

生存時間型応答の評価指標

—RMST(restricted mean survival time)を理解する—

日本製薬工業協会 医薬品評価委員会

データサイエンス部会

2017～2018 年度 タスクフォース 4

生存時間解析チーム

Ver 2.0

2019 年 3 月

作成・改訂の履歴

バージョン	作成・改訂日	内容
第 1 版	2018 年 5 月	新規作成
第 2 版	2019 年 3 月	<u>1.1 節</u> RMST の最近の報告例を追記. <u>3.2.2 項</u> SAS/STAT 15.1 より, LIFETEST プロシジャの RMST オプションを用いて, RMST の算出が正式に行えるようになった旨を追加. <u>3.4.1 項</u> SAS/STAT 15.1 より, LIFETEST プロシジャを用いて, RMST の層別解析が行えるようになった旨を追加. <u>3.4.2 項</u> SAS/STAT 15.1 より, RMSTREG プロシジャを用いて, RMST の回帰解析が行えるようになった旨を追加. <u>3.5.2 項</u> RMST の必要症例数を計算する SAS マクロで, 最終解析に必要な情報量を算出する機能を追加. <u>3.5.3 項, 3.5.4 項</u> 最終解析に必要な情報量の算出例を追加. <u>3.5.5 項</u> 境界時間τと必要症例数の関係を追加. <u>4.3.2 項</u> 生存時間データを再構築する SAS プログラムを追記. <u>5.3 節</u> シミュレーションで片側検定を用いた理由を追記. <u>7.2.4 項</u> 最終解析に必要な情報量を算出する機能が追加された RMST の必要症例数を計算する SAS マクロへ更新. <u>7.2.5 項</u> 試験実施中に測定する情報量を計算する SAS マクロを追加. <u>7.2.6 項</u> 生存時間データを再構築する SAS プログラムを追加. <u>7.4.3 項</u> 追跡期間とデータモニタリングの性質を追加.

		<u>7.4.4 項</u> 境界時間τをデータに依存して柔軟に決定する方法を提案している文献紹介を追加. <u>全体</u> 誤記修正
--	--	--

目次

略語及び用語の定義	7
1. はじめに	8
1.1. 本報告書の背景と目的.....	8
1.2. 本報告書の構成.....	9
2. 生存時間型応答の評価指標のまとめ.....	11
2.1. 生存時間型応答の評価指標の概要.....	11
2.2. 生存時間型応答の評価指標の詳細.....	12
3. RMST の紹介.....	16
3.1. RMST の定義と性質.....	16
3.1.1. パラメトリック分布を仮定した RMST.....	17
3.2. 推定	18
3.2.1. RMST の推定量.....	18
3.2.2. SAS プログラム例と留意点	18
3.3. 治療群間の比較.....	26
3.3.1. RMST の差.....	26
3.3.2. RMST の比.....	27
3.4. 調整解析	28
3.4.1. 層別解析.....	28
3.4.2. 回帰分析.....	28
3.4.3. Cox 比例ハザードモデルによる調整解析	29
3.5. 症例数設計	30
3.5.1. 計算手順.....	30
3.5.2. SAS マクロの実行	32
3.5.3. 生存時間分布に指数分布を仮定した必要症例数.....	34
3.5.4. 生存時間分布に区分指数分布を仮定した必要症例数.....	37
3.5.5. 境界時間と必要症例数の関係.....	40
4. 適用上の留意点	43
4.1. RMST の意味, 境界時間 τ の意義付け	43
4.1.1. RMST の考え方.....	43
4.1.2. RMST は何を評価しているのか.....	43
4.1.3. RMST の差とハザード比.....	43
4.2. 境界時間 τ の設定方法.....	45

4.2.1.	境界時間 τ を設定する際の留意点.....	45
4.2.2.	境界時間 τ 設定のフローチャート.....	46
4.2.3.	境界時間 τ 設定方法の一例.....	47
4.3.	RMST における群間差, 対照薬群の参照値の見積り方.....	47
4.3.1.	RMST における群間差の見積もり方.....	47
4.3.2.	RMST の参照値の求め方.....	47
4.4.	RMST の群間差に対する統計的推測の情報量.....	52
4.4.1.	情報量の計算式.....	52
4.4.2.	情報量の計算例.....	54
5.	性能評価	58
5.1.	Trinquart et al. (2016)	58
5.1.1.	Trinquart et al. (2016) の方法	58
5.1.2.	Trinquart et al. (2016) の結果	58
5.2.	Tian et al. (2017).....	60
5.2.1.	Tian et al. (2017) の方法.....	60
5.2.2.	Tian et al. (2017) の結果.....	61
5.3.	シミュレーション検討.....	61
5.3.1.	シミュレーションの概要と想定した生存関数.....	62
5.3.2.	シミュレーション結果.....	69
5.3.3.	シミュレーション検討のまとめ.....	76
6.	終わりに	78
7.	APPENDIX.....	79
7.1.	式の導出	79
7.2.	SAS プログラム例	81
7.2.1.	RMST 計算のための前処理.....	81
7.2.2.	RMST の差.....	81
7.2.3.	RMST の比.....	82
7.2.4.	必要症例数の計算.....	84
7.2.5.	試験実施中に測定する情報量の計算.....	93
7.2.6.	生存時間データ再構築の SAS マクロ	96
7.3.	実務上のノウハウ.....	102
7.3.1.	SAS/STAT(R) 9.2 User's Guide, Second Edition の注意点.....	102
7.3.2.	Newton-Raphson 法による指数パラメータの計算	103
7.4.	コラム	105

7.4.1.	症例数設計での試験期間の設定の考え方.....	105
7.4.2.	比例ハザード性の検討.....	106
7.4.3.	追跡期間とデータモニタリングの性質.....	106
7.4.4.	試験途中で境界時間 τ を変更することの是非.....	108
7.4.5.	Kaplan–Meier 曲線を数値化する際の秘訣.....	109

略語及び用語の定義

略語	英名	和名
ARE	Asymptotical relative efficiency	漸近相対効率
ASCO	American society of clinical oncology	米国臨床腫瘍学会
CI	Confidence interval	信頼区間
ERE	Empirical relative efficiency	経験相対効率
IPD	Individual patient data	各患者の個人データ
OS	Overall survival	全生存期間
PFS	Progression free survival	無増悪生存期間
RMST	Restricted mean survival time	境界内平均生存時間 制限付き平均生存時間

1. はじめに

1.1. 本報告書の背景と目的

本報告書の目的は、イベントが発現するまでの時間（生存時間型応答）の評価指標を整理し、近年注目されている RMST について理論的側面の紹介及びその特性を評価し、実際に使用する場合の留意点について紹介することである。

死亡や特定の有害事象の発現などイベントが発現するまでの時間を評価する臨床試験では、Kaplan-Meier 法で生存関数を図示し、log-rank 検定で生存関数の群間比較を行い、Cox 比例ハザードモデルで評価指標であるハザード比を推定し、治療効果の大きさを議論することが多い^[1]。生存関数の評価指標には、ハザード比に加え、半数の被験者がイベントを経験するまでの時間、ある時点の生存割合や規定した特定時点までの平均生存時間などがあり、それぞれ異なった特徴がある。一方実業務では、想定される生存関数及びその群間差がどのような状況であっても、先に示したような定型的な解析が行われ、ハザード比を主な評価指標として用いる場面が多いと思われる。

近年、免疫チェックポイント阻害剤及びがんワクチンなどのがん免疫療法の臨床試験成績が多数報告され、国内外で承認された医薬品も複数ある。これらの薬剤は、腫瘍細胞に対する特異的な免疫応答を誘導することで抗腫瘍効果を示すと考えられており、効果の発現までに一定の期間を要する可能性があると考えられている。この考えに基づくと、プラセボ群と試験薬群の生存曲線が治療開始してから重なったまま推移し、一定の時期以降に生存曲線の差が開く状況が考えられ、実際の臨床試験でもこのような結果が報告されている。また、実薬対照の場合には、生存関数が交差することが報告されている試験もある。これらの場合、「時点によらずハザード比が一定である」という比例ハザード性が成り立たず、推定されたハザード比による結果の解釈が難しくなることが生じえる点が指摘されている^{[2][3]}。2016年に公表された「がん免疫療法開発のガイダンス後期臨床試験の考え方」の報告書^[1]では III.3) 統計解析法の節において、比例ハザード性が成り立たないと想定される場合の解析方法の1つとして、規定した特定の時点までの生存曲線下の面積* (restricted mean survival time, RMST) を群間比較する手法が記載されている。また2017年3月9日に行われた第4回データサイエンスラウンドテーブル会議の抗がん剤の第III相試験における生存時間解析に関する統計的課題のテーマの中で RMST について意見が交換された^[4]。さらに国内だけでなく、Journal of clinical oncology などの雑誌でも比例ハザード性が成り立たない場合の代替指標の1つとして、RMST の利用について近年議論がされ始めている^{[2][3][5]}。2018年6月に米国で開催された American Society of Clinical Oncology (ASCO) の年会では、RMST に関連する5つの演題が報告された。特に、HER2 陽性の高齢乳癌患者を対象とした術後補助化学療法の試験では、盲検下での中間解析で観察イベント数が計画時の期待よりも少なく、ハザード比による非劣性の検出力が保証されなかったため、RMST の解析が追加されたという報告がされた^[6]。また、高齢成人に対する心疾患イベントの一次予防効果を確認する Pravastatin と通常診療の比較を行った試験への適応など、抗がん剤以外の領域においても、

RMSTの適応事例が報告され始めている^[7]。しかしながら、本報告書の第2版作成時点で我々が調査した範囲では、RMSTを主要評価指標として事前計画され、結果が報告された試験を確認できていない。このため、RMSTの実務的な利用が十分に行われているとは、まだ言い難い状況と考えられる。

RMSTによる解析は、比例ハザード性が成立していない場合における有益な方法の1つの選択肢となりえると考えているが、RMSTを中心にまとめられた成書は存在せず、実用のためにはその性質について実務的な検討が必要であると考え、本タスクフォースにより検討を行った。

1.2. 本報告書の構成

本報告書は、統計解析担当者がRMSTを使用する際に必要な情報を本タスクフォースのメンバーの観点からまとめたものである。さらに、臨床担当者にも有益な情報も一部含めた。構成は以下の通りである。

2章では生存時間型応答の評価指標をまとめた。3章ではRMSTの理論的な側面及びプログラムなどの実務的な側面の情報をまとめた。4章では適応上の留意点としてRMSTを使用する際に検討すべき点をまとめた。5章では性能評価を報告した論文の紹介及び本タスクフォースが実施したシミュレーションによる性能評価結果を示した。6章で本報告書をまとめ、さらに7章にはAppendixとして式の導出や本タスクフォースメンバーによるコラムなどを含めた。

臨床担当者には、まず2章をご一読いただきたい。各評価指標を直感的に理解できるように図を用いて整理した。また、RMSTを使用する場合には臨床担当者の一定の理解も必要となるため、4.1節及び4.2節に示したRMSTの意味や境界時間 τ に関する内容についてもご一読いただきたい。それ以外の章は統計解析担当者が実際にRMSTを実務で使用する場合に必要と思われる情報を盛り込んだ。

本報告書では、生存時間型応答の評価指標の1つとして新たに注目されているRMSTに焦点を当てたが、RMSTの利用を推奨したり、既存の評価指標を否定しているものではないことは強調したい。複数ある評価指標の各特徴を踏まえた上で、試験結果に対して臨床的・統計的に適切な解釈を与えられる評価指標を選択することが望ましい姿である。

参考文献

- [1] がん免疫療法開発のガイダンス 後期臨床試験の考え方 報告書
<https://www.pmda.go.jp/files/000221609.pdf>
- [2] Uno H, Claggett B, Tian L, Inoue E, Gallo P, Miyata T, *et al.* Moving beyond the hazard ratio in quantifying the between-group difference in survival analysis. J Clin Oncol.2014; **32**: 2380–2385.

- [3] Royston P, Parmar MKB. Restricted mean survival time: an alternative to the hazard ratio for the design and analysis of randomized trials with a time-to-event outcome. BMC Med Res Methodol. 2013; **13**: 152.
- [4] 第4回データサイエンスラウンドテーブル会議
<https://www.pmda.go.jp/review-services/symposia/0060.html>
- [5] Hasegawa T, Uno H, Wei LJ. Zoledronic Acid Dosing in Patients With Metastatic Breast Cancer. JAMA Oncol. 2017.doi:10.1001/jamaoncol.2017.0487.
- [6] Sawaki M, Saito T, Baba S, Kobayashi K, Kawashima H, Tsuneizumi M et al. Evaluation of trastuzumab without chemotherapy as a postoperative adjuvant therapy in HER2-positive elderly breast cancer patients: Randomized controlled trial (RESPECT). J Clin Oncol 2018; 36:suppl abstr 510.
- [7] Orkaby AR, *et al.* Pravastatin for Primary Prevention in Older Adults: Restricted Mean Survival Time Analysis. J Am Geriatr Soc. 2018; 66:1987-1991.

*本報告書では RMST の日本語訳としては、境界内平均生存時間を用いる。定義については 3.1 節を参照のこと

2. 生存時間型応答の評価指標のまとめ

本章では、生存時間型応答の評価指標として、主に利用が想定される時点生存割合、平均生存時間、生存時間中央値 (Median Survival Time)、ハザード、RMST を取り扱う。これらを列挙し、各指標の定義を記述する。なお、以下では、各被験者のイベント発現までの時間を「生存時間」と定義する。また、文脈の都合上、生存関数を生存曲線と呼ぶ場合がある。

2.1. 生存時間型応答の評価指標の概要

生存時間型応答の評価指標の概要を図 2-1 に図示した。ただし、ハザードを視覚的に表示することは困難であるため対象外とした。各指標の詳細は 2.2 節を参照いただきたい。ここでは各指標の定義又は図的解釈の紹介のみとした。なお、定義中に用いた「率」と「割合」という用語は意図をもって使い分けている点に留意すること。「率」は速度の概念（分母に時間の尺度を有する）を含み 0 から無限大の値をとる尺度である。一方、「割合」は速度の概念を含まず、0 から 1 の間の値をとる無次元の尺度である。

- 時点生存割合：ある特定の時点での生存割合
- 平均生存時間：生存曲線の曲線下面積全体
- 生存時間中央値：生存割合が 50%に到達するまでの期間
- ハザード：ある時点まで生存したという条件付きで次の瞬間にイベントが発現する率
- RMST：規定した特定の時点までの生存曲線下面積

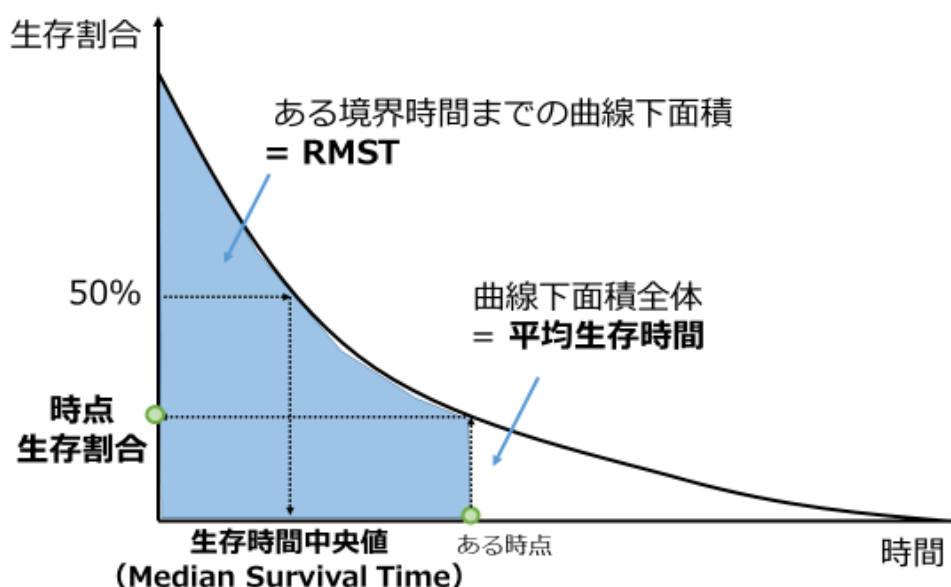


図 2-1 生存時間型応答の評価指標の概要

2.2. 生存時間型応答の評価指標の詳細

1. 時点生存割合

時点生存割合とは、ある特定の時点での生存割合のことである。 T を当該被験者の生存時間を表す確率変数とすると、時点 t での時点生存割合 $S(t)$ は

$$S(t) = P(T \geq t)$$

と表される。解析に用いられる時点 t は対象疾患や試験目的等を考慮して選択される。例えば乳癌の術後補助化学療法を対象とした臨床試験では、時点 t として3年を選択し、3年無病生存割合によって生存曲線を要約することがある。

2. 平均生存時間

平均生存時間とは、生存時間 T の期待値のことである。平均生存時間を μ と表すことにすると、 μ は生存曲線の曲線下面積に等しくなることが知られており、

$$\mu = E[T] = \int_0^{\infty} t f(t) dt = \int_0^{\infty} S(t) dt$$

として算出可能である。

いま、タイデータ（同一の生存時間データ）が存在せず、打ち切りがないことを仮定し、Kaplan–Meier 法で生存関数を推定することを考え、生存時間を短い順に $t_1 < t_2 < \dots < t_n$ (n は当該集団に属する被験者数) とすると、 μ は

$$\hat{\mu} = 1.0t_1 + \hat{S}(t_1)(t_2 - t_1) + \hat{S}(t_2)(t_3 - t_2) + \dots + \hat{S}(t_{n-1})(t_n - t_{n-1})$$

として推定される。すなわち、推定された生存曲線下の各矩形の面積の和を算出すればよい。なお、打ち切りがない状況下では、 $\hat{\mu}$ は各被験者の生存時間 t_i ($i = 1, \dots, n$) に対する単純平均

$$\bar{t} = \sum_{i=1}^n \frac{t_i}{n}$$

に帰着する。

ここで、当該集団では全被験者がイベントを発現していることを仮定しているため、 t_n は最終イベント発現までの時間である。仮に当該集団で観察された生存時間のうち最長の生存時間データが打ち切りの場合、 t_n が観測できず、Kaplan–Meier 法によって求められる生存関数が 0 に到達せず閉じた領域とならない。このため、生存曲線の曲線下面積が定義できず、平均生存時間を仮定なく推定することはできない。対処法として生存関数にモデルを仮定し、平均生存時間を推定することができるが、得られた平均生存時間は仮定したモデルに依存する。モデルに依存しない方法として、打ち切り時間を含む最長の生存時間まで、もしくは最長のイベント発現時間までの範囲で生存曲線の曲線下面積を算出し、これを平均生存時間とすることが考えられる。但し、どちらの方法でも、本来の平均生存時間 μ よりも過小に評価されることやデータに依存して生存曲線の曲線下面積の算出範囲が決定されるという問題もある。

3. 生存時間中央値 (Median Survival Time)

当該集団の生存割合が 50%に到達するまでの期間のことである。すなわち、 $\hat{S}(t) = 0.5$ を満たす t が生存時間中央値 (Median Survival Time)である。

4. ハザード

時点 t まで生存したという条件付きで次の瞬間にイベントが発現する率のことである。すなわち

$$h(t) = \lim_{\Delta t \rightarrow 0} \frac{\Pr(t \leq T \leq t + \Delta t | t \leq T)}{\Delta t}$$

である。群間差の指標として各群のハザードの比（ハザード比）が報告されることが一般的であり、通常 Cox 比例ハザードモデルにより推定される。

5. 規定した特定の時点までの生存曲線下の面積(RMST)

境界時間 τ 内での生存時間に対する平均値のことである。すなわち、境界時間 τ 内での生存時間を $X(\tau) = \min(T, \tau)$ とした場合、 $X(\tau)$ の平均値

$$\mu(\tau) = E[X(\tau)] = E[\min(T, \tau)]$$

が RMST である。詳細は 3 章を参照のこと。

上述した各指標の長所、短所を表 2-1 に整理した^{[1][2]}。

表 2-1 生存時間型応答の評価指標の長所・短所

評価指標	長所	短所
時点生存割合	- 臨床的に解釈しやすい。 - 特に生存時間中央値を推定できないようなイベント発現が稀な疾患で、要約指標として頻用されている。比較試験のサンプルサイズを設計する際、ハザード比を見積もるための対照群の参照値として利用しやすい。 - モデルに依存せず Kaplan-Meier 法によって推定された生存曲線に基づいて推定できる。	- ある 1 時点の代表値であり、全時点での生存時間の振舞いを把握することが困難である。 - 事前に、妥当な評価時点を決める必要がある。

評価指標	長所	短所
平均生存時間	臨床的に解釈しやすい.	最終生存時間が打ち切りの場合に定義できない.
生存時間 中央値	- 臨床的に解釈しやすい. - 外れ値に影響されづらい. - 臨床の現場で要約指標として頻用されている. 比較試験のサンプルサイズを設計する際, ハザード比を見積もるための対照群の参照値として利用しやすい.	- 生存割合が 50%に到達していないと推定できない. - ある 1 時点の代表値であり, 全時点での生存時間の振舞いを把握することが困難である.
ハザード (ハザード比)	比例ハザード性の仮定の下, 群間差の妥当な評価指標としてハザード比が頻用されており, データが豊富に存在する.	- ハザードは単一の指標として直観的に解釈しづらい - 比例ハザード性の仮定が崩れるとハザード比の解釈が困難となる.
RMST	- 臨床的に解釈しやすい. - 最終生存時間が打ち切りの場合の平均生存時間の問題点が解決されている. - Kaplan–Meier 曲線の境界時間 τ までの情報をすべて利用している点で, 時点生存割合よりより多くの情報を用いた推定値となっている.	事前に妥当な境界時間 τ を決める必要がある.

参考文献

- [1] Everardo D. Saad, John R. Zalcberg, Julien Pe´ron, Elisabeth Coart, Tomasz Burzykowski, Marc Buyse. Understanding and Communicating Measures of Treatment Effect on Survival: Can We Do Better? J Natl Cancer Inst. 2018; **110**: 232-240.
- [2] Uno H, Wittes J, Fu H, Solomon SD, Claggett B, Tian L, *et al.* Alternatives to Hazard Ratios for Comparing the Efficacy or Safety of Therapies in Noninferiority Studies. Ann Intern Med. 2015; **163**: 127–134.
- [3] Collett D. Modelling survival data in medical research, third edition. CRC Press; 2015.
- [4] Klein JP, Moeschberger ML. Survival Analysis: Techniques for Censored and Truncated Data second edition. Springer-Verlag: New York; 2003.

- [5] Lee ET, Wang JW. Statistical Methods for Survival Data Analysis. second edition. John Wiley & Sons: New York; 1992.

3. RMST の紹介

3.1. RMST の定義と性質

RMST は、「境界時間 τ 内でのイベント発現までの時間に対する平均値」と定義される。すなわち、イベント発現までの時間（生存時間）を T とし、境界時間 τ 内での生存時間を $X(\tau) = \min(T, \tau)$ （以下、 $X(\tau)$ は境界時間 τ 内での生存時間を表す）とした場合、 $X(\tau)$ の平均値

$$\mu(\tau) = E[X(\tau)] = E[\min(T, \tau)] \quad (3.1)$$

が RMST である。

RMST の解釈を例を用いて説明する。例えば、あるがんの治療を開始した被験者を追跡したとする。このとき、境界時間 τ を 4 年としたときの RMST が 1 年だったとする。この RMST が 1 年であるとは、治療開始後 4 年間の平均生存期間は 1 年であるということである。RMST が何を評価しているかについては、4 章に詳細に記載したので参照いただきたい。

$f(t)$, $S(t)$ をそれぞれ生存時間 T の確率密度関数、生存関数とすると、RMST は、

$$\begin{aligned} \mu(\tau) &= E[X(\tau)] = E[\min(T, \tau)] \\ &= \int_0^{\tau} t f(t) dt + \int_{\tau}^{\infty} \tau f(t) dt \\ &= \int_0^{\tau} S(t) dt \end{aligned} \quad (3.2)$$

と表現できる。つまり、RMST は、「境界時間 τ 内における生存関数の曲線下面積」としても解釈できる。式(3.2) 2 行目において、積分区間を境界時間 τ までと τ 以降に分割している。 $T \geq \tau$ のとき、境界時間 τ 以降は打ち切りとするため積分内は t ではなく τ となっている。式(3.2) の導出は本報告書 7.1 節を参照いただきたい。

必要症例数の計算で使用するため、ここで境界時間 τ 内での生存時間 $X(\tau)$ の分散を考える。まず、分散を求めるために必要な $E[X^2(\tau)]$ を算出する。

$$\begin{aligned} E[X^2(\tau)] &= \int_0^{\tau} t^2 f(t) dt + \int_{\tau}^{\infty} \tau^2 f(t) dt \\ &= 2 \int_0^{\tau} t S(t) dt \end{aligned} \quad (3.3)$$

よって、分散 $\sigma^2(\tau) (= \text{Var}[X(\tau)])$ は、

$$\begin{aligned} \sigma^2(\tau) &= \text{Var}[X(\tau)] = E[X^2(\tau)] - \{E[X(\tau)]\}^2 \\ &= 2 \int_0^{\tau} t S(t) dt - \left\{ \int_0^{\tau} S(t) dt \right\}^2 \end{aligned} \quad (3.4)$$

と得られる。式(3.3) の導出は本報告書 7.1 節を参照いただきたい。

上述のように、境界時間 τ 内での生存時間 $X(\tau)$ の平均値 (RMST)、及び分散は、生存関数で定義される。RMST とその分散の推定値はデータから生存関数を推定することで得られる。

3.1.1. パラメトリック分布を仮定した RMST

生存時間 T が指数分布に従うと仮定した場合の境界時間 τ 内の生存時間 $X(\tau)$ の平均値 (RMST), 及び分散を示す^[1]. 時間 t に対する生存関数は $S(t) = \exp(-\lambda t)$ である. この時の境界時間 τ 内の RMST 及び生存時間 $X(\tau)$ の分散は式(3.2) 及び式(3.4) より,

$$\mu(\tau) = E[X(\tau)] = \frac{1 - \exp(-\lambda\tau)}{\lambda} \quad (3.5)$$

$$\sigma^2(\tau) = \text{Var}[X(\tau)] = \frac{1 - 2\lambda\tau\exp(-\lambda\tau) - \exp(-2\lambda\tau)}{\lambda^2} \quad (3.6)$$

である.

同様に, 生存時間分布に区分指数分布を仮定した場合の RMST 及び分散を示す. 区分指数分布とは時間の区間毎に異なるある一定のハザードを仮定する分布である. 既知の時点 $t_{(1)}, \dots, t_{(k)}$ に対して, 次の区間を構成する.

$$(t_{(0)} = 0, t_{(1)}], (t_{(1)}, t_{(2)}], (t_{(2)}, t_{(3)}], \dots, (t_{(k)}, t_{(k+1)} = \infty] \quad (3.7)$$

境界時間 τ は区間 $(t_{(k)}, t_{(k+1)})$ に存在し, $\tau > t_{(k)}$ を満たすとする. 時間の区間幅を δ_{j+1} ($j = 0, \dots, k$)とする.

$$\delta_{j+1} = \begin{cases} t_{(j+1)} - t_{(j)}, & j < k \\ \tau - t_{(k)}, & j = k \end{cases} \quad (3.8)$$

区間 $(t_{(j)}, t_{(j+1)})$ のある一定のハザードを h_{j+1} とする. 時間 $t_{(j)}$ までの累積ハザードは $H_j = H(t_{(j)}) = \sum_{i=1}^j h_i \delta_i$ である. また, $h_0 = H_0 = 0$ とする. 時間 $t \in (t_{(j)}, t_{(j+1)})$ に対する生存関数は, 次のように書ける.

$$S_{j+1}(t) = \exp(-H_j) \exp(-h_{j+1}(t - t_{(j)})) \quad (3.9)$$

この区分指数分布を仮定した境界時間 τ 内の RMST 及び生存時間 $X(\tau)$ の分散は, 式(3.2) 及び式(3.4) より,

$$\mu(\tau) = E[X(\tau)] = \sum_{j=0}^k \exp(-H_j) B_{j+1} \quad (3.10)$$

$$\sigma^2(\tau) = \text{Var}[X(\tau)] = 2 \sum_{j=0}^k \exp(-H_j) (A_{j+1} + t_{(j)} B_{j+1}) - \left\{ \sum_{j=0}^k \exp(-H_j) B_{j+1} \right\}^2 \quad (3.11)$$

である. ここで,

$$\begin{aligned} A_{j+1} &= \frac{1}{h_{j+1}^2} \{1 - (1 + h_{j+1} \delta_{j+1}) \exp(-h_{j+1} \delta_{j+1})\} \\ B_{j+1} &= \frac{1 - \exp(-h_{j+1} \delta_{j+1})}{h_{j+1}} \end{aligned} \quad (3.12)$$

である.

3.2. 推定

3.2.1. RMST の推定量

RMST の推定として、Kaplan–Meier 法による生存曲線を積分する方法、ジャックナイフ法、フレキシブル・パラメトリック回帰モデリング法など、複数の方法が示されている^[1]. 生存関数を推定する際には Kaplan–Meier 法を用いることが一般に行われているため、Kaplan–Meier 法によって求められた生存関数を積分する方法について述べる. なお、本報告書では、イベント発現時間と打ち切り時間は独立であることを仮定する.

RMST は、境界時間 τ 内における生存関数の曲線下面積であることから、その推定量は、

$$\hat{\mu}(\tau) = \int_0^{\tau} \hat{S}(t) dt = \sum_{j=0}^D (t_{j+1} - t_j) \hat{S}(t_j) \quad (3.13)$$

となる. Kaplan–Meier 法による生存曲線の推定量を $\hat{S}(t)$, 境界時間 τ 内での相異なる D 個のイベント発現時点を $t_1 < t_2 < \dots < t_D$, $t_0 = 0$, $t_{D+1} = \tau$ とする. RMST の推定量 $\hat{\mu}(\tau)$ の分散は、Greenwood の公式を用いて、

$$\begin{aligned} \text{Var}[\hat{\mu}(\tau)] &= \text{Var} \left[\int_0^{\tau} \hat{S}(t) dt \right] = \text{Var} \left[\sum_{j=0}^D (t_{j+1} - t_j) \hat{S}(t_j) \right] \\ &= \sum_{j=1}^D \left[\int_{t_j}^{\tau} \hat{S}(t) dt \right]^2 \frac{d_j}{Y_j(Y_j - d_j)} \\ &= \sum_{j=1}^D \left[\sum_{i=j}^D (t_{i+1} - t_i) \hat{S}(t_i) \right]^2 \frac{d_j}{Y_j(Y_j - d_j)} \end{aligned} \quad (3.14)$$

と表せる^[2]. ここで、 Y_j , d_j はイベントが発現した時点 t_j でのリスク集合の大きさ、イベント数である. 式(3.14) の導出は本報告書 7.1 節参照いただきたい.

3.2.2. SAS プログラム例と留意点

SAS の LIFETEST プロシジャの TIMELIM オプションに境界時間 τ を指定することによって、境界時間 τ までの RMST とその標準誤差を計算することが出来る. 基本的な SAS プログラムの構文は次の通りである.

```
PROC LIFETEST DATA = DATA TIMELIM = L ;  
    TIME TIME * CENSOR(0) ;  
RUN ;
```

ここで、DATA はデータセット、TIME は時間変数、CENSOR は打ち切り変数 (0 は打ち切

りを表す) である. L には境界時間 τ を表す数値を指定する.

例えば, 以下のデータセット **SAMPLE** の場合を考える.

表 3-1 データセット **SAMPLE**

TIME	CENSOR
1	0
2	1
3	1
4	1
5	0

境界時間 $\tau = 5$ までの **RMST** を求めるためには, 以下のようにプログラミングすれば良い.

```
PROC LIFETEST DATA = SAMPLE TIMELIM = 5 ;  
  TIME TIME * CENSOR(0) ;  
RUN ;
```

RMST に関する出力結果は以下の通りである.

平均	標準誤差
3.50000	0.68465

SAS プログラミング時の留意点

1) 利用するデータセット

下記 SAS/STAT(R) 14.1 User's Guide^[3] の記載の通り, **TIMELIM** に指定した時間より後に発現したイベントがある場合, **TIMELIM** の指定に関わらず強制的に最終イベント発現時点を τ とした **RMST** が算出される. そのため, 境界時間 τ を最終イベントより前に設定したい場合は, データセット中にある境界時間 τ より後に発現したイベントをすべて打ち切りに変換しておく必要がある.

【SAS/STAT(R) 14.1 User's Guide 抜粋】

TIMELIM=*time-limit*

specifies the time limit used in the estimation of the mean survival time and its standard error.

The mean survival time can be shown to be the area under the Kaplan–Meier survival curve.

However, if the largest observed time in the data is censored, the area under the survival curve is not a closed area. In such a situation, you can choose a time limit L and estimate the mean survival curve limited to a time L (Lee 1992, pp. 72–76). This option is ignored if the largest observed time is an event time.

