ性を表し、これらパラメータの設定により4通りのシナリオを検討した。 u_i が対象者間について分布すればイベント間に相関が生じ、 $\lambda_{01} = \lambda_{02}$ であればマルコフ性、 $\lambda_{01} \neq \lambda_{02}$ であればセミマルコフ性がいえる。各シナリオのパラメータ設定値を表1に示す。

表 1: 終端イベントを考慮しない場合の各シナリオにおけるパラメータ値

シナリオ	u_i	λ_{01}	λ_{02}
異質性なし, マルコフ性	1	1	1
異質性なし, セミマルコフ性	1	1	2
異質性あり, マルコフ性	$\Gamma(0.5, 0.5)$	1	1
異質性あり, セミマルコフ性	$\Gamma(0.5, 0.5)$	1	2

表 1 における $\Gamma(0.5, 0.5)$ の場合、発生させた t_{i1} と $t_{i2} - t_{i1}$ の Kendall の順位相関係数の期待値は 0.5 となる。右側打ち切りは $\lambda_c = 0.5, 2$ の指数分布より 2 通り発生させた。各シナリオに対し、1 回あたり 100 名分のデータを発生させ、真値として打ち切りがみられなかった場合の 1 単位時間経過時における繰り返しイベント回数別の人数分布を得た。つづいて右側打ち切りがみられる生成データに対し、 $s = 0, w = 1$ として提案する Aalen-Johansen 推定量に基づいた (14) 式の動的擬似値と Kaplan-Meier 推定量に基づいた (16) 式の動的擬似値を適用し、全員分の動的擬似値の平均をとり、動的予測値を得た。ただし、データ発生の際に、1 単位時間経過より前に全員イベントもしくは打ち切りが発生することにより、いずれかのイベント発生について at risk 数が 1 単位時間以降常に 0 になってしまう場合が生じた。このような場合 (A3) の仮定を満たさず、予測を行うことはできない。そこでこのようなデータセットは棄却し、新たに 100 名分のデータを発生させ、真値、動的予測値の計算を行い、この一連の実験を 1000 回繰り返した。性能評価指標には動的予測値と真値の差の平均であるバイアスと、動

的予測値と真値の差2乗平均の平方根である Root Mean Squared Error(RMSE; 2乗平均平方根誤差)を用いた。前者はゼロに近いほど正確度が高く、後者はなるべく小さな値であると精度がよいと判断される。動的擬似値はアウトカムとしてモデル化されるため、RMSEの小さい動的擬似値のほうが動的予測確率の効率が良いことを意味する。

発生データの要約として、繰り返しイベント回数別の人数分布と、追跡期間である1単位時間に2回目イベントが発生する前に観察された右側打ち切りの割合、発生させた t_1 と $t_2 - t_1$ のKendallの順位相関係数を表2に示す。設定通りの相関のあるイベント時間データが生成され、 $\lambda_c = 0.5$ であれば3割強の対象者に、 $\lambda_c = 2$ であれば8割程度の対象者に打ち切りが観察されたことが確認された。

提案した2通りの動的擬似値について、バイアスとRMSEを評価した結果を表3に示す。いずれのシナリオについて、提案した動的擬似値はともにバイアスがほぼ0であると確認された。すなわち、マルコフ性、セミマルコフ性のいずれの場合であっても、また各イベント時点間に相関がみられる場合であっても、妥当な動的擬似値が得られることがシミュレーション実験からも示された。

RMSEに関して、Aalen-Johansen推定量に基づく動的擬似値のほうがKaplan-Meier推定量に基づくものより、同じか小さな値をとることが、いずれのシナリオについて一貫して確認された。その程度はごくわずかであり、2つの推定量間でRMSEの比をとって表せば、90%から100%であった。

表 2: 終端イベントを考慮しない場合における各シナリオ発生データの要約

u_i	設定したパラメータ		真の繰り返しイベント別の人数分布[%]				途中打ち切り割合 [%]	Kendallの相関係数
	λ_{01}	λ_{02}	λ_c	イベントなし	1回イベント	2回イベント		
1	1	1	0.5	36.7	36.8	26.5	35.7	-0.001
1	1	2	0.5	36.7	23.3	40.0	33.4	-0.001
$\Gamma(0.5, 0.5)$	1	1	0.5	57.5	19.4	23.1	35.2	0.497
$\Gamma(0.5, 0.5)$	1	2	0.5	57.5	13.1	29.4	33.7	0.497
1	1	1	2	36.7	36.8	26.5	81.2	-0.001
1	1	2	2	36.7	23.4	39.9	77.5	-0.001
$\Gamma(0.5, 0.5)$	1	1	2	57.4	19.6	23.0	79.8	0.496
$\Gamma(0.5, 0.5)$	1	2	2	57.4	13.4	29.2	77.0	0.496

表 3: 終端イベントを考慮しない場合におけるバイアスと RMSE に関するシミュレーション結果

設定したパラメータ		Aalen-Johansen推定量に基づく動的擬似値		Kaplan-Meier推定量に基づく動的擬似値					
u_i	λ_{01}	λ_{02}	λ_c	イベントなし	1回イベント	2回イベント	イベントなし	1回イベント	2回イベント
絶対バイアス*									
1	1	1	0.5	-0.0005	0.0004	0.0001	-0.0005	0.0008	-0.0004
1	1	2	0.5	-0.0005	0.0005	-0.0001	-0.0005	0.0011	-0.0006
$\Gamma(0.5, 0.5)$	1	1	0.5	-0.0006	0.0004	0.0002	-0.0006	0.0003	0.0003
$\Gamma(0.5, 0.5)$	1	2	0.5	-0.0006	0.0006	-0.00004	-0.0006	0.0007	-0.0001
1	1	1	2	-0.0034	0.0086	-0.0052	-0.0044	0.0102	-0.0063
1	1	2	2	-0.0057	0.0105	-0.0048	-0.0067	0.0127	-0.0066
$\Gamma(0.5, 0.5)$	1	1	2	-0.0066	0.0112	-0.0047	-0.0066	0.0088	-0.0022
$\Gamma(0.5, 0.5)$	1	2	2	-0.0113	0.0174	-0.0061	-0.0113	0.0129	-0.0016
Root Mean Squared Error (RMSE) [†]									
1	1	1	0.5	0.0282	0.0340	0.0270	0.0282	0.0347	0.0280
1	1	2	0.5	0.0282	0.0304	0.0281	0.0282	0.0328	0.0310
$\Gamma(0.5, 0.5)$	1	1	0.5	0.0260	0.0264	0.0220	0.0260	0.0284	0.0243
$\Gamma(0.5, 0.5)$	1	2	0.5	0.0260	0.0241	0.0245	0.0260	0.0268	0.0270
1	1	1	2	0.0837	0.0969	0.0733	0.0851	0.1004	0.0758
1	1	2	2	0.0840	0.0935	0.0813	0.0855	0.0997	0.0852
$\Gamma(0.5, 0.5)$	1	1	2	0.0702	0.0762	0.0606	0.0703	0.0797	0.0641
$\Gamma(0.5, 0.5)$	1	2	2	0.0695	0.0693	0.0628	0.0696	0.0753	0.0682

* 各イベントに対する動的擬似値の期待値 (動的予測値) と打ち切りのない正確な潜在イベント時点より計算されたイベント確率 (真値) の差の平均;

† 動的予測値と真値の差の 2 乗平均の平方根

3.2 終端イベントを検討する場合

終端イベントが存在する場合において提案する動的擬似値の性能評価を行った。データ発生方法は3.1節同様 frailty モデルに基づく方法とした。繰り返しイベントは最大2回、終端イベントは1種類発生したとし、対象者*i*の1回目のイベント時点 t_{i1} と1回目イベントから2回目のイベントの経過時間 $t_{i2} - t_{i1}$ 、終端イベント発生時点 t_{iD} はそれぞれ以下のハザード $\lambda_{i1}, \lambda_{i2}, \lambda_{iD}$ をパラメータとする指数分布に従って発生させた。

$$\lambda_{i1} = u_i \lambda_{01}, \quad \lambda_{i2} = u_i \lambda_{02}, \quad \lambda_{iD} = u_i \lambda_{0D} \quad (25)$$

ただし、 $\lambda_{01}, \lambda_{02}, \lambda_{0D}$ はベースラインハザードである。イベント間の相関の強さを規定する u_i はガンマ分布に従うとし、3つの時点情報は、互いに等しい相関の強さをもつとして発生させた。3.1節と同じ解釈をもつ4シナリオを設定し、そのパラメータの設定を表4に示す。予測評価を行えるデータ生成と真値、動的予測値の計算を1000回繰り返し、バイアスとRMSEを基準としてその性能を評価した。

表 4: 終端イベントを考慮した場合の各シナリオにおけるパラメータ値

シナリオ	u_i	λ_{01}	λ_{02}	λ_{0D}
異質性なし, マルコフ性	1	1	1	0.3
異質性なし, セミマルコフ性	1	1	2	0.3
異質性あり, マルコフ性	$\Gamma(0.5, 0.5)$	1	1	0.3
異質性あり, セミマルコフ性	$\Gamma(0.5, 0.5)$	1	2	0.3

発生データの要約として、イベント別の人数分布と、追跡期間である1単位時間に終端イベントが発生する前に観察された右側打ち切りの割合、発生させた t_1 、 $t_2 - t_1$ 、 t_D 間の Kendall の順位相関係数を表5に示す。設定通りの相関のあるイベント時間データが生成され、 $\lambda_c = 0.5$ であれば3割強の対象者に、 $\lambda_c = 2$ であれば8割弱の対象者に打ち切りが観察されたことが確認された。

提案した動的擬似値について、バイアスと RMSE を評価した結果を表6に示す。バイアスに関しては、いずれのシナリオにおいてほぼ0であることが確認された。RMSE については、イベント数や人数分布が異なるとはいえ、3.1節の終端イベントを考慮しない場合の提案する動的擬似値と同程度の大きさであり、著しく外れた値は確認されなかった。

以上より、終端イベントの有無それぞれに対し提案した動的擬似値は、ランドマーク時に100例の at risk 数程度あれば、約8割に打ち切りが観察される状況であっても、バイアスは問題とならないことを確認した。この内容はイベント間に相関がある場合や、マルコフ性が成立しない状況においても成立した。

表 5: 終端イベントを考慮する場合における各シナリオ発生データの要約

u_i	設定したパラメータ			真のイベント別人数分布[%]				途中打ち切り割合 [%]	Kendallの相関係数			
	λ_{01}	λ_{02}	λ_{0D}	λ_c	なし	1回	2回		終端	(t_1, t_2-t_1)	(t_1, t_D)	(t_2-t_1, t_D)
1	1	1	0.3	0.5	27.3	27.2	19.5	26.0	34.1	-0.0003	0.002	0.001
1	1	2	0.3	0.5	27.3	17.1	29.5	26.0	34.1	-0.0003	0.002	0.001
$\Gamma(0.5, 0.5)$	1	1	0.3	0.5	52.7	14.4	11.8	21.1	35.2	0.501	0.500	0.499
$\Gamma(0.5, 0.5)$	1	2	0.3	0.5	52.7	10.3	16.0	21.1	35.2	0.501	0.500	0.499
1	1	1	0.3	2	27.2	27.2	19.6	25.9	77.9	-0.001	0.000	0.002
1	1	2	0.3	2	27.2	17.4	29.5	25.9	77.7	-0.001	0.001	0.004
$\Gamma(0.5, 0.5)$	1	1	0.3	2	52.3	14.7	12.2	20.9	78.6	0.499	0.498	0.497
$\Gamma(0.5, 0.5)$	1	2	0.3	2	52.2	10.8	16.2	20.9	78.5	0.497	0.495	0.495

表 6: 終端イベントを考慮しない場合におけるバイアスと RMSE に関するシミュレーション結果

u_i	設定したパラメータ			提案する動的擬似値					
	λ_{01}	λ_{02}	λ_{0D}	λ_c	イベントなし	1回イベント	2回イベント	終端イベント	
絶対バイアス*									
1	1	1	0.3	0.5	0.0003	-0.00002	0.0006	-0.0008	
1	1	2	0.3	0.5	0.0003	-0.0004	0.0010	-0.0008	
$\Gamma(0.5, 0.5)$	1	1	0.3	0.5	0.0011	-0.0002	-0.0011	0.0002	
$\Gamma(0.5, 0.5)$	1	2	0.3	0.5	0.0011	-0.0002	-0.0010	0.0002	
1	1	1	0.3	2	-0.0025	-0.0017	0.0040	0.0002	
1	1	2	0.3	2	-0.0057	0.0088	-0.0032	0.0001	
$\Gamma(0.5, 0.5)$	1	1	0.3	2	-0.0081	0.0086	0.0055	-0.0060	
$\Gamma(0.5, 0.5)$	1	2	0.3	2	-0.0094	0.0137	0.0010	-0.0053	
Root Mean Squared Error (RMSE) [†]									
1	1	1	0.3	0.5	0.0264	0.0316	0.0241	0.0245	
1	1	2	0.3	0.5	0.0264	0.0270	0.0281	0.0245	
$\Gamma(0.5, 0.5)$	1	1	0.3	0.5	0.0261	0.0238	0.0207	0.0211	
$\Gamma(0.5, 0.5)$	1	2	0.3	0.5	0.0261	0.0214	0.0225	0.0209	
1	1	1	0.3	2	0.0724	0.0836	0.0697	0.0681	
1	1	2	0.3	2	0.0704	0.0749	0.0783	0.0687	
$\Gamma(0.5, 0.5)$	1	1	0.3	2	0.0703	0.0644	0.0526	0.0562	
$\Gamma(0.5, 0.5)$	1	2	0.3	2	0.0701	0.0609	0.0620	0.0575	

* 各イベントに対する動的擬似値の期待値 (動的予測値) と打ち切りのない正確な潜在イベント時点より計算されたイベント確率 (真値) の差の平均; † 動的予測値と真値の差の 2 乗平均の平方根

4 実データへの適用：大腸がん肝転移切除例の追跡データ

大腸がんはしばしば血行性転移をきたし、このうち肝臓への転移が最も頻度が高い。大腸がんの診断時において肝転移の有病割合は約 10% 存在し、大腸がんの治癒切除後に発生する再発の約 4 割が肝再発である。根治切除可能な肝転移には肝切除が推奨される。このとき、耐術可能、原発巣の制御、肝転移巣を遺残なく切除可能、肝外転移がないか制御可能、十分な残肝機能といった肝切除術の適応基準となる [84]。肝切除施行割合は、欧米より日本のほうが高く [113]、また国内でも施設により異なりが生じる。

大腸がん肝転移切除例の特徴のひとつとして、たとえ初回完全切除された症例においても、45～80% で再発がみられる。残肝再発に対しても、根治を得る可能性がある治療は外科切除のみであり、積極的に行われるものの、再切除が行われる割合は 2 割以下とされている [114, 115]。また肺転移や腹膜播腫のような再発であっても、可能であれば切除術は有効と考えられている。

肝切除後の生存に関する予後因子については、原発巣進行度、腫瘍個数、外科切除断端、肝所属リンパ節転移、肝外病巣、衛星病変が挙げられ、再切除症例についても同様の予後因子が挙げられている [113]。一方、死亡だけでなく再発を含めた総合的な予後の評価に関する研究はあまりみられない。本節では、大腸がん肝転移切除例の追跡データに提案法を適用し、再発の種類や再発時の腫瘍個数、再発回数といった情報を用いて死亡や再発の予後を動的予測した。

4.1 背景

1996年1月から2010年12月までの間に東京大学医学部附属病院肝胆膵外科にて大腸がん肝転移の先行肝切除術を受けた336名を対象とした。このうち、医師の判断や患者希望により術後補助化学療法が行われた33例、大腸癌肝転移に対する肝切除後の補助化学療法(LV+UFT)の有効性に関する比較研究(UMIN-CTR:C000000013)に登録された40例を除く、263例を解析対象集団とした。対象者の臨床情報として、初回肝切除時(以降、ベースライン時とよぶ)における年齢、性別、診断時腫瘍個数、診断時腫瘍径、術式、手術所見のほか、再発に対する再切除時における診断時腫瘍個数、診断時腫瘍径、術式、手術所見がすべて記録された。対象者のベースライン時における背景情報を表7に示す。

初回肝切除の後、再発回数の最大値は4回であった。以降、本研究では、再発後に再切除が実施された場合を切除可能再発、そうでない場合を切除不能再発と呼ぶ。再発回数を切除有無別に集計した分割表の結果を表8に示す。前回肝切除時の腫瘍個数(単発/多発)や腫瘍径(2cm以下/2cm超)によって、再発の起こりやすさ、さらにその再発が切除不能と判断されたか否かに違いがみられるかを確認すべく、サブグループ別の集計結果も併せて表8に示す。

全追跡期間において、初回か再発後かによらず肝切除を受けた対象者のうち、約四分之三の対象者は再発を(繰り返し)起こすことが確認された。前回肝切除時が単発よりも多発であるほうが、また2cm以下よりも2cm超のほうが再発を

起こしやすいことがいずれの再発回数においても一貫して確認された。

再発のうち、切除不能と判断された例は約半数を占め、3回目再発までは再発回数に依らずにこの傾向はみられた。4回目再発は全員が切除不能と判断されたが、わずか10例であった。また、1,2,3回目再発それぞれについて、前回切除時の個数(単発/多発)や腫瘍径(≤2cm/ >2cm)と再発が切除不能と判断されたか否かの間に、有意な関連はみられなかった。

