

【日本製薬工業協会シンポジウム】
生存時間解析の評価指標に関する最近の展開
—RMST (restricted mean survival time) を理解する—

7. RMSTの性能評価シミュレーション



2018年6月13日
医薬品評価委員会 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

大日本住友製薬株式会社 淀 康秀

- 本発表は、先日公開された「生存時間型応答の評価指標 -RMST(restricted mean survival)を理解する-」について、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タスクフォース4 生存時間解析チームが本シンポジウムの趣旨を踏まえ、再構成したものである

1. シミュレーションの目的
2. シミュレーションの概要
 - シミュレーションのフレームワーク
 - 区分指数分布に従う乱数の発生方法
 - 仮定した生存関数
3. シミュレーションの結果と考察
4. まとめ

1. シミュレーションの目的

Restricted Mean Survival Time (RMST) の特徴を検討する目的で、幾つかの生存曲線パターンを仮定し、シミュレーションで性能評価を実施する。

- 生存時間解析の新たな指標としてRMSTが注目されている
 - 実際の臨床試験での適用例は少ない
 - RMSTの特徴を検討した結果も少ない

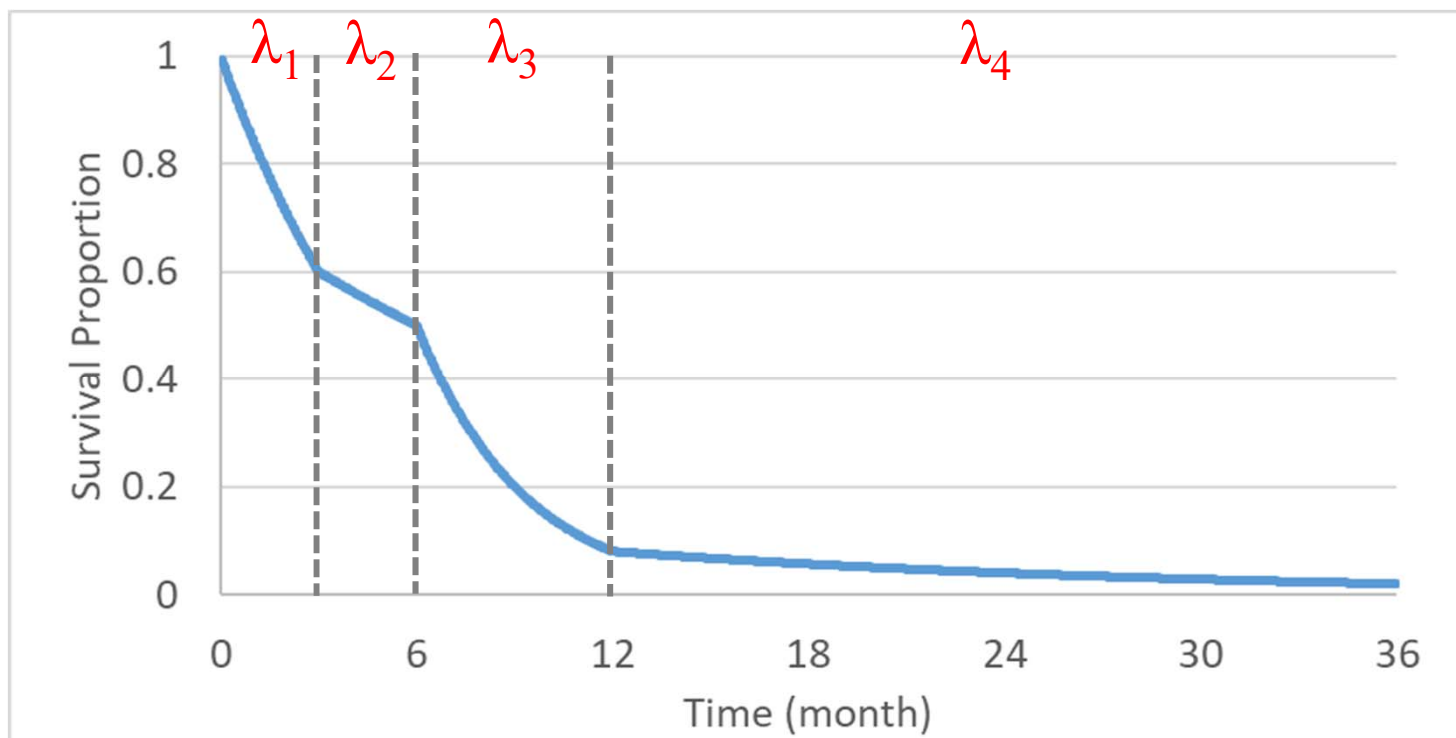
2. シミュレーションの概要

- Overall survival を指標とした生存時間解析
 - 対照薬群と実薬群の2群比較
 - 被験者の登録期間: 24ヶ月 (登録速度は一定とする)
 - 追跡期間: 最終被験者登録から12ヶ月
 - 被験者数: 500名 (250名/群)
 - 1年あたりの打ち切り発生: 5%
 - 有意水準: 片側2.5%