上記データセット SAMPLE の Kaplan–Meier 曲線と各範囲内の曲線下面積は以下の通りである。

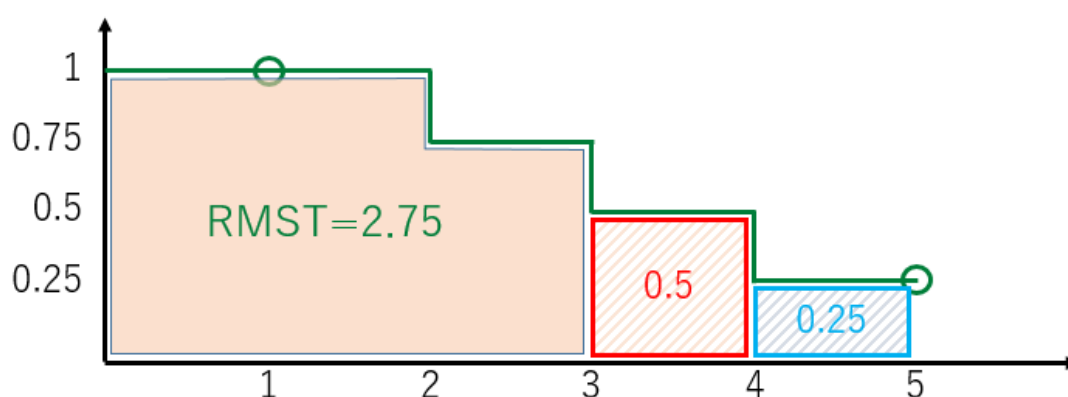


図 3-1 データセット SAMPLE の Kaplan–Meier 曲線とその曲線下面積

従って、境界時間 $\tau = 3$ までの RMST は 2.75 である。

しかし、データセット SAMPLE に対して、以下のようにプログラミングすると、出力される RMST は 3.25 となる。これは、最終イベント発現時点（時点 4）までの RMST である。

```
PROC LIFETEST DATA = SAMPLE TIMELIM = 3 ;
  TIME TIME * CENSOR(0) ;
RUN ;
```

平均	標準誤差
3.25000	0.50775

境界時間 $\tau = 3$ までの RMST を求めるためには、あらかじめ境界時間 τ より後に発生したイベントを打ち切りに変換したデータセット（ここでは SAMPLE1 とする）を作成しておく必要がある（境界時間 τ より後に発生したイベントを打ち切りに変換するプログラム例は本報告書 7.2.1 項参照）。データセット SAMPLE1, SAS プログラム及び RMST に関する出力結果は以下の通りである。

表 3-2 データセット SAMPLE1, SAS プログラム及び出力結果

	TIME	CENSOR	
	1	0	
	2	1	
	3	1	
	4	0	
	5	0	
PROC LIFETEST DATA = SAMPLE1 TIMELIM = 3 ; TIME TIME * CENSOR(0) ; RUN ;			
平均 標準誤差 2.75000 0.30619			

2) 観測された生存時間データが境界時間 τ よりも短い場合の RMST の推定

生存時間の最も長い被験者が打ち切りであり、かつその打ち切り時間が境界時間 τ よりも短い場合、SAS LIFETEST では RMST の推定値が境界時間 τ に依存して推定される。

表 3-1 に示したデータの例を用いて説明する。このデータは、時間 5 の打ち切りが最も長い生存時間として観測されている。前述の通り、境界時間 τ を 5 とした場合の RMST は、以下の SAS プログラムを実行して 3.5 と推定される。

```
PROC LIFETEST DATA = SAMPLE TIMELIM = 5 ;
  TIME TIME * CENSOR(0) ;
RUN ;
```

次に、打ち切り時間 5 よりも長い時間を境界時間 τ に設定した場合、例えば、境界時間 τ を 10 とした RMST は、以下の SAS プログラムを実行して 4.75 と推定される。この推定値は、最も長い生存時間である時間 5 の打ち切りを境界時間の 10 まで引き延ばして推定したことに相当する。

```
PROC LIFETEST DATA = SAMPLE TIMELIM = 10 ;
  TIME TIME * CENSOR(0) ;
RUN ;
```

つまり、境界時間 τ を5とした場合と10とした場合では、RMSTの推定値に異なる結果が得られる。この例のように、生存時間の最終オブザベーションが打ち切りであり、かつその打ち切り時間が境界時間 τ よりも短い生存時間データの場合、SAS LIFETESTではRMSTの推定値が境界時間 τ に依存して推定されることに注意が必要である。

3) 出力される標準誤差

SASのLIFETESTプロシジャでは、次の式を使った標準誤差が出力される。

$$\sqrt{\text{Var}[\hat{\mu}(\tau)]} = \sqrt{\frac{m}{m-1} \sum_{j=1}^D \left[\sum_{i=j}^D (t_{i+1} - t_i) \hat{S}(t_i) \right]^2 \frac{d_j}{Y_j(Y_j - d_j)}} \quad (3.15)$$

ここで、 $m = \sum_{j=1}^D d_j$ である。

式(3.15)は、式(3.14)の平方根を $\sqrt{m/(m-1)}$ 倍しているため、式(3.14)の結果を得たい場合は、SASで出力される標準誤差を $\sqrt{(m-1)/m}$ 倍する必要がある。

2つの分散の性質

前述の通り、RMSTの推定量の分散には2つの計算式があり、実務上どちらを使えば良いのかについては、成書が存在しないため、現時点で明確な指針は存在しない。実務上有益と思われるので、2つの分散について検討した。

式(3.14)をVar_Klein、式(3.15)の2乗をVar_SASとする。計算式の通り、 $\text{Var_Klein} < \text{Var_SAS}$ の関係が成り立つ。以下で、生存時間 T が指数分布に従うと仮定した場合について、2つの推定値の性質を確認する。

被験者 i ($i = 1, \dots, n$)のイベント発現までの時間（生存時間）を T_i とし、各被験者の生存時間は同一の分布に従うとする。境界時間 τ 内での打ち切りが存在しないと仮定した場合、RMST ($\mu(\tau)$)の推定量は各被験者の境界時間 τ 内での生存時間 $X_i(\tau)$ の単純平均 $\hat{\mu}(\tau) = n^{-1} \sum_{i=1}^n X_i(\tau)$ となるため、その分散は $\text{Var}[\hat{\mu}(\tau)] = \text{Var}[X_i(\tau)]/n$ である。シミュレーションにより、Var_Klein及びVar_SASと、式(3.6)をもとに計算した $\text{Var}[\hat{\mu}(\tau)]$ との変化率 $((\text{Var_Klein} - \text{Var}[\hat{\mu}(\tau)])/\text{Var}[\hat{\mu}(\tau)])$ 、 $((\text{Var_SAS} - \text{Var}[\hat{\mu}(\tau)])/\text{Var}[\hat{\mu}(\tau)])$ を求め、真値からのブレの大きさを確認した。

シミュレーション回数は100,000回、境界時間 $\tau = 2$ 年とした。被験者数は30, 50, 100名とし、2年生存割合が0.9, 0.7, 0.5, 0.3, 0.1の5パターンの指数分布に従う生存時間データを発生させ、RMSTの分散を比較した。

表 3-3 被験者数 100 名の場合

2 年 生存割合	境界時間 $\tau = 2$ 年以内の 期待イベント数	$\text{Var}[X_i(\tau)]/n$	Var_Klein の変化率	Var_SAS の変化率
0.9	10	0.001265	-0.013414 (0.001210) [3]	0.099393 (0.001263) [26]
0.7	30	0.003350	-0.011406 (0.000556) [0]	0.022994 (0.000566) [0]
0.5	50	0.004733	-0.010854 (0.000317) [0]	0.009474 (0.000321) [0]
0.3	70	0.005177	-0.010532 (0.000218) [0]	0.003894 (0.000222) [0]
0.1	90	0.003995	-0.009718 (0.000353) [0]	0.001458 (0.000358) [0]

上段：Mean (SE), 下段：計算不可回数

(計算不可回数：Var_Klein はイベント数 0 の回数, Var_SAS はイベント数 0 及び 1 の回数)

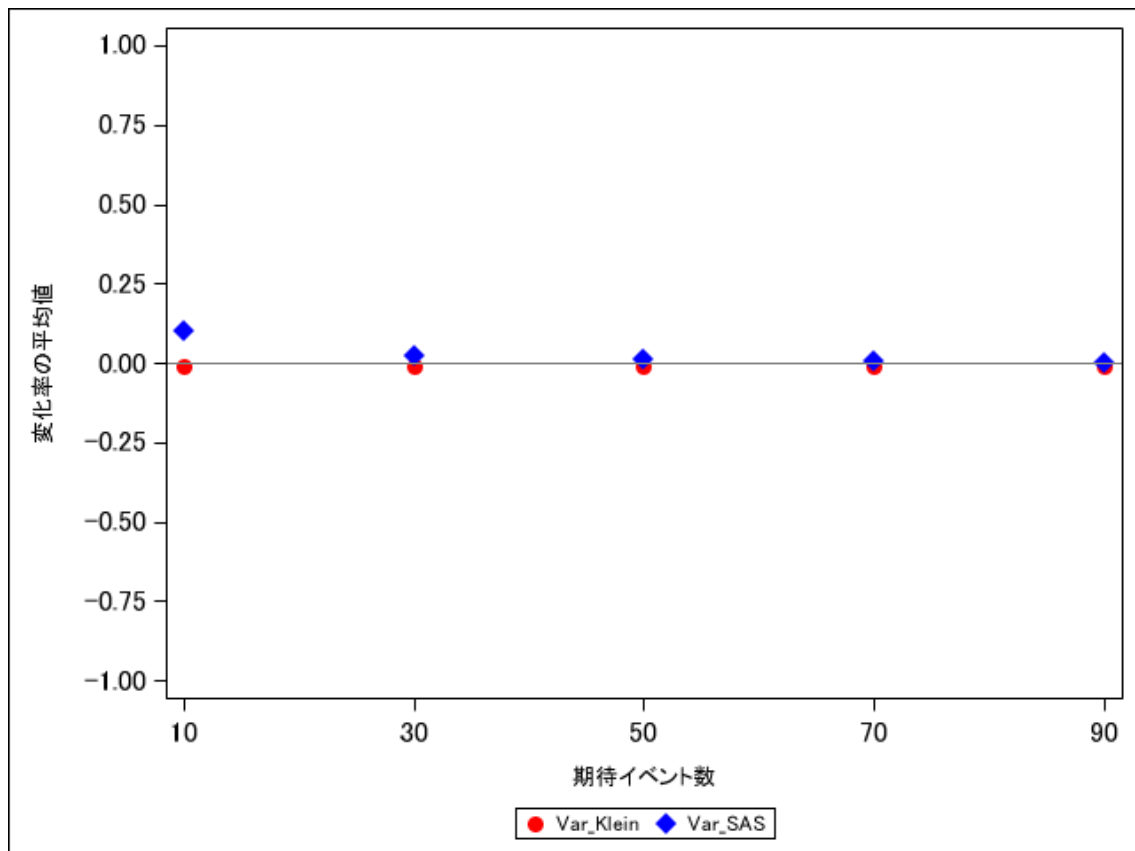


図 3-2 被験者数 100 名の場合 ($\text{Var}[X_i(\tau)]/n$ との変化率の平均値)

表 3-4 被験者数 50 名の場合

2 年 生存割合	境界時間 $\tau = 2$ 年以内の 期待イベント数	$\text{Var}[X_i(\tau)]/n$	Var_Klein の変化率	Var_SAS の変化率
0.9	5	0.002530	-0.026240 (0.001720) [501]	0.268747 (0.001870) [3456]
0.7	15	0.006700	-0.022728 (0.000783) [0]	0.048542 (0.000810) [0]
0.5	25	0.009467	-0.021683 (0.000449) [0]	0.019680 (0.000462) [0]
0.3	35	0.010354	-0.020878 (0.000311) [0]	0.008282 (0.000323) [0]
0.1	45	0.007989	-0.019464 (0.000496) [0]	0.003021 (0.000510) [0]

上段：Mean (SE), 下段：計算不可回数

(計算不可回数：Var_Klein はイベント数 0 の回数, Var_SAS はイベント数 0 及び 1 の回数)

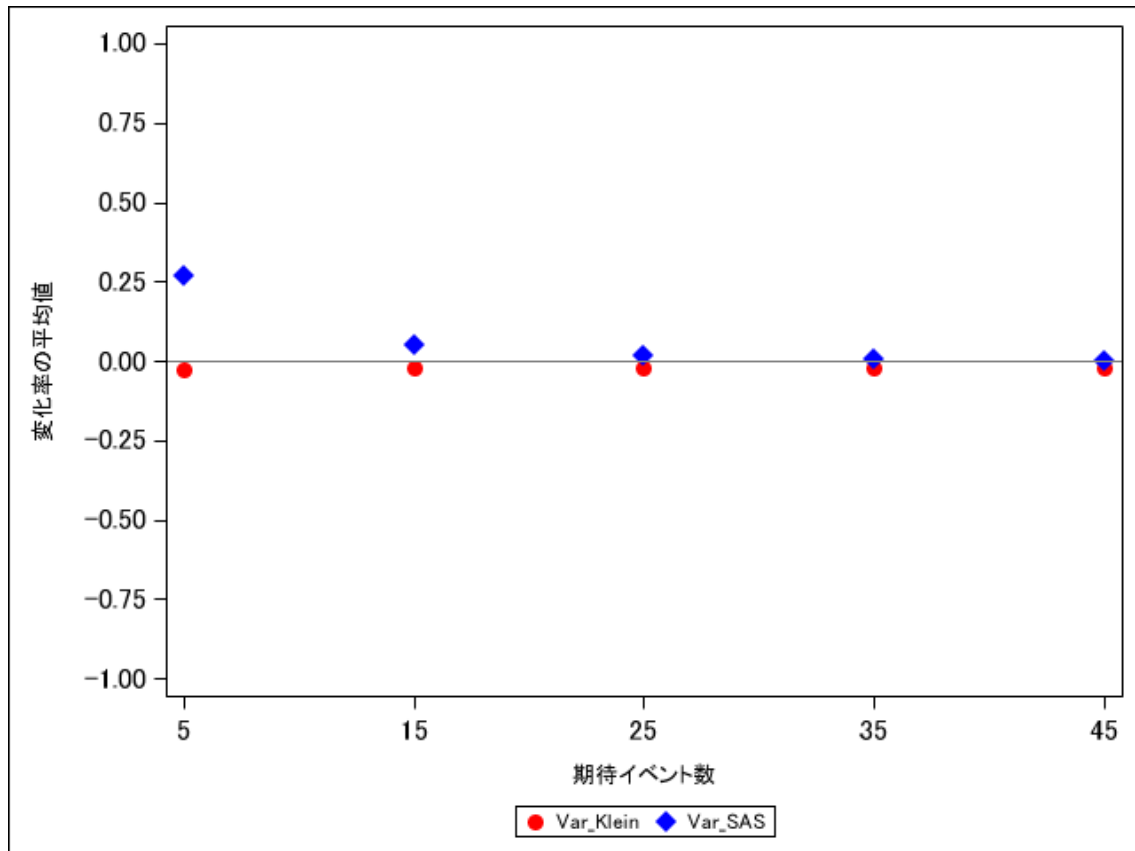


図 3-3 被験者数 50 名の場合 ($\text{Var}[X_i(\tau)]/n$ との変化率の平均値)

表 3-5 被験者数 30 名の場合

2 年 生存割合	境界時間 $\tau = 2$ 年以内の 期待イベント数	$\text{Var}[X_i(\tau)]/n$	Var_Klein の変化率	Var_SAS の変化率
0.9	3	0.004217	-0.047423 (0.002316) [4319]	0.625805 (0.002877) [18628]
0.7	9	0.011167	-0.036021 (0.001002) [3]	0.090282 (0.001061) [34]
0.5	15	0.015778	-0.034887 (0.000580) [0]	0.035939 (0.000610) [0]
0.3	21	0.017257	-0.034298 (0.000408) [0]	0.015037 (0.000435) [0]
0.1	27	0.013316	-0.033068 (0.000639) [0]	0.004680 (0.000670) [0]

上段：Mean (SE), 下段：計算不可回数

(計算不可回数：Var_Klein はイベント数 0 の回数, Var_SAS はイベント数 0 及び 1 の回数)

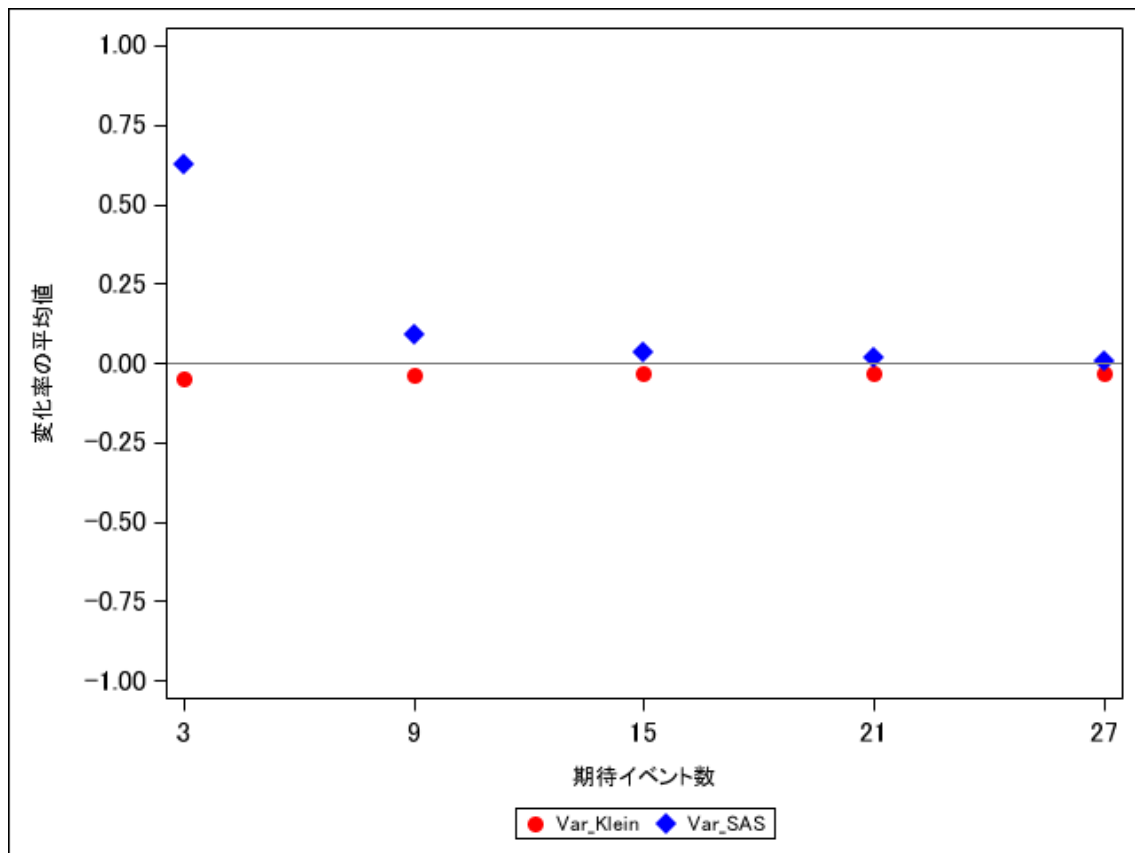


図 3-4 被験者数 30 名の場合 ($\text{Var}[X_i(\tau)]/n$ との変化率の平均値)

シミュレーションの結果，式(3.14) で求められる分散の推定値 (Var_Klein) は $\text{Var}[X_i(\tau)]/n$ より小さくなる傾向が認められた．一方，式(3.15) の推定値は，イベント数が少ないと $\text{Var}[X_i(\tau)]/n$ より大きくなり，不偏性が悪くなる傾向があるものの，イベント数が多い場合は $\text{Var}[X_i(\tau)]/n$ に近くなる傾向が認められた．

Klein et al. (2003)^[2]，Collett (2015)^[4] 及び R の `survfit` 関数の解説^[5] では，Var_Klein の式が示されている．また，SAS では，Kaplan et al. (1958)^[6] 及び Lee et al. (1992)^[7] を参考にして Var_SAS の式が用いられている．どちらの分散式を用いるべきかのコンセンサスは得られていない状況である．なお，本報告書では，Var_Klein を記載している文献が多く，文献との比較ができるという理由から，RMST の推定量に対する分散式として Var_Klein を用いた．被験者数が少ない場合に利用する際は，第一種の過誤確率に注意する必要がある．

なお，SAS/STAT 15.1 より，LIFETEST プロシジャの RMST オプションを用いて，RMST の算出が正式に行えるようになった．このため，前述の TIMELIM オプションに境界時間 τ を指定して RMST を算出する際の留意点を考慮する必要はなく，データセット中にある境界時間 τ より後に発現したイベントをすべて打ち切りに変換しておく必要がなくなった．また，LIFETEST プロシジャの RMST オプションにある BC オプションを指定することで， $m/(m-1)$ の項を付した RMST の分散が得られる．BC オプションを指定しない場合には， $m/(m-1)$ の項が含まれない RMST の分散が算出される．詳細は，以下の SAS/STAT 15.1 User's Guide を参照いただきたい．

http://documentation.sas.com/?docsetId=statug&docsetTarget=statug_lifetest_toc.htm&docsetVersion=15.1&locale=ja

3.3. 治療群間の比較

共変量を調整しない単純な RMST の群間比較を考える．群 g ($g = 0$ は対照薬群， $g = 1$ は実薬群) の RMST を $\mu_g(\tau)$ とし，RMST の推定量，及びその分散をそれぞれ $\hat{\mu}_g(\tau)$ ， $\text{Var}[\hat{\mu}_g(\tau)]$ とおく．

3.3.1. RMST の差

2 群の RMST の差を

$$\mu_1(\tau) - \mu_0(\tau) \quad (3.16)$$

と定義する．この推定量は，

$$\hat{\mu}_1(\tau) - \hat{\mu}_0(\tau) \quad (3.17)$$

である．また，推定量の分散は，

$$\text{Var}[\hat{\mu}_1(\tau) - \hat{\mu}_0(\tau)] = \text{Var}[\hat{\mu}_1(\tau)] + \text{Var}[\hat{\mu}_0(\tau)] \quad (3.18)$$

で求められる．よって，RMST の差の $100(1 - \alpha)\%$ 信頼区間は，

$$\left(\hat{\mu}_1(\tau) - \hat{\mu}_0(\tau) \pm z_{\alpha/2} \sqrt{\text{Var}[\hat{\mu}_1(\tau)] + \text{Var}[\hat{\mu}_0(\tau)]} \right) \quad (3.19)$$

となる．ここで、 z_α は標準正規分布の上側 $100\alpha\%$ 点である．

次に、以下の帰無仮説 H_0 、及び対立仮説 H_1 に対する仮説検定を考える．

$$H_0: \mu_1(\tau) - \mu_0(\tau) = 0, \quad H_1: \mu_1(\tau) - \mu_0(\tau) \neq 0 \quad (3.20)$$

これに対する検定統計量 s_D は、

$$s_D = \frac{\hat{\mu}_1(\tau) - \hat{\mu}_0(\tau)}{\sqrt{\text{Var}[\hat{\mu}_1(\tau)] + \text{Var}[\hat{\mu}_0(\tau)]}} \quad (3.21)$$

となり、これは、帰無仮説の下で漸近的に標準正規分布に従う．よって、観測された検定統計量 s_D^{obs} から、P 値を

$$2\{1 - \Phi(|s_D^{obs}|)\} \quad (3.22)$$

と算出できる．ここで、 $\Phi(\cdot)$ は標準正規分布の分布関数である．

3.3.2. RMST の比

2 群の RMST の比を

$$\frac{\mu_1(\tau)}{\mu_0(\tau)} \quad (3.23)$$

と定義する．この推定量は、

$$\frac{\hat{\mu}_1(\tau)}{\hat{\mu}_0(\tau)} \quad (3.24)$$

である．RMST の比の信頼区間は、対数変換した RMST の比の信頼区間を算出し、これを元の尺度に戻すという手順で求める．対数変換した RMST の比の推定量の分散は、各群の RMST 推定量の分散とデルタ法を用いて、

$$\begin{aligned} \text{Var} \left[\log \left\{ \frac{\hat{\mu}_1(\tau)}{\hat{\mu}_0(\tau)} \right\} \right] &= \text{Var}[\log\{\hat{\mu}_1(\tau)\} - \log\{\hat{\mu}_0(\tau)\}] \\ &= \frac{1}{\{\hat{\mu}_1(\tau)\}^2} \text{Var}[\hat{\mu}_1(\tau)] + \frac{1}{\{\hat{\mu}_0(\tau)\}^2} \text{Var}[\hat{\mu}_0(\tau)] \end{aligned} \quad (3.25)$$

と求められる．よって、RMST の比の $100(1 - \alpha)\%$ 信頼区間は、

$$\exp \left(\log\{\hat{\mu}_1(\tau)\} - \log\{\hat{\mu}_0(\tau)\} \pm z_{\alpha/2} \sqrt{\frac{1}{\{\hat{\mu}_1(\tau)\}^2} \text{Var}[\hat{\mu}_1(\tau)] + \frac{1}{\{\hat{\mu}_0(\tau)\}^2} \text{Var}[\hat{\mu}_0(\tau)]} \right) \quad (3.26)$$

となる．

次に、以下の帰無仮説 H_0 、及び対立仮説 H_1 に対する仮説検定を考える．

$$H_0: \frac{\mu_1(\tau)}{\mu_0(\tau)} = 1, \quad H_1: \frac{\mu_1(\tau)}{\mu_0(\tau)} \neq 1 \quad (3.27)$$

これらの仮説は、

$$H_0: \log\{\mu_1(\tau)\} - \log\{\mu_0(\tau)\} = 0, \quad H_1: \log\{\mu_1(\tau)\} - \log\{\mu_0(\tau)\} \neq 0 \quad (3.28)$$

と同じであることから、検定統計量 s_R は、

$$s_R = \frac{\log\{\hat{\mu}_1(\tau)\} - \log\{\hat{\mu}_0(\tau)\}}{\sqrt{\frac{1}{\{\hat{\mu}_1(\tau)\}^2} \text{Var}[\hat{\mu}_1(\tau)] + \frac{1}{\{\hat{\mu}_0(\tau)\}^2} \text{Var}[\hat{\mu}_0(\tau)]}} \quad (3.29)$$

となり、これは、帰無仮説の下で漸近的に標準正規分布に従う。よって、観測された検定統計量 s_R^{obs} から、P 値を

$$2\{1 - \Phi(|s_R^{obs}|)\} \quad (3.30)$$

と算出できる。

3.4. 調整解析

3.4.1. 層別解析

影響因子がカテゴリー型であり、共変量を用いて K 個の層が構築されるとする。ある層 k ($k = 1, \dots, K$)における群 g ($g = 0, 1$)の境界時間 τ の RMST とその推定量をそれぞれ $\mu_{kg}(\tau), \hat{\mu}_{kg}(\tau)$ とする。以下の帰無仮説 H_0 に対応する仮説検定を考える。

$$H_0: \mu_{11} - \mu_{10} = \mu_{21} - \mu_{20} = \dots = \mu_{K1} - \mu_{K0} = 0 \quad (3.31)$$

これに対する検定統計量は s_{SD} は、

$$s_{SD} = \frac{\sum_{k=1}^K \{\hat{\mu}_{k1}(\tau) - \hat{\mu}_{k0}(\tau)\}}{\sqrt{\sum_{k=1}^K \{\text{Var}[\hat{\mu}_{k1}(\tau)] + \text{Var}[\hat{\mu}_{k0}(\tau)]\}}} \quad (3.32)$$

であり、これは帰無仮説の下で漸近的に標準正規分布に従う。よって、観測された検定統計量 s_{SD}^{obs} から、P 値を

$$2\{1 - \Phi(|s_{SD}^{obs}|)\} \quad (3.33)$$

と計算することが出来る。ここで、 $\Phi(\cdot)$ は標準正規分布の分布関数である。

なお、SAS/STAT 15.1 より、LIFETEST プロシジャの STRATA ステートメントと GROUP オプションを用いて、RMST の層別解析を行えるようになった。詳細は、以下の SAS/STAT 15.1 User's Guide を参照いただきたい。

http://documentation.sas.com/?docsetId=statug&docsetTarget=statug_lifetest_toc.htm&docsetVersion=15.1&locale=ja

3.4.2. 回帰分析

Tian et al. (2014)^[8] によって、RMST の回帰分析の方法が提案されている。提案されている回帰係数 β は $E[\min(T_i, \tau) | X_i] = g(\beta'X_i)$ の想定の下で最小二乗法の考えに基づき、以下の式を解くことで推定できる。

$$W(\beta) = \sum_{i=1}^n \frac{\delta_i}{\hat{G}(T_i)} X_i \{\min(T_i, \tau) - g(\beta'X_i)\} = 0 \quad (3.34)$$

ここで、 T_i は被験者 i ($i = 1, \dots, n$)のイベント発現又は打ち切りまでの時間、 δ_i はイベント発現時に 1、それ以外で 0 を与えるイベント発現の有無、 Z_i は共変量ベクトル、 $X_i' = (1, Z_i')$ とする。打ち切りまでの時間に対する Kaplan–Meier 法による生存関数を $\hat{G}(\cdot)$ 、 $[-\infty, \infty]$ の範囲

にある値を $[0, \tau]$ の範囲へ変換するリンク関数を $g(\cdot)$ とする。回帰係数の推定方程式は、重み付きの正規方程式と捉えることができ、リンク関数を与えた重み付き最小二乗法に帰着する。

なお, SAS/STAT 15.1 より, RMSTREG プロシジャの MODEL ステートメントの METHOD オプションで IPCW と指定することで, 上述の RMST の回帰分析を行えるようになった。詳細は, 以下の SAS/STAT 15.1 User's Guide を参照いただきたい。

http://documentation.sas.com/?docsetId=statug&docsetTarget=statug_rmstreg_toc.htm&docsetVersion=15.1&locale=ja

3.4.3. Cox 比例ハザードモデルによる調整解析

Eng et al. (2015)^[9]によって, Cox 比例ハザードモデルに基づいて共変量ベクトル $\mathbf{Z}$ をもつ 1 人の被験者に対する RMST を推定する方法が提案されている。

T_i は被験者 i ($i = 1, \dots, n$)のイベント発現又は打ち切りまでの時間, δ_i はイベント発現時に 1, それ以外で 0 を与えるイベント発現の有無, $\mathbf{Z}_i = (Z_{i1}, \dots, Z_{ip})'$ は共変量ベクトルとする。 $h(t|\mathbf{Z}_i)$ が共変量ベクトル $\mathbf{Z}_i$ をもつ被験者に対する時間 t でのハザードとすると, Cox 比例ハザードモデルにより,

$$h(t|\mathbf{Z}_i) = h_0(t)\exp(\boldsymbol{\beta}'\mathbf{Z}_i) = h_0(t)\exp\left(\sum_{k=1}^p \beta_k Z_{ik}\right) \quad (3.35)$$

と表せる。これより共変量ベクトル $\mathbf{Z}_i$ をもつ被験者 i の条件付き生存関数は, ベースライン生存関数 $S_0(t)$ を使って,

$$S_i(t|\mathbf{Z}_i) = [S_0(t)]^{\exp(\boldsymbol{\beta}'\mathbf{Z}_i)} \quad (3.36)$$

と表せる。

Cox 比例ハザードモデルの回帰係数の推定値を $\hat{\boldsymbol{\beta}}$ とする。相異なる D 個のイベント発現時点を $t_1 < t_2 < \dots < t_D$, $t_0 = 0$ とし, d_i ($i = 1, \dots, D$)を時間 t_i でのイベント発生数とする。ベースライン生存関数 $S_0(t)$ の推定量は, Breslow 推定量により累積ハザード関数を求めることにより,

$$\hat{S}_0(t) = \exp\left[-\sum_{t_j \leq t} \frac{d_j}{\sum_{i \in R(t_j)} \exp(\sum_{k=1}^p \hat{\beta}_k Z_{ik})}\right] \quad (3.37)$$

と表せる。ここで, $R(t_i)$ は時点 t_i まで追跡されているすべての被験者から成るリスク集合である。これより, 共変量ベクトル $\mathbf{Z}_i$ をもつ被験者 i に対する生存関数は

$$\hat{S}_i(t|\mathbf{Z}_i) = [\hat{S}_0(t)]^{\exp(\hat{\boldsymbol{\beta}}'\mathbf{Z}_i)} \quad (3.38)$$

で推定できる。この生存関数を使うことによって, 共変量ベクトル $\mathbf{Z}_i$ をもつ被験者 i に対する RMST を以下の通り, 推定することができる。

$$\hat{\mu}(\tau|\mathbf{Z}_i) = \int_0^{\tau} \hat{S}_i(t|\mathbf{Z}_i) dt \quad (3.39)$$

ここで、 τ は境界時間である。

推定した生存関数の例を図 3-5 に示した。

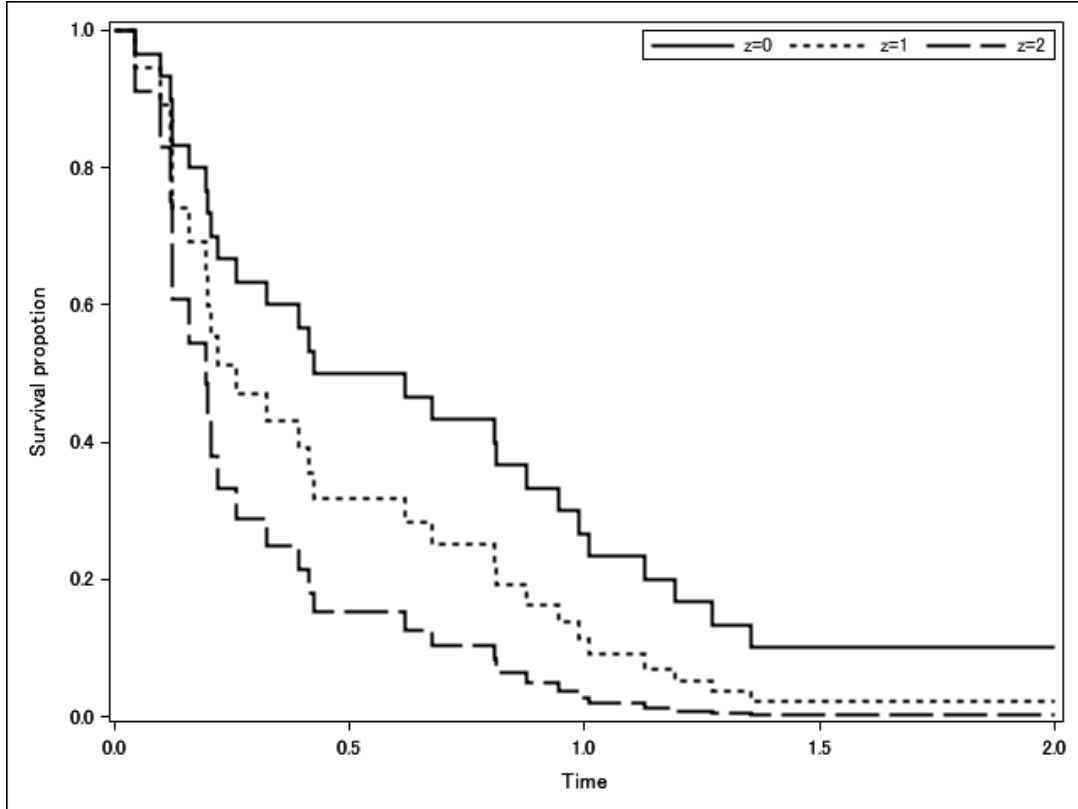


図 3-5 共変量 $\mathbf{Z} = 0, 1, 2$ をもつ被験者に対する生存曲線の例

臨床試験では、群ごとの RMST を推定することに興味がある。式(3.38)を用いて、生存関数を被験者ごとに求め、各時点での生存割合を以下の通り平均すれば、群ごとの生存曲線を推定することができる^[4]。

$$\hat{S}(t) = \frac{1}{n} \sum_{i=1}^n \hat{S}_i(t|\mathbf{Z}_i) \quad (3.40)$$

この生存曲線を用いて RMST を推定し、RMST の標準誤差は bootstrap 法を用いて推定することができる。

3.5. 症例数設計

3.5.1. 計算手順

2 群間の RMST の差を検出することを目的とした症例数設計の方法を解説する^[1]。この方法は、境界時間 τ 内の打ち切りが発生しないと仮定した必要症例数を計算し、次に打ち切り

の発生を仮定した必要症例数を計算する．実際の臨床試験では，境界時間 τ 内において一定期間の観察による打ち切りやランダムに発生する打ち切りが存在するため，打ち切りが発生することによる必要症例数の増加を考慮する必要がある．そのため，シミュレーションを用いて打ち切りを想定した分散を計算し，対応する必要症例数を計算することになる．

以下に説明する計算手順を実行する SAS マクロを 7.2.4 項に示した．同じく RMST の群間差を検出することを目的とした必要症例数を計算する方法として，R の SSRMST パッケージを利用することもできる^[10]．

手順1 症例数設計にて想定する，各群の生存時間分布，境界時間 τ ，被験者の登録期間及び追跡期間，単位時間あたりの打ち切りの発生割合，被験者の登録時期の分布，症例数の群間比，有意水準，検出力を決める．