表 7: 大腸がん肝転移切除例の追跡データ：ベースライン時における背景情報

項目	(N=263)	
年齢(歳)*	63	[29-83]
性別†		
男	177	[67.3%]
女	86	[32.7%]
診断時腫瘍個数†		
1個(単発)	135	[51.3%]
2個以上(多発)	128	[48.7%]
診断時腫瘍径†		
≤2cm	71	[27.0%]
2cm<, ≤5cm	131	[49.8%]
5cm<	61	[23.2%]
手術所見†		
R0	235	[89.4%]
R1	21	[8.0%]
R2	7	[2.7%]
術式†		
limited resection	167	[63.5%]
hemihepatectomy	53	[20.2%]
segmentectomy & sectionectomy	43	[16.3%]
追跡期間(年)*	4.5	[0.3-16.6]

* 中央値 [最小値-最大値], † 人数 [割合(%)]

表 8: 大腸がん肝転移切除例の追跡データ：再発回数集計結果

	再発なし		再発あり		切除不能再発		切除実施再発		片側 p値*
	N	[%]	N	[%]	N	(%)	N	(%)	
1回目再発	61	[23.2]	202	[76.8]	94	(46.5)	108	(53.5)	
初回肝切除時の腫瘍個数									
単発	41	[30.4]	94	[69.6]	45	(47.9)	49	(52.1)	0.42
多発	20	[15.6]	108	[84.4]	49	(45.4)	59	(54.6)	
初回肝切除時の腫瘍径									
≤2cm	21	[29.6]	50	[70.4]	23	(46.0)	27	(54.0)	0.53
>2cm	40	[20.8]	152	[79.2]	71	(46.7)	81	(53.3)	
2回目再発	29	[26.9]	79	[73.1]	39	(49.4)	40	(50.6)	
1回目再発時の腫瘍個数									
単発	24	[40.0]	36	[60.0]	17	(47.2)	19	(52.8)	0.45
多発	5	[10.4]	43	[89.6]	22	(51.2)	21	(48.8)	
1回目再発時の腫瘍径									
≤2cm	19	[30.2]	44	[69.8]	25	(56.8)	19	(43.2)	0.10
>2cm	10	[22.2]	35	[77.8]	14	(40.0)	21	(60.0)	
3回目再発	10	[25.0]	30	[75.0]	15	(50.0)	15	(50.0)	
2回目再発時の腫瘍個数									
単発	9	[36.0]	16	[64.0]	7	(43.8)	9	(56.3)	0.36
多発	1	[6.7]	14	[93.3]	8	(57.1)	6	(42.9)	
2回目再発時の腫瘍径									
≤2cm	7	[29.2]	17	[70.8]	8	(47.1)	9	(52.9)	0.50
>2cm	3	[18.8]	13	[81.3]	7	(50.0)	6	(50.0)	
4回目再発	5	[33.3]	10	[66.7]	10	(100.0)	0	(0.0)	
3回目再発時の腫瘍個数									
単発	4	[40.0]	6	[60]					
多発	1	[20.0]	4	[80]					
3回目再発時の腫瘍径									
≤2cm	5	[41.7]	7	[58.3]					
>2cm	0	[0.0]	3	[100.0]					

* 切除不能再発/切除実施再発に対するFisherの正確検定

4.2 切除不能再発発生による死亡に関する動的予測

がん臨床試験、特に術後補助療法試験では死亡までの時間 (Overall Survival; OS) が本来注目したいエンドポイントである場合が多いが、追跡の時間やコストを削減するために、代替エンドポイントの利用がしばしば検討される。大腸がん肝転移もその例に洩れず、肝切除後初回再発までの時間である RFS (Recurrence-free survival) が OS の代替エンドポイントとして検討されるものの、両者の結果にはかい離がみられるという指摘がある。その理由のひとつに、再発病変に対する再切除術が生存予後により効果を与えるという仮説がある。もしそうであれば、再切除術が行えない再発発生までの時間 (Time to Surgical Failure; TSF) が新たな代替エンドポイントとして有用と考えられる [1]。

そこで動的予測のアプローチによって RFS と TSF の間で代替エンドポイント性能を比較した。RFS や TSF は追跡途中で対象者が at risk である条件下で発生するイベントであるために、1.1 節で述べた内生変数に分類される。それゆえ通常の Cox モデルにて時間依存性共変量として扱うことはできない。そこで、適当に定めたランドマーク時点までに代替エンドポイント (初回再発あるいは切除不能再発) が発生したかにより、ランドマーク時点以降の OS を比較する動的予測アプローチにより、その代替性を検討する。

死亡のような単一イベントに対する動的予測モデルは Cox モデルを拡張した方法が提案されているが [10]、この状況は本研究で提案した動的擬似値法において、

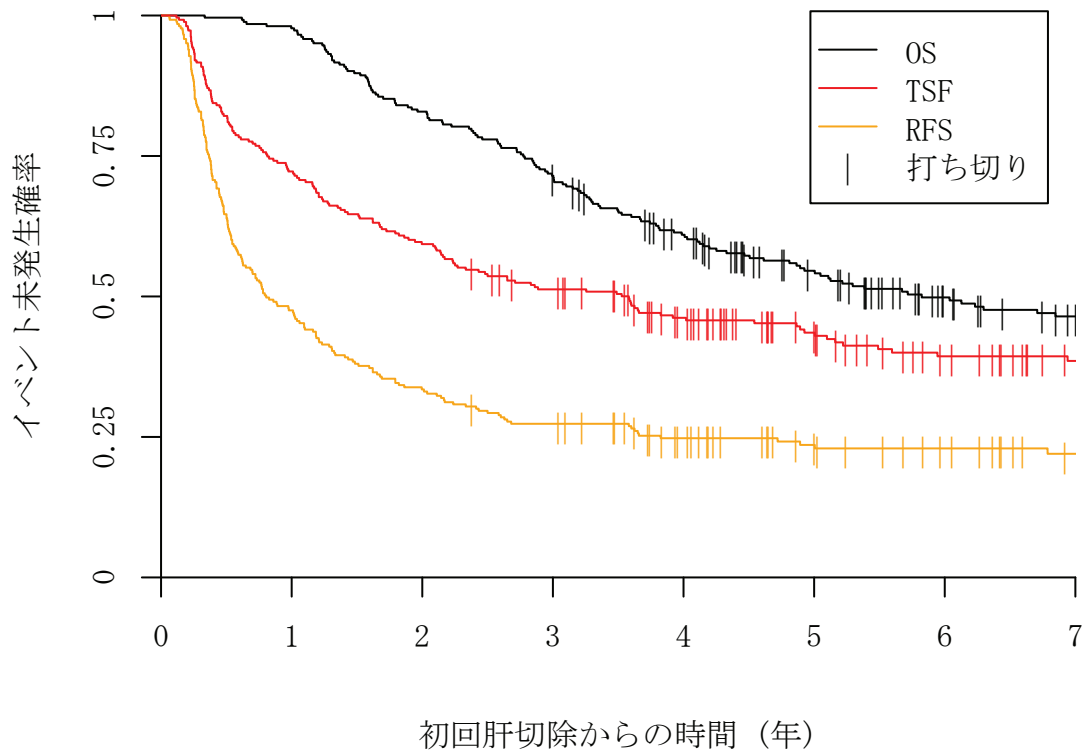
繰り返しイベント回数が1回のみという最も簡単な場合とみなすことができる。

しかもこの場合、Aalen-Johansen 推定量に基づく動的擬似値と Kaplan-Meier 推定量に基づく動的擬似値は同じになる。そこで、Cox モデルに基づく従来法と、本研究での提案法の両方を用いることとした。なお、本研究で用いたデータにおける時点の扱いは、大腸がん肝転移初回肝切除日を起算日とし、初めて再発が診断された日を RFS イベント発生時点、後につづく再切除が行われなかったような再発が診断された日を TSF イベント発生時点とした。

イベント発生状態を記述するため、OS,RFS,TSF の各エンドポイントについて Kaplan-Meier 推定を行った結果を図 5 に示す。RFS に注目すると、初回再発の多くは初回肝切除後 3 年以内に発生がみられた。一方 TSF や OS に注目すると、切除不能再発や死亡の発生は 3 年以降でもみられた。

続いて RFS イベント、TSF イベント発生有無を用いて、ランドマーク時点後 3 年死亡確率を動的予測した。まず、ランドマーク時点において死亡していないという at risk の下で、ランドマーク時点までに発生した RFS,TSF イベントを用いて対象者は以下の 3 通りに分類された。

- 初回肝切除後一度も再発を起こしていない (RFS-TSF-)
- 再発は起こしたものの、切除不能再発は起こしていない (RFS+TSF-)
- 切除不能再発を起こした (RFS+TSF+)



	初回肝切除からの時間 (年)							
	0	1	2	3	4	5	6	7
OS	263	257	218	187	151	120	92	77
TSF	263	190	156	131	103	77	59	47
RFS	263	125	88	71	53	38	31	22
	Number of at risk							

図 5: OS,TSF,RFS についてイベント未発生確率に関する Kaplan-Meier 推定値

サブグループを分類するためのランドマーク時点を 1,2,3 年の 3 通りとし、それぞれ Kaplan-Meier 推定値をプロットした結果を図 6 に示す。

図 6 より、RFS+TSF+ である患者の生存予後は極めて悪く、次いで RFS+TSF-、RFS-TSF- の順に予後はよい。ランドマーク時点 1,2,3 年とした時の、ランドマーク時点後 3 年以内の死亡確率 (1-(生存確率)) はそれぞれ、RFS-TSF- で 12.9%,9.4%,7.8%、RFS+TSF- で 31.4%,22.0%,12.4%、RFS+TSF+ で 91.3%,84.9%,86.8% であった。RFS+TSF- に属するような対象者は、ランドマーク時点が経過するごとに 3 年死亡確率が下がり、RFS-TSF- に属する患者のそれに近づく。すなわち、たとえ再

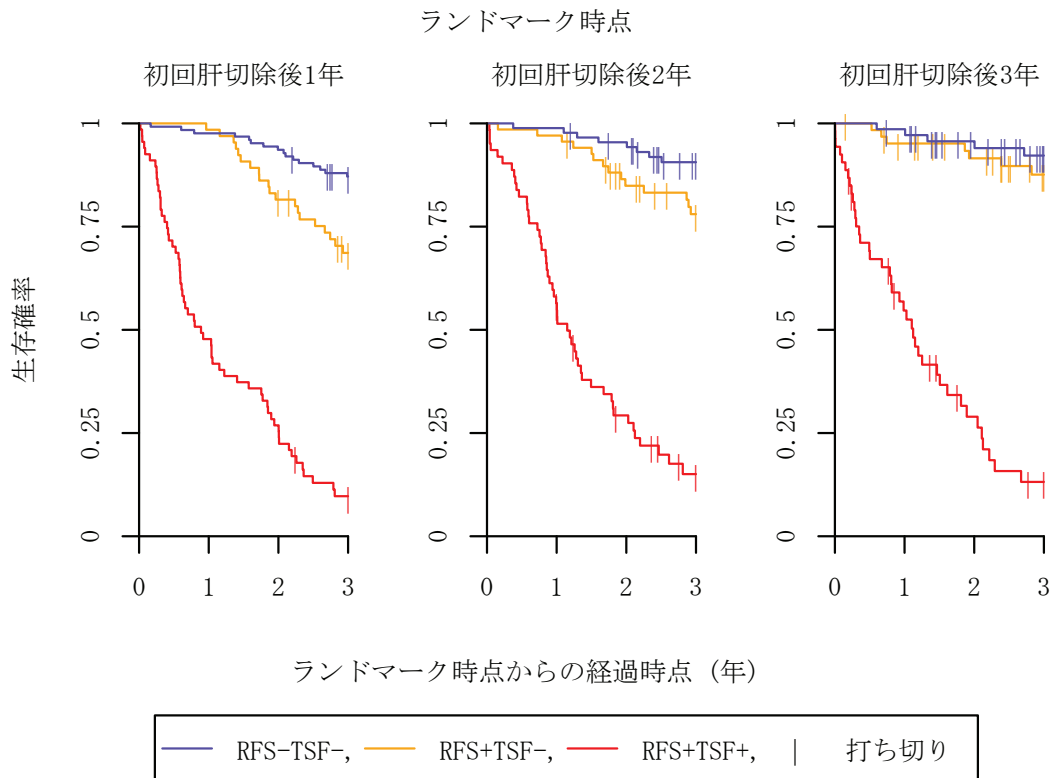


図 6: 1,2,3 年ランドマーク時における OS,TSF,RFS についてイベント未発生確率に関する Kaplan-Meier 推定値

発を起こしたとしても切除が実施されたならば、切除不能な場合に比べ生存予後は大幅によいと考えられ、さらに生存したまま初回肝切除からの時間が経過するほど、一度も再発を起こしていない場合の生存予後並みに改善されることが確認された。

ランドマーク時点を細かく変化させ、それぞれの時点について3年死亡確率を求めた。RFS,TSF イベント発生有無のダミー変数を共変量としたCoxモデルに基づく従来法 [10] と本研究で提案した動的擬似値に基づく方法で予測した確率をそれぞれ、図 7(a),(b) に示す。ランドマーク時点は初回肝切除後1年から4年にわたって0.25年おきに設定し、それぞれの時点にて得られた結果の点推定値と

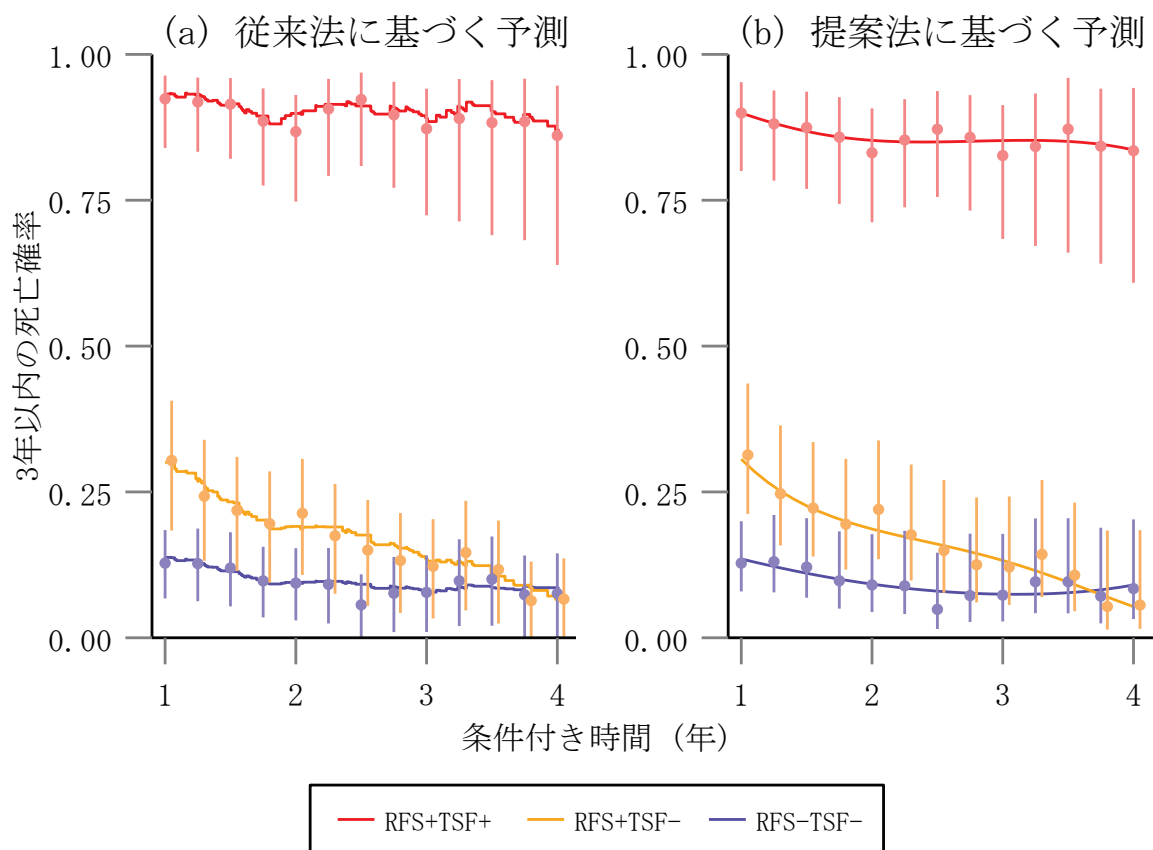


図 7: RFS,TSF イベント有無組み合わせ別の 3 年死亡確率に関する動的予測

95% 信頼区間を点とエラーバーを用いて積み上げプロットした。なお、提案法では、リンク関数 $g(\cdot)$ にロジット関数を用いた。さらに、スーパーモデルにて予測確率を 3 次スムージング、すなわち時点の多項式を $f_1(s) = s, f_2(s) = s^2, f_3(s) = s^3$ とした結果を曲線を用いて重ねて示し、その推定値を表 9(a) に示す。

図 7 から、いずれのランドマーク時点であっても RFS+TSF+ の 3 年死亡確率は 80% 以上と極めて高いことが確認された。RFS+TSF- と RFS-TSF- を比べると、初回肝切除後 1,2 年のランドマーク時点では、RFS+TSF- のほうがより 3 年死亡確率が高いものの、4 年あたりのランドマーク時点では、3 年死亡確率はほぼ等し