手順2 実薬群を $g = 1$ ，対照薬群を $g = 0$ とする．仮定した生存時間分布に基づき，各群の RMST である $\mu_g = \mu_g(\tau)$ を式(3.2)から，境界時間 τ 内での生存時間 $X(\tau)$ の分散である $\sigma_g^2 = \sigma_g^2(\tau)$ を式(3.4) から計算する．境界時間 τ 内での打ち切りが発生しないと仮定して，両群合計の必要症例数 N を以下の式(3.41)より計算する． α を両側有意水準， $1 - \beta$ を検出力， z_α を標準正規分布の上側 $100\alpha\%$ 点， f を症例数の割付比（実薬群／対照薬群）とする．

$$N = (1 + f) \frac{(z_{\alpha/2} + z_\beta)^2}{(\mu_1 - \mu_0)^2 / (\sigma_1^2 / f + \sigma_0^2)} \quad (3.41)$$

手順3 仮定した登録期間，追跡期間及び打ち切りに基づくモンテカルロ・シミュレーションにより，境界時間 τ 内の打ち切りを考慮した必要症例数を計算する．シミュレーションの繰り返し回数を M ，シミュレーションでの各群の症例数を m とする（提案の論文^[1]では $M = 50$ ， $m = 10,000$ を推奨している）．シミュレーションの各繰り返し $s (= 1, \dots, M)$ について，RMST の標準誤差 $\hat{\sigma}(\hat{\mu}_{g,s})$ を計算する．打ち切りを考慮することによる RMST の分散の調整係数 $\phi_{g,s} = \sqrt{m} \hat{\sigma}(\hat{\mu}_{g,s}) / \sigma_g$ を求め，分散 $\sigma_{g,s}^2 = \phi_{g,s}^2 \sigma_g^2$ を計算する．手順2にて計算した $X(\tau)$ の分散 σ_g^2 を $\sigma_{g,s}^2$ とし，打ち切りを考慮した両群の必要症例数 N_s を計算する．

$$N_s = (1 + f) \frac{(z_{\alpha/2} + z_\beta)^2}{(\mu_1 - \mu_0)^2 / (\sigma_{1,s}^2 / f + \sigma_{0,s}^2)} \quad (3.42)$$

手順4 打ち切りを考慮した試験全体の必要症例数 N_s の平均及び標準誤差を計算する．標準誤差は次式で求める．標準誤差は，シミュレーション結果が安定して得られているかを確認するために求められる．シミュレーションの標準誤差が実用上十分に小さいシミュレーション回数 M を選ぶ．

$$SE(N_s) = \sqrt{(N_s \text{ の標本分散}) / M} \quad (3.43)$$

3.5.2. SAS マクロの実行

前項の手順による必要症例数を計算する JPMA タスクフォース 4 のメンバーが作成した SAS マクロを紹介する．このマクロは SAS 9.4, SAS/STAT 14.2 の下で作成された．表 3-6 に SAS マクロの入力引数を示した．

表 3-6 必要症例数を計算する SAS マクロの入力引数

引数	説明
P1	引数 time に対応する実薬群の生存割合のベクトル．
P0	引数 time に対応する対照薬群の生存割合のベクトル．
time	区分指数分布の区分時間．
tau	境界時間 τ ．
enroll	被験者の登録期間．
follow	被験者の追跡期間（最後の被験者の登録時点から観察終了までの期間）．
loss	単位時間あたりのランダム打ち切りの発生割合（単位は％）．デフォルトは実薬群及び対照薬群共に 0%．
r	登録時期の分布のパラメータ．デフォルトは 1．
ratio	症例数の割付比（実薬群：対照薬群）．デフォルトは 1：1．
sig_level	両側有意水準 α ．デフォルトは 0.05．
power	検出力 $1 - \beta$ ．デフォルトは 0.8．
seed	シミュレーションでの乱数のシード．デフォルトは 123．
ss_sim	シミュレーションでの各群の症例数 m ．デフォルトは 10,000．
iteration	シミュレーションの繰り返し数 M ．デフォルトは 50

各群の生存時間が区分指数分布に従うことを仮定した．時間に関連する引数 time, tau, enroll, follow 及び loss について、時間の単位（例：年，月，週）は全て同一である．

手順 3 のモンテカルロ・シミュレーションでは、イベント発生までの時間 S ，打ち切り発生までの時間 C ，登録時期 E を乱数として発生させる．登録期間を定数 θ ，追跡期間を定数 ζ とすると，生存時間は $\min(S, C, \theta + \zeta - E, \tau)$ によって定まる． $\min(S, C, \theta + \zeta - E, \tau) = S$ のときイベント， $\min(S, C, \theta + \zeta - E, \tau) < S$ のとき打ち切りとする．つまり，次の 3 つのうち最小の期間がイベント時間よりも短い場合，その最小の期間にて打ち切りとする．

1. ランダム打ち切り発生までの期間 C
2. 追跡期間終了までの期間 $(\theta + \zeta - E)$
3. 境界時間 τ

イベント発生までの時間は引数 P1, P0 及び time に対応する区分指数分布に基づく．ランダム打ち切り発生までの期間は指数分布に従うと仮定し，この生存関数を $S(t; \lambda_{\text{loss}}) =$

$\exp(-\lambda_{\text{loss}}t)$ とする．この指数パラメータ λ_{loss} は，SAS マクロ中の引数 `loss` を用いて $-\log(1 - \text{loss})$ と計算される．登録時期は次式の分布関数に基づく^[11]．なお，分布のパラメータ r は SAS マクロ中の引数 `r` に対応する．

$$F(t; \theta, r) = \left(\frac{t}{\theta}\right)^r, (r, \theta > 0) \quad (3.44)$$

図 3-6 に被験者の登録時期の分布の例を示した． $r = 1$ の分布は一様分布に一致する． $r > 1$ の分布は，試験時期の進捗に従って登録被験者数が増加する場合に対応する．

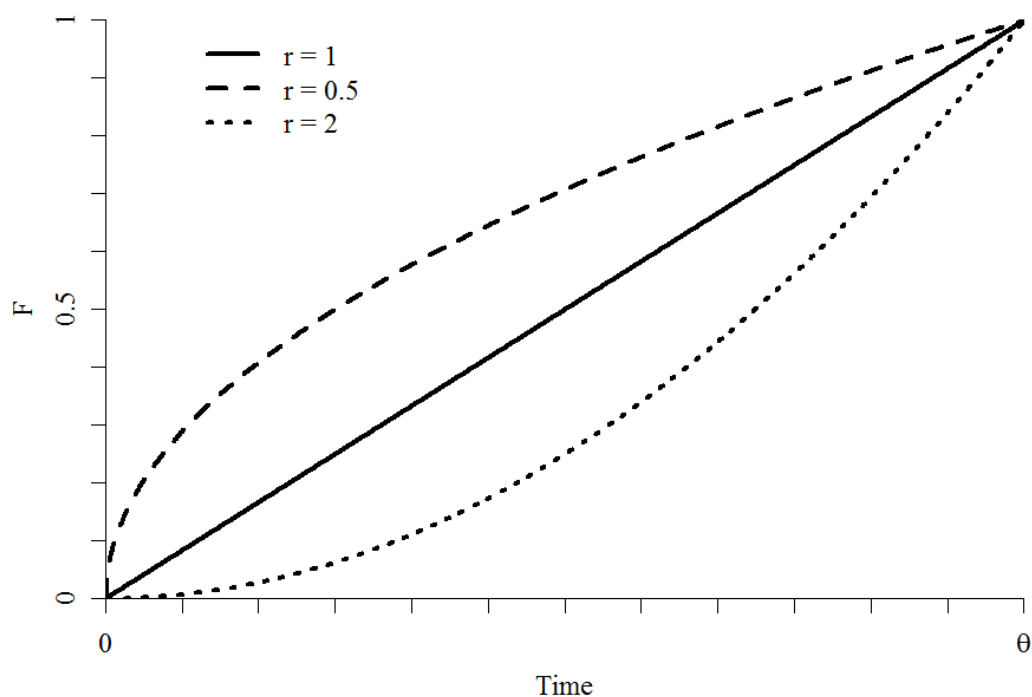


図 3-6 被験者の登録時期の分布関数

入力引数に加え，表 3-7 の結果をデータセット **RESULT** に出力する．最終解析に必要な情報量については，4.4.1 項の式(4.6)を用いて算出した値が計算される．なお，式(4.7)を用いて試験途中で計算される情報量が式(4.6)に示した情報量に到達した時点にて，事前計画の想定に基づく検出力が保証される．

表 3-7 必要症例数を計算する SAS マクロの出力

変数名	説明
MEAN1	実薬群の RMST.
MEAN0	対照薬群の RMST.
VAR1	打ち切りを考慮しない実薬群の $X(\tau)$ の分散.
VAR0	打ち切りを考慮しない対照薬群の $X(\tau)$ の分散.
N	打ち切りを考慮しない必要症例数.
MEAN_N	打ち切りを考慮した必要症例数のシミュレーション平均.
SE_N	打ち切りを考慮した必要症例数のシミュレーション標準誤差.
INFO	最終解析に必要な情報量

3.5.3. 生存時間分布に指数分布を仮定した必要症例数

1 つ目の例として、生存時間分布に指数分布を仮定した必要症例数を示す．症例数設計の想定には、非小細胞肺癌患者を対象とした CheckMate 057 試験の結果に基づく症例数設計の事例を参考とした^[12]．24 ヶ月を境界時間とした RMST の治療群間差として 3 ヶ月の延長を期待し、実薬群の RMST を 14.1 ヶ月、対照薬群の RMST を 11.1 ヶ月と想定する．式(3.5) を用いて RMST の想定値から各群の指数パラメータ λ_g を Newton-Raphson 法で求めると、 $\lambda_1 = 0.049088$, $\lambda_0 = 0.075308$ である．Newton-Raphson 法による計算は、SAS/IML の NLPNRA ルーチンまたは R の uniroot 関数を用いて容易に解くことができる．これらの Newton-Raphson 法のコードを 7.3.2 項に示した．24 ヶ月時点の生存割合は、実薬群で $\exp(-0.04908797 * 24) = 0.308$ 、対照薬群で $\exp(-0.07530796 * 24) = 0.164$ である．想定した生存時間分布の生存曲線を図 3-7 に示した．

シナリオ 1 では登録期間を 11 ヶ月、追跡期間を 15 ヶ月とした^[12]．シナリオ 2 は登録期間を 18 ヶ月、追跡期間を 8 ヶ月とした．シナリオ 3 は 1 ヶ月あたりに 1%のランダム打ち切りが 2 群の両方に発生することとした．シナリオ 4 は登録時期のパラメータ $r = 2$ とした．その他、シナリオ間で共通の想定は、境界時間 24 ヶ月、検出力 90%、他は SAS マクロのデフォルト通り ($f = 1$, $\alpha = 0.05$, $m = 10,000$, $M = 50$) とした．各シナリオの設定パラメータを表 3-8 に示した．

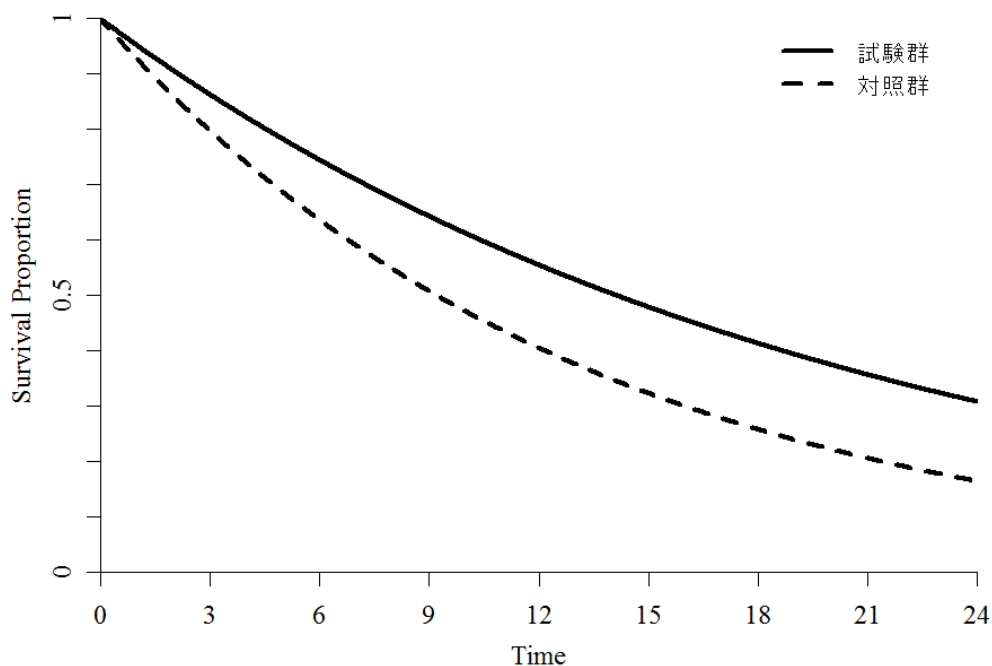


図 3-7 症例数設計で仮定した指数分布に従う生存曲線

表 3-8 各シナリオの設定パラメータ

	シナリオ 1	シナリオ 2	シナリオ 3	シナリオ 4
実薬群の指数ハザード λ_1 (／月)	0.04908797			
対照薬群の指数ハザード λ_0 (／月)	0.07530796			
境界時間 τ (月)	24			
登録期間 (月)	11	18	11	11
追跡期間 (月)	15	8	15	15
1 ヶ月あたりの打ち切り発生割合 (各群共通)	0	0	0.01	0
登録時期の分布パラメータ r	1	1	1	2
症例数の割付比	1:1			
両側有意水準	0.05			
検出力	0.9			

実行する SAS マクロを以下に示した。必要症例数の計算結果を表 3-9 に示した。

- シナリオ 1 の SAS マクロ

```
%SS_RMST(P1 = 0.3078597, P0 = 0.1640817, time = 24, tau = 24,
power = 0.9, enroll = 11, follow = 15);
```

- シナリオ 2 の SAS マクロ

```
%SS_RMST(P1 = 0.3078597, P0 = 0.1640817, time = 24, tau = 24,
           power = 0.9, enroll = 18, follow = 8);
```

- シナリオ 3 の SAS マクロ

```
%SS_RMST(P1 = 0.3078597, P0 = 0.1640817, time = 24, tau = 24,
           power = 0.9, enroll = 11, follow = 15, loss = %str(0.01, 0.01));
```

- シナリオ 4 の SAS マクロ

```
%SS_RMST(P1 = 0.3078597, P0 = 0.1640817, time = 24, tau = 24,
           power = 0.9, enroll = 11, follow = 15, r = 2);
```

表 3-9 必要症例数の計算結果

	シナリオ 1	シナリオ 2	シナリオ 3	シナリオ 4
実薬群の $RMST\mu_1$	14.1			
対照薬群の $RMST\mu_0$	11.1			
打ち切りを考慮しない 実薬群の $X(\tau)$ の分散 σ_1^2	74.6325			
打ち切りを考慮しない 対照薬群の $X(\tau)$ の分散 σ_0^2	66.9967			
打ち切りを考慮しない 必要症例数 N	332			
打ち切りを考慮した必要症例数 のシミュレーション平均	336	366	358	346
打ち切りを考慮した必要症例数 のシミュレーション標準誤差	0.21244	0.27073	0.31981	0.22079
最終解析に必要な情報量	1.16749			

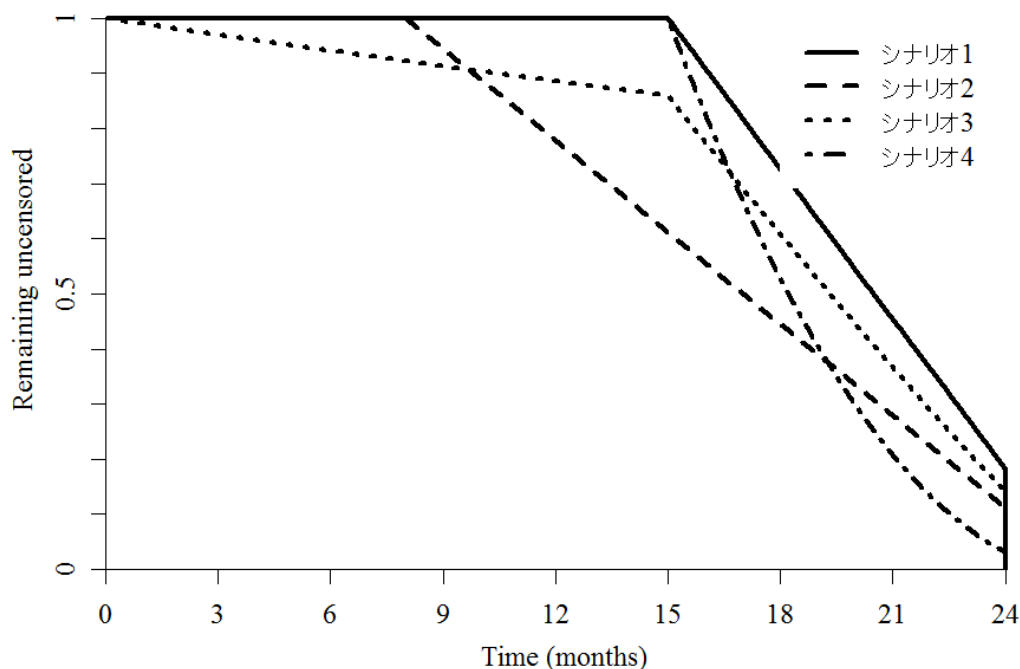


図 3-8 打ち切りの生存時間分布

シナリオ 1 の必要症例数 336 例は既報告と一致した^[12]。シナリオ 2～4 はシナリオ 1 と比較して打ち切りが増加し、分散が大きくなる想定であるため、打ち切りを考慮した必要症例数は増加した。シナリオ 2 はシナリオ 1 と比較して登録期間は長いが追跡期間が短いため、観察期間が短い被験者が増加することから打ち切りは増加する。シナリオ 3 及び 4 はシナリオ 1 と比較して登録期間及び追跡期間は同じだが、シナリオ 3 はランダムに発生する打ち切りを想定しているため打ち切りは増加し、シナリオ 4 は登録スピードが徐々に速くなる分布を想定しており観察期間が短い被験者が増加することから打ち切りは増加する。2 つの群を併合した打ち切りの生存時間分布を図 3-8 に示した。シミュレーション標準誤差はいずれのシナリオについても 0.2～0.3 程度であり必要症例数に対して十分に小さいため、シミュレーションによる必要症例数は十分な精度で推定できていると考えられる。

実際に臨床試験で必要症例数を決定する際には、打ち切り発生に関わる情報（登録期間、登録スピード、追跡期間、ランダム打ち切りの発生）を複数のシナリオで検討し、安定して結論を得られる必要症例数を設定することが重要であると考えられる。

3.5.4. 生存時間分布に区分指数分布を仮定した必要症例数

2 つ目の例として、生存時間に区分指数分布を仮定した場合の必要症例数を示す。がん免疫療法のように投与開始から治療効果の発現までに一定の期間を必要とする場合、時間によって治療効果の差が異なる。これを表現するため、区分指数分布を仮定することが考えられる^[13]。

時点 0 から 3 ヶ月時点までは実薬群と対照薬群で同一の生存時間分布を仮定する．指数分布を仮定した前項の例と同様，対照薬群の生存時間分布に指数分布を仮定し，境界時間を 24 ヶ月とした RMST を 11.1 ヶ月と想定する．3 ヶ月時点の対照薬群の生存割合は $\exp(-0.07530796 * 3) = 0.798$ である．実薬群は対照薬群に対して RMST の差として 3 ヶ月の改善を期待し，実薬群の RMST を 14.1 ヶ月と想定する．式(3.10) から，実薬群の 3 ヶ月から 24 ヶ月までの指数パラメータは Newton-Raphson 法を用いて 0.0392188 と計算される．この Newton-Raphson 法のコードを 7.3.2 項に示した．24 ヶ月時点の対照薬群の生存割合は $\exp(-0.075308 * 24) = 0.164$ ，実薬群は $\exp(-0.075308 * 3) * \exp(-0.0392188 * 21) = 0.350$ であり，24 ヶ月生存割合の差として 18.6%を想定することに対応する．想定した生存時間分布の生存曲線を図 3-9 に示した．

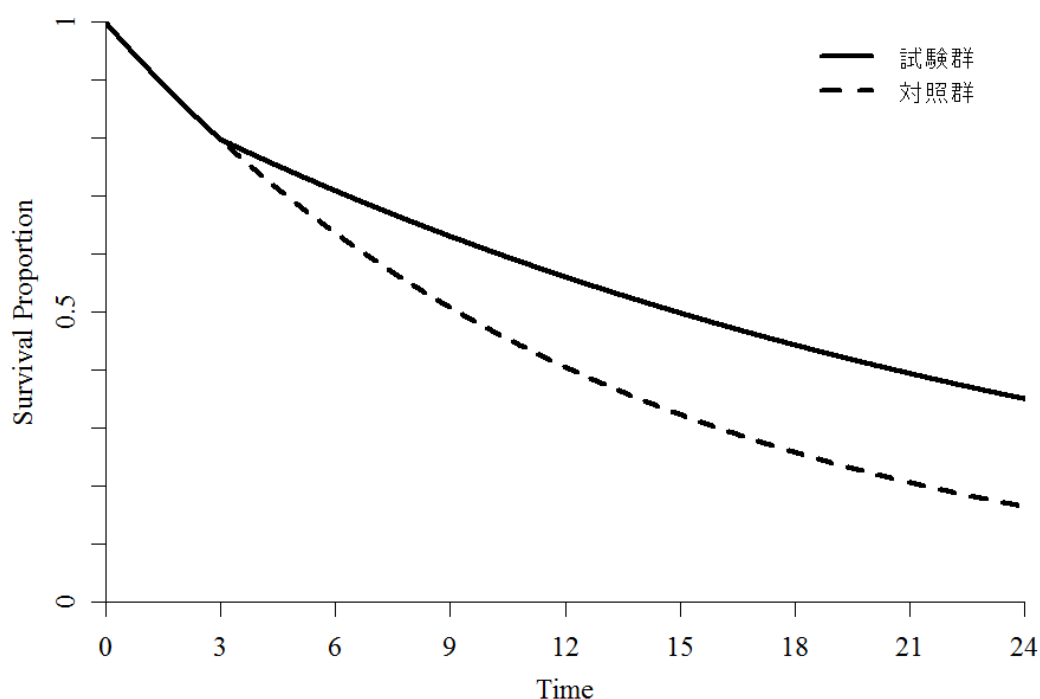


図 3-9 症例数設計で仮定した区分指数分布に従う生存曲線

生存時間分布以外のパラメータは，前項の指数分布と同じシナリオ 1~4 を想定した．各シナリオの設定パラメータを表 3-10 に示した．

表 3-10 各シナリオの設定パラメータ

	シナリオ 1	シナリオ 2	シナリオ 3	シナリオ 4
3 ヶ月までの指数ハザード(／月, 各群共通)	0.07530796			
3 ヶ月から 24 ヶ月までの実薬群の指数ハザード (／月)	0.0392188			
3 ヶ月から 24 ヶ月までの対照薬群の指数ハザード (／月)	0.07530796			
境界時間 τ (月)	24			
登録期間 (月)	11	18	11	11
追跡期間 (月)	15	8	15	15
1 ヶ月あたりの打ち切り発生割合 (各群共通)	0	0	0.01	0
登録時期の分布パラメータ r	1	1	1	2
症例数の群間比 f	1			
両側有意水準	0.05			
検出力	0.9			

実行する SAS マクロは各シナリオで次の通りである．必要症例数の計算結果を表 3-11 に示した．

- シナリオ 1 の SAS マクロ

```
%SS_RMST(P1 = %str(0.7977788, 0.3501062), P0 = %str(0.7977788, 0.1640817),
time = %str(3, 24), tau = 24, power = 0.9, enroll = 11, follow = 15);
```

- シナリオ 2 の SAS マクロ

```
%SS_RMST(P1 = %str(0.7977788, 0.3501062), P0 = %str(0.7977788, 0.1640817),
time = %str(3, 24), tau = 24, power = 0.9, enroll = 18, follow = 8);
```

- シナリオ 3 の SAS マクロ

```
%SS_RMST(P1 = %str(0.7977788, 0.3501062), P0 = %str(0.7977788, 0.1640817),
time = %str(3, 24), tau = 24, power = 0.9, enroll = 11, follow = 15, loss = %str(0.01, 0.01));
```

- シナリオ 4 の SAS マクロ

<code>%SS_RMST(P1 = %str(0.7977788, 0.3501062), P0 = %str(0.7977788, 0.1640817), time = %str(3, 24), tau = 24, power = 0.9, enroll = 11, follow = 15, r = 2);</code>
--

表 3-11 必要症例数の計算結果

	シナリオ 1	シナリオ 2	シナリオ 3	シナリオ 4
実薬群の RMST μ_1	14.1			
対照薬群の RMST μ_0	11.1			
打ち切りを考慮しない 実薬群の $X(\tau)$ の分散 σ_1^2	84.6029			
打ち切りを考慮しない 対照薬群の $X(\tau)$ の分散 σ_0^2	66.9967			
打ち切りを考慮しない 必要症例数 N	354			
打ち切りを考慮した必要症例数 のシミュレーション平均	360	388	380	368
打ち切りを考慮した必要症例数 のシミュレーション標準誤差	0.20846	0.28428	0.35165	0.2336
最終解析に必要な情報量	1.16749			

前項と同様，シナリオ 2～4 の打ち切りを考慮した必要症例数はシナリオ 1 と比較して増加した．生存時間分布に指数分布を仮定した前項の症例数設計と比較すると，RMST の群間差として同じ 3 ヶ月の差を想定しているが，区分指数分布を仮定したシナリオは，指数分布を仮定したシナリオと比較して分散が大きいため，区分指数分布を仮定した場合の方が指数分布を仮定した場合と比較して必要症例数は多かった．最終解析に必要な情報量は RMST の群間差，有意水準及び検出力で決定されるため，指数分布を仮定した前項と変わらない．

3.5.5. 境界時間と必要症例数の関係

境界時間 τ を変化させた場合に，試験全体の必要症例数がどのように変化するかを調べた．3.5.3 項に示した指数分布に基づく症例数設計のシナリオ（表 3-10 各シナリオの設定パラメータ）の下で，境界時間 τ を 12 ヶ月から 36 ヶ月まで変化させた場合の試験全体の必要症例数及び RMST の群間差を，それぞれ図 3-10 境界時間 τ に対する必要症例数及び図 3-11 に示した．境界時間 τ 以外の設定は表 3-10 各シナリオの設定パラメータと同じとした．

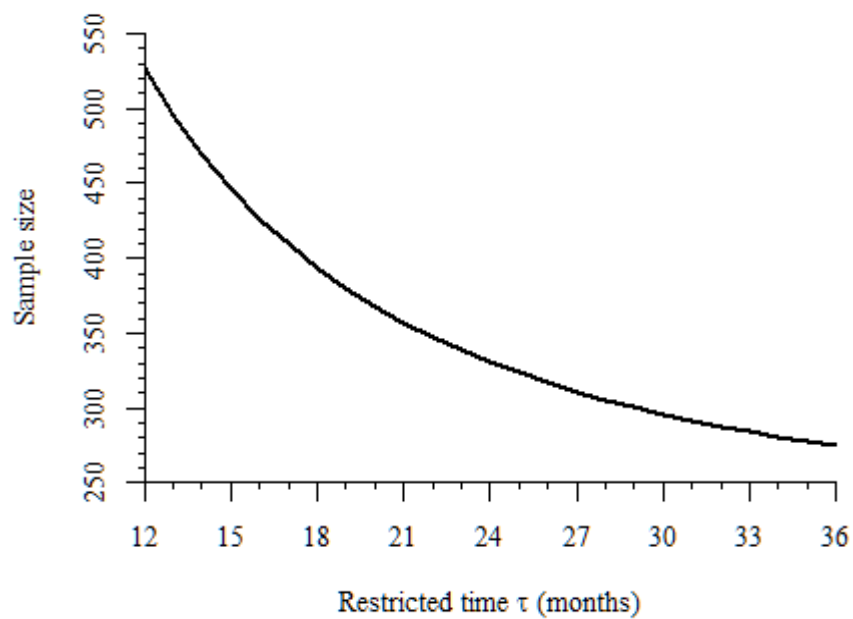


図 3-10 境界時間 τ に対する必要症例数

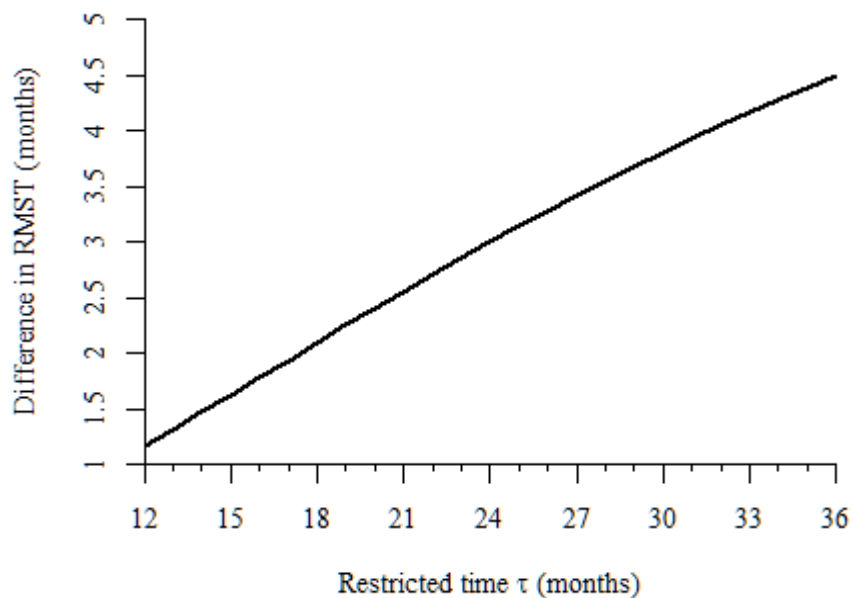


図 3-11 境界時間 τ に対する RMST の群間差

境界時間 τ が延びるに従い RMST の群間差は大きくなるものの、その標準誤差も大きくなる。結果として、今回は検定統計量である式(3.21)の値が大きくなる状況であったため、境界時間 τ が延びるに従い必要症例数は単調に減少した。本項では指数分布を仮定した場合を示したが、生存時間分布の想定によっては、境界時間 τ が延びるに従い必要症例数が増加する可能性もあることに留意する必要がある。

参考文献

- [1] Royston P, Parmar MKB. Restricted mean survival time: an alternative to the hazard ratio for the design and analysis of randomized trials with a time-to-event outcome. *BMC Med Res Methodol.* 2013; **13**: 152.
- [2] Klein JP, Moeschberger ML. *Survival Analysis: Techniques for Censored and Truncated Data* second edition. Springer-Verlag: New York; 2003.
- [3] SAS/STATA(R) 14.1 User's Guide
http://support.sas.com/documentation/cdl/en/statug/68162/HTML/default/viewer.htm#statug_lifetest_syntax01.htm
- [4] Collett D. *Modelling survival data in medical research*, third edition. CRC Press; 2015.
- [5] Diez DM. *Survival Analysis in R*; 2013.
https://www.openintro.org/download.php?file=survival_analysis_in_R
- [6] Kaplan EL, Meier P. Nonparametric Estimation From Incomplete Observations. *Journal of the American Statistical Association.* 1958; **53**(282): 457-481.
- [7] Lee ET, Wang JW. *Statistical Methods for Survival Data Analysis*. second edition. John Wiley & Sons: New York; 1992.
- [8] Tian L, Zhao L, Wei LJ. Predicting the restricted mean event time with the subject's baseline covariates in survival analysis. *Biostatistics.* 2014; **15**: 222-233.
- [9] Eng KH, Schiller E, Morrell K. On representing the prognostic value of continuous gene expression biomarkers with the restricted mean survival curve. *Oncotarget.* 2015; **6**(34): 36308-36318.
- [10] Horiguchi M, Uno H. CRAN – Package SSRMST.
<https://cran.r-project.org/web/packages/SSRMST/index.html>
- [11] Hasegawa T. Group sequential monitoring based on the weighted log-rank test statistic with the Fleming–Harrington class of weights in cancer vaccine studies. *Pharm Stat.* 2016; **15**:412–419.
- [12] Pak K, Uno H, Kim DH, Tian L, Kane RC, Takeuchi M, *et al.* Interpretability of cancer clinical trial result using restricted mean survival time as an alternative to the hazard ratio. *JAMA Oncol.* 2017; **3**(12): 1692–1696.
- [13] Fine GD. Consequences of delayed treatment effects on analysis of time-to-event endpoints. *Drug Inf J.* 2007; **41**: 535–539.