表 9: 死亡を動的予測するための提案したスパーモデルの推定結果

	(a) 両イベントの組み合わせ		(b) 再発のみ		(c) 切除不能再発のみ	
	推定値	標準誤差	推定値	標準誤差	推定値	標準誤差
intercept	-1.44	1.79	-1.44	1.79	-0.72	1.03
再発 (あり/なし)						
intercept	2.37	2.42	2.80	1.98		
s	-2.24	3.68	-1.03	3.07		
s^2	1.09	1.66	0.61	1.41		
s^3	-0.18	0.23	-0.11	0.20		
切除不能再発 (あり/なし)						
intercept	2.69	2.40			4.34	2.15
s	0.46	3.52			-1.22	3.19
s^2	-0.19	1.57			0.57	1.46
s^3	0.05	0.21			-0.07	0.20
ランドマーク時点						
s	-0.32	2.81	-0.32	2.81	-0.89	1.62
s^2	-0.13	1.30	-0.13	1.30	0.21	0.75
s^3	0.04	0.18	0.04	0.18	-0.02	0.10

くなることが確認された。図 7(a) と (b) の間では、スムージングした曲線のなめらかさに違いはあるものの、予測された死亡確率に大きな違いはみられなかった。

次に、RFS イベント発生有無のみで死亡予測を行った場合と、TSF イベント発生有無のみで同様に行った場合を比較することで、どちらのエンドポイントが死亡予測性能が高いかを評価した。図 7 と同様の解析を行い、プロットした結果を図 8、図 9 に示す。また提案法の推定結果を表 9(b),(c) に示す。

図 8 にて、RFS+ である場合 3 年死亡確率は 60% から 40% 程度であり、時間経過とともに低下する傾向がみられた。一方、RFS- であれば、3 年死亡確率は 10% 台と低い。図 9 にて、TSF+ である場合は 3 年死亡確率は 80% 以上であれば、TSF- である場合、死亡確率は 20% 前後と低い。

図 8,9 を比較すると TSF 有無のほうが RFS 有無より 3 年死亡確率の別れ方が大きい。そこで予測性能を確認するため、Harrell's c-index [116, 117] を用いて評価した。これは 2 値判別性能評価にて頻用される ROC 曲線下面積 (AUC) を、生存時間データに拡張したものと捉えることができ、0.5 のとき予測性能なし、1 のときに完全な予測性能をもつと解釈できる。実際の c-index は、打ち切り分布に依らずに一致推定されるよう改良された Uno らの方法 [118] で計算した。図 7-9 の各 (b) に示した、提案法に基づくランドマーク時点別の予測結果について、c-index を計算した結果を図 10 に示す。横軸にランドマーク時点、縦軸に c-index の値をプロットし折れ線で結んだ。

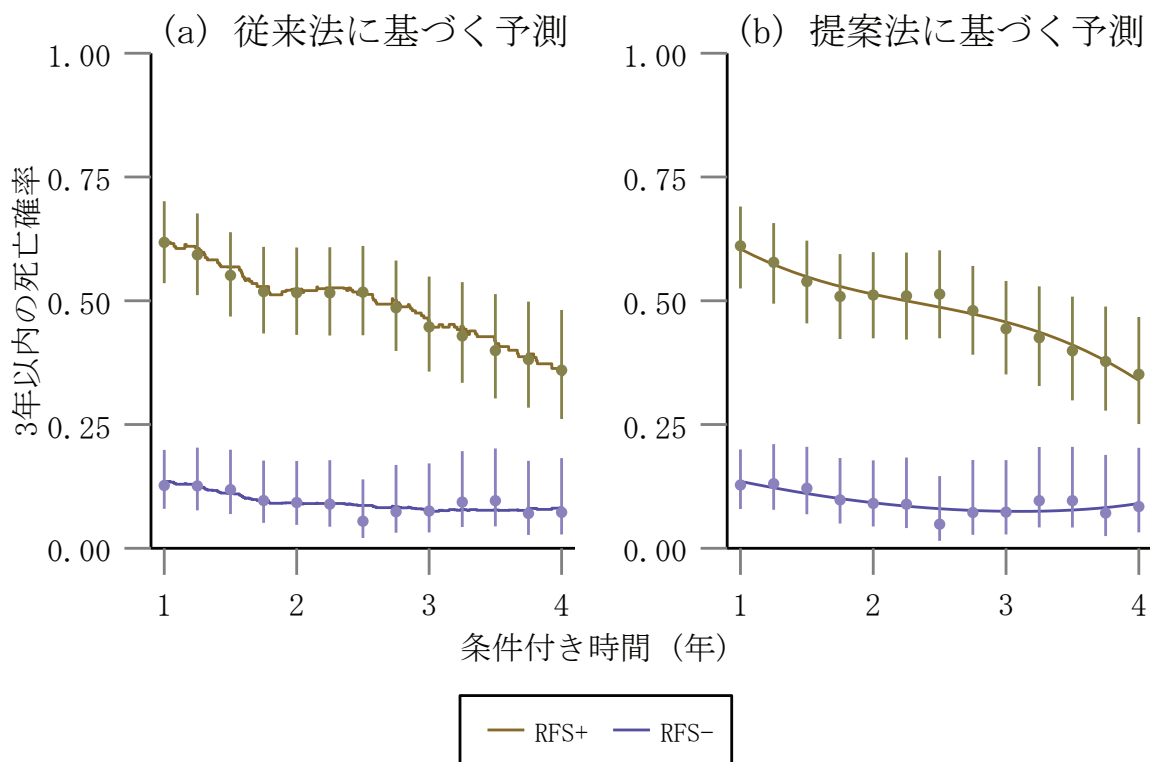


図 8: RFS 有無別の 3 年死亡確率に関する動的予測

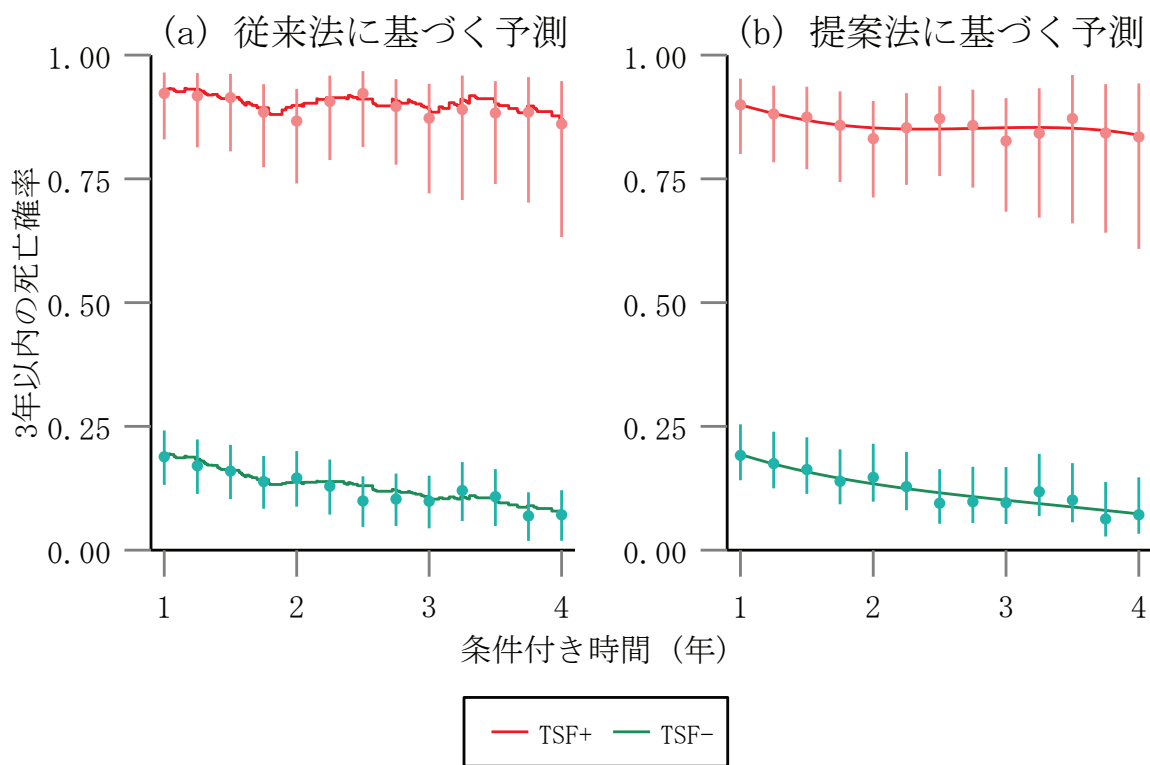


図 9: TSF 有無別の 3 年死亡確率に関する動的予測

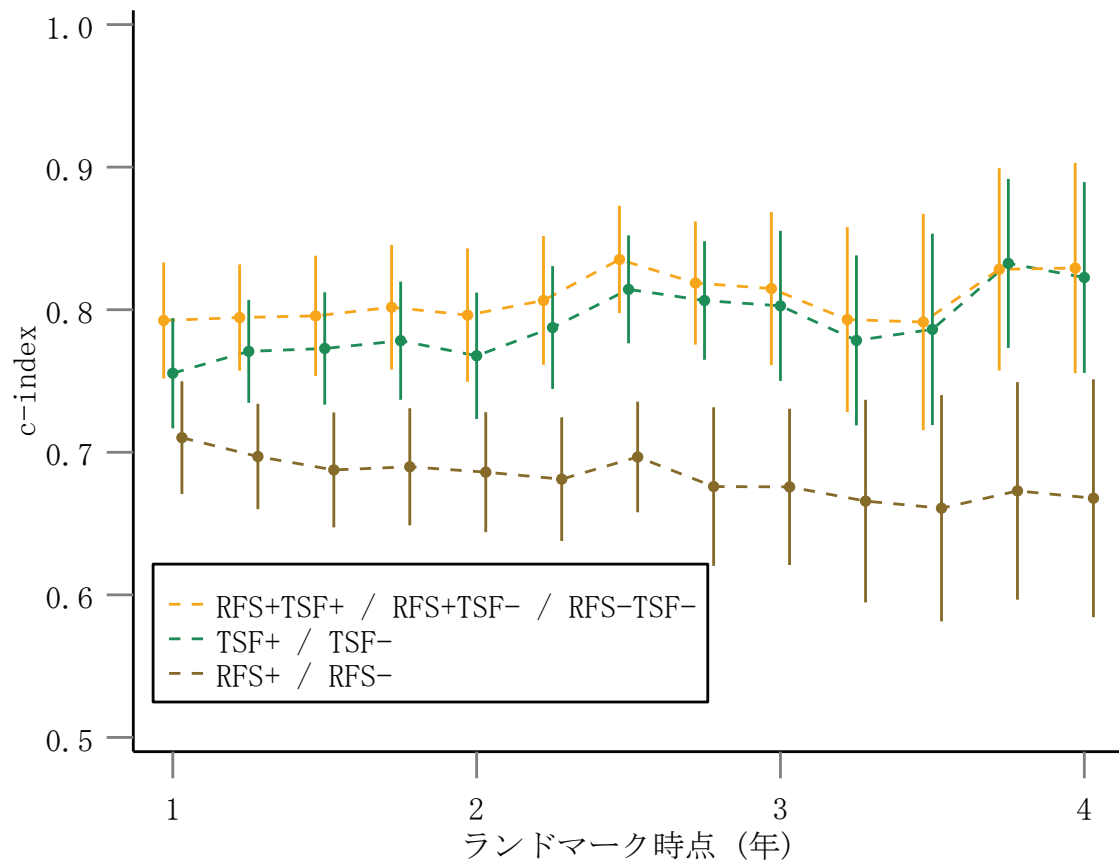


図 10: RFS,TSF 有無別の c-index を用いた OS 予測性能評価

点が c-index の点推定値、縦棒がその 95% ブートストラップ信頼区間を表す。

図 10 より、RFS+/-で死亡予測を行うと c-index は 0.7 程度であるが、TSF+/-で行えば c-index は 0.75 から 0.85 程度と、いかなるランドマーク時点においても性能がよいことが確認された。橙色で示した RFS と TSF の組み合わせによって最も細かくサブグループに分類した場合は最も予測性能がよいと考えられるが、これとほぼ同等な予測性能を TSF+/-の場合は、3 年ランドマーク時点以降持っていた。

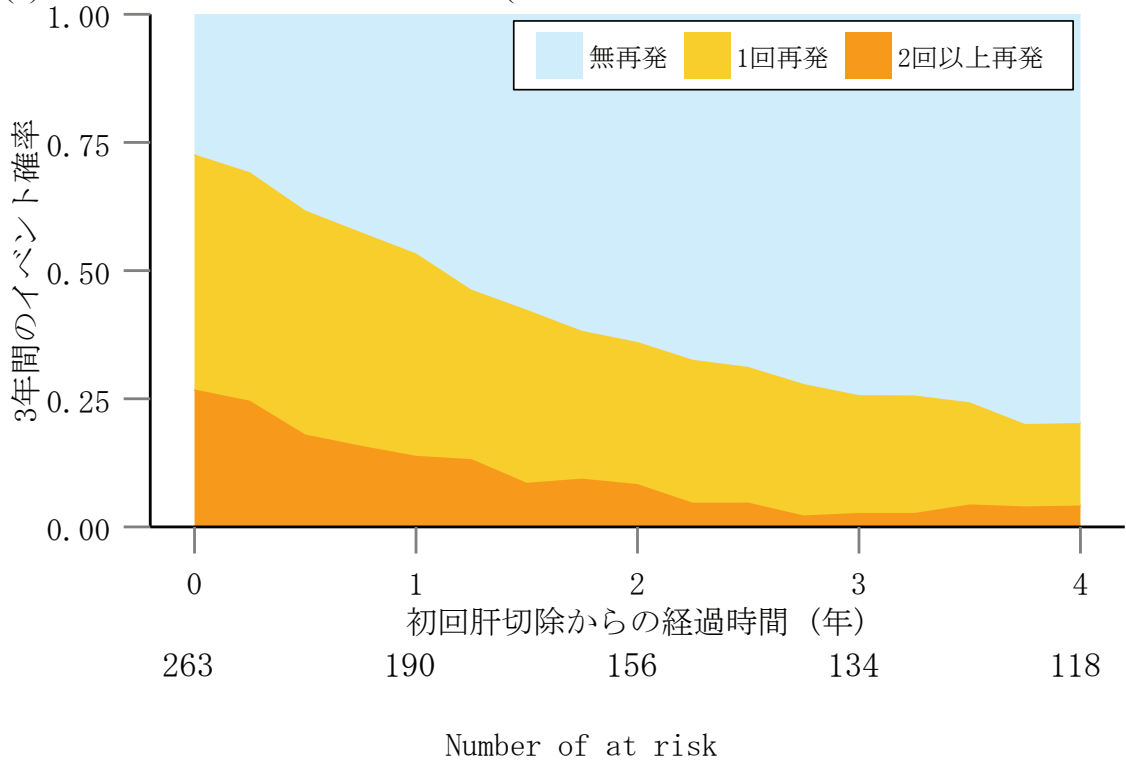
4.3 切除不能再発が起こっていない下での再発に関する動的予測

切除不能再発や死亡が起こっていない限り、対象者は再発を起こしうる。どのような対象者が再発を起こしやすいかを、前回肝切除時に判明した情報を用いて予測することができれば、肝切除後のフォローアップに有用な情報となりうる。そこで、前回肝切除時の腫瘍個数や再発回数を用いて3年イベント確率を予測した。ここでの at risk であるとは、切除不能再発と死亡が起きていない、すなわち再発を起こしうる状態と定義し、ランドマーク時点は3か月ごとに設定した。

まず、対象データの3年イベント確率に関する積み上げ面グラフを図11に示す。図11のうち、死亡を打ち切りとみなした場合が(a),(b)、終端イベントとして扱った場合が(c)である。提案した Aalen-Johansen 推定量に基づく動的擬似値を利用した結果を(a)、Kaplan-Meier 推定量に基づく動的擬似値を利用した結果を(b)に示した。横軸にランドマーク時点、縦軸にランドマーク時点以降3年間に発生するイベントの確率を死亡、2回以上再発、1回再発、無再発の順に積み上げて示した。ここでは再発の有無に依らず3年以内に死亡すれば死亡イベントに分類し、他の3つのイベントはすべて3年以内の死亡は起こらない場合にいずれかに分類した。またイベント分類に用いた再発回数は、累積再発回数ではなく、ランドマーク時点以降に発生した再発回数として定義した。

図11より、初回肝切除直後から3年間において無再発である確率は1/4程度であるが、時間が経過するにつれ、at riskのうち3年間に再発が起こらない集団

(a) 死亡を打ち切りとみなした場合 (Aalen-Johansen 推定量に基づく動的擬似値)



(b) 死亡を打ち切りとみなした場合 (Kaplan-Meier 推定量に基づく動的擬似値)