4. 適用上の留意点

4.1. RMST の意味, 境界時間 τ の意義付け

4.1.1. RMST の考え方

臨床試験では、試験期間中に全ての被験者のイベントを得ることが困難なことが多い。平均生存時間を評価指標とした場合、生存関数が生存割合 0%に達せず、算出できないことがある。この場合、最終イベント発現時点、または最終追跡時点までの範囲の Kaplan–Meier 曲線の曲線下面積を平均生存時間と定義する場合がある。しかし、どの時点までを解析に含めるかによって平均生存時間の推定値が変わってしまう。同一試験内で 2 群間の平均生存時間を比較することは可能だが、追跡期間が異なる他試験との正確な比較は難しい。

RMST は、境界時間 τ 内という条件を付けることにより、被験者毎の追跡期間が一定となり、他の試験とも公平に比較することができる。また、境界時間 τ を臨床的観点から設定することで、臨床的に解釈できる評価指標として RMST を意義付けられると考えられる。

4.1.2. RMST は何を評価しているのか

境界時間 τ の値を検討するためには、RMST を用いて何を評価するかを明確にする必要がある。RMST の使用目的として下記の 2 つが考えられる。

- ① 臨床的に意味のある時点までの平均生存時間を評価
- ② 全期間の平均生存時間の近似、もしくは代替指標として評価

それぞれの境界時間 τ の考え方として、①は境界時間 τ 以降のデータを取得せずとも薬剤の効果・特性を十分に評価・解釈できる時点となる。したがって、境界時間 τ を変化させることは RMST の値の頑健性を評価することを目的とした感度分析とはならないかもしれない。一方、②の境界時間 τ は、全ての被験者のイベントが観察された場合の平均生存時間の近似、もしくは代替指標となる時点であるため、境界時間 τ を変化させることが全期間の平均生存時間としての頑健性を評価する感度分析になる可能性がある。

本報告書では、RMST は主に①臨床的に意味のある時点までの平均生存時間を評価することを想定して記載する。

4.1.3. RMST の差とハザード比

比較的新しい評価指標である RMST の特徴を理解するために、汎用されているハザード比との違いについて整理した。なお、RMST の差と比は統計的性質が同じであり、差を用いた方が RMST の利点の一つである単位がある形で示すことができるため、本項では RMST の差について述べる。

表 4-1 RMST の差とハザード比の評価対象, 使用状況の違い

	RMST の差	ハザード比
何を評価しているのか	境界時間 τ 内の治療効果（期間） ✓ 境界時間 τ の意味：効果があるか	試験期間を通しての治療効果（リスク）

	どうかを臨床的に判断できる期間	
解釈	対照薬群と比較して、イベント発現までの期間を平均的にどれほど延長するか	対照薬群と比較して、リスクをどれほど抑えられるか
使用が適している状況	境界時間 τ が臨床的に決まる	比例ハザード性が成り立つと思われる状況
推定値と観察期間の関係	推定値が境界時間 τ に依存する	比例ハザード性が成り立っている場合、推定値は観察期間に依存しない
その他の特徴	RMST の差と比は統計的な特徴（検出力等）は基本的には同じ RMST の差だけではなく、RMST そのもので解釈可能（例：12 ヶ月間追跡した場合の平均的な生存時間）	ハザードのみでは解釈しにくいことが多い 一般的には、ハザード比という相対リスクにより解釈される

ハザード比を解釈する際には、比例ハザード性が成り立っていることが条件として必要である。比例ハザード性が成り立っている場合には、RMST の差と比べ、ハザード比の検出力は同等か高くなることが報告されている^[1]。一方で RMST の差は比例ハザード性が成り立たない場合にも使用可能で、対照薬群の参照値が明示され、かつ時間という単位があることから、臨床的に解釈しやすい指標と考えられる^[2]。ただし、試験計画時に境界時間 τ が明確に決まる、もしくは平均生存時間の近似と考えられるほど大きい境界時間 τ が設定できる必要がある。

評価指標に RMST を用いるかどうかを判断するには、臨床メンバーと議論する必要がある。適切な判断を行うための材料として、例えば、下記の情報を臨床メンバーにも共有することが考えられる。

表 4-2 臨床メンバーに提供する情報

	RMST の差	ハザード比
評価するもの	境界内平均生存期間がどれほど延長するか	リスクがどれほど減るか
メリット	● 比例ハザード性が成り立たない状況でも使える ● 臨床的に解釈・説明がしやすい ● 境界時間τが一致している場合、他試験との比較が容易	● 比例ハザード性が成り立っていれば検出力が高い ● RMST はτ以降のデータを使用しないが、ハザード比は得られた全てのデータを使える
デメリット	● 境界時間τの明確な根拠が必要	● 比例ハザード性が成り立っていない時の解釈には注意が必要

4.2. 境界時間 τ の設定方法

RMST では、境界時間 τ の設定に伴い、結果及びその解釈が変わることから、境界時間 τ の設定には注意が必要である。本節では、主に①臨床的に意味のある時点までの平均生存時間を評価することを想定して、境界時間 τ を設定する際の留意点を説明し、本タスクフォースメンバーが考えるこれらの留意点を踏まえた境界時間 τ 設定のフローチャートを提示する。最後に、境界時間 τ の設定方法の一例を紹介する。

4.2.1. 境界時間 τ を設定する際の留意点

以下に、境界時間 τ を設定する際の留意点を記載する。

(1) 「臨床的に意味のある時点」が明確なとき

境界時間 τ を設定する際、初めに「試験目的に対応する臨床的に意味のある時点」が存在するかを確認する。同領域において、RMST を用いた類似の試験がある、時点評価（時点生存割合など）に一定のコンセンサスが得られている、複数の臨床家から推奨される評価時点があるときなどは、これらの情報に基づいて境界時間 τ を設定することが考えられる。なお、この方法で境界時間 τ を設定することは他試験との比較が可能になるという利点がある。

(2) 「臨床的に意味のある時点」が明確でないとき

臨床家の間で意見が異なり「臨床的に意味のある時点」を定められない、もしくは明確でないときは、平均生存時間の近似となるように、十分に長い期間を境界時間 τ として設定することが考えられる。生存時間が異なる 2 つの薬剤の比較（メカニズムが異なる 2 つの薬剤の比較など）では、境界時間 τ をあまりに短く設定してしまうと、適切な比較が行えない可能性がある。

(3) 境界時間 τ を決定するタイミング

境界時間 τ は症例数設計よりも前に決定されるべきであり、主要評価項目の評価指標の設定時に決定される。目標症例数を鑑みながら後付けで境界時間 τ を決定することは推奨されない。

(4) 境界時間 τ を試験途中で変更することの是非

境界時間 τ を試験途中で変更することは、主要評価項目の評価指標を変更することと同等である。よって、境界時間 τ 決定後に、境界時間 τ を変更することは一般的には薦められない。

(5) その他

境界時間 τ の設定は、主要評価項目の評価指標、症例数設計などの多くの要素に影響を与えるため、必要に応じて、試験開始前に規制当局との相談・合意しておくことも一案である。

4.2.2. 境界時間 τ 設定のフローチャート

4.2.1 項で示した留意点を踏まえて、 τ 設定のフローチャートを以下に示す。なお、以下のフローはあくまで一例であるため、詳細等に関しては、各状況に応じて適宜変更頂きたい。

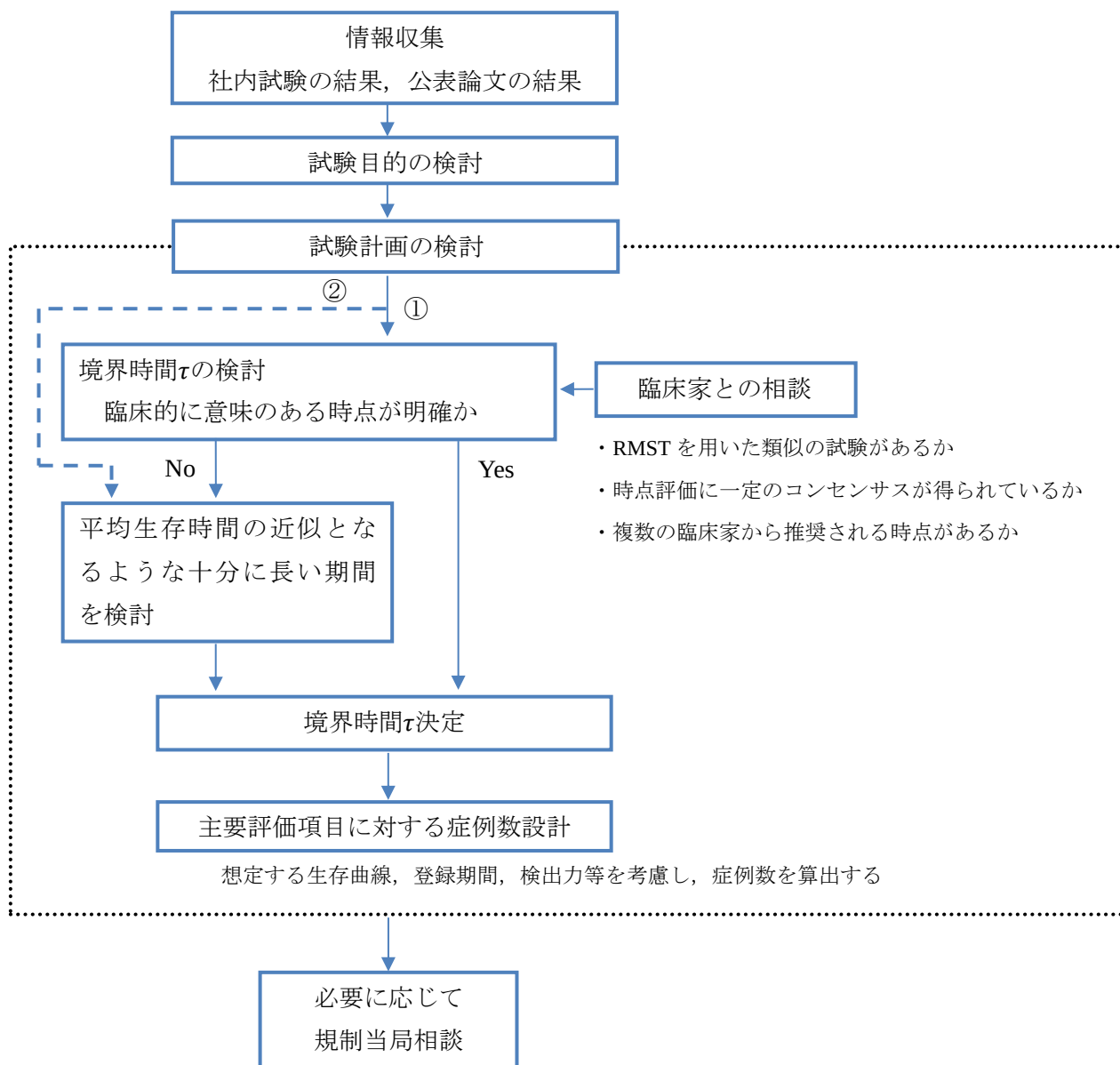


図 4-1 境界時間 τ 設定のフローチャート

4.2.3. 境界時間 τ 設定方法の一例

Murray and Tsiatis (1999)^[3] に以下のような記載があったため、境界時間 τ 設定方法の一例として紹介する。

For example, patients who have experienced a myocardial infarction may be given anticoagulants for a period of 30 days with survival benefit not expected to extend beyond an initial 6-month period of high risk for the patients in the study. After 6 months, survivors should have similar risk regardless of initial treatment. In such a case, the data may be monitored periodically after 6 months using $\tau = 6$.

意識すると「心筋梗塞の患者さんへ 30 日間、抗凝血薬を投与する試験において、初期の 6 ヶ月間を超えて生存ベネフィットが得られることが期待されないと考えられたため、 $\tau = 6$ を設定して評価を行なう」といった内容である。このように特定の期間を超えてベネフィットが期待できない場合には、ベネフィットが期待できる期間までを τ として設定し、評価することが紹介されている。

4.3. RMST における群間差、対照薬群の参照値の見積り方

4.3.1. RMST における群間差の見積もり方

試験デザインの検討において被験者数を決定する際には、実薬群及び対照薬群の RMST の真値を見積もる必要がある。これらの値の見積もり方としては、以下の方法が考えられる。

- (1) 実薬群及び対照薬群に対して、適切にパラメータを設定したモデル（例えば、ワイブル分布に基づくモデル、区分指数モデルなど）を仮定した生存曲線より見積もる。
- (2) 過去に行われた自社の試験データの生存曲線から RMST を見積もる。
- (3) 公表されている論文の生存曲線から RMST を見積もる。

4.3.2. RMST の参照値の求め方

本項では、前項で挙げた「(3) 公表されている論文の生存曲線から RMST を見積もる」の例を示す。

現状では、公表されている論文で RMST が報告されていない場合が多いが、その場合であっても、図を数値化するソフトウェア（例えば、Ingo Bormann による DigitizeIt）を利用して、論文中の Kaplan–Meier 曲線から RMST の近似値を求めることができる。

上述の DigitizeIt を利用して、公表された論文中の Kaplan–Meier 曲線の各点の x 座標及び y 座標を数値データに変換して RMST を計算する例を以下に示す。なお、本項では、Robert et al. (2015)^[4] で報告されている Figure 1 A の Kaplan–Meier 曲線の図を参考とした。なお、図中の打ち切りのシンボルについては、以下の計算では使用しないため省略した。

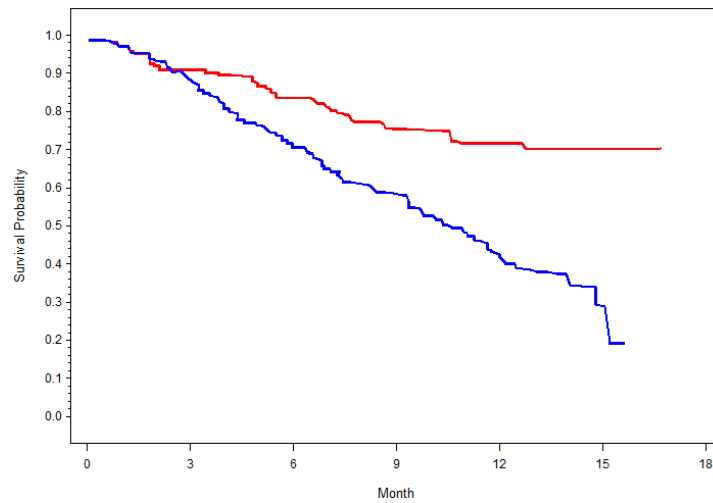


図 4-2 例題として用いる Kaplan-Meier 曲線

手順 1

DigitizeIt を起動する。「File」の「Open」により、Kaplan-Meier 曲線を含む画像ファイル (gif, bmp, png 等) を読み込む。

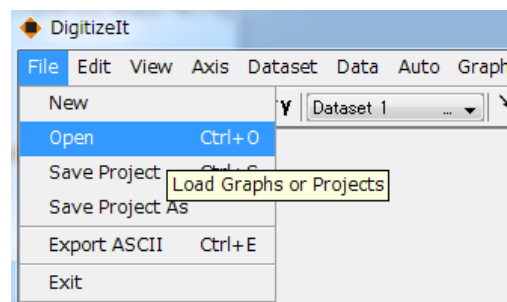
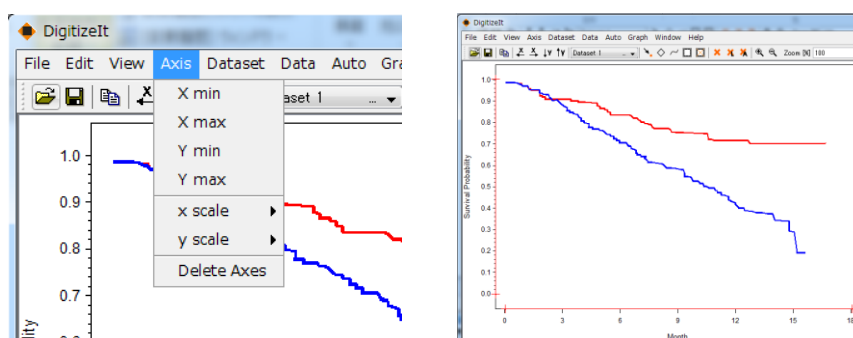


図 4-3 ファイルの読み込み

手順 2

「Axis」を用いて、x 座標の最大値及び最小値、y 座標の最大値及び最小値の位置を、図 4-4 の当該部分をクリックすることで指定する。



(x 軸及び y 軸の位置が赤線で指定されている)

図 4-4 x 軸及び y 軸の指定

手順 3

「Axis」の「x scale」及び「y scale」により、x 軸及び y 軸は「linear scale」であることを指定する。

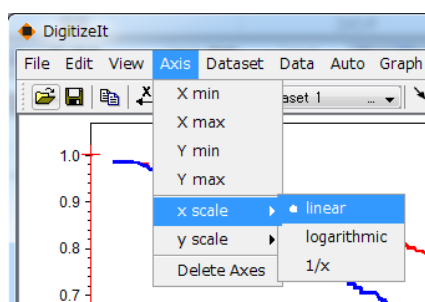
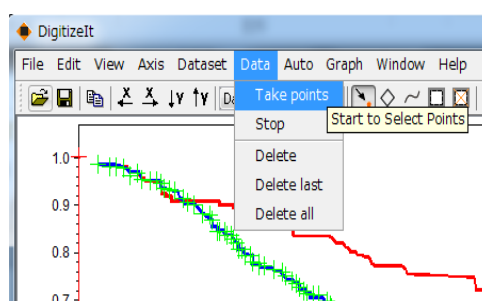


図 4-5 x 軸及び y 軸のスケールの指定

手順 4

「Data」の「Take points」により、Kaplan-Meier 曲線上を複数クリックすることで、Kaplan-Meier 曲線の座標の位置を指定する。



(黄緑色の部分が Kaplan-Meier 曲線の線上をクリックして位置を指定した部分である)

図 4-6 各点の座標のクリック

この際に、「View」の「Zoom in」により図を拡大し、Kaplan-Meier 曲線上を y 座標が単調非増加となるように丁寧にクリックして、座標の位置を指定していくことが大切である。

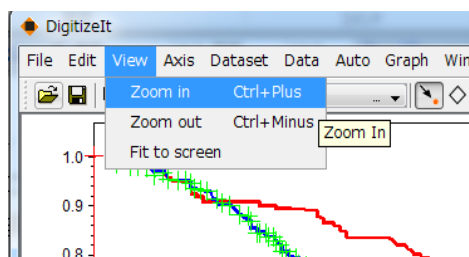
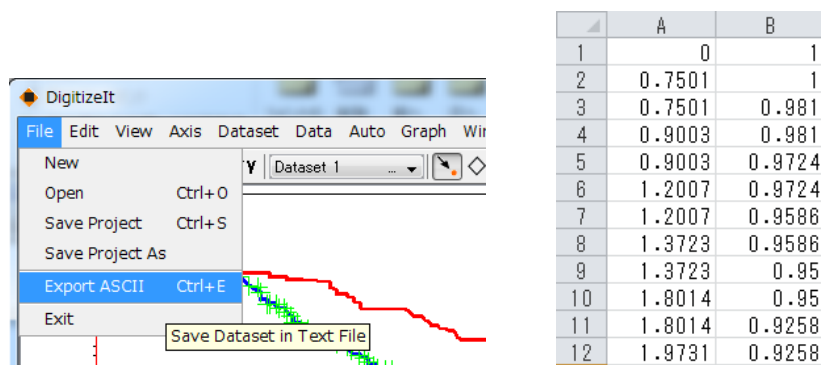


図 4-7 Zoom 機能による画面の拡大

手順 5

「File」の「Export ASCII」により、x 座標及び y 座標の位置の数値データを CSV ファイル等の形式のファイルに保存する。



(Spread sheet の A 列が x 座標:時点, B 列が y 座標:生存確率である)

図 4-8 座標ファイルの保存

手順 6

x 座標及び y 座標のデータから、境界時間 t までの曲線下面積を求め、RMST の近似値とする。

なお, DigitizeIt で作成した x-y 座標のデータ及び複数時点の at risk の被験者数及びイベント数のデータから, Guyot et al. (2013)^[5] で報告されている R のプログラムにより, 各被験者の近似的な生存時間を再構築することができる。また, SAS が利用可能な環境であれば, R のプログラムの代わりに, 7.2.6 項に掲載した SAS のプログラムを利用することもできる。これらの再構築データを用いることで, RMST 及びその信頼区間を近似的に計算することが

可能である．さらに，再構築データから算出される中央値やハザード比等を，文献で報告されている数値と比較することで，再構築データの再現性を確認することも可能である．

各被験者の近似的な生存時間を再構築する手順を以下に示す．

手順 7

DigitizeIt で作成した x-y 座標のデータに，observation 番号を付与した後テキストファイルとして保存する．

	A	B	C
1	1	0	1
2	2	0.7501	1
3	3	0.7501	0.981
4	4	0.9003	0.981
5	5	0.9003	0.9724
6	6	1.2007	0.9724
7	7	1.2007	0.9586
8	8	1.3723	0.9586
9	9	1.3723	0.95
10	10	1.8014	0.95
11	11	1.8014	0.9258
12	12	1.9731	0.9258
13	13	1.9731	0.9206
14	14	2.1018	0.9206
15	15	2.1018	0.9086
16	16	3.4535	0.9086
17	17	3.4535	0.9017
18	18	3.8397	0.9017
19	19	3.8397	0.8947
20	20	4.2044	0.8947

(Spread sheet の A 列は Observation 番号，B 列は x 軸:時間，C 列は y 座標:生存確率である．R のファイルの読み込み時に Header=True とする場合には，一行目に変数名を入力しておく必要がある．)

図 4-9 R に読み込む座標のファイル作成

各時点のリスクセットの大きさを表わすテキストファイルを作成する．

	A	B	C	D	E
1	1	0	1	15	210
2	2	3	16	31	185
3	3	6	32	49	150
4	4	9	50	55	105
5	5	12	56	57	45
6	6	15	58	59	8

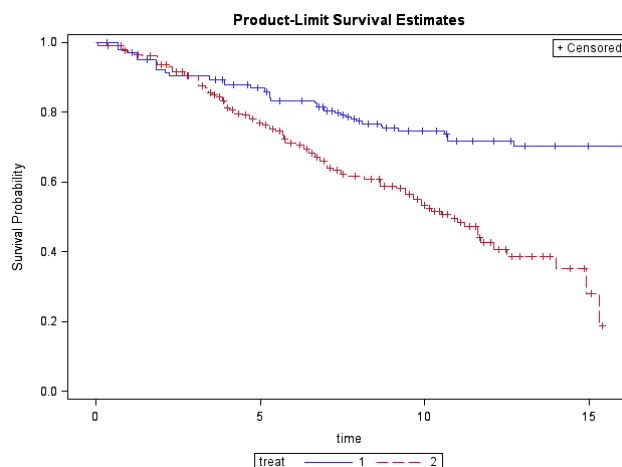
(A 列は順番のラベル，B 列は x 軸をいくつかの区間に分割した際の各区間の最小値 (月)，C 列及び D 列は，それぞれ，各区間の最初及び最終の observation 番号 [この例では，1～15 番目の observation が 0～3 ヶ月の区間に入っている] を示し，E 列は各区間の最初の時点におけるリスクセットの大きさである．R のファイルの読み込み時に Header=True とする場合には，一行目に変数名を入力しておく必要がある．)

図 4-10 R に読み込むリスクセットのファイルの作成

手順 8

座標のファイル及びリスクセットの大きさのファイルを群ごとに読み込み，Guyot et al. (2013)^[5] で報告されている R のプログラムもしくは 7.2.6 項に掲載した SAS のプログラムを

実行すると、各被験者の生存期間を再構築したデータが再構築される。なお、再構築された各被験者の生存期間データから SAS を用いて Kaplan–Meier 曲線を作成すると図 4-11 のようになった。 $\tau = 12$ ヶ月とした推定された RMST (標準誤差) は、2 つの群それぞれ、8.7 ヶ月 (0.28) 及び 10.1 ヶ月 (0.25) となった。



(Treat が 1 の場合には Nivolumab 群、2 の場合には Dacarbazine 群を表す)

図 4-11 再構築した被験者のデータに基づく Kaplan–Meier 曲線

4.4. RMST の群間差に対する統計的推測の情報量

生存時間型応答を主要評価項目とする試験では、実際に観測される生存関数や打ち切りの分布が、計画時の想定と異なる場合がある。このため、事前に規定した検出したい差を与えた下で、検出力を確保できるように、試験期間を調整し、最終解析の時期を決定する場合がある。例えば、log-rank 検定、または Cox 比例ハザードモデルを解析方法として用いる場合、試験計画時に算出される必要イベント数に基づき、最終解析の時期が決定される。本節では、RMST の群間差で評価する場合を想定し、これに対応する統計的推測の情報量について述べる。この情報量をモニタリングすることで、想定する差に対する検出力を確保できるように最終解析の時期を決定することが可能となる。また、 α 消費関数を用いた群逐次検定を中間解析で用いる場合にも、この情報量を利用することができる。なお、4.4.1 項で述べるが、log-rank 検定、及び Cox 比例ハザードモデルとは異なり、RMST の群間差に対する情報量はイベント数のみでは決定されず、リスク集合の大きさの推移にも依存して決定されることに留意する必要がある。

4.4.1. 情報量の計算式

一般に、統計的推測の情報量は、パラメータ推定の分散の逆数として定義される。例えば、両側有意水準 α で臨床的に意味のある差 Δ を検出力 $1 - \beta$ で検出したいならば、

$\Delta^2/\text{Var}(\Delta) = (z_{\alpha/2} + z_\beta)^2$ である必要がある．ここで、 z_α は標準正規分布の上側 100 α %点である．つまり、 $[\text{Var}(\Delta)]^{-1} = \Delta^{-2}(z_{\alpha/2} + z_\beta)^2$ を満たす必要があることになる．

さて、log-rank 検定に用いられる情報量がイベント数によって決まるということを以下で示す．イベントが発現した時点 T_j ($j = 1, \dots, D$)での被験者の内訳を表 4-3 のようにおく．

表 4-3 時点 T_j ($j = 1, \dots, D$)での被験者の内訳

	イベント	生存	リスク集合
対照薬群	d_{0j}	$n_{0j} - d_{0j}$	n_{0j}
実薬群	d_{1j}	$n_{1j} - d_{1j}$	n_{1j}
合計	d_j	$n_j - d_j$	n_j

このとき、log-rank 検定の検定統計量は以下のように表せる．

$$\frac{\sum_{j=1}^D (d_{0j} - E[d_{0j}])}{\sqrt{\sum_{j=1}^D \text{Var}[d_{0j}]}} \quad (4.1)$$

ここで、

$$E[d_{0j}] = \frac{n_{0j}d_j}{n_j}, \quad \text{Var}[d_{0j}] = \frac{n_{0j}n_{1j}d_j(n_j - d_j)}{n_j^2(n_j - 1)}$$

である．そして、これに対する情報量は、近似的に対数ハザード比の情報量、すなわち log-rank 検定統計量の分母にある $\sum_{j=1}^D \text{Var}[d_{0j}]$ に従うことが知られている．このため、log-rank 検定を用いた場合の情報量は、

$$\begin{aligned} \sum_{j=1}^D \text{Var}[d_{0j}] &= \sum_{j=1}^D \frac{n_{0j}n_{1j}d_j(n_j - d_j)}{n_j^2(n_j - 1)} \\ &\approx w_0w_1 \sum_{j=1}^D d_j, \quad (\because n_j - d_j \approx n_j - 1, n_{0j} \approx w_0n_j, n_{1j} \approx w_1n_j) \end{aligned} \quad (4.2)$$

と近似することが出来る．ここで、 w_g ($g = 0, 1$)は、解析対象集団の各群の例数割合を表し、帰無仮説の下では、計画時の各群の例数割合と等しいと考えられる．このため、log-rank 検定に対応する情報量はイベント数で近似できる．

ここでは、RMST の群間差を評価する試験を考え、統計的推測の情報量に基づき最終解析の時期を決定する際に確認する指標について述べる．群 g ($g = 0, 1$)の生存関数を $S_g(t)$ 、境界時間 τ までの RMST を $\mu_g(\tau)$ とすると、境界時間 τ までの RMST の群間差の推定量は、

$$\hat{\mu}_1(\tau) - \hat{\mu}_0(\tau) = \int_0^\tau \hat{S}_1(t)dt - \int_0^\tau \hat{S}_0(t)dt \quad (4.3)$$

と表せる． $\mu_1(\tau) - \mu_0(\tau)$ に対する情報量は、分散 $\text{Var}[\hat{\mu}_1(\tau) - \hat{\mu}_0(\tau)]$ の逆数で推定されるので、

$$\left\{ \text{Var} \left[\int_0^\tau \hat{S}_0(t) dt \right] + \text{Var} \left[\int_0^\tau \hat{S}_1(t) dt \right] \right\}^{-1} \quad (4.4)$$

と表せる．また，

$$\text{Var} \left[\int_0^\tau \hat{S}_g(t) dt \right] = \frac{1}{nw_g} \int_0^\tau \frac{\left\{ \int_u^\tau \tilde{S}(u) du \right\}^2}{\tilde{S}(t) \hat{G}_g(t) \tilde{Y}(t)} d\tilde{N}(t) \quad (4.5)$$

であることが，Murray and Tsatis (1999)^[3] で示されている．ここで， w_g ($g = 0, 1$)は，大きさ n の解析対象集団の各群の例数割合， $\tilde{S}(t)$ は併合群の Kaplan–Meier 法による生存関数， $\hat{G}_g(t)$ は群 g の打ち切りをイベントとする生存関数， $\tilde{N}(t)$ は時点 t で併合群において観測されたイベント数， $\tilde{Y}(t)$ は時点 t における併合群のリスク集合の大きさをそれぞれ表す．打ち切りの発生は割付けと独立であることを仮定することが多く，盲検下であっても， $\hat{G}_g(t) = \tilde{G}(t)$ として推定できる．したがって，計画時に算出する必要情報量は，作業対立仮説下で想定する各群の RMST を用いて，

$$\frac{(z_{\alpha/2} + z_\beta)^2}{\{\mu_1(\tau) - \mu_0(\tau)\}^2} \quad (4.6)$$

と計算する．そして，この必要情報量については，プロトコールに記載しておくことが望ましい．また，試験実施中に測定する情報量は，併合群データを用いて，

$$\begin{aligned} & \left\{ \frac{n}{nw_0} \sum_{j=1}^D \left[\sum_{k=j}^D (t_{k+1} - t_k) \tilde{S}(t_k) \right]^2 \frac{d_j}{n_j(n_j - d_j)} + \frac{n}{nw_1} \sum_{j=1}^D \left[\sum_{k=j}^D (t_{k+1} - t_k) \tilde{S}(t_k) \right]^2 \frac{d_j}{n_j(n_j - d_j)} \right\}^{-1} \\ &= \left\{ \frac{1}{w_0 w_1} \sum_{j=1}^D \left[\sum_{k=j}^D (t_{k+1} - t_k) \tilde{S}(t_k) \right]^2 \frac{d_j}{n_j(n_j - d_j)} \right\}^{-1} \end{aligned} \quad (4.7)$$

と算出する．すなわち，併合群データにおける境界時間 τ までの RMST の分散に $w_0 w_1$ で除した値の逆数が，評価時点の情報量となる．ここで，併合群における境界時間 τ までの相異なる D 個のイベント発現時点を $t_1 < t_2 < \dots < t_D, t_0 = 0, t_{D+1} = \tau$ とする．この情報量には，各時点 t_j における併合群のイベント数である d_j ，及びリスク集合の大きさである n_j が含まれていることから，イベント数だけでなく打ち切りの状況も考慮されていると解釈ができる．RMST の群間差に対する統計的推測の情報量が，必要情報量を超えた時点，もしくは情報量を必要情報量で割った情報時間が 1 を超えた時点で最終解析を実施すれば良いことになる．

なお，Murray and Tsatis (1999)^[3] では，RMST の群間差に対する統計量は，独立増分性が成り立つことを証明している．したがって，上記の情報量を用いることで，RMST の群間差に対して， α 消費関数を用いた群逐次検定を適用することができる．

4.4.2. 情報量の計算例

RMST の群間差に対する統計的推測の情報量の計算例を示す．ここでは，Robert et al. (2015)^[4] で報告されている Nivolumab のメラノーマ試験を参考に，情報量計算の具体的な手

順を述べる．Robert et al. (2015)^[4] の Figure 1A で示される全生存期間の Kaplan–Meier 曲線は，遅発効果と長期生存例の存在が特徴として挙げられる．このため，Hasegawa (2016)^[6] で提案されているこれら 2 つの特徴を考慮できる以下の生存関数モデルを各群で仮定した．

$$\begin{aligned} \text{対照薬群} : S_0(t) &= \kappa_0 + (1 - \kappa_0)\exp[-(\lambda t)^k] \\ \text{実薬群} : S_1(t) &= \begin{cases} \kappa_0 + (1 - \kappa_0)\exp[-(\lambda t)^k], & 0 \leq t < \varepsilon \\ \kappa_1 + c'\exp[-\{\psi\lambda(t - \varepsilon)\}^k], & \varepsilon \leq t \end{cases} \end{aligned} \quad (4.8)$$

ここで， $c' = \kappa_0 - \kappa_1 + (1 - \kappa_0)\exp[-(\lambda\varepsilon)^k]$ である．そして，各パラメータに $\kappa_0 = 0.0, \kappa_1 = 0.45, \varepsilon = 2.5, \lambda = 0.075, \psi = 0.9, k = 1.4$ と与え，図 4-12 に示す生存曲線を想定した．主要評価項目は，境界時間 τ を 12 ヶ月とする RMST とし，RMST の差により群間差を評価するものとした．このとき，実薬群，及び対照薬群の RMST は，それぞれ 10.20 ヶ月と 8.65 ヶ月であり，群間差は 1.55 ヶ月となる．一定の速さで登録が進むものとし，試験の登録期間は 13 ヶ月，そして登録終了から試験終了までの追跡期間を 5 ヶ月とした．すなわち，試験終了時点では，一部の被験者の追跡期間が 12 ヶ月に達していない状況を想定した．このような想定で，有意水準を両側 5%，検出力を 90%とした時，3.5 節で述べた Royston and Parmar (2013)^[7] の方法を用いて，各群 119 例が必要と算出された．更に，式(4.6) から必要情報量は 4.35 となる．

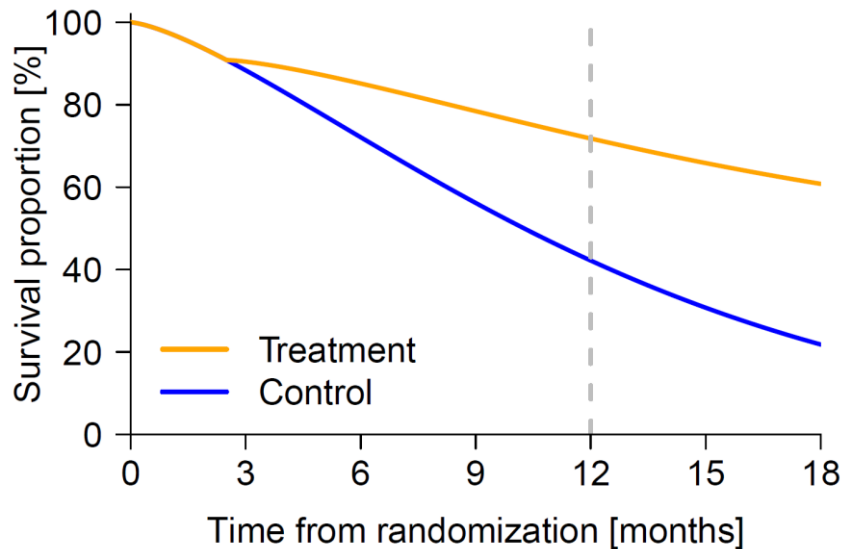


図 4-12 想定した生存曲線

境界時間 τ である 12 ヶ月を越えて被験者が追跡されている状況下で，RMST に対する情報は算出できる．このため，試験開始から 12 ヶ月以降の各時点の情報量を式(4.7) で算出し，これを必要情報量 4.35 で除した情報時間の推移図を描いた．10 回の臨床試験シミュレーション

ョンデータに基づき、10本の情報時間の推移を描いたものが図4-13の左図である。試験終了と計画されている18ヶ月目で情報時間が1を超えている場合と超えていない場合それぞれの状況が確認された。想定よりもイベント数が少なく、また打ち切りが少ない場合には、RMSTの群間差に対する分散が小さくなることで、その逆数で定義される情報量が大きくなり、結果として情報時間が1を超えたと考えられた。逆に、想定よりもイベント数が多く、また打ち切りが多い場合には、RMSTの群間差に対する分散が大きくなることで、情報時間が1に達しなかったと考えられた。また、多くの被験者の追跡期間が境界時間 τ である12ヶ月を越えていると考えられる21ヶ月目以降では情報時間の推移がほぼ頭打ちとなっていた。特に、情報時間が1未満で頭打ちとなる場合には、被験者を追加しない限り、必要情報量へ到達することが出来ないと考えられた。図4-13の右図は、同一の10回のシミュレーションデータを用いて、各評価時点におけるlog-rank検定、及びCox比例ハザードモデルの情報量に対応するイベント数（横軸）とRMSTの群間差に対する情報時間（縦軸）の関係を表した。イベント数が増えるとRMSTの群間差に対する情報時間も増加していたものの、イベント数とRMSTの群間差に対する情報時間の間で、比例関係はなかった。また、同じイベント数の中でも、RMSTの群間差に対する情報時間はデータによって異なっていたことから、情報時間が1に達したかどうかをイベント数で代用して判断することは出来ないと解釈された。このため、併合群データを用いて式(4.7)から情報量を算出し、情報時間をモニタリングする必要があると考えられた。

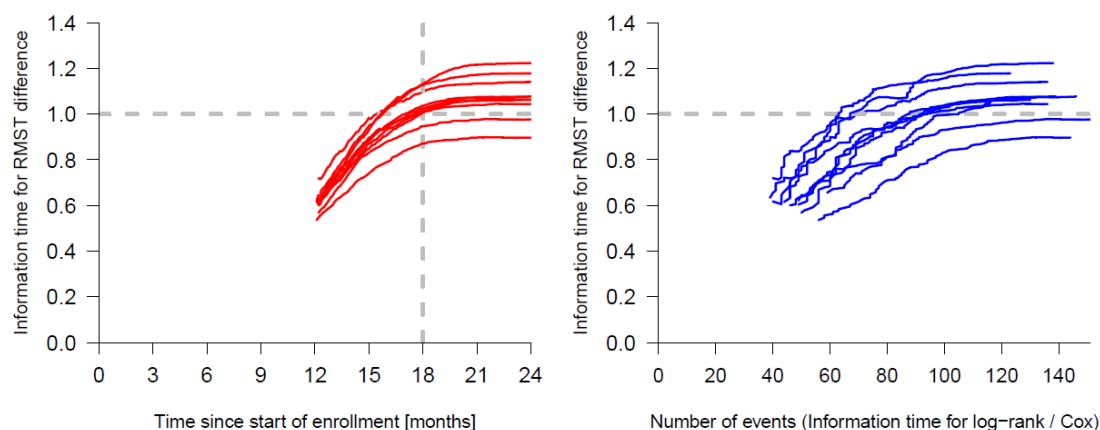


図 4-13 10回のシミュレーションデータにおける情報時間の推移

参考文献

- [1] Tian L, Fu H, Ruberg SJ, Uno H, Wei LJ. Efficiency of two sample tests via the restricted mean survival time for analyzing event time observations. *Biometrics*. 2017; doi:10.1111/biom.12770.
- [2] Uno H, Claggett B, Tian L, Fu H, Huang B, Kim D H, *et.al*. Adding a new analytical procedure with clinical interpretation in the tool box of survival analysis. *Ann Oncol*. 2018. <https://doi.org/10.1093/annonc/mdy109>

- [3] Murray S, Tsiatis AA. Sequential methods for comparing years of life saved in the two-sample censored data problem. *Biometrics*. 1999; **55**(4): 1085–1092.
- [4] Robert C, Long GV, Brady B, Dutriaux C, Maio M, Mortier L, *et.al*. Nivolumab in previously untreated melanoma without BRAF mutation. *N Engl J Med*. 2015; **372**: 320-330.
- [5] Guyot P, Ades AE, Ouwens MJ, Welton NJ. Enhanced secondary analysis of survival data: reconstructing the data from published Kaplan–Meier survival curves. *BMC Med Res Methodol*. 2012; **12**:9.
- [6] Hasegawa T. Group sequential monitoring based on the weighted log-rank test statistic with the Fleming–Harrington class of weights in cancer vaccine studies. *Pharm Stat*. 2016; **15**:412–419.
- [7] Royston P, Parmar MKB. Restricted mean survival time: an alternative to the hazard ratio for the design and analysis of randomized trials with a time-to-event outcome. *BMC Med Res Methodol*. 2013; **13**: 152.

5. 性能評価

本章では生存時間解析における評価指標としての RMST の性能評価を行った。まず、性能評価の検討事例として、本報告書作成時点でハザード比と比較した結果が纏められていた Trinquart et al. (2016)^[1] 及び Tian et al. (2017)^[2] を著者らの意見に則って概説する。次に本タスクフォースで実施したシミュレーションによる検討を紹介する。論文等で公開されている臨床試験結果を参考に幾つかの生存曲線パターンを想定し、それぞれで log-rank 検定及び RMST の差に対する検定を実施、検出力の観点から結果を考察した。

5.1. Trinquart et al. (2016)

5.1.1. Trinquart et al. (2016) の方法

Trinquart et al. (2016)^[1] は、がん臨床試験を対象に、ハザード比と RMST の差及び比によって測定された治療効果を複数の事例に基づいて比較した結果を報告している。5 つの主要学術誌から 2014 年 7 月 1 日～12 月 31 日に掲載されたがんを対象とするランダム化比較試験を抽出した。抽出に際しては、2 人の review author が独立にタイトルやアブストラクトを精査した。試験の抽出にあたり、具体的な選択基準等は Trinquart et al. (2016)^[1] を参照いただきたい。

本研究では、各試験で報告された Kaplan–Meier 曲線から各患者の個人データ (individual patient data, IPD) を試験毎に Digitizeit software を利用して再構築した。その上で試験毎にハザード比と RMST の差 (及び比) を算出し、各指標による治療効果の値を比較した。RMST を算出する際に必要となる境界時間 τ は、各群で観察されるイベント発現までの最長期間のうち、短い方の期間と定義した。なお、各評価指標は対照薬群に対する実薬群の薬効として定義した。解析に際し、ハザード比も RMST の比も 1 より大きい場合に実薬群が優れていると解釈できるようにした。また、RMST の差が 0 より大きければ実薬群の治療効果が大きいことを示す。ハザード比と RMST の差 (及び比) は 2 つの生存曲線の差の定量的な指標として異なる意味を持つが、本論文では、いずれも治療効果の大きさを群間で比較する指標であることに着目し、一方の指標が系統的に他方と異なっているかどうかを評価することを目的とした。著者らは、本論文の評価指標として「RMST の比に対するハザード比」の比の平均を利用した。すなわち、当該指標が 1 より大きければ、RMST の比よりもハザード比が大きく、ハザード比に基づく評価は相対的に治療効果を大きめに見積もっていると解釈できる。

5.1.2. Trinquart et al. (2016) の結果

以下、本論文に示された結果を記載する。54 のランダム化比較試験が選択され、33,212 例のデータが解析された。選択された試験のうち、全生存期間 (overall survival, OS) を評価指標とした試験は 21 試験 (39%)、PFS が 23 試験 (43%) であった。選択された試験の詳細は本論文の Table 1 を参照いただきたい。log-rank 検定の結果、24 試験 (44%) で生存曲線間

に有意差が認められ、いずれも実薬群が優る結果であった。Grambsch-Therneau test（有意水準 10%）を用いて比例ハザード性を検討したところ、比例ハザード性が成立していないと考えられる試験が 13 試験（24%）含まれていた。

次に RMST の差（及び比）とハザード比を比較した結果を記載する。なお、本報告書では著作権上の理由により図の掲載を割愛する。RMST の差及び比とハザード比をプロットした結果が本論文の Figure 1 と 2 にそれぞれ描画されている。4 試験を除き、ハザード比と RMST（差、比）で治療効果の方向性は一貫していた。本論文の Figure 3 には 54 試験の RMST の比とハザード比及び信頼区間を並べて表示した結果が示されている。1 試験を除き、いずれの試験でも RMST の比とハザード比で統計学的有意性は一貫していた。1 試験のみ RMST の差及び比では実薬群が統計学的に有意であったが、ハザード比では有意とならなかった。また、54 試験中 41 試験（76%）でハザード比が RMST の比よりも実薬群で大きな治療効果を示し、20 試験（37%）で有意な結果であった。また、4 試験（7%）でハザード比の値が RMST の比の 2 倍以上を示した。一方、13 試験で RMST の比がハザード比より大きな値を示したが、統計学的有意差を示した試験はなかった。

「RMST の比に対するハザード比」の比の平均値は 1.11（95%CI：1.07, 1.15）で、系統的にハザード比に基づく評価がより楽観的な値を示すことが示唆された。なお、本論文の Figure 4 で示されているように、試験間変動はかなり大きかった（ $I^2 = 86\%$ 、 I^2 は試験間の不均一性を測る指標）。以上のように、治療効果の統計学的有意性が示されるか否かは RMST の差（及び比）とハザード比で一貫していたが、RMST に基づく治療効果の推定はハザード比よりも保守的な推定値を返すことが経験的に示唆された。この傾向は OS でも他の評価指標でも一貫していた。また、比例ハザード性が成立しているかどうかにも依存しない結果であった。この結果は治療効果の指標として RMST とハザード比が交換不能なものであると解釈することができる。もしハザード比が RMST の比より系統的に大きな治療効果を返す指標であると一般化して解釈できるとすると、特に実臨床の現場では誤った解釈を引き起こすリスクが示唆される。

本調査では実に 24%の試験で比例ハザード性が成立していないと考えられたため、HR だけではなく、RMST の差も併せて報告されるべきであると考えられる。

最後に本研究の limitation をいくつか挙げる。まず、本研究は再構築された IPD に基づいた調査結果であり、各試験の元データから算出される結果との一貫性を主張することは困難である。しかしながら、すべての試験の元データを入手することは現実的に不可能であり、さらに検証されたアルゴリズムに基づいて IPD を算出しており、正確性や再現性に優れた手法を用いている。また、再構築されたデータから算出した中央値やハザード比はオリジナルの論文に近い推定値を与えている。2 つ目に、本論文では、Kaplan-Meier 曲線を直接積分することにより RMST を推定している。Kaplan-Meier 推定値は at risk が小さい時に不安定になる。多くの場合、RMST を推定する際の境界時間 τ を超えたあたりで Kaplan-Meier 推定値が不安定になっている可能性がある。このような懸念がある場合には、RMST

の推定に際し、例えばパラメトリックモデルを用いる方法等が考えられるかもしれない。3つ目として、本研究では公表された影響力の大きいがんランダム化比較試験を対象としているため、一般化可能性に欠けるかもしれない。

結論として、ハザード比と RMST の差（及び比）に基づく治療効果の指標は統計学的有意性という観点では一致するが、ハザード比は平均的に RMST の比よりも大きな治療効果を示す傾向にあった。したがって、time-to-event のランダム化試験では RMST に基づく治療効果の推定結果も定期的に報告されるべきである。

5.2. Tian et al. (2017)

本項では Tian et al. (2017)^[2] で実施されたシミュレーションによる RMST とハザード比の比較について述べる。Tian et al. (2017)^[2] は前述の Trinquart et al. (2016)^[1] により経験的に示唆された結果を受けて、より理論的な状況で RMST とハザード比の比較を検討した。なお、Tian et al. (2017)^[2] では、RMST を t-year mean survival time として記載しているが、同一の統計量であるため、本報告書では t-year mean survival time を RMST として読み替えて記載する。

5.2.1. Tian et al. (2017) の方法

Tian et al. (2017)^[2] では理論的かつ実用的な状況下で、比例ハザード性が成り立つ場合と成り立たない場合において、ハザード比に基づく検定と RMST の差に基づく検定について比較を行い、相対的な効率性について評価した。それぞれの検定が漸近的に同等であることを示し、シミュレーション実験を実施した。シミュレーション実験における評価指標として、Asymptotical relative efficiency (ARE) と Empirical relative efficiency (ERE) という2種類の指標を用いて評価している。ARE は RMST の差に基づく検定とハザード比に基づく検定、それぞれについて漸近的な効率性 ξ_1 と ξ_2 を導出し ξ_1/ξ_2 を指標として用いた。この指標は各検定で検出力を保持するために必要なサンプルサイズの比の逆数と解釈することができる。標準的な log-rank 検定（タイデータがない）は、2つの生存関数が等しいという仮説に対する検定であるスコア検定と一致し、スコア検定はワルド型の検定と漸近的に等価である。そのため、 ξ_1/ξ_2 は log-rank 検定と比較した RMST の検定の ARE となる。しかし、ARE は効果の差がかなり大きい場合にはサンプルサイズが小さくなるため、精度の高い近似とならない可能性があり、その場合には ERE を用いる。ERE は効果の大きさの期待値をそれぞれの標準誤差で割ることで標準化した effect size を2乗した比である。これらの指標の詳細は Tian et al. (2017)^[2] を参照いただきたい。

Tian et al. (2017)^[2] では理論的に考えられる特性を確認するために、Rajkumar et al. (2010)^[3] の多発性骨髄腫患者を対象とした低用量と高用量の dexamethasone を比較した試験 (ECOG-ACRIN) を用いた。この試験にはランダム化した222名の被験者が低用量群に割り付けられ、223名の被験者が高用量群に割り付けられた OS の結果が示されている。OS の実際の Kaplan-Meier 曲線については Tian et al. (2017)^[2] を参照していただきたいが、見かけ

上低用量群の OS の方が延長しているように見え、ハザード比に基づく検定と RMST の差に基づく検定の P 値はそれぞれ 0.47, 0.04 であった。この違いは Kaplan-Meier 曲線が後半で交差していることによるものと考えられる。さらに ECOG-ACRIN 試験の結果を参考に、新たな試験を立案する状況を想定し、条件を変えてシミュレーションを実施し、ハザード比に基づく検定と RMST の差に基づく検定の ARE, ERE が示す特性について検討及び評価した。その後、条件を変更した追加のシミュレーションも合わせて実施し検討を行った。

5.2.2. Tian et al. (2017) の結果

比例ハザード性が成り立つ状況下において、ハザード比に基づく検定と RMST の差に基づく検定の ARE の結果はおおよそ同様であった。比例ハザード性が成り立っていない状況下では、フォローアップ期間中の打ち切りが少なく、試験終盤近くの生存関数の差が大きい場合を除き、RMST の差に基づく検定はハザード比に基づく検定よりも効率が良かった。その他の結果も含め要約すると、RMST の差に基づく検定は早期に効果を期待できる場合、ハザード比に基づく検定は近年の免疫治療のように遅発性の効果を検討する場合に検出力が高いと言えた。しかしながら試験終盤近くの生存関数の差が大きい場合にはハザード比に基づく検定のほうがより検出力が高くなるかもしれないが、その状況下では比例ハザード性が成り立っていないため、ハザード比の推定値自体の単純な臨床的解釈ができないかもしれないと本論文では述べている。

5.3. シミュレーション検討

死亡や特定の有害事象の発現などイベントが発現するまでの時間を評価する臨床試験では、Kaplan-Meier 法で生存関数を図示し、log-rank 検定で生存関数の群間比較を行い、Cox 比例ハザードモデルで評価指標であるハザード比を推定し、治療効果の大きさを議論することが多い。一方で、生存関数の評価指標には、ハザード比に加え、半数の被験者がイベントを経験するまでの時間、ある時点の生存割合や規定した特定時点までの平均生存時間などがあり、それぞれの観点から結果の考察に用いられている。近年、生存時間解析の新たな指標として RMST が注目されているが、現実問題として実際の臨床試験での適用例が多くなく、妥当な結果の解釈や適用場面といった RMST の特徴を検討した結果も少ない。そこで本節では RMST の特徴を検討する目的で、実際の臨床試験結果を参考に幾つかの生存関数パターンを仮定した上で、シミュレーションにより性能評価の検討を実施した。ここでは、RMST と他の評価指標で評価しているものが異なることから、有意な差を検出するという観点のみに着目し、RMST の差を用いた検定と log-rank 検定の検出力の比較を行った。対照薬群に対する実薬群の優越性を検証するというよく見られる臨床試験の目的にあわせて、本シミュレーションでは検出力の算出に片側有意水準 2.5%を用いた。なお、評価指標を選択する際には、検出力のみに着目するのではなく、評価指標の各特徴を踏まえた上で、試験結果に対して臨床的・統計的に適切な解釈を与えられる評価指標を選択することが望

ましい姿であると考えられる。

5.3.1. シミュレーションの概要と想定した生存関数

比例ハザード性が成り立っていない生存時間データを想定して、log-rank 検定と RMST の差を用いた検定を実施するシミュレーション検討を実施した。生存時間データの発生には区分指数分布を用い、想定した生存関数に応じて区間ごとにハザードを仮定した。

5.3.1.1. 区分指数分布に従う乱数の発生方法

時点 t_{j-1} から時点 t_j の区間 j ($j = 1, \dots, J$)におけるハザードを一定の λ_j とすると、時点 t のハザード関数 $\lambda(t)$ 、及び累積ハザード関数 $\Lambda(t)$ は以下のように書ける。

$$\lambda(t) = \begin{cases} \lambda_1, & \text{if } 0 < t \leq t_1 \\ \lambda_2, & \text{if } t_1 < t \leq t_2 \\ \vdots & \vdots \\ \lambda_j, & \text{if } t_{j-1} < t \leq t_j \end{cases}$$

$$\Lambda(t) = \begin{cases} \lambda_1 t, & \text{if } 0 < t \leq t_1 \\ \lambda_1 t_1 + \lambda_2 (t - t_1), & \text{if } t_1 < t \leq t_2 \\ \vdots & \vdots \\ \sum_{i=1}^{j-1} \lambda_i (t_i - t_{i-1}) + \lambda_j (t - t_{j-1}), & \text{if } t_{j-1} < t \leq t_j \end{cases}$$

以上から、個々の被験者の生存時間 T は $\lambda = 1$ の指数分布から発生させた乱数 x を用いて以下から得ることができる。

$$T = \Lambda^{-1}(x) = \begin{cases} \frac{x}{\lambda_1}, & \text{if } 0 < x \leq \lambda_1 t_1 \\ t_1 + \frac{x - \lambda_1 t_1}{\lambda_2}, & \text{if } \lambda_1 t_1 < x \leq \lambda_1 t_1 + \lambda_2 (t_2 - t_1) \\ \vdots & \vdots \\ t_{j-1} + \frac{x - \sum_{i=1}^{j-1} \lambda_i (t_i - t_{i-1})}{\lambda_j}, & \text{if } \sum_{i=1}^{j-1} \lambda_i (t_i - t_{i-1}) < x \leq \sum_{i=1}^j \lambda_i (t_i - t_{i-1}) \end{cases}$$

5.3.1.2. 仮定した生存関数

実際の臨床試験では比例ハザード性を仮定して試験デザインを決定することが多いが、試験結果として得られる生存時間データに比例ハザード性が見られないといったケースは少なくない。本シミュレーションでも比例ハザード性の成否に関わらず、log-rank 検定と RMST の差を用いた検定を実施した場合の結果を比較検討した。すなわち、試験期間を通じて比例ハザード性が成立しているケースに加え、時間区分ごとに異なるハザードを設定し比例ハザード性が成立していない条件で生存時間データを発生させた。

本報告書では対照薬群と実薬群の 2 群比較とし、以下の 5 パターンによる仮定を用いてモンテカルロ・シミュレーションにより生存時間データを発生させた。なお、以下に挙げる項目を全てのパターンにおける共通の設定とした。

- ・登録期間：24 ヶ月
- ・登録の速さ：一定

- ・ 追跡期間：12 ヶ月
- ・ 被験者数：500 名（2 群合計）
- ・ 1 年あたりの打ち切り発生割合：5%
- ・ 有意水準：片側 2.5%
- ・ シミュレーション回数：10,000 回

想定した 5 つの生存関数を以下にまとめる.

	生存関数の特徴
パターン 1	比例ハザード性が成立している場合
パターン 2	特定時点までは差がなく、それ以降に差が開く場合
パターン 3	試験期間を通じてイベントの発現数が少ない場合（比例ハザード性は成立）
パターン 4	特定時点までは差が大きい、試験終了時には差がなくなる場合
パターン 5	生存曲線がクロスする場合

【パターン 1】 比例ハザード性が成立している場合

対照薬群：6 ヶ月生存割合 70%，36 ヶ月生存割合 11.8%（試験期間中のハザードは一定）

実薬群：6 ヶ月生存割合 78%，36 ヶ月生存割合 22.5%（試験期間中のハザードは一定）

生存関数：

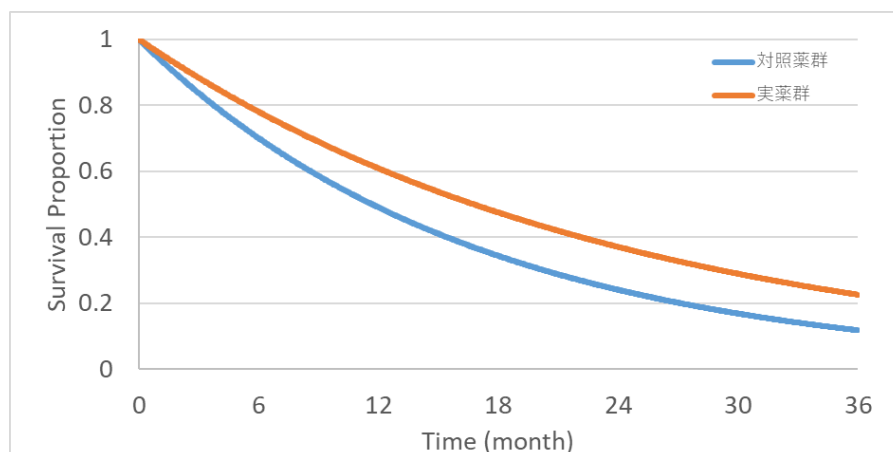


図 5-1 パターン 1 の生存関数

表 5-1 パターン 1 の各時点の生存割合

Time (month)	生存割合（対照薬群）	生存割合（実薬群）
6	70.0%	78.0%
12	49.0%	60.8%
18	34.3%	47.5%
24	24.0%	37.0%
30	16.8%	28.9%
36	11.8%	22.5%

ハザード比：

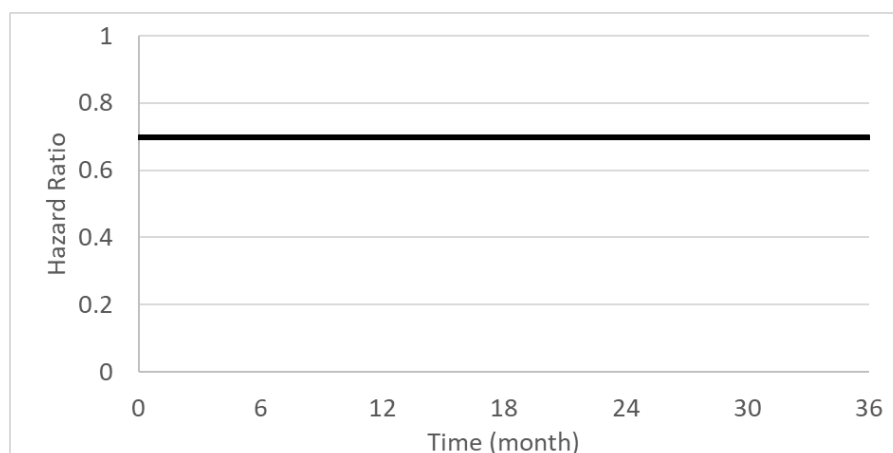


図 5-2 パターン 1 のハザード比の推移

【パターン2】 特定時点までは差がなく、それ以降に差が開く場合

対照薬群：6ヶ月生存割合 70%，36ヶ月生存割合 11.8%（試験期間中のハザードは一定）

実薬群：6ヶ月生存割合 70%，36ヶ月生存割合 30%（6ヶ月以降ハザードが変化）

生存関数：

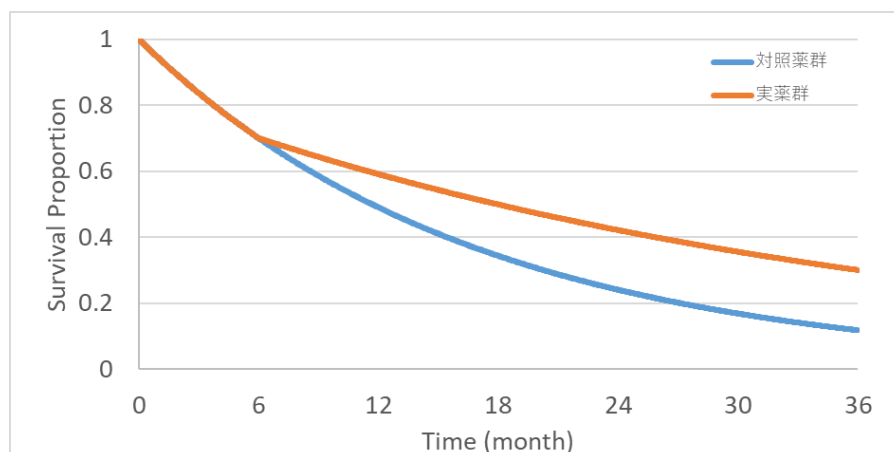


図 5-3 パターン2の生存関数

表 5-2 パターン2の各時点の生存割合

Time (month)	生存割合（対照薬群）	生存割合（実薬群）
6	70.0%	70.0%
12	49.0%	59.1%
18	34.3%	49.9%
24	24.0%	42.1%
30	16.8%	35.5%
36	11.8%	30.0%

ハザード比：

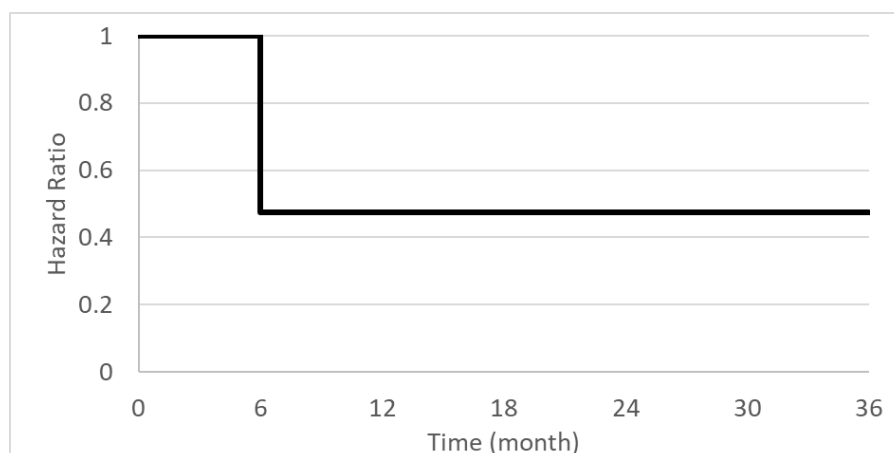


図 5-4 パターン2のハザード比の推移

【パターン 3】 試験期間を通じてイベントの発現数が少ない場合（比例ハザード性は成立）

対照薬群：36 ヶ月生存割合 85%

実薬群：36 ヶ月生存割合 90%

生存関数：

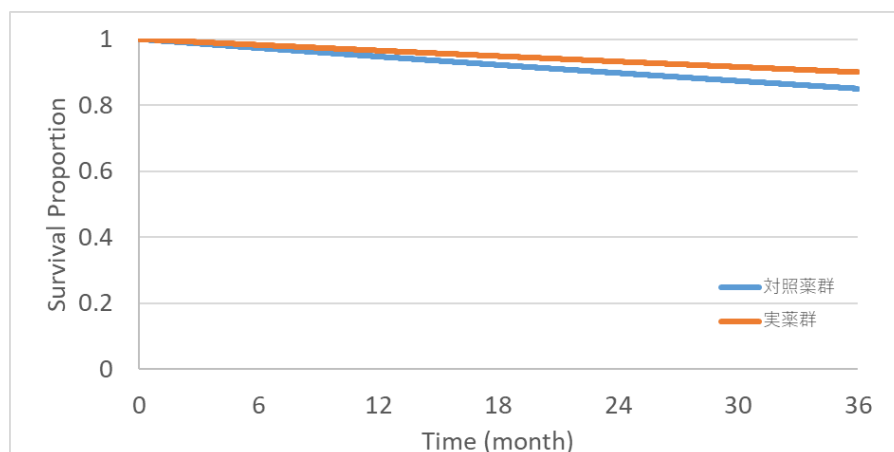


図 5-5 パターン 3 の生存関数

表 5-3 パターン 3 の各時点の生存割合

Time (month)	生存割合（対照薬群）	生存割合（実薬群）
6	97.3%	98.3%
12	94.7%	96.5%
18	92.2%	94.9%
24	89.7%	93.2%
30	87.3%	91.6%
36	85.0%	90.0%

ハザード関数：

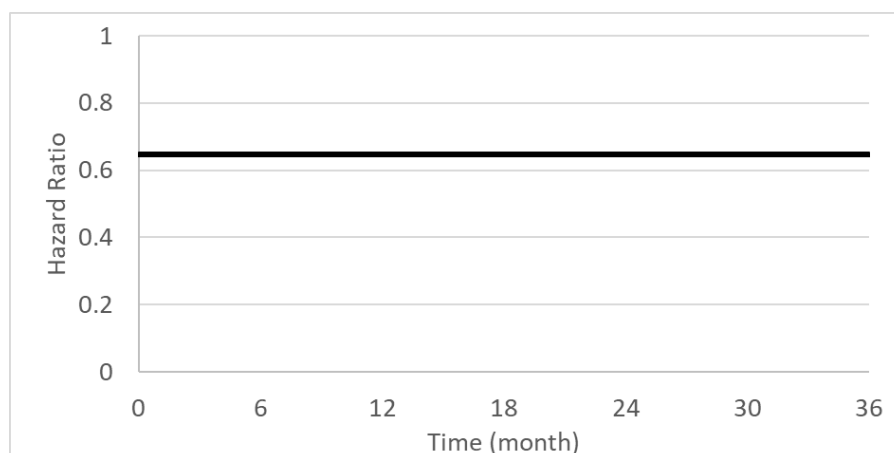


図 5-6 パターン 3 のハザード比の推移

【パターン4】 特定時点までは差が大きいですが、試験終了時には差がなくなる場合

対照薬群：36ヶ月生存割合 11.8%（試験期間中のハザードは一定）

実薬群：9ヶ月生存割合 85%，15ヶ月生存割合 45%，36ヶ月生存割合 11.8%

生存関数：

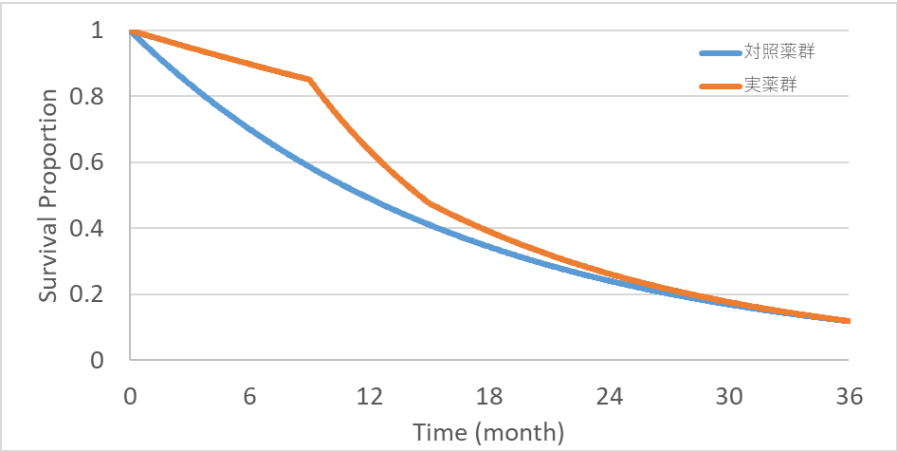


図 5-7 パターン4の生存関数

表 5-4 パターン4の各時点の生存割合

Time (month)	生存割合（対照薬群）	生存割合（実薬群）
6	70.0%	89.7%
12	49.0%	61.8%
18	34.3%	37.2%
24	24.0%	25.4%
30	16.8%	17.3%
36	11.8%	11.8%

ハザード比：

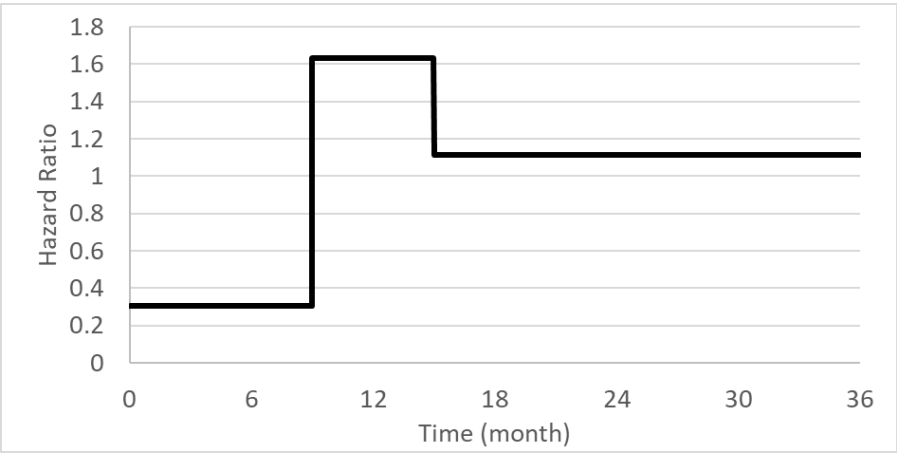


図 5-8 パターン4のハザード比の推移

【パターン 5】 生存曲線がクロスする場合

対照薬群：3 ヶ月生存割合 60%，6 ヶ月生存割合 50%，12 ヶ月生存割合 8%，
36 ヶ月生存割合 2%

実薬群：3 ヶ月生存割合 55%，8 ヶ月生存割合 25%，12 ヶ月生存割合 19%，
36 ヶ月生存割合 10%

生存関数：

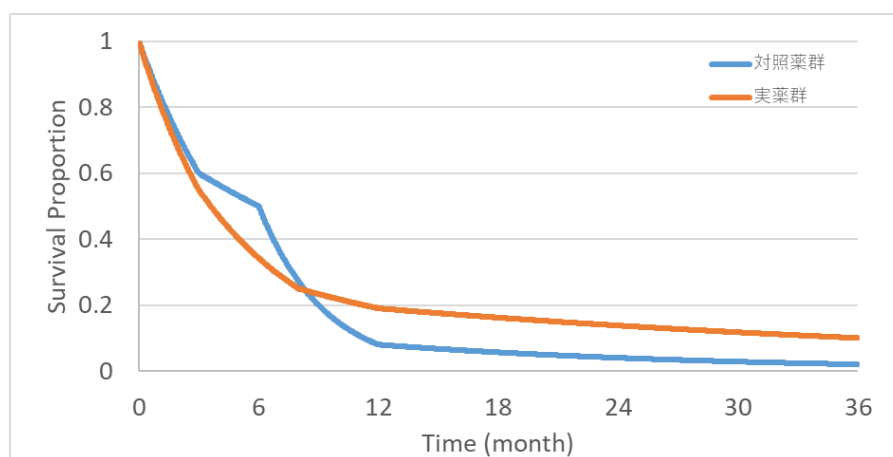


図 5-9 パターン 5 の生存関数

表 5-5 パターン 5 の各時点の生存割合

Time (month)	生存割合（対照薬群）	生存割合（実薬群）
6	50.0%	34.3%
12	8.0%	19.0%
18	5.7%	16.2%
24	4.0%	13.8%
30	2.8%	11.7%
36	2.0%	10.0%