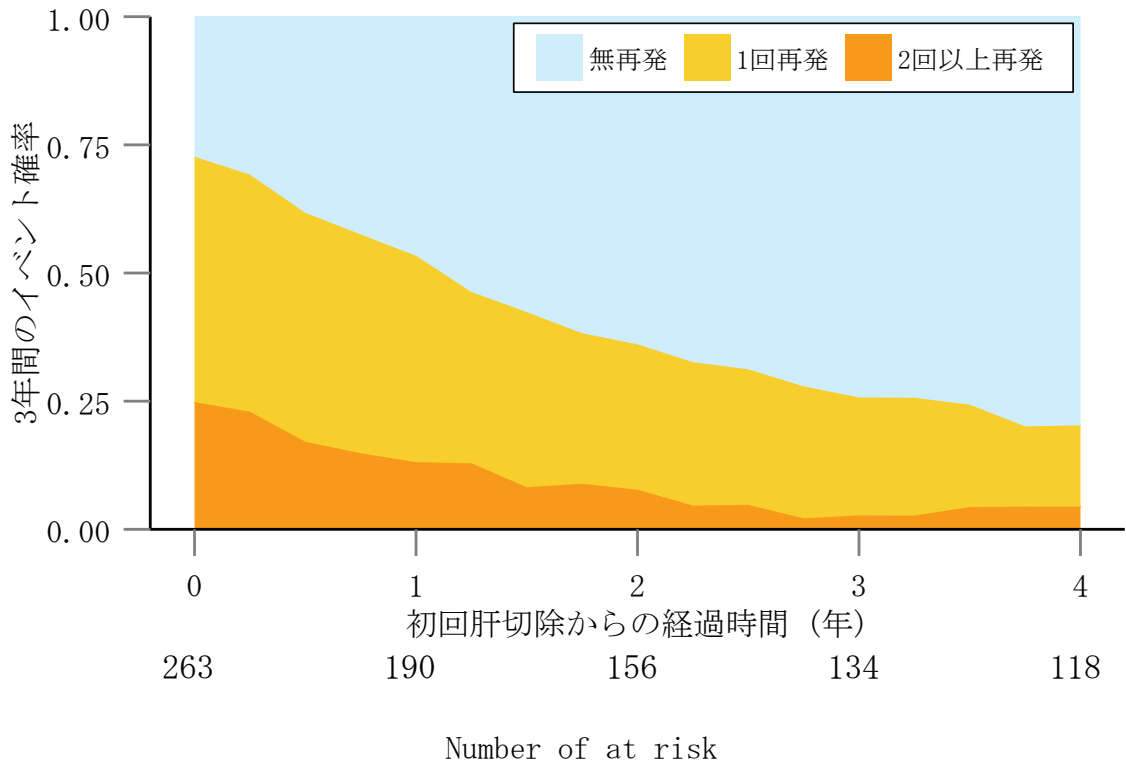


図 11: 大腸がん肝転移切除例の追跡データにおける切除不能再発と死亡が起こっていない条件下での3年イベント確率に関する積み上げ面グラフ

(c) 死亡を終端イベントとして扱った場合

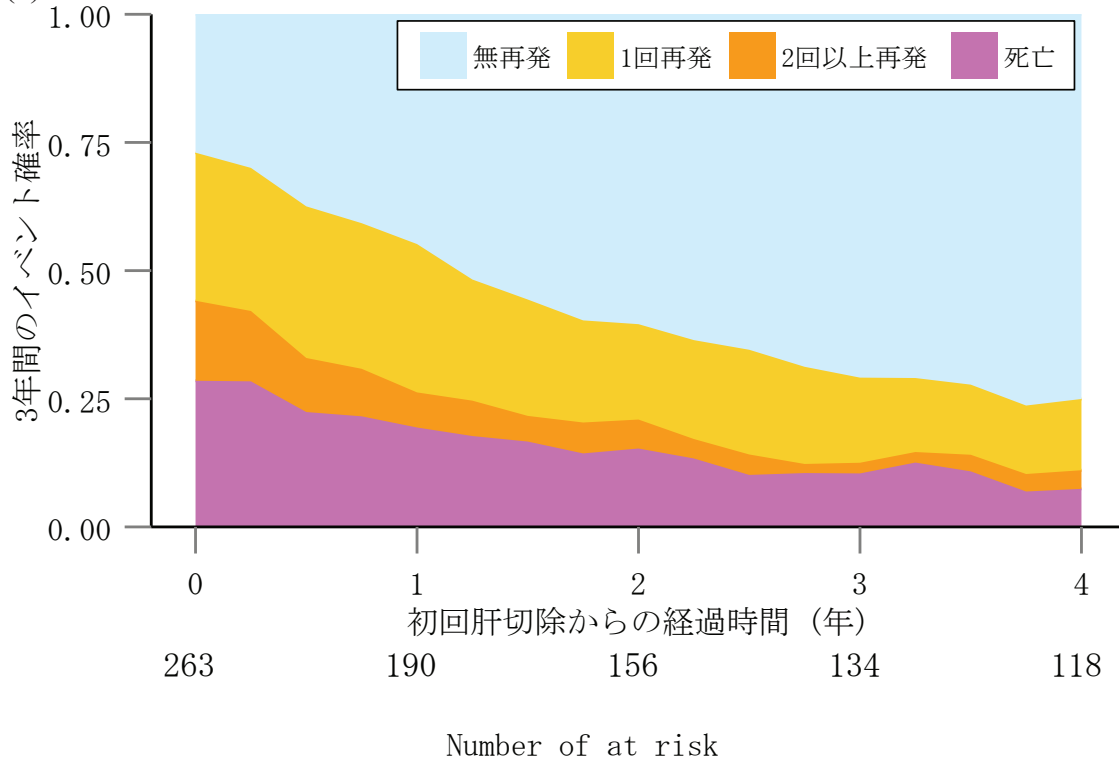


図 11: 大腸がん肝転移切除例の追跡データにおける切除不能再発と死亡が起っていない条件下での3年イベント確率に関する積み上げ面グラフ (続き)

の割合は増加し、初回肝切除後4年経過において、その後3年間の無再発確率は3/4程度となることが確認された。再発を起こしうる条件の下では、初回肝切除直後にて再発や死亡が最も発生しやすく、時間経過とともにこれらイベントは発生しにくくなるといえた。

図 11(a) と (b) を比べると、すべての時点に関して予測されるイベント確率はほぼ等しかった。(a) と (c) を比べると、いずれの時点において無再発確率は両者に違いがみられず、1回再発、2回以上再発確率だけが(a)よりも(c)の場合に大きくなった。本追跡データで死亡が確認された141例中、137例(97.2%)が死亡

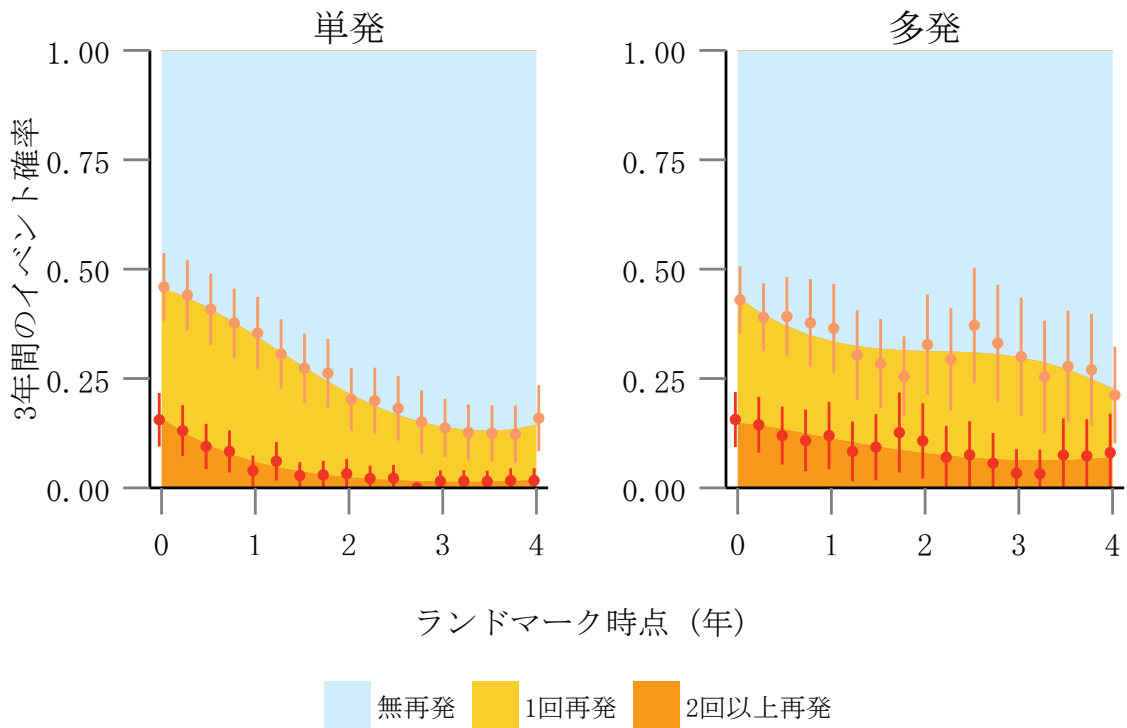
の前に少なくとも1回以上の再発を起こしたことから、(c)にて死亡に分類される対象者の多くが(a)では1回再発ないし2回以上再発を起こした下で死亡による打ち切りを受けたことによる。

以降、提案手法を適用した結果を示す。イベント確率の予測にはすべて、提案した動的擬似値をアウトカムとした一般化ロジットモデルを用いた。死亡を終端イベントせず打ち切りとして扱う場合は、Aalen-Johansen 推定量に基づいた動的擬似値の結果のみ示した。この動的擬似値と Kaplan-Meier 推定量に基づいた動的擬似値の性質は極めて似ていたことがこれまでで示されたためである。ランドマークスーパーモデルによるスムージングを行う際、時点の多項式は3次式、すなわち $f_1(s) = s, f_2(s) = s^2, f_3(s) = s^3$ とし、(21)式のように共変量 \tilde{Z}_i^* を加工した。

まず、前回切除時の腫瘍個数を共変量として3年イベント確率を予測した。腫瘍個数は1個すなわち単発であれば0、2個以上すなわち多発であれば1をとるようダミー変数に加工し、モデルに組み込んだ。積み上げグラフを用いて、ランドマーク時点別に得られた結果の点推定値と95%信頼区間を点とエラーバーを用いて積み上げプロットし、スーパーモデルにてスムージングした点推定値を面グラフとして重ねて描いたものを図12、スーパーモデルでの推定結果を表10に示す。

まず、表10よりすべての説明変数に対し回帰係数が推定され、少なくとも著しく不安定な結果ではないことが確認された。図12よりすべてのランドマーク時点を通して、単発に比べ多発であると、3年後の再発や死亡リスクが高いこと

(a) 死亡を打ち切りとみなした場合



(b) 死亡を終端イベントとして扱った場合

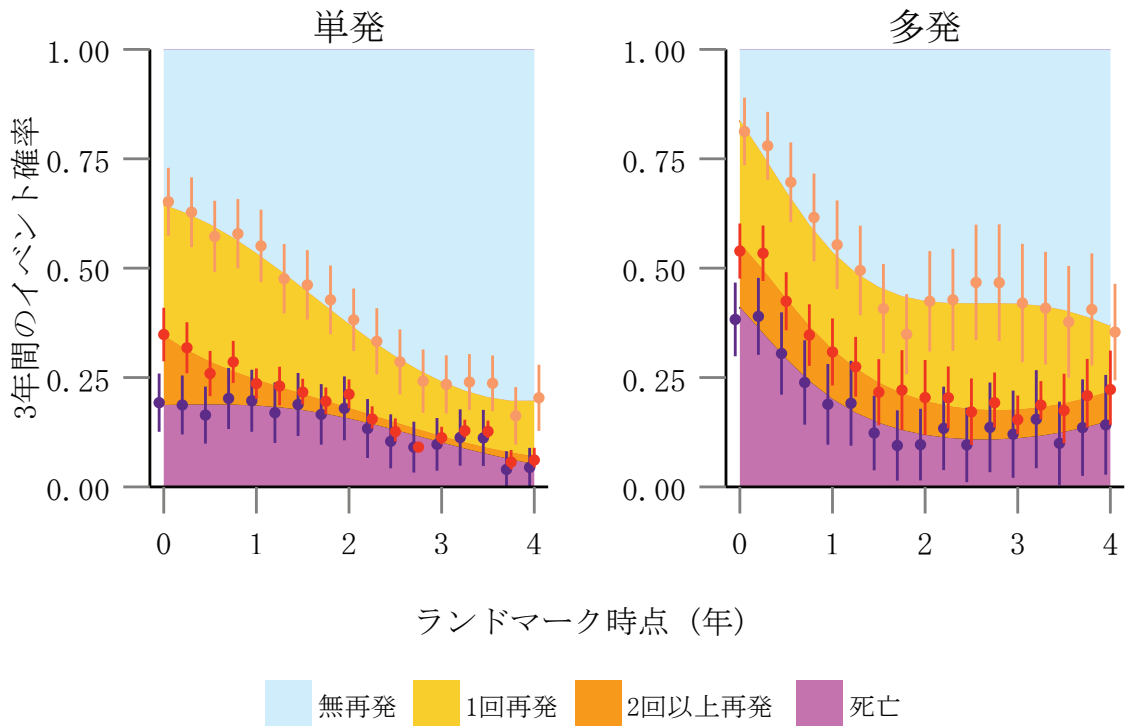


図 12: 前回切除時腫瘍個数別の 3 年イベント確率に関する積み上げグラフ

表 10: 前回切除時腫瘍個数によって再発を動的予測するためのスーパーモデルの推定結果

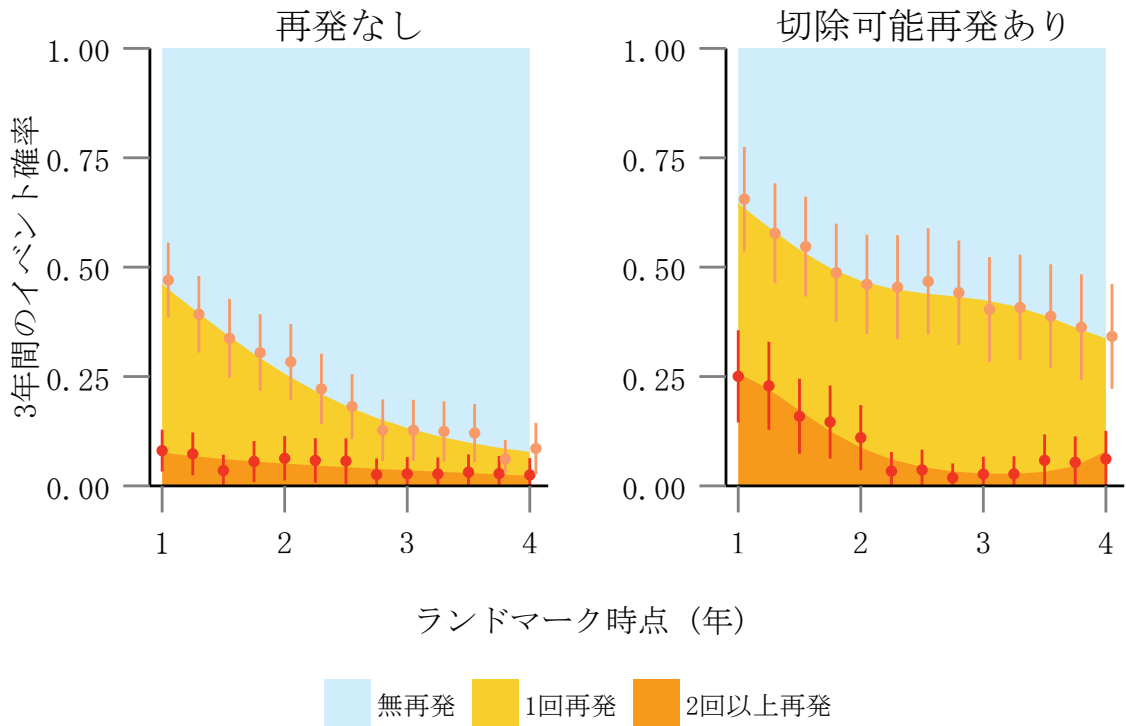
イベント種類 項目	終端イベントを含まない		終端イベントを含む	
	推定値	ロバスト 標準誤差	推定値	ロバスト 標準誤差
1回のみ再発				
intercept	-0.59	0.21	-0.17	0.23
時点 (年)				
s	0.13	0.41	0.06	0.40
s^2	-0.42	0.27	-0.44	0.27
s^3	0.08	0.05	0.08	0.05
腫瘍個数 (多発/単発)				
intercept	-0.08	0.30	0.72	0.36
s	-0.92	0.67	-2.17	0.67
s^2	0.87	0.45	1.42	0.45
s^3	-0.15	0.08	-0.22	0.08
2回以上再発				
intercept	-1.23	0.26	-0.80	0.28
時点 (年)				
s	-1.08	0.60	-1.14	0.59
s^2	-0.16	0.40	-0.19	0.40
s^3	0.07	0.08	0.07	0.08
腫瘍個数 (多発/単発)				
intercept	-0.10	0.40	0.70	0.45
s	0.64	1.03	-0.63	1.02
s^2	0.13	0.69	0.69	0.69
s^3	-0.05	0.13	-0.12	0.13
死亡				
intercept			-0.64	0.26
時点 (年)				
s			-0.17	0.48
s^2			-0.12	0.30
s^3			0.01	0.05
腫瘍個数 (多発/単発)				
intercept			1.57	0.37
s			-2.24	0.79
s^2			0.81	0.57
s^3			-0.07	0.10

が確認された。(a)のグラフを比べると、ベースライン時のランドマーク時点で各イベント確率はほぼ同等といえるが、時間が経過するにつれ、多発であると再発確率はほぼ変化しない一方、単発であれば再発確率は低下した。特に単発では、2年ランドマーク時点以降は2回以上の頻回再発を3年以内に起こす確率はほぼゼロとなり、1回再発確率も約15%程度と低い。多発ではランドマーク時点によらず、3年以内に1回再発を起こす確率は約20%強、頻回再発を起こす確率は約10%前後と、合わせて30%~40%あたりの確率での再発が起こりうる。(b)のグラフを比べると、1年ランドマーク時までは単発、多発とも3年無再発確率は50%を割っているものの、その後多発では2年ランドマーク時あたりから無再発確率は60%弱であまり変化しなくなるが、単発では無再発確率は増え続け、3年ランドマーク時点以降では75%を超えていた。2年ランドマーク時以降の死亡確率に注目すると、多発では11~15%の間で推移するが、単発では15%から5%へ向かって減少し続けていた。