ハザード比：

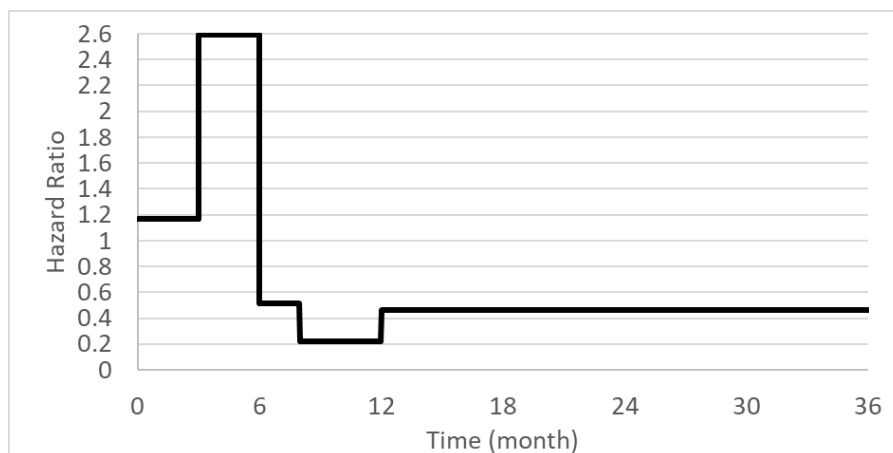


図 5-10 パターン 5 のハザード比の推移

5.3.2. シミュレーション結果

本検討ではモンテカルロ・シミュレーションにより発生させた生存時間データに対して 3 種類の検定を実施した。すなわち、境界時間 τ における RMST の差の検定、log-rank [all]検定及び log-rank [τ]検定である。すべての検定は試験期間（登録期間+追跡期間）完了時に得られている生存時間データを用いて解析するが、RMST の差の検定及び log-rank [τ]検定では境界時間 τ 以降に発現したイベントを打ち切りとして処理したデータを用いた。通常の log-rank [all]検定では境界時間 τ 以降に発現したイベント情報も解析に用いるが、RMST の差の検定では境界時間 τ 以降に発現したイベントは検定実施時点で観察されていたとしても解析には用いない。よって、解析に用いるデータを同じとした場合の検出力の違いを確認するために log-rank [τ]検定を実施した。しかしながら、log-rank [τ]検定では解析実施時には得られている境界時間 τ 以降に発現したイベントを用いないことから結果の過小評価になっていることに留意すべきである。

シミュレーション結果として各検定の検出力、RMST の差の点推定値及び標準誤差、境界時間 τ までの発現イベント数、打ち切り数を出力した。更に、RMST と log-rank [all]のシミュレーションごとの P 値の散布図を作成することで、想定した 5 つの生存関数パターン別に、検定間の P 値の関係を視覚的に検討した。なお、境界時間 τ は 12 ヶ月から 36 ヶ月まで 3 ヶ月ごとに設定した。検定間の比較は検出力を用いて検討し、境界時間 τ までの At risk 割合を見ることで RMST の差の検定と log-rank [all]検定で検定に用いる情報量の違いを確認することができる。

5.3.2.1. 比例ハザード性が成立している場合（パターン 1）

実薬群と対照薬群の 6 ヶ月生存割合をそれぞれ 78%、70%（ハザード比 0.7）として試験期間を通じて比例ハザード性が成立する場合を想定した。

$\tau = 12$ で打ち切りとなった被験者の割合を見ると実薬群，対照薬群で 3.9%, 3.6%と低く，12 ヶ月時点のリスク集合に含まれる被験者（12 ヶ月時点までにイベントも打ち切りも発生しなかった被験者）の割合は 57.8%, 46.6%と高かった．これは，試験の追跡期間を 12 ヶ月とした為に， $\tau = 12$ ではほぼ全ての被験者で 12 ヶ月の観察が終了していたことが結果に反映され，純粋に途中脱落の被験者が数字に表れていると考えられた．一方で， $\tau = 36$ では全ての被験者の追跡期間が終了しているためリスク集団の割合が 0%になっていた．登録期間終盤で登録された被験者は 36 ヶ月の観察期間が確保できないため，境界時間 τ よりも前の時点で打ち切りとして処理された．

比例ハザード性が成立している下では，境界時間 τ の設定に関わらず，log-rank 検定の方が高い検出力を示したが，両手法とも境界時間 τ が大きくなるにつれて検出力も増大した．P 値の散布図を見ると，全ての境界時間 τ を通じて log-rank 検定の方が小さい P 値を示し，境界時間 τ が長くなるにつれて両手法の P 値の差が少なくなっていた．

表 5-6 RMST と log-rank 検定の性能評価
(パターン 1：比例ハザード性が成立している場合)

τ	Power			Δ RMST (Mean $\pm$ se)	実薬群 [#]			対照薬群 [#]		
	RMST	log-rank [τ]	log-rank [all]		Event	Censor ^{##}	At risk	Event	Censor ^{##}	At risk
12	68.4	76.0	90.1	0.88 $\pm$ 0.36	38.3%	3.9%	57.8%	49.8%	3.6%	46.6%
15	75.8	82.4		1.25 $\pm$ 0.47	44.5%	11.3%	44.1%	56.9%	9.4%	33.7%
18	81.1	85.6		1.64 $\pm$ 0.58	49.3%	17.7%	33.0%	62.0%	14.1%	23.8%
21	84.7	87.6		2.03 $\pm$ 0.68	52.8%	23.2%	24.0%	65.6%	18.0%	16.4%
24	87.3	88.4		2.43 $\pm$ 0.78	55.3%	28.0%	16.7%	68.0%	21.1%	10.9%
27	88.8	89.4		2.81 $\pm$ 0.88	57.0%	32.0%	10.9%	69.5%	23.7%	6.7%
30	89.6	89.7		3.18 $\pm$ 0.98	58.1%	35.6%	6.4%	70.5%	25.8%	3.7%
33	90.2	90.0		3.54 $\pm$ 1.08	58.6%	38.6%	2.8%	70.9%	27.6%	1.5%
36	90.1	90.1		3.87 $\pm$ 1.19	58.8%	41.2%	0.0%	71.0%	29.0%	0.0%

時点 τ でのイベント発現，打ち切り及びリスク集合に含まれる被験者割合．

観察期間が時点 τ に到達せずに打ち切りとなった被験者を含める．

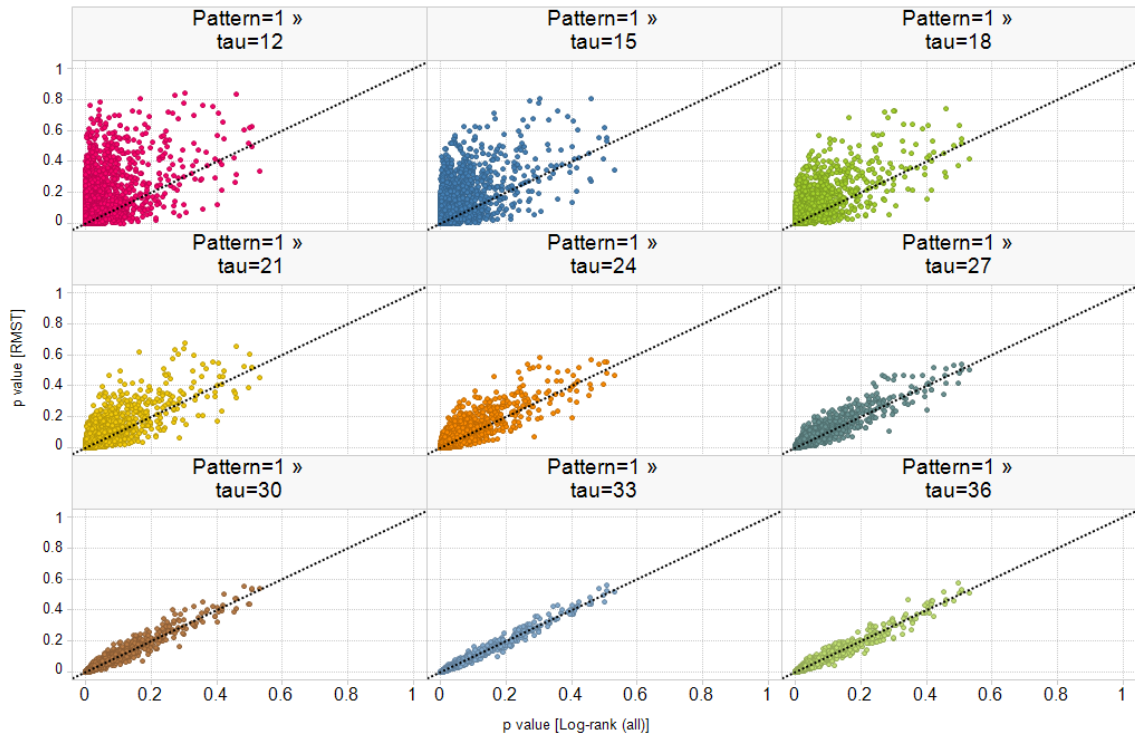


図 5-11 RMST の差を用いた検定と log-rank 検定の結果 (P 値) の散布図
(パターン 1 : 比例ハザード性が成立している場合)

5.3.2.2. 特定時点までは差がなく、それ以降に差が開く場合 (パターン 2)

抗がん剤の免疫チェックポイント阻害剤などに特徴的にみられる生存関数として、試験序盤では生存割合に大きな差は見られないがある特定時点から生存曲線が分離していくような場合を想定した。ここでは投与開始から 6 ヶ月までは両群に差はなく (6 ヶ月生存割合は共に 70%)、6 ヶ月以降は実薬群の効果が大きく発現するとした。

試験序盤で生存関数に差がないため、 $\tau = 24$ までは RMST の検出力は低く 80%を下回った。後半で生存曲線の差が広がるため、両手法とも検出力が境界時間 τ の増加と共に増大した。P 値を比較すると、 $\tau < 30$ までは RMST で log-rank 検定よりも P 値が大きく、境界時間 τ が大きくなるにつれて両手法の P 値の差は小さくなっていった。

表 5-7 RMST と log-rank 検定の性能評価

(パターン 2：特定時点までは差がなく、それ以降に差が開く場合)

τ	Power			Δ RMST (Mean $\pm$ se)	実薬群 [#]			対照薬群 [#]		
	RMST	log-rank [τ]	log-rank [all]		Event	Censor ^{##}	At risk	Event	Censor ^{##}	At risk
12	14.5	45.3	92.7	0.33 $\pm$ 0.38	40.1%	3.7%	56.1%	49.8%	3.6%	46.6%
15	29.3	66.7		0.68 $\pm$ 0.49	44.4%	11.0%	44.6%	56.9%	9.4%	33.7%
18	45.9	79.4		1.12 $\pm$ 0.60	47.7%	17.6%	34.7%	62.0%	14.1%	23.8%
21	62.6	85.7		1.61 $\pm$ 0.71	50.3%	23.5%	26.2%	65.6%	18.0%	16.4%
24	75.2	89.2		2.14 $\pm$ 0.81	52.2%	28.8%	19.0%	68.0%	21.1%	10.9%
27	83.4	91.1		2.69 $\pm$ 0.92	53.5%	33.5%	12.9%	69.5%	23.7%	6.7%
30	88.5	92.1		3.26 $\pm$ 1.02	54.4%	37.7%	7.8%	70.5%	25.8%	3.7%
33	91.8	92.6		3.82 $\pm$ 1.13	54.9%	41.6%	3.6%	70.9%	27.6%	1.5%
36	93.8	92.7		4.37 $\pm$ 1.24	55.0%	45.0%	0.0%	71.0%	29.0%	0.0%

時点 τ でのイベント発現，打ち切り及びリスク集合に含まれる被験者割合．

観察期間が時点 τ に到達せずに打ち切りとなった被験者を含める．

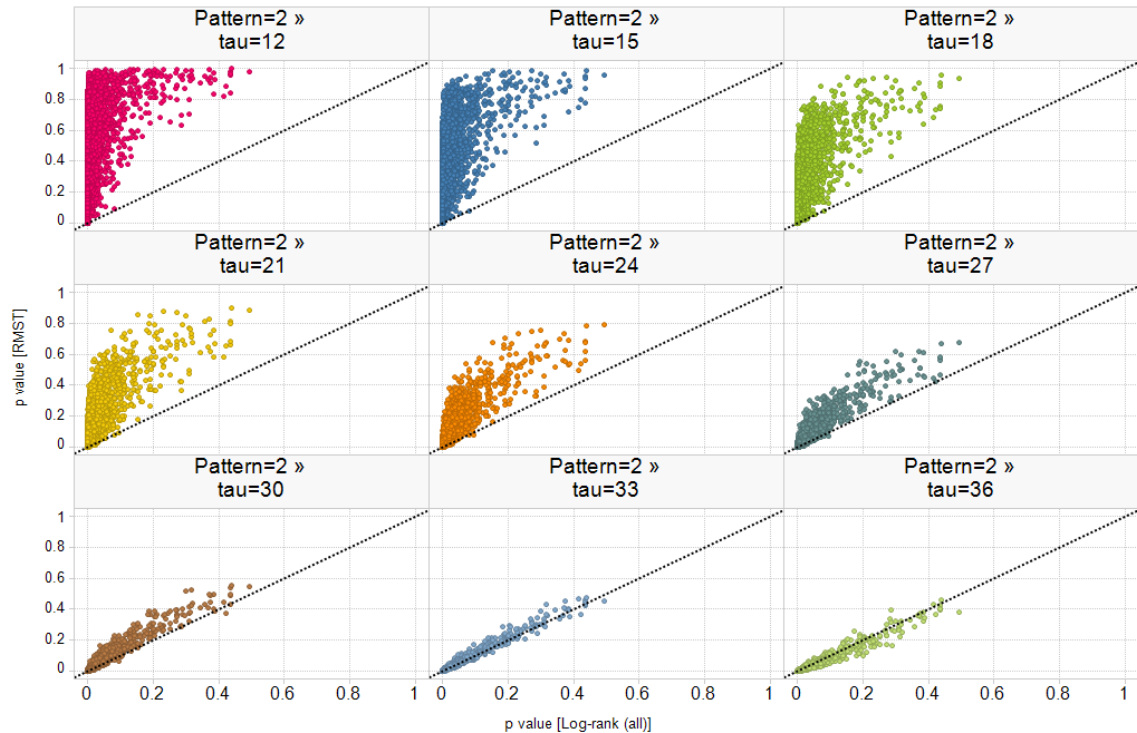


図 5-12 RMST の差を用いた検定と log-rank 検定の結果 (P 値) の散布図

(パターン 2：特定時点までは差がなく、それ以降に差が開く場合)

5.3.2.3. 試験期間を通じてイベントの発現数が少ない場合（比例ハザード性は成立）

生存時間解析では抗がん剤試験のように 1～3 年程度の追跡で生存割合が 50%を下回るものもあるが，対象疾患が変わればイベント発現率が非常に低い試験もある．パターン 3 ではイベント発現が少ないもの，3 年の追跡で生存割合が 90%（実薬群），85%（対照薬群）というケースを想定してシミュレーションを実施した．

時点 τ でイベントがほとんど発現していない状況では、RMST を用いても log-rank 検定を用いても検出力は低く、両者の結果に違いは認められなかった。傾向としてはパターン 1 で示した比例ハザード性が成立している場合と同様であり、境界時間 τ に関わらず log-rank 検定の検出力が高かった。P 値の比較では境界時間 τ が大きくなるほど両手法での値の差も小さくなった。

表 5-8 RMST と log-rank 検定の性能評価

(パターン 3：試験期間を通じてイベントの発現数が少ない場合)

τ	Power			Δ RMST (Mean $\pm$ se)	実薬群 [#]			対照薬群 [#]		
	RMST	log-rank [τ]	log-rank [all]		Event	Censor ^{##}	At risk	Event	Censor ^{##}	At risk
12	13.5	16.6	27.1	0.11 $\pm$ 0.13	3.4%	4.9%	91.7%	5.1%	4.9%	90.0%
15	16.0	19.1		0.17 $\pm$ 0.18	4.1%	17.3%	78.5%	6.3%	17.1%	76.7%
18	18.7	21.0		0.25 $\pm$ 0.23	4.7%	29.4%	65.9%	7.2%	28.8%	64.0%
21	21.1	23.3		0.33 $\pm$ 0.29	5.3%	41.0%	53.7%	8.0%	40.0%	52.0%
24	23.1	25.0		0.43 $\pm$ 0.36	5.7%	52.2%	42.1%	8.6%	50.9%	40.5%
27	24.5	25.8		0.54 $\pm$ 0.43	6.0%	63.1%	30.9%	9.1%	61.3%	29.6%
30	25.6	26.9		0.66 $\pm$ 0.51	6.2%	73.6%	20.2%	9.4%	71.4%	19.2%
33	26.7	27.1		0.80 $\pm$ 0.60	6.4%	83.7%	9.9%	9.6%	81.0%	9.4%
36	27.5	27.1		0.94 $\pm$ 0.69	6.4%	93.6%	0.0%	9.7%	90.3%	0.0%

時点 τ でのイベント発現，打ち切り及びリスク集合に含まれる被験者割合。

観察期間が時点 τ に到達せずに打ち切りとなった被験者を含める。

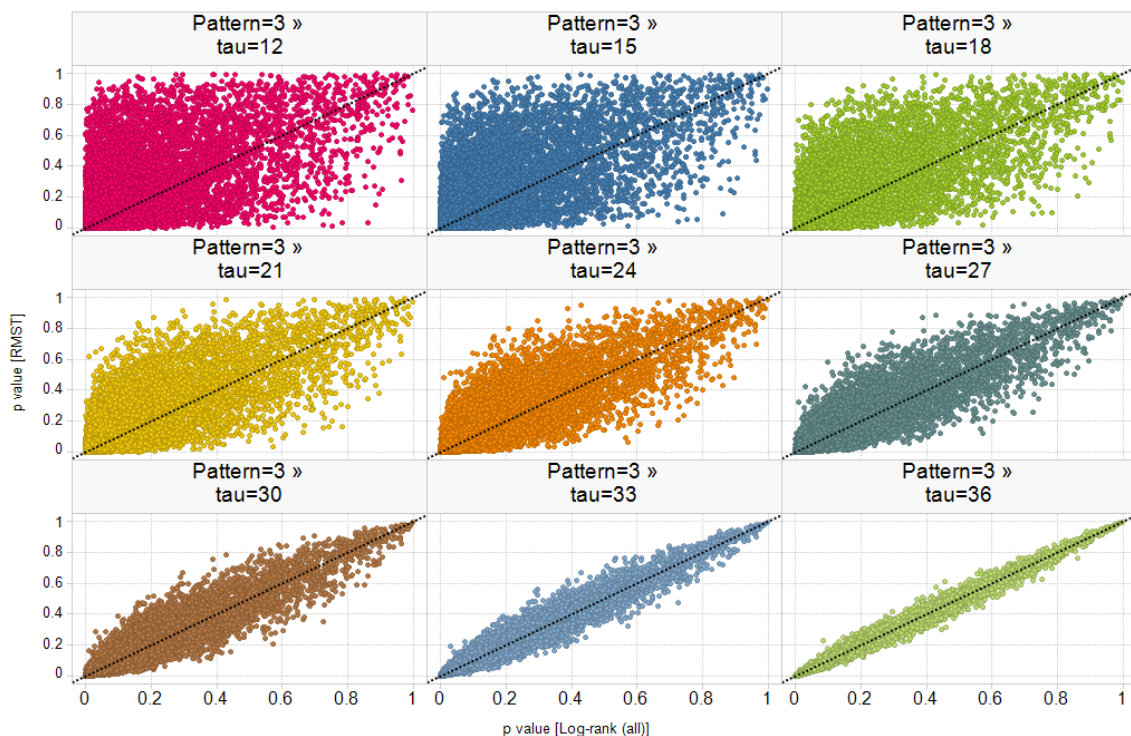


図 5-13 RMST の差を用いた検定と log-rank 検定の結果 (P 値) の散布図

(パターン 3：試験期間を通じてイベントの発現数が少ない場合)

5.3.2.4. 特定時点までは差が大きい、試験終了時には差がなくなる場合

臨床試験での追跡期間は、最後の被験者の登録 (Last patient in) からイベント発現を追跡する期間を示すことが多く、例えば、追跡期間を 12 ヶ月とした試験であっても、実際には 12 ヶ月以上観察される被験者がほとんどである。つまり、追跡期間が 12 ヶ月であっても登録期間と併せて 24 ヶ月以上観察されていることも少なくない。一般的に用いられる log-rank 検定では追跡期間終了時までには得られている全ての生存時間データを解析に用いるが、境界時間 τ を用いた RMST の差の検定では境界時間 τ 以降に観察されたイベントは結果に寄与しない。パターン 4 では追跡期間である投与 12 ヶ月までは生存割合に差があるが、12 ヶ月以降は徐々に差がなくなっていくという生存曲線を想定してシミュレーションを実施した。

投与開始から 9 ヶ月後まではハザード比が 0.3 と非常に大きな差があるため、生存割合にほとんど差がなくなる 24 ヶ月を境界時間 τ とした場合でも RMST では検出力が 95%を保っていた。一方で、通常の解析同様、解析時点で観察されているすべてのデータを用いた log-rank 検定では検出力が約 65%まで落ちている。さらに、境界時間 τ までの生存時間データを用いた log-rank [τ] の結果を見ると、投与開始から 9 ヶ月以降ではハザード比が 1 を超えることから境界時間 τ が大きくなるほど検出力大きく低下した。両手法の P 値の比較をすると、境界時間 τ に関わらず RMST の方が小さな P 値となっていることが散布図から分かる。また、境界時間 τ が大きくなるにつれて両手法とも P 値が大きくなる傾向にあるが、log-rank 検定の方が小さな境界時間 τ で大きな P 値となった。

表 5-9 RMST と log-rank 検定の性能評価

(パターン 4：特定時点までは差が大きい、試験終了時には差がなくなる場合)

τ	Power			Δ RMST (Mean $\pm$ se)	実薬群 [#]			対照薬群 [#]		
	RMST	log-rank [τ]	log-rank [all]		Event	Censor ^{##}	At risk	Event	Censor ^{##}	At risk
12	100.0	98.3	65.1	1.94 $\pm$ 0.32	35.3%	4.4%	60.4%	49.8%	3.6%	46.6%
15	100.0	83.3		2.24 $\pm$ 0.42	49.5%	11.5%	39.0%	56.9%	9.4%	33.7%
18	99.5	76.5		2.40 $\pm$ 0.52	56.0%	17.0%	27.1%	62.0%	14.2%	23.9%
21	97.9	72.1		2.52 $\pm$ 0.62	60.4%	21.3%	18.2%	65.6%	18.0%	16.4%
24	95.0	68.6		2.60 $\pm$ 0.72	63.4%	24.8%	11.8%	67.9%	21.2%	10.9%
27	90.6	66.7		2.65 $\pm$ 0.81	65.3%	27.6%	7.2%	69.5%	23.8%	6.7%
30	84.3	65.8		2.68 $\pm$ 0.90	66.4%	29.8%	3.9%	70.4%	25.9%	3.7%
33	77.7	65.1		2.69 $\pm$ 0.99	66.9%	31.5%	1.6%	70.9%	27.6%	1.5%
36	70.6	65.1		2.70 $\pm$ 1.08	67.0%	33.0%	0.0%	71.0%	29.0%	0.0%

時点 τ でのイベント発現、打ち切り及びリスク集合に含まれる被験者割合。

観察期間が時点 τ に到達せずに打ち切りとなった被験者を含める。

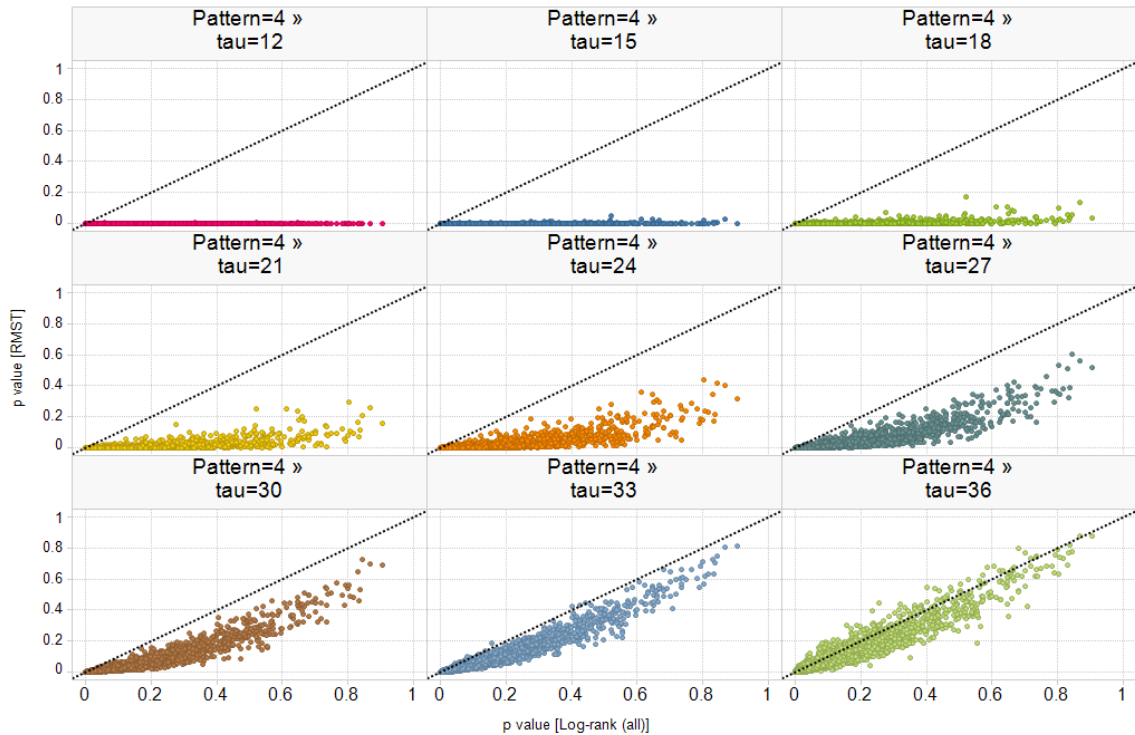


図 5-14 RMST の差を用いた検定と log-rank 検定の結果 (P 値) の散布図
(パターン 4 : 特定時点までは差が大きいが, 試験終了時には差がなくなる場合)

5.3.2.5. 生存曲線がクロスする場合

未知のバイオマーカーや予後因子などが存在する場合, 全体集団に対して Kaplan-Meier 曲線を作成すると曲線が明らかに交差するようなケースも多い. また, 対照薬群を化学療法としたがんワクチン試験でもクロスした生存曲線が報告されている^{[4][5]}. 実際の解析では部分集団解析を行う等の対応が必要になるが, 本シミュレーションでは生存曲線がクロスする場合に対して RMST の差を用いた検定がどのような結果になるかを検討した.

生存曲線がクロスするのが投与開始から 8.3 ヶ月であり, 試験序盤では対照薬群の方が高い生存割合を示している. RMST の群間差 (Δ RMST) から投与 15 ヶ月までは対照薬群の RMST の方が大きいという結果になっている. 総じて検出力は低い, RMST では境界時間 τ を 24 ヶ月以上と設定した場合に log-rank 検定の 2 倍以上の検出力となった. ただし, 生存曲線がクロスするような場合では得られた生存曲線から何を示すことで有効性の違い等を述べるかが重要となり, 例えば RMST で有意差が見られたとしても, より詳細な解析が必要となることには変わりがないと考えられる.

表 5-10 RMST と log-rank 検定の性能評価
(パターン5：生存曲線がクロスする場合)

τ	Power			Δ RMST (Mean $\pm$ se)	実薬群 [#]			対照薬群 [#]		
	RMST	log-rank [τ]	log-rank [all]		Event	Censor ^{##}	At risk	Event	Censor ^{##}	At risk
12	0.2	9.0	14.8	-0.33 $\pm$ 0.36	79.8%	2.1%	18.0%	90.1%	2.3%	7.6%
15	2.3	10.7		-0.01 $\pm$ 0.43	81.1%	4.5%	14.4%	91.3%	3.2%	5.5%
18	8.9	12.0		0.31 $\pm$ 0.49	82.2%	6.6%	11.2%	92.1%	4.0%	3.9%
21	19.2	13.1		0.62 $\pm$ 0.56	82.9%	8.5%	8.5%	92.6%	4.6%	2.7%
24	30.8	13.9		0.92 $\pm$ 0.62	83.5%	10.3%	6.2%	93.0%	5.2%	1.8%
27	42.2	14.4		1.21 $\pm$ 0.68	83.9%	11.8%	4.3%	93.3%	5.6%	1.1%
30	51.6	14.7		1.48 $\pm$ 0.74	84.2%	13.2%	2.6%	93.4%	5.9%	0.6%
33	58.7	14.8		1.74 $\pm$ 0.80	84.4%	14.5%	1.2%	93.5%	6.2%	0.3%
36	63.7	14.8		1.99 $\pm$ 0.86	84.4%	15.6%	0.0%	93.5%	6.5%	0.0%

時点 τ でのイベント発現，打ち切り及びリスク集合に含まれる被験者割合．

観察期間が時点 τ に到達せずに打ち切りとなった被験者を含める．

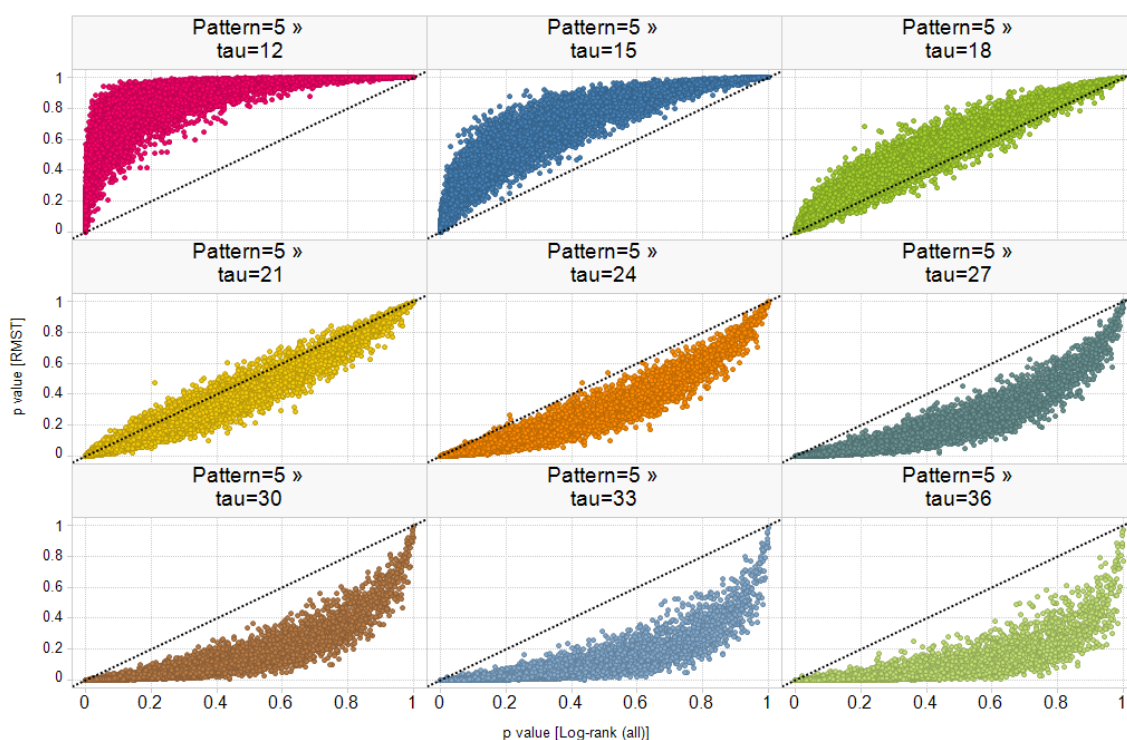


図 5-15 RMST の差を用いた検定と log-rank 検定の結果 (P 値) の散布図
(パターン5：生存曲線がクロスする場合)

5.3.3. シミュレーション検討のまとめ

シミュレーション検討として実際の臨床試験でありえる比例ハザード性が成立しないような生存関数を想定してRMSTの差による検定とlog-rank検定の違いや挙動を検出力及びP値を指標に検討した．シミュレーションによって生成された生存時間データが，比例ハザード性が成立している，もしくはそれに近いような形状を取った場合は総じてlog-rank検定

の検出力が高かった。一方で、試験序盤で実薬群の生存割合が対照薬群と比べて十分に高いが、観察期間後半では生存割合が同程度もしくは逆転するような場合では RMST と log-rank 検定で検出力が逆転する傾向にあった。RMST を用いた検定と log-rank 検定の大きな違いは境界時間 τ の導入である。試験開始前に試験薬の特徴を慎重に検討し、想定される生存曲線及び意味のある τ を定義することで、RMST を用いた検定による検出力の高い解析となる可能性があると考えられた。

参考文献

- [1] Trinquart L, Jacot J, Conner SC, Porcher R. Comparison of treatment effects measured by the hazard ratio and by the ratio of restricted mean survival times in oncology randomized controlled trials. *J Clin Oncol*. 2016; **34**:1813–19
- [2] Tian L, Fu H, Ruberg SJ, Uno H, Wei LJ. Efficiency of two sample tests via the restricted mean survival time for analyzing event time observations. *Biometrics*. 2017; doi:10.1111/biom.12770.
- [3] Rajkumar SV, Jacobus S, Callander NS, Fonseca R, Vesole DH, Williams ME, *et al*. Lenalidomide plus high-dose dexamethasone versus lenalidomide plus low-dose dexamethasone as initial therapy for newly diagnosed multiple myeloma: an open-label randomised controlled trial. *Lancet Oncol*. 2010; **11**(1):29–37.
- [4] Borghaei H, Paz-Ares L, Horn L, Spigel DR, Steins M, Ready NE, *et al*. Nivolumab versus docetaxel in advanced nonsquamous non–small-cell lung cancer. *N Engl J Med*. 2015; **373**:1627–39.
- [5] Bellmunt J, Wit RD, Vaughn DJ, Fradet Y, Lee JL, Fong L, *et al*. Pembrolizumab as second-line therapy for advanced urothelial carcinoma. *N Engl J Med*. 2017; **376**: 1015–1026.

6. 終わりに

本報告書では生存時間型応答の評価指標を整理し、近年注目されている RMST について理論的側面の紹介及びその特性について整理・評価した。RMST は比例ハザード性の仮定が成り立たない場合の代替指標の 1 つとして近年注目されており、各論文における検討で、従来の手法と性能が大幅に劣ることはないことが報告されている。RMST は境界時間 τ の設定が重要であり、境界時間 τ が臨床的観点から設定することができれば、臨床的に解釈できる評価指標として RMST を意義付けられると考えられた。RMST に限らず、臨床試験の結果を解釈する際には、解析結果の統計的解釈を臨床的解釈へ翻訳することが大切である。そのためには状況に応じて、適切に解釈できる評価指標を選択し、評価していくことが重要と考えられる。本報告書がその一助になれば幸いである。

なお、本報告書は RMST の利用を推奨したり、これまでの方法を否定するものではない。RMST を利用する際には、最新の論文などを確認し、これから実施しようとしている臨床試験に対して最適なものであるかを自ら確認していただきたい。また、本報告書では非劣性試験や中間解析を実施する試験で、RMST を評価する場合の検討はまだ行っていない。これらについては今後 RMST に関する知識及び報告が増えてきた場合の検討としたい。

臨床担当の方々へ

本報告書では生存時間型応答の評価指標として、主に近年注目されている RMST について理論的側面の紹介、及びその特性について整理・評価した。生存時間型応答の評価指標の選択、また RMST の実適用を考える際には、これまで以上に臨床担当者と統計担当者の連携が重要である。特に、臨床担当の方々へご理解・ご協力いただきたい点は以下のとおりである。

臨床試験の結果を解釈する際には、解析結果の統計的解釈を臨床的解釈へ翻訳することが大切である。特に、RMST を適用する際には、境界時間 τ の設定が重要である。臨床的に解釈できる評価指標として RMST を意義付けるためには、境界時間 τ を臨床的観点から設定することが望まれる。そのため、実際の臨床試験において RMST を適応する場合、RMST がどのような指標で、どのような性質を有しているのか臨床担当の方々にも理解していただくと共に、妥当な境界時間 τ 設定の議論へ関与いただくことが、実務を行う上で必要であると本タスクフォースでは考えている。そして、臨床試験で想定される状況に応じて、適切に解釈できる評価指標を選択し、評価していくことが重要と考えられる。本報告書がその一助になれば幸いである。

7. Appendix

7.1. 式の導出

本節では、本報告書 3.1 節及び 3.2.1 項に記載した $\mu(\tau)$, $E[X^2(\tau)]$ 及び $\text{Var}[\hat{\mu}(\tau)]$ の導出式を示す.

式(3.2) の導出は以下の通りである.

$$\begin{aligned}\mu(\tau) &= E[X(\tau)] = E[\min(T, \tau)] \\ &= \int_0^\tau t f(t) dt + \int_\tau^\infty \tau f(t) dt \\ &= [tF(t)]_0^\tau - \int_0^\tau F(t) dt + \tau[F(t)]_\tau^\infty \\ &= \tau F(\tau) - \int_0^\tau \{1 - S(t)\} dt + \tau\{1 - F(\tau)\} \\ &= -[t]_0^\tau + \int_0^\tau S(t) dt + \tau \\ &= -\tau + \int_0^\tau S(t) dt + \tau \\ &= \int_0^\tau S(t) dt\end{aligned}$$

式(3.3) の導出は以下の通りである.

$$\begin{aligned}E[X^2(\tau)] &= \int_0^\tau t^2 f(t) dt + \int_\tau^\infty \tau^2 f(t) dt \\ &= [t^2 F(t)]_0^\tau - \int_0^\tau 2tF(t) dt + \tau^2[F(t)]_\tau^\infty \\ &= \tau^2 F(\tau) - \int_0^\tau 2t(1 - S(t)) dt + \tau^2\{1 - F(\tau)\} \\ &= -[t^2]_0^\tau + 2 \int_0^\tau tS(t) dt + \tau^2 \\ &= -\tau^2 + 2 \int_0^\tau tS(t) dt + \tau^2 \\ &= 2 \int_0^\tau tS(t) dt\end{aligned}$$

式(3.14) の導出は以下の通りである.

$$\begin{aligned}
\text{Var}[\hat{\mu}(\tau)] &= \text{Var}\left[\int_0^\tau \hat{S}(t)dt\right] = \text{Var}\left[\sum_{j=0}^D (t_{j+1} - t_j) \hat{S}(t_j)\right] \\
&= (t_1 - t_0)^2 \text{Var}[\hat{S}(t_0)] + \sum_{j=1}^D (t_{j+1} - t_j)^2 \text{Var}[\hat{S}(t_j)] \\
&\quad + 2 \sum_{j < k} (t_{j+1} - t_j)(t_{k+1} - t_k) \text{Cov}[\hat{S}(t_j), \hat{S}(t_k)] \\
&= \sum_{j=1}^D (t_{j+1} - t_j)^2 \hat{S}^2(t_j) \sum_{l=1}^j \frac{d_l}{Y_l(Y_l - d_l)} \\
&\quad + 2 \sum_{j < k} (t_{j+1} - t_j)(t_{k+1} - t_k) \hat{S}(t_j) \hat{S}(t_k) \sum_{l=1}^j \frac{d_l}{Y_l(Y_l - d_l)} \\
&= \sum_{j=1}^D \left[\sum_{l=j}^D (t_{l+1} - t_l)^2 \hat{S}^2(t_l) \right. \\
&\quad \left. + 2 \sum_{j \leq l < k} (t_{j+1} - t_j)(t_{k+1} - t_k) \hat{S}(t_j) \hat{S}(t_k) \right] \frac{d_j}{Y_j(Y_j - d_j)} \\
&= \sum_{j=1}^D \left[\sum_{l=j}^D (t_{l+1} - t_l) \hat{S}(t_l) \right]^2 \frac{d_j}{Y_j(Y_j - d_j)} \\
&= \sum_{j=1}^D \left[\int_{t_j}^\tau \hat{S}(t)dt \right]^2 \frac{d_j}{Y_j(Y_j - d_j)} \\
&= \sum_{j=1}^D \left[\sum_{i=j}^D (t_{i+1} - t_i) \hat{S}(t_i) \right]^2 \frac{d_j}{Y_j(Y_j - d_j)}
\end{aligned}$$

7.2. SAS プログラム例

7.2.1. RMST 計算のための前処理

本報告書 3.2.2 項に記載した通り、境界時間 τ より後に発現したイベントがある場合、データセット中にある境界時間 τ 以降のイベントをすべて打ち切りに変換しておく必要がある。本報告書 3.2.2 項に示した例の場合の変換プログラム例を以下に示す。

```
%LET TAU = 3;  
  
DATA SAMPLE1;  
  SET SAMPLE;  
  IF TIME > &TAU. THEN CENSOR = 0;  
RUN;
```

7.2.2. RMST の差

本報告書 3.3.1 項に示した 2 群の RMST の差とその信頼区間及び P 値を求めるための SAS プログラム例を示す。使用するデータセット DATA の例は以下の通りである。

表 7-1 データセット DATA の例

GROUP	TIME	CENSOR
0	1	0
0	3	1
0	5	1
1	2	1
1	4	1
1	6	1

GROUP は群変数（実薬群=1, 対照薬群=0）、TIME は時間変数、CENSOR は打ち切り変数（0 は打ち切りを表す）である。TAU には境界時間を表す数値を指定する。ALPHA は両側有意水準であり、次の例では 0.05 としている。また、RMST の分散は式(3.14) を用いている。

データセット中に境界時間より後に発現したイベントが含まれている場合は、本報告書 7.2.1 項のプログラム例を参考にして、境界時間より後のイベントをすべて打ち切りに変換しておく必要がある。

```
%LET ALPHA = 0.05;      /* two-sided */  
  
PROC LIFETEST DATA = DATA TIMELIM = TAU;  
  TIME TIME * CENSOR(0);  
  STRATA GROUP;  
  ODS OUTPUT Means = _MEAN CensoredSummary = _EVENT;
```

```

RUN ;

PROC SORT DATA = _MEAN ;
  BY GROUP ;
RUN ;

PROC SORT DATA = _EVENT ;
  BY GROUP ;
RUN ;

DATA _OUT ;
  MERGE _MEAN(IN = A) _EVENT ;
  BY GROUP ;
  SE = StdErr * SQRT((Failed - 1) / Failed) ;
  IF A ;
RUN ;

PROC TRANSPOSE DATA = _OUT OUT = _OUT_M PREFIX = MEAN ;
  VAR Mean ;
  ID GROUP ;
RUN ;

PROC TRANSPOSE DATA = _OUT OUT = _OUT_S PREFIX = SE ;
  VAR SE ;
  ID GROUP ;
RUN ;

DATA RESULT ;
  MERGE _OUT_M _OUT_S ;
  DIF_M = MEAN1 - MEAN0 ;      /* RMST の差 */
  DIF_V = SE0**2 + SE1**2 ;    /* RMST の差の分散 */
  CI_L = DIF_M - PROBIT(1 - &ALPHA./2) * SQRT(DIF_V) ;
                                     /* RMST 差の信頼区間下限 */
  CI_U = DIF_M + PROBIT(1 - &ALPHA./2) * SQRT(DIF_V) ;
                                     /* RMST 差の信頼区間上限 */

  T = DIF_M / SQRT(DIF_V) ;      /* 検定統計量 */
  P = 2 * (1 - PROBNORM(ABS(T))) ; /* P 値 */
RUN ;

```

7.2.3. RMST の比

本報告書 3.3.2 項に示した 2 群の RMST の比とその信頼区間及び P 値を求めるための SAS プログラム例を示す。データセット及び変数の指定は、本報告書 7.2.2 項と同様である。RMST の分散は式(3.14) を用いている。

データセット中に境界時間より後に発現したイベントが含まれている場合は、本報告書 7.2.1 項のプログラム例を参考にして、境界時間より後のイベントをすべて打ち切りに変換しておく必要がある。

```

%LET ALPHA = 0.05 ;          /* two-sided */

PROC LIFETEST DATA = DATA TIMELIM = TAU ;
  TIME TIME * CENSOR(0) ;
  STRATA GROUP ;
  ODS OUTPUT Means = _MEAN CensoredSummary = _EVENT ;
RUN ;

PROC SORT DATA = _MEAN ;
  BY GROUP ;
RUN ;

PROC SORT DATA = _EVENT ;
  BY GROUP ;
RUN ;

DATA _OUT ;
  MERGE _MEAN(IN = A) _EVENT ;
  BY GROUP ;
  SE = StdErr * SQRT((Failed - 1) / Failed) ;
  IF A ;
RUN ;

PROC TRANSPOSE DATA = _OUT OUT = _OUT_M PREFIX = MEAN ;
  VAR Mean ;
  ID GROUP ;
RUN ;

PROC TRANSPOSE DATA = _OUT OUT = _OUT_S PREFIX = SE ;
  VAR SE ;
  ID GROUP ;
RUN ;

DATA RESULT ;
  MERGE _OUT_M _OUT_S ;
  R_M = MEAN1 / MEAN0 ;          /* RMST の比 */
  LOG_R_V = SE1**2 / MEAN1**2 + SE0**2 / MEAN0**2 ;
                                     /* 対数変換した RMST 比の分散 */
  CI_L = EXP(LOG(MEAN1 / MEAN0) - PROBIT(1 - &ALPHA./2) * SQRT(LOG_R_V)) ;
                                     /* RMST 比の信頼区間下限 */
  CI_U = EXP(LOG(MEAN1 / MEAN0) + PROBIT(1 - &ALPHA./2) * SQRT(LOG_R_V)) ;
                                     /* RMST 比の信頼区間上限 */
  T = LOG(R_M) / SQRT(LOG_R_V) ;    /* 検定統計量 */
  P = 2 * (1 - PROBNORM(ABS(T))) ;  /* P 値 */
RUN ;

```

7.2.4. 必要症例数の計算

3.5 節にて用いた, RMST の群間差を検出することを目的とした必要症例数を計算する SAS マクロを以下に示す.

時間に関連する引数 time, tau, enroll, follow 及び loss について, 時間の単位 (例: 年, 月, 週) は全て同一とする. 引数 P1, P0, time, loss 及び ratio に 2 以上の長さのベクトルを指定する場合, %str()関数を用いてコンマ区切りで入力する. 例えば, 実薬群及び対照薬群の症例数の割付比を 2 : 1 とする場合, 「ratio = %str(2, 1)」と入力する.

他, 引数に以下の入力規定を設けている.

- 引数 P1, P0 及び time は全て同じ長さのベクトルとする.
- 引数 P1 及び P0 は 0 より大きく 1 以下, かつ単調減少のベクトルとする.
- 引数 time は 0 以上, かつ単調増加, かつその最大値が tau 以上とする.
- 引数 loss は 0 以上 1 未満, かつ長さの 2 のベクトルとする.
- 引数 r は 0 より大とする.
- 引数 ratio は整数かつ, 長さの 2 のベクトルとする.
- 引数 sig_level 及び power は 0 より大きく 1 未満とする.

必要症例数の計算結果は, データセット RESULT に出力される.

```
%MACRO SS_RMST(P1, P0, time, tau, enroll, follow, loss=%str(0, 0), r=1, ratio=%str(1, 1),  
sig_level=0.05, power=0.8, seed=123, ss_sim=10000, iteration=50);  
  
%LET FLG1 = 1;  
%LET FLG2 = 1;  
  
DATA DATA1;  
  LENGTH X1 X2 X3 $500;  
  RETAIN P1_RETAIN P0_RETAIN TIME_RETAIN;  
  X1 = "&P1.";   
  X2 = "&P0.";   
  X3 = "&time.";   
  X4 = "&loss.";   
  X5 = "&ratio.";   
  LENGTH1 = LENGTH(X1) - LENGTH(COMPRESS(X1, ',')) + 1;  
  LENGTH2 = LENGTH(X2) - LENGTH(COMPRESS(X2, ',')) + 1;  
  LENGTH3 = LENGTH(X3) - LENGTH(COMPRESS(X3, ',')) + 1;  
  LENGTH4 = LENGTH(X4) - LENGTH(COMPRESS(X4, ',')) + 1;  
  LENGTH5 = LENGTH(X5) - LENGTH(COMPRESS(X5, ',')) + 1;  
  IF (LENGTH1 ^= LENGTH2 OR LENGTH1 ^= LENGTH3) THEN DO;  
    CALL SYMPUT('FLG1', '0');  
    PUT 'ERROR: 引数P1, P0, timeは全て同じ長さのベクトルである必要があります. ';  
    STOP;  
  END;  
  IF (LENGTH4 ^= 2) THEN DO;  
    CALL SYMPUT('FLG1', '0');  
    PUT 'ERROR: 引数lossは長さの2のベクトルである必要があります. ';  
    STOP;  
  END;  
END;
```

```

END ;
IF (LENGTH5 ^= 2) THEN DO ;
  CALL SYMPUT('FLG1', '0');
  PUT 'ERROR: 引数ratioは長さの2のベクトルである必要があります. ';
  STOP ;
END ;
DO i = 1 TO LENGTH1 ;
  P1 = INPUT(SCAN(X1, i, ','), BEST12.);
  P0 = INPUT(SCAN(X2, i, ','), BEST12.);
  TIME = INPUT(SCAN(X3, i, ','), BEST12.);
  TAU = INPUT("&tau.", BEST12.);
  ENROLL = INPUT("&enroll.", BEST12.);
  FOLLOW = INPUT("&follow.", BEST12.);
  LOSS1 = INPUT(SCAN(X4, 1, ','), BEST12.);
  LOSS0 = INPUT(SCAN(X4, 2, ','), BEST12.);
  R = INPUT(&r., BEST12.);
  RATIO1 = INPUT(SCAN(X5, 1, ','), BEST12.);
  RATIO0 = INPUT(SCAN(X5, 2, ','), BEST12.);
  SIG_LEVEL = INPUT(&sig_level., BEST12.);
  POWER = INPUT(&power., BEST12.);
  SEED = INPUT(&seed., BEST12.);
  SS_SIM = INPUT(&ss_sim., BEST12.);
  ITERATION = INPUT(&iteration., BEST12.);
  IF (i = 1 AND TIME ^= 0) THEN DO ;
    CALL SYMPUT('FLG2', '0');
  END ;
  IF (P1 = .) THEN DO ;
    CALL SYMPUT('FLG1', '0');
    PUT 'ERROR: 引数P1の入力にエラーがあります. ';
    STOP ;
  END ;
  IF (P0 = .) THEN DO ;
    CALL SYMPUT('FLG1', '0');
    PUT 'ERROR: 引数P0の入力にエラーがあります. ';
    STOP ;
  END ;
  IF (P1 > 1 OR P0 > 1 OR P1 <= 0 OR P0 <= 0) THEN DO ;
    CALL SYMPUT('FLG1', '0');
    PUT 'ERROR: 引数P1及びP0は0より大きく1以下である必要があります. ';
    STOP ;
  END ;
  IF (i >= 2 AND (P1 >= P1_RETAIN OR P0 >= P0_RETAIN)) THEN DO ;
    CALL SYMPUT('FLG1', '0');
    PUT 'ERROR: 引数P1及びP0は単調減少である必要があります. ';
    STOP ;
  END ;
  IF (TIME = .) THEN DO ;
    CALL SYMPUT('FLG1', '0');
    PUT 'ERROR: 引数timeの入力にエラーがあります. ';
    STOP ;
  END ;
  IF (TIME < 0) THEN DO ;
    CALL SYMPUT('FLG1', '0');

```

```

    PUT 'ERROR: 引数timeは0以上である必要があります. ';
    STOP;
END;
IF (i = LENGTH1 AND TIME < TAU) THEN DO;
    CALL SYMPUT('FLG1', '0');
    PUT 'ERROR: 引数timeの最大値はtau以上の大きい値である必要があります. ';
    STOP;
END;
IF (i >= 2 AND TIME <= TIME_RETAIN) THEN DO;
    CALL SYMPUT('FLG1', '0');
    PUT 'ERROR: 引数timeは単調増加である必要があります. ';
    STOP;
END;
IF (ENROLL = .) THEN DO;
    CALL SYMPUT('FLG1', '0');
    PUT 'ERROR: 引数enrollの入力にエラーがあります. ';
    STOP;
END;
IF (FOLLOW = .) THEN DO;
    CALL SYMPUT('FLG1', '0');
    PUT 'ERROR: 引数followの入力にエラーがあります. ';
    STOP;
END;
IF (LOSS1 < 0 OR LOSS0 < 0 OR LOSS1 >= 1 OR LOSS0 >= 1) THEN DO;
    CALL SYMPUT('FLG1', '0');
    PUT 'ERROR: 引数lossは0以上1未満である必要があります. ';
    STOP;
END;
IF (LOSS1 = . OR LOSS0 = .) THEN DO;
    CALL SYMPUT('FLG1', '0');
    PUT 'ERROR: 引数lossの入力にエラーがあります. ';
    STOP;
END;
RATIO_INT1 = SCAN(PUT(RATIO1, BEST12.), 2, '.');
RATIO_INT0 = SCAN(PUT(RATIO0, BEST12.), 2, '.');
IF (RATIO_INT1 ^= " OR RATIO_INT0 ^= ") THEN DO;
    CALL SYMPUT('FLG1', '0');
    PUT 'ERROR: 引数ratioは整数のベクトルである必要があります. ';
    STOP;
END;
IF (RATIO1 = . OR RATIO0 = .) THEN DO;
    CALL SYMPUT('FLG1', '0');
    PUT 'ERROR: 引数ratioの入力にエラーがあります. ';
    STOP;
END;
IF (R <= 0) THEN DO;
    CALL SYMPUT('FLG1', '0');
    PUT 'ERROR: 引数rは0より大である必要があります. ';
    STOP;
END;
IF (SIG_LEVEL <= 0 OR SIG_LEVEL >= 1) THEN DO;
    CALL SYMPUT('FLG1', '0');

```

```

        PUT 'ERROR: 引数sig_levelは0より大きく1未満である必要があります. ';
        STOP ;
    END ;
    IF (POWER <= 0 OR POWER >= 1) THEN DO ;
        CALL SYMPUT('FLG1', '0') ;
        PUT 'ERROR: 引数powerは0より大きく1未満である必要があります. ';
        STOP ;
    END ;
    IF (LENGTH1 > 1 AND TIME_RETAIN < TAU AND TAU < TIME) THEN DELETE ;
    TIME_RETAIN = TIME ;
    P1_RETAIN = P1 ;
    P0_RETAIN = P0 ;
    OUTPUT ;
END ;
CALL SYMPUTX('LOSS1', LOSS1) ;
CALL SYMPUTX('LOSS0', LOSS0) ;
CALL SYMPUTX('RATIO1', RATIO1) ;
CALL SYMPUTX('RATIO0', RATIO0) ;
DROP X1 X2 X3 X4 X5 LENGTH1 LENGTH2 LENGTH3 LENGTH4 LENGTH5
P1_RETAIN P0_RETAIN TIME_RETAIN RATIO_INT1 RATIO_INT0 ;
RUN ;

%IF &FLG1. = 1 %THEN %DO ;

%IF &FLG2. = 0 %THEN %DO ;

DATA TEMP1 ;
    P1 = 1 ;
    P0 = 1 ;
    TIME = 0 ;
    ENROLL = &enroll. ;
    FOLLOW = &follow. ;
    LOSS1 = &LOSS1. ;
    LOSS0 = &LOSS0. ;
    TAU = &tau. ;
    R = &r. ;
    RATIO1 = &RATIO1. ;
    RATIO0 = &RATIO0. ;
    SIG_LEVEL = &sig_level. ;
    POWER = &power. ;
    SEED = &seed. ;
    SS_SIM = &ss_sim. ;
    ITERATION = &iteration. ;
    i = 0 ;
RUN ;

DATA DATA1 ;
    SET TEMP1 DATA1 ;
RUN ;

%END ;

DATA TEMP2 ;
    SET DATA1 ;

```

```

i = i - 1 ;
TIME_UPPER = TIME ;
P1_UPPER = P1 ;
P0_UPPER = P0 ;
KEEP i TIME_UPPER P1_UPPER P0_UPPER ;
RUN ;

DATA DATA2 ;
MERGE DATA1(IN = IN1) TEMP2 ;
BY i ;
IF IN1 ;
P1_LOWER = P1 ;
P0_LOWER = P0 ;
TIME_LOWER = TIME ;
IF (TIME_LOWER >= TAU) THEN DELETE ;
HAZARD1 = -LOG(1 - (P1_LOWER - P1_UPPER) / P1_LOWER) / (TIME_UPPER -
TIME_LOWER) ;
HAZARD0 = -LOG(1 - (P0_LOWER - P0_UPPER) / P0_LOWER) / (TIME_UPPER -
TIME_LOWER) ;
RUN ;

DATA TEMP3 ;
SET DATA2 ;
RETAIN CUM_HAZARD1 CUM_HAZARD0 H1 H0 DELTA ;
IF TIME_LOWER = 0 THEN DO ;
CUM_HAZARD1 = 0 ;
CUM_HAZARD0 = 0 ;
H1 = 0 ;
H0 = 0 ;
DELTA = 0 ;
END ;
CUM_HAZARD1 = CUM_HAZARD1 + DELTA * H1 ;
CUM_HAZARD0 = CUM_HAZARD0 + DELTA * H0 ;
DELTA = TIME_UPPER - TIME_LOWER ;
H1 = HAZARD1 ;
H0 = HAZARD0 ;
KEEP TIME_LOWER TIME_UPPER P1_LOWER P1_UPPER P0_LOWER P0_UPPER
HAZARD1 HAZARD0 CUM_HAZARD1 CUM_HAZARD0 DELTA ;
RUN ;

DATA RESULT ;
MERGE DATA2(IN = IN1) TEMP3 ;
BY TIME_LOWER ;
LENGTH MEAN1 MEAN0 VAR1 VAR0 8. ;
RETAIN MEAN1 MEAN0 VAR1 VAR0 ;
IF IN1 ;
IF (TIME_UPPER > TAU) THEN DO ;
TIME_UPPER = TAU ;
DELTA = TIME_UPPER - TIME_LOWER ;
END ;
IF TIME_LOWER = 0 THEN DO ;
MEAN1 = 0 ;
MEAN0 = 0 ;
VAR1 = 0 ;

```



```

    VAR0 = 0 ;
END ;
MEAN1 = MEAN1 + EXP(-CUM_HAZARD1) * (1 - EXP(-HAZARD1 * DELTA)) /
HAZARD1 ;
MEAN0 = MEAN0 + EXP(-CUM_HAZARD0) * (1 - EXP(-HAZARD0 * DELTA)) /
HAZARD0 ;
VAR1 = VAR1 + 2 * EXP(-CUM_HAZARD1) * (((1 - EXP(-HAZARD1 * DELTA)) * (1 +
HAZARD1 * DELTA)) / (HAZARD1 ** 2)) + TIME_LOWER * ((1 - EXP(-HAZARD1 *
DELTA)) / HAZARD1)) ;
VAR0 = VAR0 + 2 * EXP(-CUM_HAZARD0) * (((1 - EXP(-HAZARD0 * DELTA)) * (1 +
HAZARD0 * DELTA)) / (HAZARD0 ** 2)) + TIME_LOWER * ((1 - EXP(-HAZARD0 *
DELTA)) / HAZARD0)) ;
IF TIME_UPPER = TAU THEN DO ;
    VAR1 = VAR1 - MEAN1 ** 2 ;
    VAR0 = VAR0 - MEAN0 ** 2 ;
    N = (1 + (&RATIO1. / &RATIO0.)) * (PROBIT(1 - SIG_LEVEL / 2) + PROBIT(POWER))
** 2 / ((MEAN1 - MEAN0) ** 2 / (VAR1 / (&RATIO1. / &RATIO0.) + VAR0)) ;
    N = i (&RATIO1. + &RATIO0.) * CEIL(N / (&RATIO1. + &RATIO0.)) ;
    INFO = (PROBIT(1 - SIG_LEVEL / 2) + PROBIT(POWER)) ** 2 / (MEAN1 -
MEAN0) ** 2 ;
    OUTPUT ;
END ;
DROP TIME TIME_LOWER TIME_UPPER P1 P1_LOWER P1_UPPER P0 P0_LOWER
P0_UPPER HAZARD1 HAZARD0 CUM_HAZARD1 CUM_HAZARD0 DELTA i ;
RUN ;

DATA RN1 RN0 ;
    CALL STREAMINIT(&seed.);
    DO m = 1 TO &iteration. ;
        DO i = 1 TO &ss_sim. ;
            RN_P = RAND('UNIFORM') ;
            RN_ENROLL = &enroll. * RAND('UNIFORM') ** (1 / &r.);
            %IF &LOSS1. > 0 %THEN %DO ;
                LOSS = RAND('EXPONENTIAL') / (-LOG(1 - &LOSS1.)) ;
            %END ;
            OUTPUT RN1 ;
            RN_P = RAND('UNIFORM') ;
            RN_ENROLL = &enroll. * RAND('UNIFORM') ** (1 / &r.);
            %IF &LOSS0. > 0 %THEN %DO ;
                LOSS = RAND('EXPONENTIAL') / (-LOG(1 - &LOSS0.)) ;
            %END ;
            OUTPUT RN0 ;
        END ;
    END ;
RUN ;

PROC SORT DATA = RN1 OUT = RN1 ;
    BY m i ;
RUN ;

PROC SORT DATA = RN0 OUT = RN0 ;
    BY m i ;
RUN ;

```

```

DATA TEMP4 ;
  SET TEMP3 ;
  IF (TIME_UPPER > &tau.) THEN DO ;
    TIME_UPPER = &tau. ;
    DELTA = TIME_UPPER - TIME_LOWER ;
  END ;
  DO m = 1 TO &iteration. ;
    DO i = 1 TO &ss_sim. ;
      OUTPUT ;
    END ;
  END ;
RUN ;

PROC SORT DATA = TEMP4 OUT = TEMP4 ;
  BY m i DESCENDING TIME_LOWER ;
RUN ;

DATA DATA3_1 ;
  MERGE RN1 TEMP4 ;
  BY m i ;
  IF RN_P < P1_UPPER THEN DO ;
    RN_TIME = TIME_UPPER ;
    RN_EVENT = 0 ;
  END ;
  ELSE IF (RN_P >= P1_UPPER AND RN_P < P1_LOWER) THEN DO ;
    RN_TIME = TIME_LOWER - LOG(RN_P / P1_LOWER) / HAZARD1 ;
    RN_EVENT = 1 ;
  END ;
RUN ;

DATA DATA3_1 ;
  SET DATA3_1 ;
  WHERE RN_TIME ^= . ;
  BY m i ;
  IF FIRST.i ;
    GROUP = 1 ;
    STUDY = &enroll. + &follow. - RN_ENROLL ;
  %IF &LOSS1. > 0 %THEN %DO ;
    IF RN_TIME > MIN(STUDY, &tau., LOSS) THEN DO ;
      RN_TIME = MIN(STUDY, &tau., LOSS) ;
      RN_EVENT = 0 ;
    END ;
  %END ;
  %ELSE %IF &LOSS1. = 0 %THEN %DO ;
    IF RN_TIME > MIN(STUDY, &tau.) THEN DO ;
      RN_TIME = MIN(STUDY, &tau.) ;
      RN_EVENT = 0 ;
    END ;
  %END ;
RUN ;

DATA DATA3_0 ;
  MERGE RN0 TEMP4 ;
  BY m i ;

```

```

IF RN_P < P0_UPPER THEN DO ;
    RN_TIME = TIME_UPPER ;
    RN_EVENT = 0 ;
END ;
ELSE IF (RN_P >= P0_UPPER AND RN_P < P0_LOWER) THEN DO ;
    RN_TIME = TIME_LOWER - LOG(RN_P / P0_LOWER) / HAZARD0 ;
    RN_EVENT = 1 ;
END ;
RUN ;

DATA DATA3_0 ;
    SET DATA3_0 ;
    WHERE RN_TIME ^= . ;
    BY m i ;
    IF FIRST.i ;
        GROUP = 0 ;
        STUDY = &enroll. + &follow. - RN_ENROLL ;
%IF &LOSS0. > 0 %THEN %DO ;
    IF RN_TIME > MIN(STUDY, &tau., LOSS) THEN DO ;
        RN_TIME = MIN(STUDY, &tau., LOSS) ;
        RN_EVENT = 0 ;
    END ;
%END ;
%ELSE %IF &LOSS0. = 0 %THEN %DO ;
    IF RN_TIME > MIN(STUDY, &tau.) THEN DO ;
        RN_TIME = MIN(STUDY, &tau.) ;
        RN_EVENT = 0 ;
    END ;
%END ;
RUN ;

DATA DATA3 ;
    SET DATA3_1 DATA3_0 ;
    KEEP m i RN_TIME RN_EVENT GROUP ;
RUN ;

PROC SORT DATA = DATA3 ;
    BY m ;
RUN ;

ODS EXCLUDE ALL ;

PROC LIFETEST DATA = DATA3 TIMELIM = &tau. ;
    TIME RN_TIME * RN_EVENT(0) ;
    STRATA GROUP ;
    BY m ;
    ODS OUTPUT Means = SE CensoredSummary = EVENT ;
RUN ;

DATA SE ;
    SET SE ;
    WHERE GROUP IN (1, 0) ;
    LABEL = 'SE' || PUT(GROUP, 1.) ;
RUN ;

```

```

PROC TRANSPOSE DATA = SE OUT = SE ;
  VAR StdErr ;
  BY m ;
  ID LABEL ;
RUN ;

DATA EVENT ;
  SET EVENT ;
  WHERE GROUP IN (1, 0) ;
  LABEL = 'EVENT' || PUT(GROUP, 1.) ;
RUN ;

PROC TRANSPOSE DATA = EVENT OUT = EVENT ;
  VAR Failed ;
  BY m ;
  ID LABEL ;
RUN ;

DATA TEMP5 ;
  MERGE SE(KEEP = m SE1 SE0) EVENT(KEEP = m EVENT1 EVENT0) ;
  BY m ;
RUN ;

DATA _NULL_ ;
  SET RESULT ;
  CALL SYMPUT('MEAN1', PUT(MEAN1, 8.5)) ;
  CALL SYMPUT('MEAN0', PUT(MEAN0, 8.5)) ;
  CALL SYMPUT('VAR1', PUT(VAR1, 8.5)) ;
  CALL SYMPUT('VAR0', PUT(VAR0, 8.5)) ;
RUN ;

DATA DATA4 ;
  SET TEMP5 ;
  PHI1 = SE1 * SQRT(&ss_sim. * ((EVENT1 - 1) / EVENT1) / &VAR1.) ;
  PHI0 = SE0 * SQRT(&ss_sim. * ((EVENT0 - 1) / EVENT0) / &VAR0.) ;
  VAR1_SIM = PHI1 ** 2 * &VAR1. ;
  VAR0_SIM = PHI0 ** 2 * &VAR0. ;
  N_SIM = (1 + (&RATIO1. / &RATIO0.)) * (PROBIT(1 - &sig_level. / 2) +
  PROBIT(&power.) ** 2 / ((&MEAN1. - &MEAN0.) ** 2 / (VAR1_SIM / (&RATIO1. /
  &RATIO0.) + VAR0_SIM))) ;
RUN ;

PROC MEANS DATA = DATA4 VARDEF = N MEAN VAR ;
  VAR N_SIM ;
  OUTPUT OUT = DATA5 N = N MEAN = MEAN VAR = VAR ;
RUN ;

DATA _NULL_ ;
  SET DATA5 ;
  CALL SYMPUT('MEAN_N', PUT(MEAN, 8.5)) ;
  CALL SYMPUT('SE_N', PUT(SQRT(VAR / N), 8.5)) ;
RUN ;

```

```

DATA RESULT ;
  LABEL P1 = 'Survival proportion in experimental group'
        P0 = 'Survival proportion in control group'
        TIME = 'Time'
        TAU = 'Restricted time'
        ENROLL = 'Enrollment period'
        FOLLOW = 'Follow-up period from enrollment of the last patient'
        LOSS1 = 'Proportion of lost follow-up in a unit of time in experimental group'
        LOSS0 = 'Proportion of lost follow-up in a unit of time in control group'
        R = 'Parameter of enrollment distribution'
        RATIO1 = 'Allocation ratio in experimental group'
        RATIO0 = 'Allocation ratio in control group'
        SIG_LEVEL = 'Significance level (two-sided)'
        POWER = 'Power'
        SEED = 'Seed of random number'
        SS_SIM = 'Sample size for each group in simulation'
        ITERATION = 'Number of iterations in simulation'
        MEAN1 = 'Restricted mean survival time in experimental group'
        MEAN0 = 'Restricted mean survival time in control group'
        VAR1 = 'Restricted variance of survival time in experimental group'
        VAR0 = 'Restricted variance of survival time in control group'
        N = 'Sample size without censoring'
        MEAN_N = 'Mean of sample size with censoring by simulation'
        SE_N = 'Standard error of sample size by simulation'
        INFO = 'Required information' ;
  SET RESULT ;
  P1 = PUT("&P1.", $100.);
  P0 = PUT("&P0.", $100.);
  TIME = PUT("&time.", $100.);
  MEAN_N = (&RATIO1. + &RATIO0.) * CEIL(&MEAN_N. / (&RATIO1. + &RATIO0.));
  SE_N = &SE_N.;
RUN ;

ODS EXCLUDE NONE ;

PROC PRINT DATA = RESULT LABEL ;
RUN ;

PROC DATASETS LIB = WORK NOLIST ;
  SAVE RESULT ;
RUN ;

QUIT ;

%END ;

%MEND SS_RMST ;

```

7.2.5. 試験実施中に測定する情報量の計算

RMST の群間差を評価する試験において、計画時に算出する必要情報量は、式(4.6) で求められる。これは、7.2.4 項で示した必要症例数を計算する SAS マクロの中で計算することが出来る。一方で、試験実施中に測定する情報量は、式(4.7) で求められる。本項では、こ

れを計算する SAS マクロ OINF_RMST を示す。本マクロでは、7 個の引数が用意され、それぞれで指定する内容を表 7-2 に示した。また、試験実施中に測定する情報量の計算結果は、データセット RESULT に出力される。