次に、再発回数を共変量として3年イベント確率を予測した。再発回数は1回以上、すなわち切除可能再発を起こしていれば1、0回すなわち初回肝切除後1度も再発を起こしていなければ0をとるダミー変数に加工し、モデルに組み込んだ。上述の腫瘍個数の場合同様、積み上げグラフを用いてプロットした結果を図13、スーパーモデルの推定結果を表11に示す。

表11に示した通り、推定結果が著しく不安定になることはなかった。図13よ

(a) 死亡を打ち切りとみなした場合



(b) 死亡を終端イベントとして扱った場合

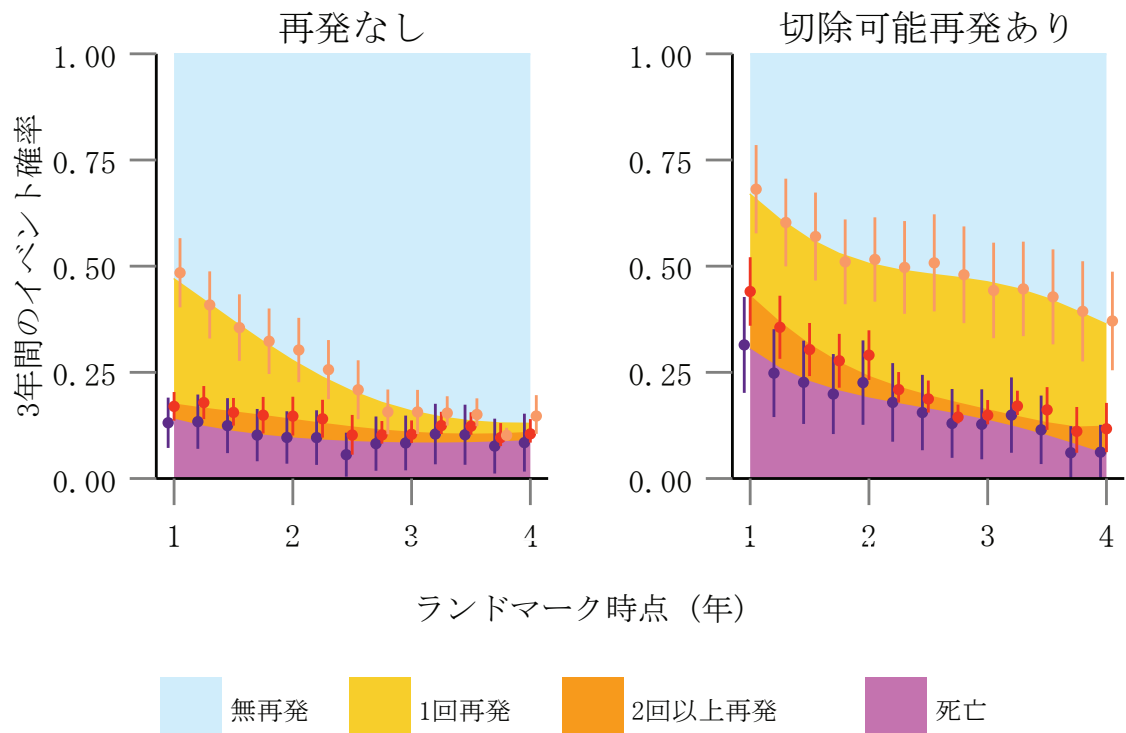


図 13: 再発回数別の3年イベント確率に関する積み上げ面グラフ

表 11: 再発回数によって再発を動的予測するためのスーパーモデルの推定結果

イベント種類 項目	終端イベントを含まない		終端イベントを含む	
	推定値	ロバスト 標準誤差	推定値	ロバスト 標準誤差
1回のみ再発				
intercept	0.43	1.31	-0.02	1.75
時点 (年)				
s	-0.61	2.16	-0.18	2.92
s^2	-0.20	1.08	-0.45	1.50
s^3	0.04	0.16	0.07	0.23
再発回数 (1回以上/0回)				
intercept	1.38	1.80	1.11	2.21
s	-2.07	2.78	-2.04	3.49
s^2	1.31	1.33	1.39	1.72
s^3	-0.19	0.19	-0.19	0.26
2回以上再発				
intercept	-0.84	2.94	-4.18	3.03
時点 (年)				
s	-1.38	4.50	2.61	4.50
s^2	0.29	1.94	-1.29	1.94
s^3	-0.03	0.25	0.17	0.24
再発回数 (1回以上/0回)				
intercept	0.47	4.97	3.90	5.06
s	2.80	7.55	-2.61	7.52
s^2	-1.97	3.27	0.42	3.19
s^3	0.33	0.41	0.02	0.40
死亡				
intercept			-0.14	1.65
時点 (年)				
s			-1.46	2.58
s^2			0.29	1.19
s^3			-0.01	0.16
再発回数 (1回以上/0回)				
intercept			2.52	2.41
s			-2.14	3.58
s^2			1.04	1.59
s^3			-0.17	0.22

り切除可能再発がみられた対象者が、再発のない対象者よりも、さらなる再発や死亡リスクが高くなることが確認された。(a)を比べると、初回肝切除以降の再発の有無によらず、3年無再発確率は時間経過とともに高くなる傾向がいえ。特に、再発のみられない対象者では、1年ランドマーク時点で54%の無再発確率であるが、4年ランドマーク時点では92%まで改善された。(b)より初回肝切除以降再発のみられない対象者であれば、3年死亡確率は10%程度であるものの、3年ランドマーク時点あたりから再発確率は5%程度と低く推移した。初回肝切除以降、切除可能再発がみられた対象者であると、3年再発確率が30%強とほぼ一定で推移し、再発のみられない対象者に比べ、再発が起りやすくなっていることが確認された。

最後に、切除可能再発がみられた対象者が再発のない対象者よりも再発や死亡リスクが高いという結果の、腫瘍個数や腫瘍径を調整した下での一貫性を確認するため、サブグループ解析を行った。腫瘍個数は多発であれば1、単発であれば0とし、腫瘍径は2cm超であれば1、2cm以下であれば0をとるようなダミー変数とし、モデルに組み込んだ。各サブグループでの推定結果を図14,15に、スーパーモデルのあてはめ結果を表12に示す。これら結果はすべて、腫瘍回数と腫瘍径を調整した下での腫瘍個数による動的予測結果に一貫性があるかを確認することにも利用できる。

この解析で用いたスーパーモデルは本解析の中で最も説明変数を多く含んだ

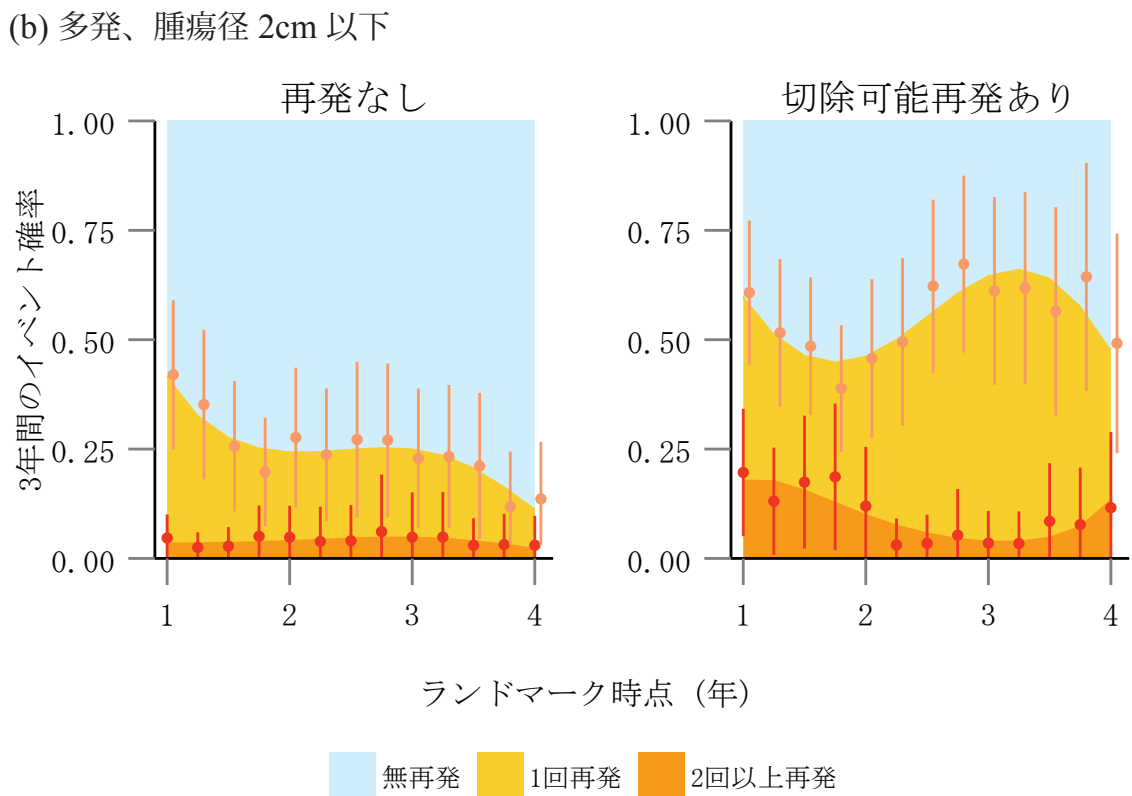
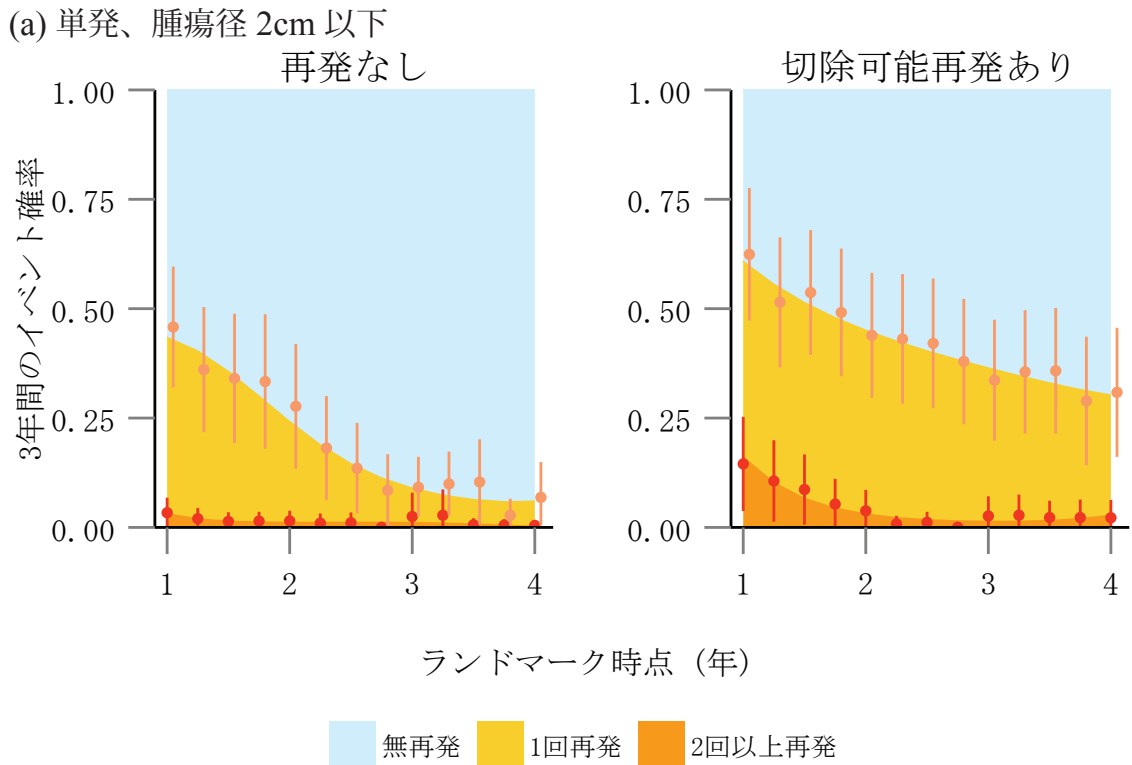
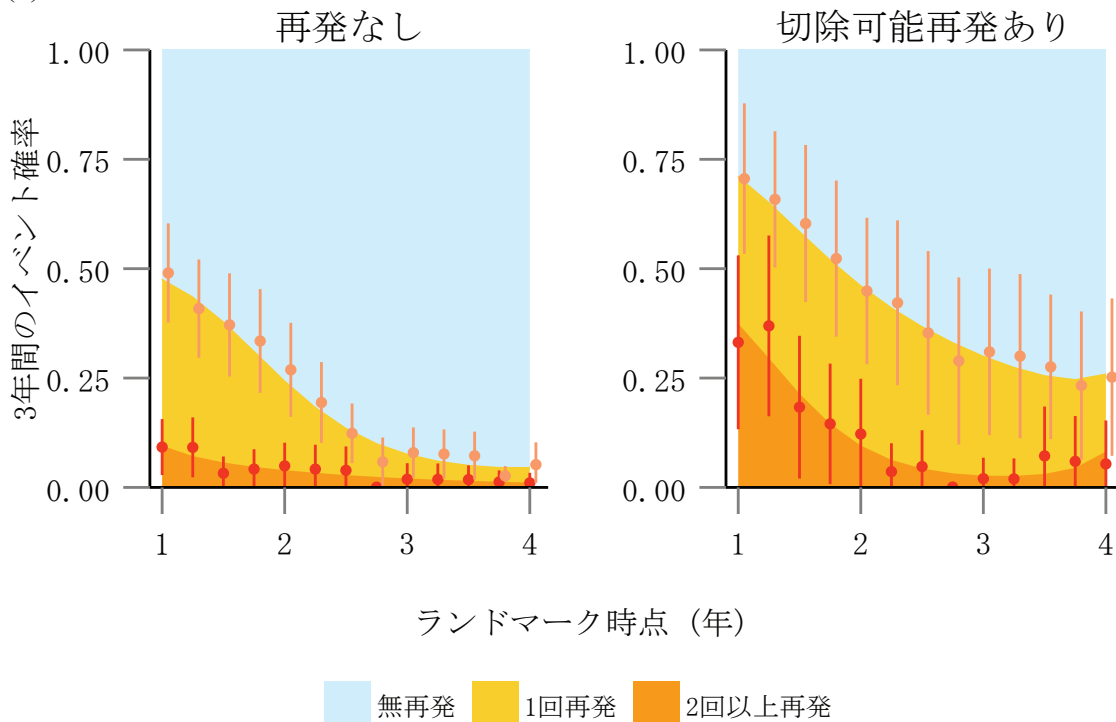


図 14: 腫瘍個数と腫瘍径でサブグループ化した再発回数別の 3 年イベント確率に関する積み上げ面グラフ (死亡を打ち切りとして扱った場合)

(c) 単発、腫瘍径 2cm 超



(d) 多発、腫瘍径 2cm 超

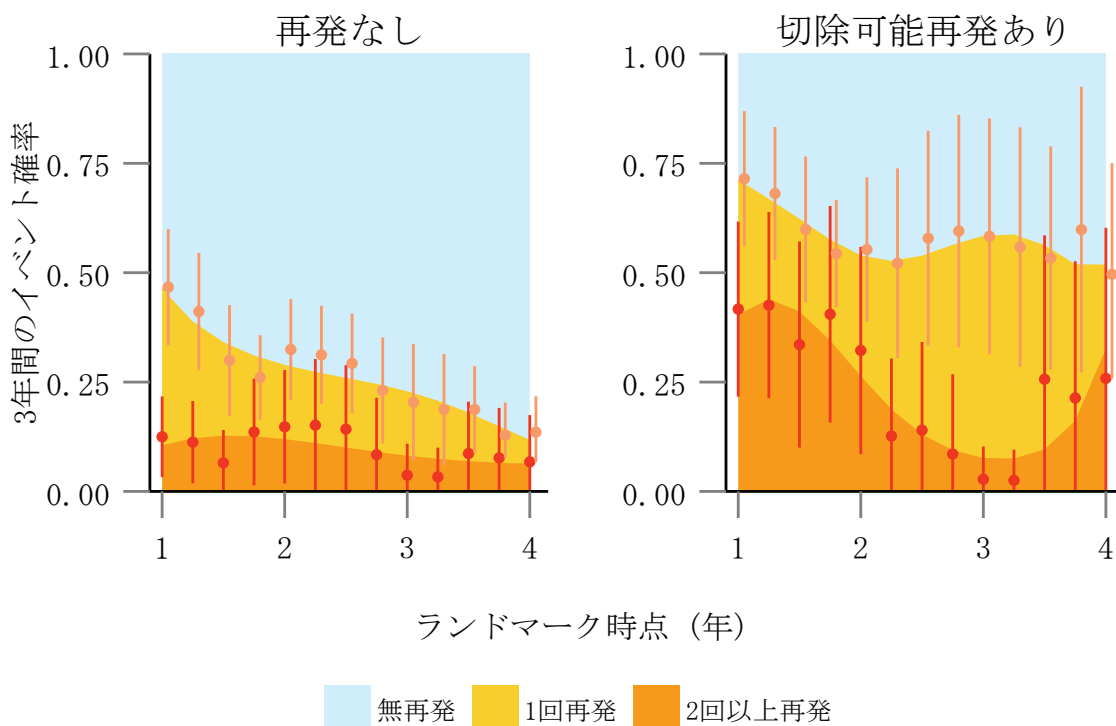


図 14: 腫瘍个数と腫瘍径でサブグループ化した再発回数別の 3 年イベント確率に関する積み上げ面グラフ (死亡を打ち切りとして扱った場合, 続き)

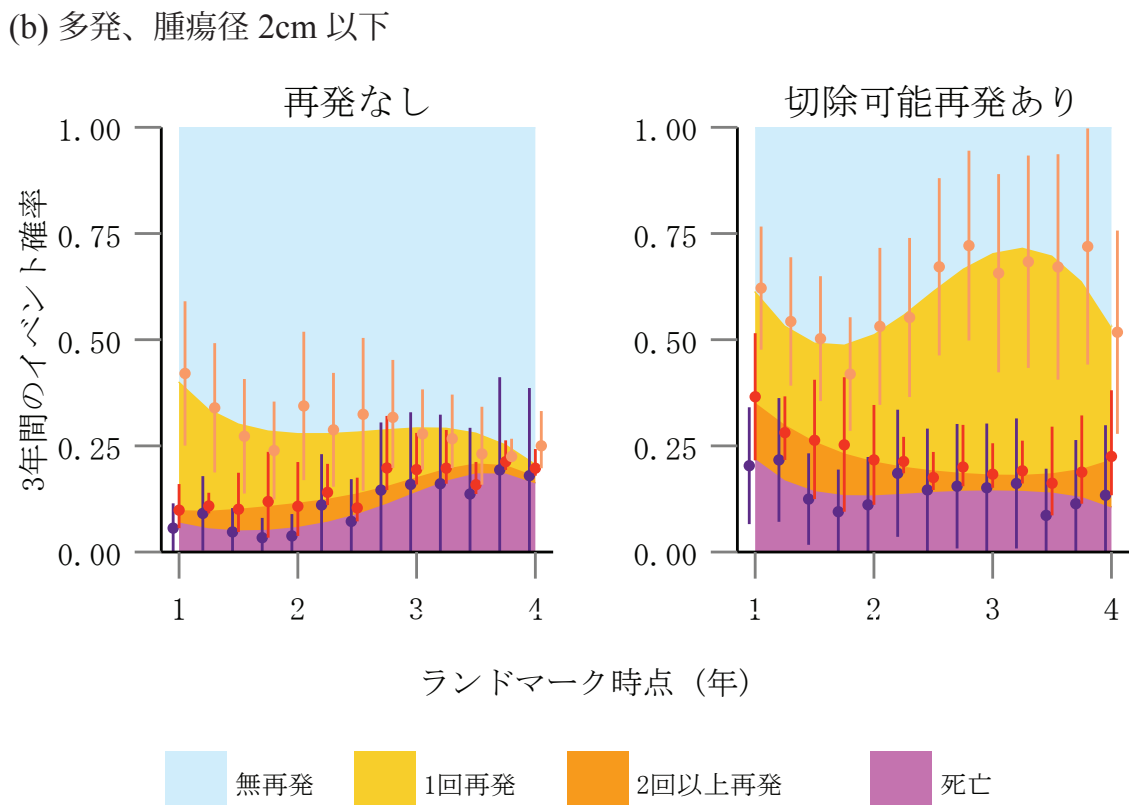
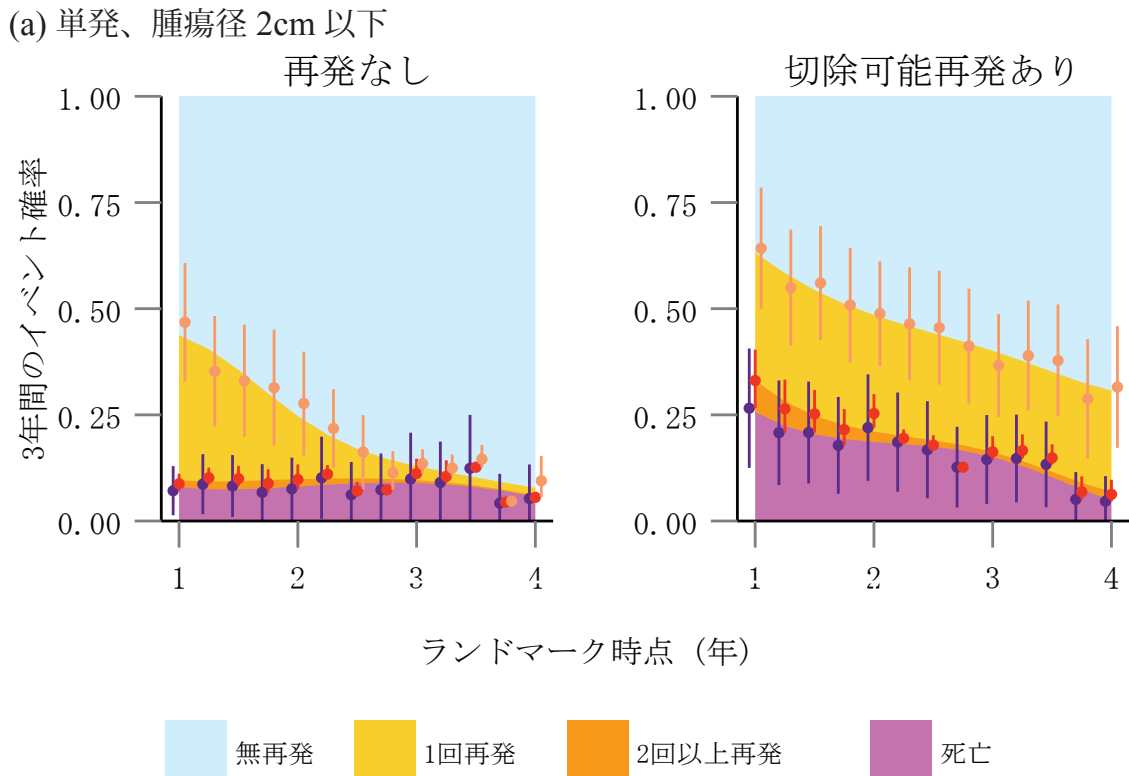
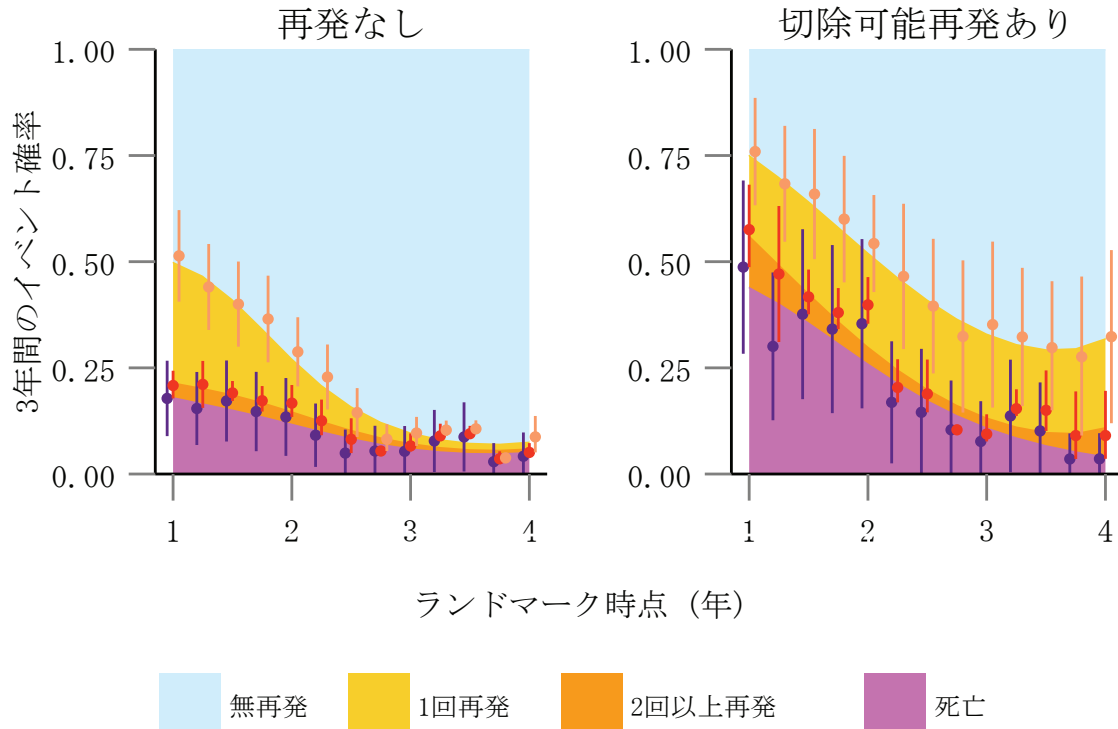


図 15: 腫瘍個数と腫瘍径でサブグループ化した再発回数別の 3 年イベント確率に関する積み上げ面グラフ (死亡を終端イベントとして扱った場合)

(c) 単発、腫瘍径 2cm 超



(d) 多発、腫瘍径 2cm 超

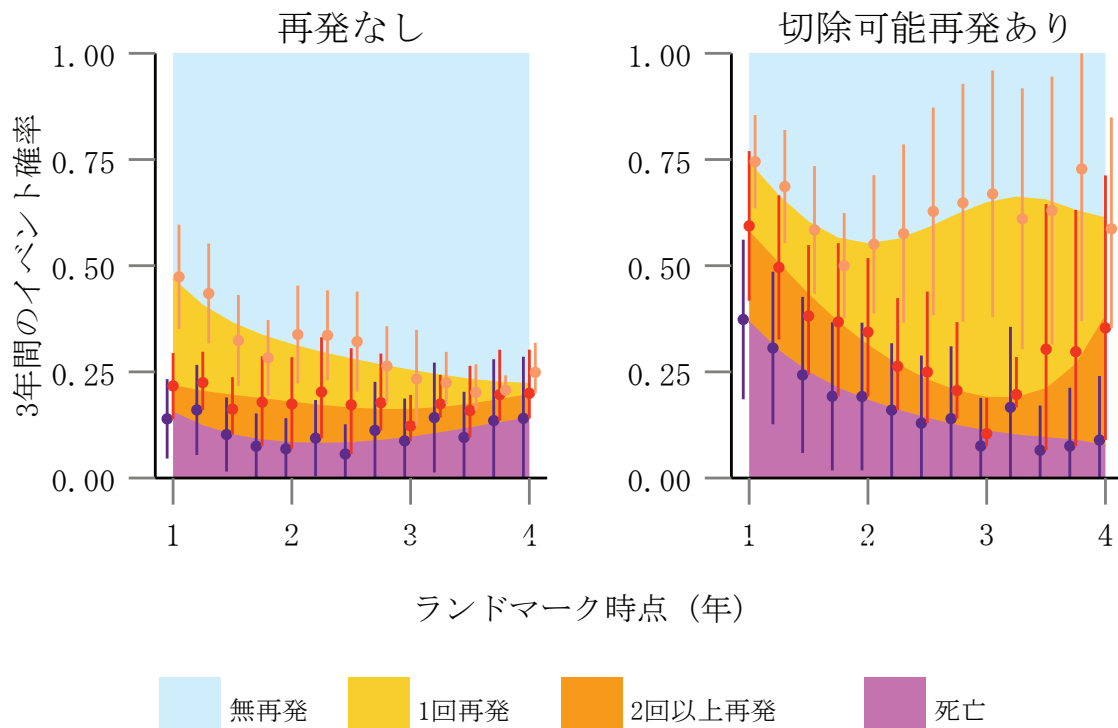


図 15: 腫瘍個数と腫瘍径でサブグループ化した再発回数別の 3 年イベント確率に関する積み上げ面グラフ (死亡を終端イベントとして扱った場合, 続き)

表 12: 腫瘍個数と腫瘍径でサブグループ化した再発回数によって再発を動的予測するためのスーパーモデルの推定結果

イベント種類 項目	終端イベントを含まない		終端イベントを含む	
	推定値	ロバスト 標準誤差	推定値	ロバスト 標準誤差
1回のみ再発				
intercept	-1.22	1.99	-1.43	2.60
時点 (年)				
s	2.19	3.28	2.46	4.30
s^2	-1.53	1.64	-1.79	2.21
s^3	0.22	0.24	0.25	0.33
再発回数 (1回以上/0回)				
intercept	1.92	2.12	1.77	2.68
s	-2.87	3.26	-3.06	4.19
s^2	1.65	1.55	1.84	2.05
s^3	-0.24	0.22	-0.25	0.30
腫瘍個数 (多発/単発)				
intercept	4.46	2.09	3.71	2.33
s	-7.87	3.21	-7.06	3.59
s^2	3.86	1.50	3.70	1.67
s^3	-0.53	0.21	-0.53	0.23
腫瘍径 (2cm超/2cm以下)				
intercept	-0.10	1.87	-0.50	2.17
s	0.31	2.88	0.83	3.31
s^2	-0.21	1.34	-0.47	1.52
s^3	0.03	0.19	0.07	0.21
2回以上再発				
intercept	1.30	4.27	-5.14	4.20
時点 (年)				
s	-6.24	6.28	2.82	5.98
s^2	2.42	2.69	-1.38	2.55
s^3	-0.32	0.34	0.16	0.32
再発回数 (1回以上/0回)				
intercept	0.49	4.72	4.80	4.50
s	3.36	7.01	-3.72	6.48
s^2	-2.26	2.99	0.90	2.71
s^3	0.38	0.38	-0.03	0.34

(次項につづく)

表 12: 腫瘍個数と腫瘍径でサブグループ化した再発回数によって再発を動的予測するためのスーパーモデルの推定結果 (続き)

イベント種類 項目	終端イベントを含まない		終端イベントを含む	
	推定値	ロバスト 標準誤差	推定値	ロバスト 標準誤差
2回以上再発				
腫瘍個数 (多発/単発)				
intercept	-2.27	4.32	0.04	4.65
s	3.16	6.64	0.28	6.99
s^2	-0.91	2.89	0.27	2.96
s^3	0.09	0.37	-0.05	0.37
腫瘍径 (2cm超/2cm以下)				
intercept	-1.19	4.07	1.29	4.08
s	4.08	6.21	-0.37	6.12
s^2	-2.05	2.70	-0.08	2.62
s^3	0.29	0.35	0.04	0.33
死亡				
intercept			-0.78	3.47
時点 (年)				
s			-1.84	5.10
s^2			0.79	2.17
s^3			-0.11	0.28
再発回数 (1回以上/0回)				
intercept			2.60	2.90
s			-1.52	4.25
s^2			0.60	1.83
s^3			-0.09	0.24
腫瘍個数 (多発/単発)				
intercept			2.34	2.60
s			-4.26	3.99
s^2			1.96	1.83
s^3			-0.24	0.25
腫瘍径 (2cm超/2cm以下)				
intercept			-0.33	3.00
s			2.65	4.39
s^2			-1.63	1.89
s^3			0.25	0.25

が、表 12 に示した通り、反復計算は収束し、推定結果は得られた。図 14,15 を通して、上述の腫瘍個数や腫瘍回数を用いた動的予測は、より細かく分類したサブグループ間でも同様の結果として得られた。ただし、単発かつ腫瘍径 2cm 以下以外のサブグループでは、切除可能再発ありの at risk 数は 10 名前後と少ないため、特に図に示したランドマーク時点別に予測したイベント確率の信頼区間はかなり広いことに注意せねばならない。例えば、図 15(b) の再発なしにおいて、後期のランドマーク時点において 3 年死亡確率が高くなる結果となったが、2.5 年ランドマーク時点以降 at risk 数が 10 名未満ゆえ、その信頼区間の下限は 0 を含んでいた。

5 考察

本研究では、動的擬似値を用いたランドマークモデルによって繰り返しイベントデータの動的予測モデルを提案した。動的予測する確率を一致推定するために必要な動的擬似値の仮定を満たすような方法を提案し、またシミュレーション実験を通して、有限サンプルにおいてバイアスの少ない結果が得られることを確認した。これまで繰り返しイベントデータに対する動的予測モデルは開発されておらず、本研究によって、既に提案された単一イベント、競合リスクイベントに対する動的予測モデルと併せ、実際の臨床疫学研究で観察されるほぼすべての種類の time-to-event データに対し、ランドマークモデルに基づいた動的予測を行うことが可能となった。

まず、提案法の特長のひとつとして、終端イベントを想定しない単純な繰り返しイベントデータの場合、Kaplan-Meier 推定量に基づく提案法は、統計解析パッケージにて簡便に実装することができる実用性が極めて高いことが挙げられる。既に提案された競合リスクを念頭にした Aalen-Johansen 推定量に基づく動的擬似値 [65] に比べ、行列計算が不要であることから、解析プログラムは単純であり計算負荷が少ない。

緒言で述べたように、ランドマークモデルの特長は経時データの情報を強い仮定を置かずに取り込めることと、比例ハザード性の仮定が不要なモデルの柔軟さである。このアイデアに基づく提案法は、特にモデルの柔軟さが必要とされる予

測を行う上で有用な解析方法といえる。また、繰り返しイベントではなく単一イベントデータであっても、既に提案されている Cox モデルに基づく従来法 [10] より、動的擬似値を使用する利点が 4.2 節で示した結果から考察される。4.2 節では、死亡予測を行うために、初回再発もしくは切除不能再発の発生を経時データとして取り込んだ。従来法ではスーパーモデルによってスムージングを行う際、ベースラインハザード関数を時点の多項式とノンパラメトリックな関数の積として推定する。それゆえ、得られるベースラインハザード関数は階段状の関数となり、予測確率をプロットしても階段状の形をとる。一方提案法は直接予測確率を時点の多項式でスムージングするため、予測確率をプロットするとなめらかな形状として得ることができる。また、ランドマーク時点 s から w だけ経過した時点 $s + w$ でのイベント確率を推定することを目的とする動的予測では、従来法で推定されている $(s, s + w]$ に関するハザードの情報は不要である。むしろノンパラメトリックに推定された動的擬似値を用いて、時点や共変量で条件付けた期待値がそのままイベント確率である提案法のほうがより柔軟なモデル化可能と考えられる。しかしながら、どちらの手法がよりバイアスの少ない予測を行えるかは必ずしも明らかではない。本研究で使用した c-index のほかに、2 値判別問題の 2 乗平均誤差である Brier スコア [119] を指標とした比較が考えられるが、膨大な数のシナリオに対するシミュレーション実験が必要であり、これは今後の課題としたい。