表 7-2 SAS マクロ OINF_RMST の引数

引数	内容
dsn	情報量算出対象の SAS データセット名
time	イベント発現、または打ち切りまでの時間を表す変数名
status	イベント発現を 1、打ち切りを 0 とする指示変数名
tau	与える境界時間 τ の値
group	群を表す変数名 (各被験者の群が既知の場合に指定)
gw1	実薬群への割付け比 (各被験者の群が未知の場合に指定)
gw0	対照薬群への割付け比 (各被験者の群が未知の場合に指定)

```
%MACRO OINF_RMST(dsn, time, status, tau, group, gw1 = ., gw0 = .);

DATA DATA1;
  LENGTH X1 $500.;

  X1 = "&group.";
  GW1 = INPUT("&gw1.", BEST12.);
  GW0 = INPUT("&gw0.", BEST12.);

  IF LENGTHN(X1) > 0 AND (GW1 = . OR GW0 = .) THEN DO;
    CALL SYMPUT('FLG1', '1');
    PUT 'NOTE: Information for RMST difference is calculated based on the observed number
of participants.';
  END;
  ELSE DO;
    IF LENGTHN(X1) = 0 AND GW1 ^= . AND GW0 ^= . THEN DO;
      CALL SYMPUT('FLG1', '2');
      PUT 'NOTE: Information for RMST difference is calculated based on the randomization
ratio.';
    END;
    ELSE DO;
      CALL SYMPUT('FLG1', '0');
      PUT 'ERROR: Input parameters are not acceptable.';
    END;
  END;
  RUN;

%IF &FLG1. ^= 0 %THEN %DO;

DATA _TEMP1;
  SET &dsn.;

  IF &time. > &tau. THEN &status. = 0;
```

```

RUN ;

PROC LIFETEST DATA = _TEMP1 TIMELIM = &tau. ;
  TIME &time. * &status.(0) ;
  ODS OUTPUT Means = _MEAN CensoredSummary = _EVENT ;
RUN ;

%IF &FLG1. = 1 %THEN %DO ;

PROC MEANS DATA = &dsn. NOPRINT NWAY ;
  CLASS &group. ;
  VAR &status. ;
  OUTPUT OUT = _N1 N = N ;
RUN ;

PROC TRANSPOSE DATA = _N1 OUT = _N2 PREFIX = GN ;
  VAR N ;
  ID &group. ;
RUN ;

%END ;

%IF &FLG1. = 2 %THEN %DO ;

DATA _N2 ;
  SET DATA1;

  GN1 = GW1;
  GN0 = GW0;
RUN;

%END ;

DATA RESULT ;
  MERGE _MEAN (IN = A) _EVENT _N2 ;

  W0 = GN0 / (GN0 + GN1) ;
  W1 = GN1 / (GN0 + GN1) ;
  SE = StdErr * SQRT((Failed - 1) / Failed) ;

  OINF = (1 / W0 / W1 * SE ** 2) ** -1 ;

  IF A ;

  KEEP Mean SE Total Failed OINF ;

  LABEL OINF = 'Observed information' ;
RUN ;

PROC PRINT DATA = RESULT LABEL NOOBS ;
  VAR Total Failed Mean SE OINF ;
RUN ;

%END ;

```

```
%MEND OINF_RMST;
```

SAS 9.4 の SASHELP ライブラリに含まれている SAS データセット BMT を用い、SAS マクロ OINF_RMST の適用例を示す。各被験者の群が既知の場合、例えば二重盲検下試験で中間解析を実施する際には、引数 dsn で指定した SAS データセットの中にある群を表す変数名を引数 group で指定する。この際、各群への割付け比を指定する引数 gw1, gw0 には指定を行わないことに留意する。この例では、情報量が 0.219 と算出された。また、各被験者の群が未知の場合、例えば二重盲検下試験で中間解析や最終解析の実施時期を検討する際には、引数 group には指定を行わず、各群への割付け比を引数 gw1, gw0 で指定する。等割付けとした場合の例では、情報量が 0.221 と算出された。

```
DATA SAMPLE ;
SET SASHELP.BMT ;

TIME = T / 365.25 * 4 ;

SELECT (Group) ;
  WHEN ('AML-Low Risk') GNO = 0 ;
  WHEN ('AML-High Risk') GNO = 1 ;
  OTHERWISE GNO = . ;
END ;

IF GNO IN (0, 1) THEN OUTPUT ;
RUN ;

/* Group information is known */
%OINF_RMST(dsn = SAMPLE, time = TIME, status = Status, tau = 24, group = GNO) ;

/* Group information is unknown */
%OINF_RMST(dsn = SAMPLE, time = TIME, status = Status, tau = 24, gw1 = 1, gw0 = 1) ;

/* Error */
%OINF_RMST(dsn = SAMPLE, time = TIME, status = Status, tau = 24) ;

%OINF_RMST(dsn = SAMPLE, time = TIME, status = Status, tau = 24, group = GNO, gw1 = 1, gw0 = 1) ;
```

7.2.6. 生存時間データ再構築の SAS マクロ

4.3.2 項で用いた、個々の被験者の生存時間データを再構築するマクロを以下に示す。4.3.2 項で説明した、DigitizeIt で読み込んだ生存時間のファイル及びリスクセットのファイルを CSV の形式で準備しておくことが必要である。

```
/* ----- */
/* constructing survival data from Kaplan-Meier estimates */
/* ----- */
```



```

/* Make sure if survival probability is non-increasing with time.
Macro parameters
survivalprobability: csv file for survival probability extracted by degitizeit
risksetsize:csv file for riskset size
totaleventnumber:number of total event, if unknown please input . (missing)
treatcode:treatment code number (numeric)
outdataset: output SAS dataset name */
/* ----- */
%macro reconst(survivalprobability,risksetsize,totaleventnumber,treatcode,outdataset);
  filename fil1 "&survivalprobability..csv";
  filename fil2 "&risksetsize..csv";
  data d0;
    totevents=&totaleventnumber; /* if unknown, set . */
    armid=&treatcode;
  run;
  options nocenter;
  proc import out=d1 datafile=fil1 dbms=csv replace;
    getnames=no;
    datarow=1;
    guessingrows=500;
  run;
  proc sort data=d1 out=d1;by var1 var2 descending var3;run;
  data d2(drop=var1 var2 var3);
    set d1;
    id=var1;
    ts=var2;
    s=var3;
  run;
  proc import out=d3 datafile=fil2 dbms=csv replace;
    getnames=no;
    datarow=1;
    guessingrows=10;
  run;
  data d4(drop=var2-var5);
    set d3;
    trisk=var2;
    lower=var3;
    upper=var4;
    nrisk=var5;
  run;
  proc iml;
    /* 条件設定、初期値設定等 */
    use d2;
    read all into digizeit;
    ts=digizeit[,2];
    s=digizeit[,3];
    use d4;
    read all into pubrisk;
    trisk=pubrisk[,2];
    lower=pubrisk[,3];
    upper=pubrisk[,4];
    nrisk=pubrisk[,5];
    nint=nrow(nrisk);
    nt=upper[nint];

```

```

use d0;
  read all into size;
  arm=shape(size[2],1,nrisk[1]);
  totevents=size[1];
  ncensor=shape(0,nint-1,1);
  nhathat=shape(nrisk[1]+1,nt,1);
  cen=shape(0,nt,1);
  d=shape(0,nt,1);
  kmhat=shape(1,nt,1);
  lasti=shape(1,nint,1);
  sumdl=0;
  /* 各時点のイベント数 cen, 打ち切り数 d を計算（最後の区間を除く） */
  if (nint>1) then do;
    do i=1 to nint-1;
      ncensor[i]=round(nrisk[i]#s[lower[i+1]]/s[lower[i]]-nrisk[i+1]);
    do
  while((nhathat[lower[i+1]]>nrisk[i+1])&((nhathat[lower[i+1]]<nrisk[i+1])&(ncensor[i]>0)));
    if (ncensor[i]<=0) then do;
      cen[(lower[i]):(upper[i])]=0;
      ncensor[i]=0;
    end;
    if (ncensor[i]>0) then do;
      cent=shape(0,ncensor[i],1);
      do j=1 to ncensor[i];
        cent[j]=ts[lower[i]]+j#(ts[lower[i+1]]-ts[lower[i]])/(ncensor[i]+1);
      end;
      cut=ts[(lower[i]):(lower[i+1])];
      do q=1 to nrow(cut)-1;
        if (cut[q]=cut[q+1]) then cut[q+1]=cut[q+1]+0.00000001;
      end;
      b=bin(cen,cut);
      count=shape(0,nrow(cut),1);
      do q=1 to nrow(cut);
        coun=element(b,q);
        count[q]=sum(coun);
      end;
      do q=lower[i] to upper[i];
        cen[q]=count[q-lower[i]+1];
      end;
    end;
    nhathat[lower[i]]=nrisk[i];
    lasti=lasti[i];
    do k=lower[i] to upper[i];
      if ((i=1)&(k=lower[i])) then do;
        d[k]=0;
        kmhat[k]=1;
      end;
      else do;
        d[k]=round(nhathat[k]#(1-(s[k]/kmhat[last])));
        kmhat[k]=kmhat[last]#(1-(d[k]/nhathat[k]));
      end;
      nhathat[k+1]=nhathat[k]-d[k]-cen[k];
      if (d[k] ^= 0) then last=k;
    end;
  end;

```

```

        ncensor[i]=ncensor[i]+(nhat[lower[i+1]]-nrisk[i+1]);
    end;
    if (nhat[lower[i+1]]<nrisk[i+1]) then nrisk[i+1]=nhat[lower[i+1]];
    lasti[i+1]=last;
end;
end;
/* 各時点のイベント数 d, 打ち切り数 cen を計算（最後の区間）*/
if (nint>1) then do;
    ncensorn=min(round((sum(ncensor[1:(nint-1)]))#(ts[upper[nint]]
        -ts[lower[nint]])/(ts[upper[(nint-1)]]-ts[lower[1]])),nrisk[nint]);
    ncensor=ncensor//ncensorn;
end;
if (nint=1) then do;
    ncensorn=0;
    ncensor=ncensor//ncensorn;
end;
if (ncensor[nint]<=0) then do;
    cen[lower[nint]:(upper[nint]-1)]=0;
    ncensorn=0;
    ncensor=ncensor//ncensorn;
end;
if (ncensor[nint]>0) then do;
    cent=shape(0,ncensor[nint],1);
    do j=1 to ncensor[nint];
        cent[j]=ts[lower[nint]]+j#(ts[upper[nint]]-ts[lower[nint]])/(ncensor[nint]+1);
    end;
    cut=ts[(lower[nint]):(upper[nint])];
    do q=1 to nrow(cut)-1;
        if (cut[q]=cut[q+1]) then cut[q+1]=cut[q+1]+0.00000001;
    end;
    b=bin(cen,cut);
    count=shape(0,nrow(cut),1);
    do q=1 to nrow(cut);
        coun=element(b,q);
        count[q]=sum(coun);
    end;
    do q=lower[nint] to (upper[nint]-1);
        cen[q]=count[q-lower[i]+1];
    end;
end;
nhat[lower[nint]]=nrisk[nint];
last=lasti[nint];
do k=lower[nint] to upper[nint];
    if (kmhat[last] ^= 0) then d[k]=round(nhat[k]#(1-(s[k]/kmhat[last])));
    else d[k]=0;
    kmhat[k]=kmhat[last]#(1-(d[k]/nhat[k]));
    if (k ^= upper[nint]) then nhat[k+1]=nhat[k]-d[k]-cen[k];
    else do;
        nhatn=nhat[k]-d[k]-cen[k];
        nhat=nhat//nhatn;
    end;
    if (nhat[k+1]<0) then do;
        nhat[k+1]=0;
        cen[k]=nhat[k]-d[k];
    end;
end;

```

```

end;
if (d[k] ^= 0) then last=k;
end;
/* あらかじめ指定したイベント数をもとに調整 */
if (totevents ^= .) then do;
if (nint>1) then sumdl=sum(d[1:upper[(nint-1)]]);
if (sumdl>= totevents) then do;
d[lower[nint]:upper[nint]]=shape(0,(upper[nint]-lower[nint]+1),1);
cen[lower[nint]:(upper[nint]-1)]=shape(0,(upper[nint]-lower[nint]),1);
nhat[(lower[nint]+1):(upper[nint]+1)]=shape(nrisk[nint],(upper[nint]+1-lower[nint]),1);
end;
end;
if ((sumdl<totevents)|(nint=1)) then do;
sumd=sum(d[1:upper[nint]]);
do while((sumd>totevents)|((sumd<totevents)&(ncensor[nint]>0)));
ncensor[nint]=ncensor[nint]+(sumd-totevents);
if (ncensor[nint]<=0) then do;
cen[lower[nint]:(upper[nint]-1)]=0;
ncensor[nint]=0;
end;
if (ncensor[nint]>0) then do;
cent=shape(0,ncensor[nint],1);
do j=1 to ncensor[nint];
cent[j]=ts[lower[nint]]+j*(ts[upper[nint]]-ts[lower[nint]])/(ncensor[nint]+1);
end;
cut=ts[(lower[nint]):(upper[nint])];
do q=1 to nrow(cut)-1;
if (cut[q]=cut[q+1]) then cut[q+1]=cut[q+1]+0.00000001;
end;
b=bin(cent,cut);
count=shape(0,nrow(cut),1);
do q=1 to nrow(cut);
coun=element(b,q);
count[q]=sum(coun);
end;
do q=lower[nint] to (upper[nint]-1);
cen[q]=count[q-lower[i]+1];
end;
nhat[lower[nint]]=nrisk[nint];
last=lasti[nint];
do k=lower[nint] to upper[nint];
d[k]=round(nhat[k]#(1-(s[k]/kmhat[last])));
kmhat[k]=kmhat[last]#(1-(d[k]/nhat[k]));
if (k ^= upper[nint]) then do;
nhat[k+1]=nhat[k]-d[k]-cen[k];
if (nhat[k+1]<0) then do;
nhat[k+1]=0;
cen[k]=nhat[k]-d[k];
end;
end;
if (d[k] ^= 0) then last=k;
end;
sumd=sum(d[1:upper[nint]]);
end;

```

```

        end;
    end;
    survprob=shape(.,nrow(ts),4);
    survprob[1]=ts;
    survprob[2]=nhat[1:nt];
    survprob[3]=d;
    survprob[4]=cen;
    print survprob;
    tipd=shape(ts[nt],nrisk[1],1);
    eventipd=shape(0,nrisk[1],1);
    h=1;
    do j=1 to nt;
        if (d[j] ^= 0) then do;
            do u=h to (h+d[j]-1);
                tipd[u]=ts[j];
            end;
            do u=h to (h+d[j]-1);
                eventipd[u]=1;
            end;
            h=h+d[j];
        end;
    end;
    ipd=shape(.,nrow(tipd),3);
    ipd[1]=tipd;
    ipd[2]=eventipd;
    ipd[3]=arm`;
    do j=1 to nt-1;
        if (cen[j] ^=0) then do;
            do u=h to (h+cen[j]-1);
                tipd[u]=(ts[j]+ts[j+1])/2;
            end;
            do u=h to (h+cen[j]-1);
                eventipd[u]=0;
            end;
            h=h+cen[j];
        end;
    end;
    ipd[1]=tipd;
    ipd[2]=eventipd;
    ipd[3]=arm`;
    print ipd;
    create idata0 from ipd;
    append from ipd;
    close idata0;
run;quit;
proc sort data=idata0 out=idata0;by col1 col2 col3;run;
data &outdataset(drop=col1-col3);
    set idata0;
    time=col1;event=col2;group=col3;
run;
%mend reconst;

/* example ----- */
%reconst(c:¥nivo1,c:¥nivorisk1,50,1,kaplan);

```

```
proc print data=kaplan;run;
proc lifetest data=kaplan plots=(s);
  time time*event(0);
run;
```

7.3. 実務上のノウハウ

本節では、RMST を使用する際の実務上のノウハウをまとめた。まず、SAS/STAT(R) 9.2 User's Guide, Second Edition の注意点を紹介する。次に、症例数設計に仮定する指数分布のパラメータを Newton-Raphson 法により計算するための SAS/IML 及び R のコードを示す。

7.3.1. SAS/STAT(R) 9.2 User's Guide, Second Edition の注意点

SAS/STAT(R) 9.2 User's Guide, Second Edition の LIFETEST procedure の Details: Computational Formulas には、RMST の推定量 $\hat{\mu}$ の推定量を導出する式が以下の通り記載されている（部分抜粋）が、 $\hat{\sigma}(\hat{\mu})$ の算出式は誤記である。

Let $t_1 < t_2 < \dots < t_D$ represent the distinct event times. For each $i = 1, \dots, D$, let n_i be the number of surviving units, the size of the risk set, just prior to t_i . Let d_i be the number of units that fail at t_i , and let $s_i = n_i - d_i$.

The standard error of $\hat{\mu}$ is estimated as

$$\hat{\sigma}(\hat{\mu}) = \sqrt{\frac{m}{m-1} \sum_{i=1}^{D-1} \frac{A_i^2}{n_i s_i}}$$

where

$$A_i = \sum_{j=i}^{D-1} \hat{S}(t_j)(t_{j+1} - t_j)$$

$$m = \sum_{j=1}^D d_j$$

If the largest observed time is not an event, you can use the TIMELIM= option to specify a time limit L and estimate the mean survival time limited to the time L and its standard error by replacing k by $k+1$ with $t_{k+1} = L$.

正確な導出式は以下の通りであり、SAS は以下の式によって算出される結果を出力している。

$$\hat{\sigma}(\hat{\mu}) = \sqrt{\frac{m}{m-1} \sum_{i=1}^{D-1} \frac{d_i A_i^2}{n_i S_i}}$$

また、部分抜粋した最後の段落にある TIMELIM= option に関する記載中の k を D に置き換えて読む必要がある。この記載は、SAS/STAT 14.2 から修正されている。

7.3.2. Newton-Raphson 法による指数パラメータの計算

3.5.3 項及び 3.5.4 項にて、症例数設計に仮定する指数分布のパラメータを計算するため、Newton-Raphson 法を用いた。本項では、この Newton-Raphson 法を実行するための SAS/IML 及び R のコードを示す。

3.5.3 項では生存時間に指数分布を想定した症例数設計の方法を述べた。式(3.5) の指数パラメータ λ を解くために、方程式について右辺をイコールゼロとした以下の式をコード中に定義する。

$$\frac{1 - \exp(-\lambda\tau)}{\lambda} - \mu(\tau) = 0 \quad (7.1)$$

SAS/IML では以下のコードで求めることが出来る。NLPNRA ルーチンは Newton-Raphson 法を用いて関数が最小となる値を求めるため、式(7.1) の左辺の 2 乗を関数として定義する。実行結果の XRES が、目的の指数パラメータである。

- 実薬群の指数パラメータを求める SAS/IML コードと実行結果

```
PROC IML ;
  START FUNC(X) ;
    TAU = 24 ;
    RMST = 14.1 ;
    F = ((1 - EXP(-X * TAU)) / X - RMST) ** 2 ;
    RETURN(F) ;
  FINISH FUNC ;
  X0 = {0.1} ;
  CALL NLPNRA(RC, XRES, "FUNC", X0) ;
  F = FUNC(XRES) ;
  PRINT XRES, F ;
QUIT ;
```

XRES

0.049088

F

6.78E-13

- 対照薬群の指数ハザードを求める SAS/IML コードと実行結果

```
PROC IML ;
  START FUNC(X);
    TAU = 24;
    RMST = 11.1;
    F = ((1 - EXP(-X * TAU)) / X - RMST) ** 2;
    RETURN(F);
  FINISH FUNC;
  X0 = {0.1};
  CALL NLPNRA(RC, XRES, "FUNC", X0);
  F = FUNC(XRES);
  PRINT XRES, F;
QUIT;
```

XRES

0.075308

F

3.636E-13

R では uniroot 関数を用いて以下のコードで求めることが出来る．計算した結果オブジェクトの \$root が，目的の指数パラメータである．

- 実薬群の指数パラメータを求める R コードと実行結果

```
f <- function(x) (1 - exp(-x * 24)) / x - 14.1
uniroot(f, c(0.01, 1), tol = 1e-10)$root
```

```
> f <- function(x) (1 - exp(-x * 24)) / x - 14.1
> uniroot(f, c(0.01, 1), tol = 1e-10) $root
[1] 0.04908797
```

- 対照薬群の指数ハザードを求める R コードと実行結果

```
f <- function(x) (1 - exp(-x * 24)) / x - 11.1
uniroot(f, c(0.01, 1), tol = 1e-10)$root
```

```
> f <- function(x) (1 - exp(-x * 24)) / x - 11.1
> uniroot(f, c(0.01, 1), tol = 1e-10) $root
[1] 0.07530796
```

同様に，3.5.4 項にて示した区分指数分布を仮定した実薬群の指数パラメータを計算するための SAS/IML 及び R のコードを示す．

- 実薬群の区分指数パラメータを求める SAS/IML コードと実行結果

```
PROC IML ;
  START FUNC(X);
    TAU = 24;
    RMST = 14.1;
    H = 0.07530796;
    F = (((1 - EXP(-H * 3)) / H) + EXP(-H * 3) * (1 - EXP(-X * (TAU - 3))) / X - RMST) ** 2;
    RETURN(F);
  FINISH FUNC;
  X0 = {0.1};
  CALL NLPNRA(RC, XRES, "FUNC", X0);
  F = FUNC(XRES);
  PRINT XRES, F;
QUIT;
```

XRES

0.0392188

F

9.714E-14

- 実薬群の区分指数パラメータを求める R コードと実行結果

```
f <- function(x) {
  h <- 0.07530796
  ((1 - exp(-h * 3)) / h) + exp(-h * 3) * (1 - exp(-x * 21)) / x - 14.1
}
uniroot(f, c(0.01, 1), tol = 1e-10)$root
```

```
> f <- function(x) {
+   h <- 0.07530796
+   ((1 - exp(-h * 3)) / h) + exp(-h * 3) * (1 - exp(-x * 21)) / x - 14.1
+ }
> uniroot(f, c(0.01, 1), tol = 1e-10) $root
[1] 0.03921881
```

7.4. コラム

本タスクフォース活動の中で議論され、実務で直面するかもしれない課題を検討する際に有用と考えられた話題を、備忘録として本節にまとめた。

7.4.1. 症例数設計での試験期間の設定の考え方

3.5 節では症例数設計を例示した。いずれのシナリオでも境界時間 τ 前に打ち切りが存在する想定であり、打ち切りの発生に応じて、境界時間 τ までの生存時間 $X(\tau)$ の分散が大きくなることで、打ち切りを考慮した必要症例数が増える。これは、登録期間及び追跡期間の設定

定が検出力に影響を及ぼすことを意味する。

例えば、試験途中で盲検下レビューを行った結果、生存時間が計画時の想定よりも長かった場合、計画当初に基づく追跡期間では多くの打ち切りが発生し、検出力が不足することが考えられる。検出力の低下を補う方法として追跡期間の延長が考えられる。他には、主要変数である RMST の群間差の仮定を保ちつつ各群の RMST の仮定を盲検下の情報に基づいて変更する方法が考えられる。つまり、 $\mu_1(\tau) - \mu_0(\tau)$ を変更せずに $\mu_0(\tau)$, $\mu_1(\tau)$, $\sigma_1^2(\tau)$ 及び $\sigma_0^2(\tau)$ を変更するため、式(3.41)及び式(3.42)による必要症例数が変化する。ただし、RMST の差を主要変数とし、盲検下での情報に基づいて症例数を変更することの性質はまだ明らかでなく、第一種の過誤確率に影響を与える可能性があるため、注意が必要である。

7.4.2. 比例ハザード性の検討

Cox 比例ハザードモデルにより推定されるハザード比は、比例ハザード性の仮定に基づいていることは前述の通りである。比例ハザード性の検証には、二重対数プロット、時間依存性共変量の解析、Schoenfeld 残差のチェックなど、数値的及び視覚的な様々な方法がある^{[1][2]}。しかし、比例ハザード性の検証は治療群間の比較に基づくため、盲検下で行うことができない。従って、比例ハザード性の仮定が成立しているかどうかを実際のデータに基づいて検証し、Cox 比例ハザードモデルを用いることの是非は、盲検化解除後に初めて議論が可能となる。

そこで、比例ハザード性の検証と治療効果の治療群間の違いの推定という二つの解析をどのように事前計画するかという問題が発生する。一つの方法として、比例ハザード性の検証を予備検定として行い、その結果に応じて比例ハザード性が成立している場合は Cox 比例ハザードモデルを用い、成立していない場合は RMST など他の評価指標による推定を用いるかを選択するという検定手順が考えられる。しかし、連続変数に対する t 検定での等分散性の問題^[3]と同様、予備検定の検証結果に応じた検定手順は、第一種の過誤確率を上昇させる懸念がある。

7.4.3. 追跡期間とデータモニタリングの性質

4.1.3 項で RMST とハザード比の特性について述べた。本項では、実際に臨床試験を実施するにあたっての追跡期間の考え方やデータモニタリングの性質について述べる。ここでは、RMST を主要評価指標とする試験デザインとして、全被験者を境界時間 τ まで追跡する場合と一部の被験者を境界時間 τ まで追跡しない場合に分けて考えた。

従来のハザード比を主要評価指標とした試験デザインでは、必要イベント数に到達した時点で解析することを踏まえ、必要イベント数が観測されるように登録期間と追跡期間を定めることが多く、対象患者の評価すべき観察期間を明確に定めることは少なかった。この場合、観察期間を短くして被験者数を増やすことや、観察期間を長くして被験者数を減らすことは問題ないことを意味する。一方、RMST の場合、観察期間を事前に議論すること

が考えられるため、全被験者を境界時間 τ まで追跡するか、又は打ち切りの発生が生存時間と独立である仮定に基づいて一部の被験者を境界時間 τ まで追跡しないとするかを試験計画時に選択することになる。例えば、境界時間 τ を3年としたとき、登録期間を2年、その後の追跡期間を3年とした場合が前者に該当する。一方で、登録期間を2年、その後の追跡期間を2年とした場合が後者に該当する。したがって、前者では τ 以前に計画的な打ち切りはないが、後者では τ 以前に計画的な打ち切りが発生することになる。その他、解析時期の決定方法、必要情報量確保の対応、RMST 評価に必要な生存時間データ固定可能な単位、試験期間の長さ及び生存時間データのモニタリングについて比較した内容を表 7-3 追跡期間とデータモニタリングの性質の比較にまとめた。

表 7-3 追跡期間とデータモニタリングの性質の比較

	RMST		ハザード比
	全被験者を境界時間 τ 以上追跡する	一部の被験者で境界時間 τ まで追跡しない	
観察継続中の打ち切り	τ 以前で計画的な打ち切りはない	τ 以前で計画的な打ち切りが発生する	打ち切りが発生してもよい
解析時期の決定	最後の被験者が登録されてから τ 時間以降	事前計画した情報量を満たした時点	事前計画したイベント数を満たした時点
必要情報量確保の対応	被験者を追加する	追跡期間が境界時間 τ 以前の被験者の追跡を継続する 全被験者の境界時間 τ 以上の追跡が完了した時点で必要情報量未達の場合には、被験者を追加する	追跡期間を延長する、もしくは被験者を追加する
RMST 評価に必要な生存時間データ固定可能な単位	追跡期間が境界時間 τ を超えた被験者ごとに随時データ固定可能		必要情報量達成時点で被験者全体で同時にデータ固定
試験期間の長さ	一部の被験者を境界時間 τ まで追跡しない場合は、全被験者を境界時間 τ まで追跡する場合と比較して試験期間が短くなる		RMST に基づく解析と必要情報量が異なるため、一概に比較できない
生存時間データのモニタリング	最後の被験者が登録されてから境界時間 τ 後	情報量を確認するために、各被験者の生存時	情報量を確認するために、総イベント数

	の時点にて 1 回調査すれば十分	間データを定期的に確認することが必要	を定期的に確認することが必要
--	------------------	--------------------	----------------

臨床試験において重要なことの一つとして、検出力の担保がある。ハザード比の場合、検出力の担保は必要イベント数を達成しているかどうかで判断ができる。実際のイベント数が予定よりも少なく、検出力を保てない場合には、試験期間を延ばすことや、被験者を追加することなどで対応が可能である。一方で RMST の場合、検出力を確認するには、情報量を継続的に計算する必要がある。計算により検出力が足りないことが発覚した場合、境界時間 τ が予め決まっているため、全被験者で境界時間 τ までの追跡が完了している場合には、追跡期間を延ばしても検出力は高まらない。その場合には、被験者を追加する対応となる。一方で、一部の被験者の追跡期間が境界時間 τ よりも短い場合には、追跡期間を延ばすことで検出力を担保できる可能性がある。上記の理由により、試験期間中の検出力の担保においてはハザード比の方が柔軟な対応が可能と考えられる。

RMST の実施上の利点の一つに、追跡期間が境界時間 τ を超えた被験者ごとに随時データ固定可能である点が挙げられる。ハザード比のように、各被験者の追跡期間を固定しない試験デザインでは、ある時点で一斉に被験者全体に対してデータ固定を行うことになる。このため、データ固定に関する作業が短期間に集中してしまう。しかし、RMST では登録期間の初期に登録した被験者は、試験期間中に境界時間 τ に達成するため、境界時間 τ に達した被験者のデータ固定を随時行うことができる。これにより試験終了時の作業が分散されるという実務上の利点がある。ただし、これはあくまで RMST のみを観察する場合であり、副次評価指標にハザード比等がある場合には生存時間データの固定を被験者単位にすることは、副次評価指標の推定精度低下に繋がると考えられる。

7.4.4. 試験途中で境界時間 τ を変更することの是非

「試験途中において、盲検化レビューの結果や最新の類薬試験の結果等を考慮し、境界時間 τ を変更することは許容されるか？」という疑問については、4.2.1 項で記載した通り、基本的には試験途中で境界時間 τ を変更することは許容されないと考えられる。しかし、非常に稀な状況として、試験途中で、関連学会等で治療環境の変化に対する効果の影響が検討され、その情報が境界時間 τ の設定に影響を与えるときは、試験途中でも境界時間 τ を変更することが可能となるケースがあると考えられる。ただし、境界時間 τ を変更する場合には、境界時間 τ の意義、必要症例数等を再考するとともに、規制当局への確認も必要になると考えられる。

Horiguchi et al. (2018)^[4] は、複数の境界時間 τ の候補の中から観察されたデータに基づき、最も検定統計量が大きな τ を選択するという手法を提案しており、検出力の観点で通常頻用されている log-rank 検定よりも有用であると主張している。具体的には、 $\tau^* = \{\tau_1, \tau_2, \dots, \tau_K\}$ として複数の境界時間の候補を用意し、検定統計量として $Z^* = \max_{\tau \in \tau^*} Z(\tau)$ を利用するもので

ある (RMST の群間差の場合、 $Z(\tau)$ は式(3.21)を意味する)。

ただし、複数の境界時間 τ を臨床的な観点から事前に規定することがどの程度現実的なのか、サンプルサイズをどのように設計するのか、解析時に選択された境界時間 τ での RMST の臨床的解釈をどうすればよいのか等、実適用する際にはいくつか議論の余地があると考えられる。

7.4.5. Kaplan–Meier 曲線を数値化する際の秘訣

4.3.2 項で述べたように、公表されている論文の Kaplan–Meier 曲線から RMST を見積もる際、生存曲線の図を数値化する必要がある。つまり、図を数値化するソフトウェアを用いて、生存曲線が通る x 座標、及び y 座標からなる数値データを得ることになる。ソフトウェアで自動的に生存曲線が通る箇所を選択できる機能があるものの、例えば以下のような状況では、座標を正確に読み取ることができず、 x 座標の値が大きくなるにつれて、 y 座標の値が単調非増加となるように数値データを構築できない場合がある。

- 1つの図に複数の生存曲線があり、これらが重なっている個所がある場合
- 打ち切りの印が描かれている場合
- 生存曲線が破線や点線で描かれている場合

このような状況で、適切な数値データを得るためには、生存曲線が通る箇所をクリックしながら、手動で数値データを構築する必要がある。しかしながら、生存曲線を細かくクリックし、選択していくことは、大変な労力である。ここでは、Kaplan–Meier 曲線が階段関数であるという特徴を踏まえ、効率的に数値データを得る秘訣を紹介する。

まず、ソフトウェアへ読み込んだ Kaplan–Meier 曲線の図を、生存曲線が階段上に見えるまで拡大する。例えば、図 7-1 のようになる。次に、階段が落ちた箇所、すなわち図 7-1 の③と⑤をクリックして選択する。これに加えて、生存曲線の終点である⑥をクリックする。このとき、⑤と⑥の y 座標の値が等しくなるように留意する。なお、 y 座標の値を等しく選択できない場合には、数値データ化した際に修正可能である。

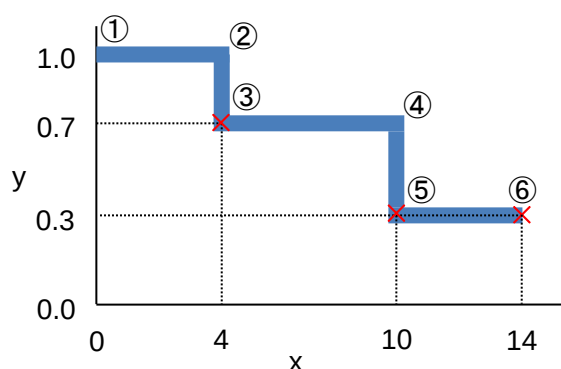


図 7-1 拡大した Kaplan–Meier 曲線の例

表 7-4 数値データ

a. ソフトウェアから作成されるもの

	x 座標	y 座標
③	4	0.7
⑤	10	0.3
⑥	14	0.3

b. 補完したもの

	x 座標	y 座標
①	<u>0</u>	<u>1</u>
②	<u>4</u>	<u>1</u>
③	4	0.7
④	<u>10</u>	<u>0.7</u>
⑤	10	0.3
⑥	14	0.3

次に、ソフトウェア上で図 7-1 の③、⑤、⑥を選択した状態で、数値データを作成する。その結果は、表 7-4a のようになる。これを Excel などの表計算ソフトを用いて、①、②、④に対応する数値データを表 7-4b で示す下線部のように補完する。すなわち、①の座標は、必ず (0, 1) であることから、そのまま補完する。②と③、及び④と⑤の x 座標は等しいことから、ソフトウェアで読み取られた x 座標の値を補完する。同様に、①と②、及び③と④の y 座標は等しいことから、ソフトウェアで読み取られた y 座標の値を補完する。

ソフトウェアで自動的に生存曲線が通る箇所を選択できない場合には、上記の手順を用いることで、数値データを取得することが出来る。そして、表 7-4b のような数値データを用いることで、RMST の近似値を求めたり、各被験者の生存時間データの近似値を再構築したりして、生存時間データの様々な評価指標の推定値やその信頼区間を算出することが出来る。

参考文献

- [1] 大橋靖雄・浜田知久馬 (1995) 生存時間解析 SAS による生物統計, 東京大学出版会.
- [2] 大橋靖雄・浜田知久馬・魚住龍史 (2016) 生存時間解析 [応用編] SAS による生物統計, 東京大学出版会.
- [3] Zimmerman DW, A note on preliminary tests of equality of variances. Br J Math Stat Psychol. 2004;**57**:173–181.
- [4] Horiguchi M, Cronin AM, Takeuchi M, Uno H. A flexible and coherent test/estimation procedure based on restricted mean survival times for censored time-to-event data in randomized clinical trials. Statistics in Medicine. 2018;**37**:2307–2320.

2019 年 3 月

日本製薬工業協会 データサイエンス部会

タスクフォース 4 生存時間解析チーム

飯森 孝行 (杏林製薬株式会社) [2018 年 11 月まで]

鵜飼 裕之 (日本ベーリンガーインゲルハイム株式会社)

田中 慎一 (日本新薬株式会社)

棚瀬 貴紀 (大鵬薬品工業株式会社)

仲川 慎太郎 (中外製薬株式会社)

長谷川 貴大* (塩野義製薬株式会社)

三澤 早織 (日本化薬株式会社)

淀 康秀 (大日本住友製薬株式会社)

若菜 明 (MSD 株式会社)

* チームリーダー