繰り返しイベントデータは 1950 年代頃から研究がなされ [120]、1980 年代には代表的な回帰モデルが 3 通り提案されたが [85–87]、いまだに実際の医学データにおいての適用例は少ない。この理由として、治療効果としてのハザード比の解釈がモデルにより異なることや、終端イベントの扱いの難しさが考えられる。本研究で提案した動的予測は、再発しやすさを確率で表現し、また終端イベントにも対応可能な解釈の容易な方法である。がんをはじめとする慢性疾患では、生存から死亡のような単純なものではなく、患者は例えば繰り返しイベントのような様々な病態を経る。患者ひとりひとりが将来どのような病態を経るのかの予測結果を分かりやすく示すことは、個別化医療へ向けた現在の潮流において必要なことであり、医師-患者間のコミュニケーションにおいて大いに役立つ可能性があることから、提案法をはじめとする動的予測を用いた解析が広まっていくと思われる。

シミュレーション実験について考察を行う。これまで動的擬似値の性能評価のためのシミュレーション実験は報告されておらず、簡単ではあるが、本研究にて初めて実施された。これまで動的擬似値は、大標本において動的予測する確率を一致推定するための条件が示されてきたが [65, 100]、現実には観察しうる可能性のある有限サンプルであまり対象者数が多くない 100 名のデータであっても、バイアスを無視できる推定が行えることが確かめられたことには大きな意義がある。

3節を通して、(A1)-(A3)の仮定を満たす限り、対象者間のイベントの起こりやすさに関する異質性の有無によらず、また繰り返しイベント発生過程がマルコフ性、セミマルコフ性であるかによらず、動的擬似値はバイアスがないことが確かめられた。がんのような慢性疾患では、イベントが起こった対象者であるほど、潜在的なリスクが高く、イベントを繰り返しやすいという仮説は自然であり、対象者間でイベントの起こりやすさは一定という仮定は現実的ではない。イベントによって生理学的メカニズムに変化が生じることで、次のイベントが起こりやすくなるという仮説も自然であり、マルコフ性の仮定も現実的ではない。提案法は、対象者間のイベントの起こりやすさに関して異質性がみられず、かつマルコフ性も成立するような数学的に単純な場面はもちろん、異質性があり、マルコフ性も成立しないような現実的な場面にも対応できる手法である。

繰り返しイベントデータにおいて、繰り返しイベントと相関のある終端イベントを打ち切りとして扱うことは情報のある打ち切りとなり、予測結果にバイアスを招く。この場合、提案した終端イベントを考慮しない2通りの動的擬似値はバイアスがあることが、3節で行ったような簡単なシミュレーション実験より確認された。終端イベントを考慮しない方法は、歯科疾患の再発や、多発性硬化症における神経症状の再発 [121] のような、追跡期間において死亡がほとんどみられない状況での利用がすすめられる。

シミュレーション実験において、データ生成過程には frailty モデルを採用し

た。繰り返しイベントにはこのようなデータ生成は適すると考えられる一方、繰り返しイベントと死亡のような終端イベントの間にはイベント自体に明らかな相関があり、frailtyで条件付ければ互いのイベントは独立という仮定は適さないという指摘がある [110]。このようなイベント自体に相関があると考えられる場合は、それぞれの周辺生存関数の相関構造を与えるコピュラ [122, 123] に基づくデータ生成のほうが適していると考えられる。ただし、本研究ではイベント間に相関がみられる下で、時点 $s + w$ においていずれのイベントを起こした状態にあるかについての打ち切りが存在しない場合に得られる経験分布を、打ち切りがみられる下で、動的擬似値によって一致推定できるかに注目したシミュレーション内容であるため、この目的のための基本的な性能評価は行えたと考える。なお、シミュレーション実験では、他に競合する手法がないため、共変量のモデルを仮定して得られる動的予測結果の正確性に関する評価は行わなかった。

次に、大腸がん肝転移切除例の追跡データへの適用について考察する。4.2節では、ランドマーク解析により、初回再発の有無よりも切除不能再発発生の有無でサブグループ化したほうが生存関数は大きく異なり、死亡の予測性能が高いことを示した。すなわち、初回肝切除を受けた患者にとって、再発が起こるかよりも切除不能再発として起こるかのほうが、生存に対してより重要な情報であることを意味する。

しかしながら、本研究では TSF イベントを表す切除不能再発を、再発後に実際に切除がされたか否かで定義した。切除が可能であると医師によって判断されたものの、実際には切除が行われなかった場合は切除不能再発に該当し、本来興味のある切除不能と判断された再発か否かに対して誤分類が生じているが、データから検討することは難しい。さらに、切除不能の判断は外科医の技量や患者の状態に依る部分が大きく、恣意的なものであるおそれがある。実際に、本研究で使用したデータは、表 8 に示したように、再発症例のうち、ほぼ半数が切除可能再発、すなわち再切除された。これは先行研究 [114, 115] で再切除割合について示された 2 割という結果に比べ圧倒的に高く、追跡データの対象集団は他施設よりも積極的な切除を受けたことが分かる。本研究結果の一般化可能性を高めるためには、術後の残肝割合のような客観的な指標を用いて定義した TSF イベントに類似する代替エンドポイントでの評価を併せて行うべきと考える。

4.3 節では再発や死亡発生に対し、前回切除時の再発回数やそれまでの再発回数による予測を行い、これら因子が予測するために有用であることが確認された。イベントの要約部分に注目すると、死亡の前に 97% の対象者が再発を起こしていることから、死亡を終端イベントとして扱うか打ち切りとして扱うかによらず、無再発生存確率は似たような結果になったと考えられる。図 12(b) より腫瘍回数について、単発では 3 年無再発生存確率が増加し続けるのに対し、多発では 2 年ランドマーク時点あたりから 60% 程度と定常的になることが確認された。

また、図 12(a)(b) から、3 年再発確率は単発では減少し続ける一方、多発ではほぼ横ばいである。このことから、初回肝切除から時間がある程度経過したとしても少なくとも 4 年間に関して、前回切除時多発であった患者は一定の再発リスクを持ち続けるといえ、予測に対し腫瘍個数が有用であると考えられる。

図 13 より再発回数について、初回肝切除後再発が起こった対象者のほうが、再び再発を起こしやすいことが確認された。この結果は、図 14、15 より、腫瘍個数や腫瘍径を調整しても、再発回数別の 3 年イベント確率について、同様の傾向がいった。再発を起こした対象者は潜在的に再発リスクが高かったのか、再発を起こしたことで生理学的メカニズムが変化し、再発を繰り返しやすくなったのかをこの結果から言及することはできないが、少なくとも再発を起こした対象者は一定の再発リスクを持ち続けることから、特に注意して予後をフォローアップする必要があるといえよう。一方、初回肝切除後 3 年程度、再発を起こさなかった対象者であれば、その後の 3 年間の無再発生存確率は 80% 以上におよび、再発が起こった対象者ほどの密な検査間隔は必要でないと考えられる。このようなリスクに応じて検査間隔を変えることは、医師・患者の負担軽減という点で有用になると考えられる。ただし、本研究では、7 年以上追跡している症例データが少なかったため、ランドマーク時点を 4 年までとした点が限界であり、リスクの低い対象者のより長期に及ぶ予後予測は、よりデータが集積した後の課題としたい。

図 14、15 それぞれにおいて (a) と (c)、(b) と (d) を比べることで、腫瘍径の大

きさによって3年イベント確率に大きな違いが生じるかが検討したところ、腫瘍径の大小によらず、3年イベント確率は似た値をとる、つまり、3年イベント確率はあまり変わらないということが確認された。腫瘍個数よりも腫瘍径が生存に関する予後因子であるという報告があるが [124]、再発と死亡を総合的に検討した本研究からは、再発や死亡リスク上昇は、多発や再発回数があった場合にはみられたものの、腫瘍径が大きいことはあまり影響しないといえる。

実データ適用の限界として、動的予測モデルの妥当性評価が十分に行えなかった点が挙げられる。適用データのみならず多くの医学データにおいて、限られた例数のデータから、予測モデルの過適合を防ぐためには、例えば交差検証法の利用が考えられる。評価指標として、本研究で検討したような多値アウトカムの場合には、2値アウトカムに対する c-index や上述の Brier スコア [119] を拡張した指標を用いることが考えられる。動的予測を含む、予測モデルの構築が様々な研究で行われる将来に向け、こうした評価指標の開発も重要であると考えられる。

実データ適用に際し、あてはめた一般化線形モデルはすべて一般化ロジットを用いた。生存時間データではこの他に complementary log-log リンクが用いられることもあるものの [107]、この2つのリンク関数を用いた場合に大きな違いはみられづらいことや、多値アウトカムに対し、一般化ロジットリンクを使用した場合は推定方程式が単純な形で書き下せることから、本研究では一般化ロジットリンクを用いた。もちろん、本解析結果の感度解析として他のリンク関数を用い

ることは重要であり、例えば、Klein [72] では状況によって恒等リンクを用いることも述べている。

また、実データ適用における今後の課題として、前回再発部位を用いた再発・死亡の動的予測が挙げられる。再発部位が肝内あるいは肺内のみであれば、再手術されやすい傾向にあり、比較的予後もよいといわれている。一方、再発部位が遠隔であれば予後不良とされているため、それらの情報を予後予測に利用することは有用と思われる。

以下、提案手法に関する限界を2つ述べる。まず、終端イベントを検討する場合の動的擬似値として Aalen-Johansen 推定量に基づく方法のみ提案し、Kaplan-Meier 推定量に基づくようなより簡便な方法は提案できなかった。終端イベントと繰り返しイベントは相関すると考えるのが自然であり、また、終端イベント発生により繰り返しイベントは発生しなくなるため、競合リスク同様それぞれの Kaplan-Meier 推定量を基に発生確率を表すことはできないからである。

二つ目として、多値アウトカムの擬似値に対し、多項分布に基づく一般化線形モデルのあてはめは、統計解析パッケージを利用した適用が難しいことが挙げられる。本研究では SAS/IML を用いて計算プログラムを書く必要があったように、多少のプログラミング能力が求められる。多値アウトカムは積二項アウトカムに変換することで、例えば SAS であれば `genmod` プロシジャを用いて、点推

定は正しく得られる。一方、アウトカム間の共分散をゼロとみなす二項分布に基づく分散推定となるため、多項分布の場合とは異なる分散推定となってしまう。しかし、現実のデータ解析では、信頼区間の構成にはリンク関数の選択も影響するため、二項アウトカムとして分散を構成したとしても、結果を解釈する上で問題になるとは考えづらい。それゆえ、二項アウトカムとして解析しても実際上の問題は少ないと考える。

提案法では (A2) で示したような、無情報打ち切り (non-informative censoring) を仮定した。これは多くの生存時間解析法にて仮定されるものであるが、現実には種々の理由により破たんする場面がみられる。解決法のひとつに、逆打ち切り確率重み付け法、いわゆる inverse probability of censoring weighted (以下、IPCW) 法による補正が提案されている [125–127]。最近、動的擬似値ではないが、単純な擬似値については IPCW 法による情報のある打ち切りの補正法が提案され [76]、基本的には擬似値と同じアイデアをもつ提案法にも拡張ができると考えられ、今後の課題としたい。

ランドマークモデルはモデルの仮定が少ない一方、効率が決して良いとはいえず、例えば予測確率の精度は必ずしも高くない場合がある。同時モデルによって経時データをモデル化すればより効率の高い方法は提案でき、実際に単一イベントの場合はいくつか提案されている [2, 11, 44–46]。この同時モデルアプローチは繰り返しイベントに対しても適用可能である。その場合は、WLW モデルの拡

張が簡便であり、WLW モデルや Kaplan-Meier 推定量に基づく提案法のように、イベントの順序性を利用するアイデアが重要となると思われる。

6 結論

動的擬似値を応用したランドマークモデルによって、繰り返しイベントデータに対する動的予測モデルを提案した。提案法がバイアスの少ない推定を行えることをシミュレーション実験を通じた確認を行った。提案手法を大腸がん肝転移切除例の追跡データに適用し、切除不能再発のほうが初回再発よりも死亡イベントに対する予測能が高いことを示した。また、腫瘍個数や再発回数が、対象者の再発・死亡確率を予測するために有用であることを示唆した。

7 謝辞

本研究の発想に至った臨床研究のデータ解析に携わらせて頂き、また方向性に関しての指導ならびに議論をして下さった東京大学大学院医学系研究科 生物統計学分野 松山裕 教授に心より感謝申し上げます。

本研究のために大変貴重なデータをご提供下さった、共同研究者の東京大学医学部附属病院肝胆膵外科 國土典宏 教授、同 大場大 助教に深く御礼申し上げます。また、データ提供においてご尽力くださった同 山下俊 助教にも感謝申し上げます。

研究早期より数々の有用な助言を下された中央大学工学部人間総合理工学科 生物統計学 大橋靖雄 教授に深く感謝申し上げます。ならびに東京大学大学院 生物統計学分野の教室員みなさまに深く感謝申し上げます。最後に支えてくださった家族に感謝を表します。

文献

- 1 Oba M, Hasegawa K, Matsuyama Y, Shindoh J, Mise Y, Aoki T, et al. Discrepancy between recurrence-free survival and overall survival in patients with resectable colorectal liver metastases: a potential surrogate endpoint for time to surgical failure. *Ann Surg Oncol*. 21:1817–24. 2014.
- 2 Taylor JMG, Park Y, Ankerst DP, Proust-Lima C, Williams S, Kestin L, et al. Real-Time Individual Predictions of Prostate Cancer Recurrence Using Joint Models. *Biometrics*. 69:206–13. 2013.
- 3 Cox DR, Oakes D. *Analysis of Survival Data*. London: Chapman and Hall/CRC; 1984.
- 4 Kalbfleish JD, Prentice RL. *The Statistical Analysis of Failure Time Data, Second Edition*. Hoboken: Wiley; 2002.
- 5 Klein JP, Moeschberger ML. *Survival Analysis :Techniques for Censored and Truncated Data, 2nd ed*. NewYork: Springer; 2003.
- 6 Nelson W. Theory and applications of hazard plotting for censored failure data. *Technometrics*. 14:945–65. 1972.
- 7 Aalen OO. Nonparametric inference for a family of counting processes. *Ann Stat*. 6:701–26. 1978.
- 8 Andersen PK, Borgan O, Gill RD, Keiding N. *Statistical Models Based on*

- Counting Processes*. NewYork: Springer-Verlag; 1993.
- 9 Cox DR. Regression models and lifetables (with discussion). *J R Stat Soc Series B Stat Methodol*. 34:187–220. 1972.
 - 10 van Houwelingen HC. Dynamic prediction by landmarking in event history analysis. *Scand Stat Theory Appl*. 34:70–85. 2007.
 - 11 Rizopoulos D. *Joint Models for Longitudinal and Time-to-Event Data, with Applications in R*. Boca Raton: Chapman and Hall/CRC; 2012.
 - 12 Skuladottir H, Olsen JH. Conditional survival of patients with the four major histologic subgroups of lung cancer in Denmark. *J Clin Oncol*. 21:3035–40. 2003.
 - 13 Zamboni BA, Yothers G, Choi M, Fuller CD, Dignam JJ, Raich PC, et al. Conditional survival and the choice of conditioning set for patients with colon cancer: an analysis of NSABP trials C-03 Through C-07. *J Clin Oncol*. 28:2544–8. 2010.
 - 14 Harshman LC, Xie W, Bjarnason GA, Knox JJ, MacKenzie M, Wood L, et al. Conditional survival of patients with metastatic renal-cell carcinoma treated with VEGF-targeted therapy: a population-based study. *Lancet Oncol*. 13:927–35. 2012.
 - 15 Sabatier R, Eymard JC, Walz J, Deville JL, Narbonne H, Boher JM, et al.

- Could thyroid dysfunction influence outcome in sunitinib-treated metastatic renal cell carcinoma? *Ann Oncol.* 23:714–21. 2012.
- 16 Yu XQ, Baade PD, O’Connell DL. Conditional survival of cancer patients: an Australian perspective. *BMC Cancer.* 12:460. 2012.
 - 17 Katz MHG, Hu CY, Fleming JB, Pisters PWT, Lee JE, Chang GJ. Clinical calculator of conditional survival estimates for resected and unresected survivors of pancreatic cancer. *Arch Surg.* 147:513–9. 2012.
 - 18 Ito Y, Nakayama T, Miyashiro I, Ioka A, Tsukuma H. Conditional survival for longer-term survivors from 2000-2004 using population-based cancer registry data in Osaka, Japan. *BMC Cancer.* 13:304. 2013.
 - 19 Chang GJ, Hu CY, Eng C, Skibber JM, Rodriguez-Bigas MA. Practical application of a calculator for conditional survival in colon cancer. *J Clin Oncol.* 27:5938–43. 2009.
 - 20 Baade PD, Youlden DR, Chambers SK. When do I know I am cured? Using conditional estimates to provide better information about cancer survival prospects. *Med J Aust.* 194:73–7. 2011.
 - 21 Zabor EC, Gonen M, Chapman PB, Panageas KS. Dynamic prognostication using conditional survival estimates. *Cancer.* 3589–92. 2013.
 - 22 Prentice RL, Kalbfleisch JD. Hazard rate models with covariates. *Biometrics.*

- 35:25–39. 1979.
- 23 Hougaard P. Multi-state models: a review. *Lifetime Data Anal.* 5:239–64. 1999.
- 24 Andersen PK, Keiding N. Multi-state models for event history analysis. *Stat Methods Med Res.* 11:91–115. 2002.
- 25 Putter H, Fiocco M, Geskus RM. Tutorial in biostatistics: competing risks and multi-state models. *Stat Med.* 26:2389–430. 2007.
- 26 Clayton D. The analysis of event history data - a review of progress and outstanding problems. *Stat Med.* 7:819–41. 1988.
- 27 Hougaard P. *Analysis of Multivariate Survival Data.* New York: Springer-Verlag; 2000.
- 28 Aalen OO, Borgan O, Gjessing HK. *Survival and Event History Analysis.* New York: Springer; 2008.
- 29 Gail M. A review and critique of some models used in competing risk analysis. *Biometrics.* 31:209–22. 1975.
- 30 Andersen PK, Abildstrom SZ, Rosthøj S. Competing risks as a multi-state model. *Stat Methods Med Res.* 11:203–15. 2002.
- 31 Pintilie M. Analysing and interpreting competing risk data. *Stat Med.* 26:1360–7. 2007.

- 32 Kay R. A Markov model for analysing cancer markers and disease states in survival studies. *Biometrics*. 42:855–65. 1986.
- 33 Andersen PK, Hansen LS, Keiding N. Assessing the influence of reversible disease indicators on survival. *Stat Med*. 10:1061–7. 1991.
- 34 Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *J Am Stat Assoc*. 53:457–81. 1958.
- 35 Aalen OO, Johansen S. Empirical transition matrix for nonhomogeneous Markov-chains based on censored observations. *Scand J Stat*. 5:141–50. 1978.
- 36 van Houwelingen HC. Dynamic predicting by landmarking as an alternative for multi-state modeling: an application to acute lymphoid leukemia data. *Lifetime Data Anal*. 14:447–63. 2008.
- 37 Henderson R, Diggle P, Dobson A. Joint modelling of longitudinal measurements and event time data. *Biostatistics*. 1:465–80. 2000.
- 38 Diggle PJ, Heagerty P, Liang KY, Zeger SL. *Analysis of Longitudinal Data, Second edition*. New York: Oxford University Press; 2002.
- 39 Wu MC, Carroll RJ. Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. *Biometrics*. 44:175–88. 1988.

- 40 Wu MC, Bailey KR. Estimation and comparison of changes in the presence of informative right censoring: conditional linear model. *Biometrics*. 45:939–55. 1989.
- 41 Tsiatis AA, Degruittola V, Wulfsohn MS. Modeling the relationship of survival to longitudinal data measured with error - applications to survival and CD4 Counts in patients with AIDS. *J Am Stat Assoc*. 90:27–37. 1995.
- 42 Wulfsohn MS, Tsiatis AA. A joint model for survival and longitudinal data measured with error. *Biometrics*. 53:330–39. 1997.
- 43 Diggle P, Henderson R, Philippon P. Random-effects models for joint analysis of repeated-measurement and time-to-event outcomes. In: Fitzmaurice G, Davidian M, Verbeke G, Molenberghs G, editors. *Longitudinal Data Analysis*. Boca Raton: CRC Press; 2008. 349–66.
- 44 Proust-Lima C, Taylor JMG. Development and validation of a dynamic prognostic tool for prostate cancer recurrence using repeated measures of post-treatment PSA: a joint modeling approach. *Biostatistics*. 10:535–49. 2009.
- 45 Rizopoulos D. Dynamic predictions and prospective accuracy in joint models for longitudinal and time-to-event data. *Biometrics*. 67:819–29. 2011.
- 46 Njagi EN, Rizopoulos D, Molenberghs G, Dendale P, Willekens K. A joint survival-longitudinal modelling approach for the dynamic prediction of rehos-

- pitalization in telemonitored chronic heart failure patients. *Statistical Modelling*. 13:179–98. 2013.
- 47 van Houwelingen HC, Putter H. *Dynamic Prediction in Clinical Survival Analysis*. Boca Raton: CRC Press; 2011.
- 48 Zheng Y, Heagerty PJ. Partly conditional survival models for longitudinal data. *Biometrics*. 61:371–91. 2005.
- 49 van Houwelingen HC. From model building to validation and back: a plea for robustness. *Stat Med*. [epub ahead of print].
- 50 Anderson J, Cain K, Gelber R. Analysis of survival by tumor response. *J Clin Oncol*. 1:710–9. 1983.
- 51 Green S, Benedetti J, Crowley J. *Clinical Trials in Oncology, 2nd edition*. Boca Raton: Chapman and Hall; 2003.
- 52 Putter H. Landmarking. In: Klein JP, vanHouwelingen HC, Ibrahim JG, Scheike TH, editors. *Handbook of Survival Analysis*. Boca Raton: CRC Press; 2014. 441–56.
- 53 Murphy SA, Sen PK. Time-dependent coefficients in a Cox-type regression model. *Stoch Process Applic*. 39:153–80. 1991.
- 54 Hastie T, Tibshirani R. Varying-coefficient Models (with discussion). *J R Stat Soc Series B Stat Methodol*. 55:757–96. 1993.

- 55 Putter H. dynpred: Companion package to "Dynamic Prediction in Clinical Survival Analysis"; 2011. R package version 0.1.1. Available from: <http://CRAN.R-project.org/package=dynpred>.
- 56 Aihara T, Yokota I, Hozumi Y, Aogi K, Iwata H, Tamura M, et al. Anastrozole versus tamoxifen as adjuvant therapy for Japanese postmenopausal patients with hormone-responsive breast cancer: efficacy results of long-term follow-up data from the N-SAS BC 03 trial. *Breast Cancer Res Treat.* 148:337–43. 2014.
- 57 Pepe MS, Mori M. Kaplan-Meier, marginal or conditional probability curves in summarizing competing risks failure time data? *Stat Med.* 12:737–51. 1993.
- 58 Prentice RL, Kalbfleisch JD, A V Peterson J, Flornoy N, Farewell VT, Breslow NE. The analysis of failure times in the presence of competing risks. *Biometrics.* 34:541–54. 1978.
- 59 Haller B, Schmidt G, Ulm K. Applying competing risks regression models: an overview. *Lifetime Data Anal.* 19:33–58. 2013.
- 60 Tsiatis AA. A nonidentifiability aspect of the problem of competing risks. *Proc Natl Acad Sci USA.* 72:20–2. 1975.
- 61 Lau B, Cole SR, Gange SJ. Competing risk regression models for epidemio-

- logic data. *Am J Epidemiol.* 170:244–56. 2009.
- 62 Fine JP, Gray RJ. A proportional hazards model for the sub distribution of a competing risk. *J Am Stat Assoc.* 94:496–509. 1999.
- 63 Latouche A, Allignol A, Beyersmann J, Labopin M, Fine JP. A competing risks analysis should report results on all cause-specific hazards and cumulative incidence functions. *J Clin Epidemiol.* 66:648–53. 2013.
- 64 Nicolaie MA, van Houwelingen JC, de Witte TM, Putter H. Dynamic prediction by landmarking in competing risks. *Stat Med.* 32:2031–47. 2013.
- 65 Nicolaie MA, van Houwelingen JC, de Witte TM, Putter H. Dynamic pseudo-observations: A robust approach to dynamic prediction in competing risks. *Biometrics.* 69:1043–52. 2013.
- 66 Quenouille MH. Approximate tests of correlation in time-series. *J R Stat Soc Series B Stat Methodol.* 11:68–84. 1949.
- 67 Tukey JW. Bias and confidence in not-quite large samples (Abstract). *Ann Math Statist.* 29:614. 1958.
- 68 Miller RG. The jackknife—a review. *Biometrika.* 61:1–15. 1974.
- 69 Efron B, Tibshirani RJ. *An Introduction to the Bootstrap.* Boca Raton: Chapman and Hall/CRC; 1994.
- 70 Andersen PK, Klein JP, Rosthøj S. Generalised linear models for correlated

- pseudo-observations, with applications to multi-state models. *Biometrika*. 90:15–27. 2003.
- 71 Klein JP, Andersen PK. Regression modeling of competing risks data based on pseudovalues of the cumulative incidence function. *Biometrics*. 61:223–9. 2005.
- 72 Klein JP. Modelling competing risks in cancer studies. *Stat Med*. 25:1015–34. 2006.
- 73 Ambrogi F, Biganzoli E, Boracchi P. Estimates of clinically useful measures in competing risks survival analysis. *Stat Med*. 27:6407–25. 2008.
- 74 Perme MP, Andersen PK. Checking hazard regression models using pseudo-observations. *Stat Med*. 27:5309–28. 2008.
- 75 Mogensen UB, Gerds TA. A random forest approach for competing risks based on pseudo-values. *Stat Med*. 32:3102–14. 2013.
- 76 Binder N, Gerds TA, Andersen PK. Pseudo-observations for competing risks with covariate dependent censoring. *Lifetime Data Anal*. 20:303–15. 2014.
- 77 Hansen SN, Andersen PK, Parner ET. Events per variable for risk differences and relative risks using pseudo-observations. *Lifetime Data Anal*. 2014 [Epub ahead of print].
- 78 Lawless JF. The analysis of recurrent events for multiple subjects. *J R Stat*

- Soc Ser C Appl Stat.* 44:487–98. 1995.
- 79 Wei LJ, Glidden DV. An overview of statistical methods for multiple failure time data in clinical trials. *Stat Med.* 16:833–9. 1997.
- 80 Cook RJ, Lawless JF. Analysis of repeated events. *Stat Methods Med Res.* 11:141–66. 2002.
- 81 Naito S, Imamura H, Tukada A, Matsuyama Y, Yoshimoto J, Sugo H, et al. Postoperative recurrence pattern and prognosis of patients with hepatocellular carcinoma, with particular reference to the hepatitis viral infection status. *Liver Int.* 34:802–13. 2014.
- 82 Rosen LS, Gordon D, Tchekmedyian S, Yanagihara R, Hirsh V, Krzakowski M, et al. Zoledronic acid versus placebo in the treatment of skeletal metastases in patients with lung cancer and other solid tumors: a phase III, double-blind, randomized trial—the zoledronic acid lung cancer and other solid tumors study group. *J Clin Oncol.* 21:3150–7. 2003.
- 83 Atkinson K, Storb R, Prentice RL, Weiden PL, Witherspoon RP, Sullivan K, et al. Analysis of late infections in 89 long-term survivors of bone marrow transplantation. *Blood.* 53:720–31. 1979.
- 84 Japanese Society for Cancer of the Colon and Rectum, editor. *JSCCR Guidelines 2014 for the Treatment of Colorectal Cancer.* Tokyo: Kanehara; 2014.

- 85 Andersen PK, Gill RD. Cox's regression model for counting processes: a large sample study. *Ann Stat.* 10:1100–20. 1982.
- 86 Prentice RL, Williams BJ, Peterson AV. On the regression analysis of multivariate failure time data. *Biometrika.* 68:373–9. 1981.
- 87 Wei LJ, Lin DY, Weissfeld L. Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *J Am Stat Assoc.* 84:1065–73. 1989.
- 88 Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika.* 73:13–22. 1986.
- 89 Lin DY, Wei LJ. The robust inference for the Cox proportional hazards model. *J Am Stat Assoc.* 84:1074–8. 1989.
- 90 Kelly PJ, Lim LLY. Survival analysis for recurrent event data: an application to childhood infectious diseases. *Stat Med.* 19:13–33. 2000.
- 91 Lin DY. Cox regression analysis of multivariate failure time data: the marginal approach. *Stat Med.* 13:2233–47. 1994.
- 92 Metcalfe C, Thompson SG. Wei, Lin and Weissfeld's marginal analysis of multivariate failure time data: should it be applied to a recurrent events outcome? *Stat Methods Med Res.* 16:103–22. 2007.
- 93 Cook RJ, Lawless JF. Marginal analysis of recurrent events and a terminating

- event. *Stat Med.* 16:911–24. 1997.
- 94 Ghosh D, Lin DY. Nonparametric analysis of recurrent events and death. *Biometrics.* 56:554–62. 2000.
- 95 Wang MC, Qin J, Chiang CT. Analyzing recurrent event data with informative censoring. *J Am Stat Assoc.* 96:1057–65. 2001.
- 96 Liu L, Wolfe RA, Huang XL. Shared frailty models for recurrent events and a terminal event. *Biometrics.* 60:747–56. 2004.
- 97 Ye Y, Kalbfleisch JD, Schaubel DE. Semiparametric Analysis of Correlated Recurrent and Terminal Events. *Biometrics.* 63:78–87. 2007.
- 98 Fine JP, Jiang H, Chappell R. On semi-competing risks data. *Biometrika.* 88:907–19. 2001.
- 99 Xu JF, Kalbfleisch JD, Tai BC. Statistical Analysis of illness-death processes and semicompeting risks data. *Biometrics.* 66:716–25. 2010.
- 100 Graw F, Gerds TA, Schumacher M. On pseudo-values for regression analysis in competing risks models. *Lifetime Data Anal.* 15:241–255. 2009.
- 101 Datta S, Satten GA. Validity of the Aalen-Johansen estimators of stage occupation probabilities and Nelson-Aalen estimators of integrated transition hazards for non-Markov models. *Stat Probabil Lett.* 55:403–11. 2001.
- 102 Hartzel J, Agresti A, Caffo B. Multinomial logit random effects models. *Stat*

- Modelling*. 1:81–102. 2001.
- 103 Kuss O, McLerran D. A note on the estimation of the multinomial logistic model with correlated responses in SAS. *Comput Methods Programs Biomed*. 87:262–9. 2007.
- 104 Allison PD. *Logistic Regression Using SAS®: Theory and Application, Second Edition*. Cary: SAS Institute Inc.; 2012.
- 105 Agresti A. *Analysis of Ordinal Categorical Data, Second Edition*. Hoboken: Wiley; 2010.
- 106 Agresti A. *Categorical Data Analysis, Third Edition*. Hoboken: Wiley; 2013.
- 107 McCullagh P, Nelder JA. *Generalized linear models, Second edition*. Boca Raton: Chapman and Hall/CRC; 1989.
- 108 Kurland BF, Heagerty PJ. Directly parameterized regression conditioning on being alive: analysis of longitudinal data truncated by deaths. *Biostatistics*. 6:241–58. 2005.
- 109 Goethals K, Janssen P, Duchateau L. Frailty models and copulas: similarities and differences. *J Appl Stat*. 35:1071–9. 2008.
- 110 Hougaard P. Frailty Models. In: Klein JP, vanHouwelingen HC, Ibrahim JG, Scheike TH, editors. *Handbook of Survival Analysis*. Boca Raton: CRC Press; 2014. 457–73.

- 111 Metcalfe C, Thompson SG. The importance of varying the event generation process in simulation studies of statistical methods for recurrent events. *Stat Med.* 25:165–79. 2006.
- 112 Box-Steffensmeier JM, De Boef S. Repeated events survival models: the conditional frailty model. *Stat Med.* 25:3518–33. 2006.
- 113 山本 順司 阪本 良弘. 肝転移の治療. In: 杉原 健一 編. 再発大腸癌治療ガイドブック . 東京: 南江堂; 2003. 89–101.
- 114 Pinson CW, Wright JK, Chapman WC, Garrard CL, Blair TK, Sawyers JL. Repeat hepatic surgery for colorectal cancer metastasis to the liver. *Ann Surg.* 223:765–76. 1996.
- 115 Yamamoto J, Kosuge T, Shimada K, Yamasaki S, Moriya Y, Sugihara K. Repeat liver resection for recurrent colorectal liver metastases. *Am J Surg.* 178:275–81. 1999.
- 116 Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adewuacy, and measuring and reducing errors. *Stat Med.* 15:361–87. 1996.
- 117 Harrell FE Jr. *Regression modeling strategies : with application to linear models, logistic regression, and survival analysis.* New York: Springer-Verlag; 2001.

- 118 Uno H, Cai T, Pencina MJ, D'Agostino RB, Wei LJ. On the C-statistics for evaluating overall adequacy of risk prediction procedures with censored survival data. *Stat Med.* 30. 2011.
- 119 Graf E, Schmoor C, Sauerbrei W, Schumacher M. Assessment and comparison of prognostic classification schemes for survival data. *Stat Med.* 18:2529–45. 1999.
- 120 Cox DR. *Renewal Theory.* London: Methuen; 1962.
- 121 Mieno M, Yamaguchi T, Ohashi Y. Alternative statistical methods for estimating efficacy of interferon beta-1b for multiple sclerosis clinical trials. *BMC Med Res Methodol.* 11:80. 2011.
- 122 Clayton DG. A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika.* 65:141–51. 1978.
- 123 Shih JH. Copula Models. In: Klein JP, vanHouwelingen HC, Ibrahim JG, Scheike TH, editors. *Handbook of Survival Analysis.* Boca Raton: CRC Press; 2014. 489–510.
- 124 Adair RA, Young AL, Cockbain AJ, Malde D, Prasad KR, Lodge JPA, et al. Repeat hepatic resection for colorectal liver metastases. *Br J Surg.* 99:1278–83. 2012.

- 125 Robins JM, Finkelstein DM. Correcting for noncompliance and dependent censoring in an AIDS clinical trial with inverse probability of censoring weighted (IPCW) log-rank tests. *Biometrics*. 56:779–88. 2000.
- 126 Scharfstein DO, Robins JM. Estimation of the failure time distribution in the presence of informative censoring. *Biometrika*. 89:617–34. 2002.
- 127 Matsuyama Y, Yamaguchi T. Estimation of the marginal survival time in the presence of dependent competing risks using inverse probability of censoring weighted (IPCW) methods. *Pharm Stat*. 7:202–14. 2008.