- 事前に仮定した生存関数に合致する区分指数分布から個々の生存時間を発生させ, RMSTの差を用いた検定と log-rank検定の結果を検出力で比較する.
 - シミュレーション回数: 10000回

区分指数分布に従う乱数の発生方法

- 試験期間を通じてハザードが一定となる(一定とみなせる)結果が得られることは少ない
- 観察期間を幾つかの区間に区切り, その区間の中では一定のハザードが保たれると考える



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 $\lambda(t)$, 累積ハザード関数 $\Lambda(t)$ は以下のように書ける

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

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

区分指数分布に従う乱数の発生方法

- 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られる

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

- 乱数 x の値に応じて T を求める式を選択する

本シミュレーションにおけるハザードの決め方

- 想定する生存曲線から特定時点での生存割合を抽出し、区間ごとのハザードに変換する
- 例えば:
 - 1年生存割合 70%, 2年生存割合, 20%, 3年生存割合2%

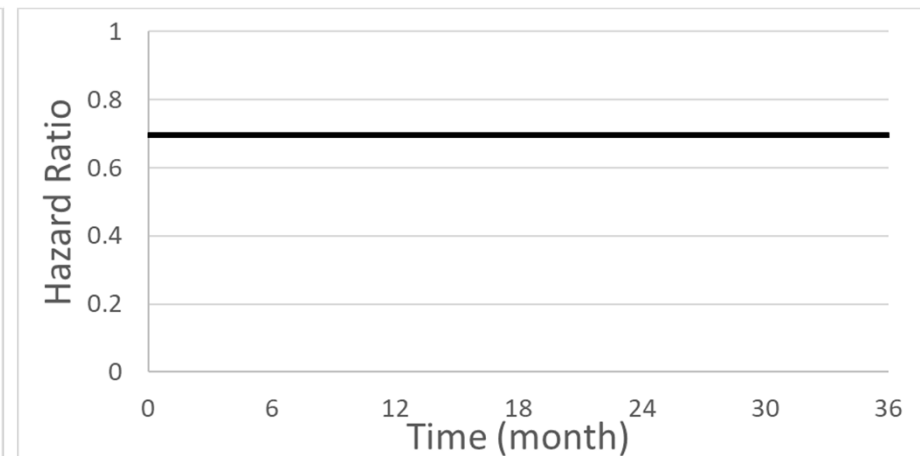
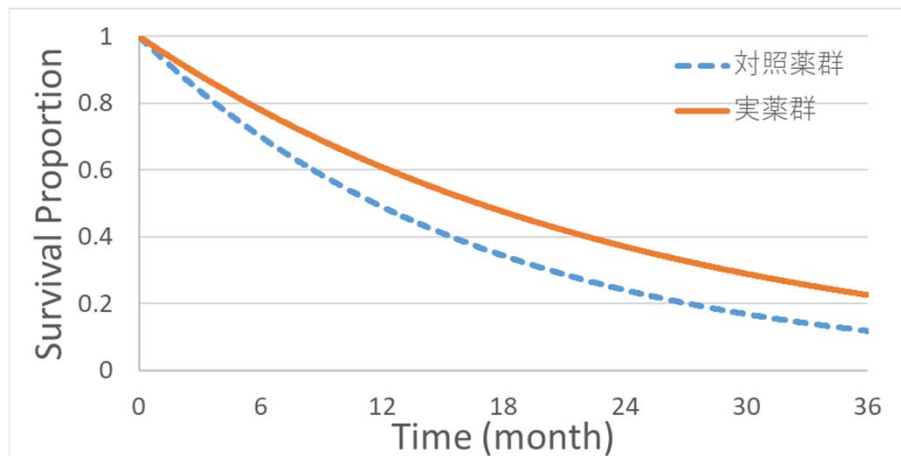
$$\begin{cases} \lambda_1 = -LN(0.7)/12 \\ \lambda_2 = -LN(0.2/0.7)/(24 - 12) \\ \lambda_3 = -LN(0.02/0.2)/(36 - 24) \\ t_1 = 12, t_2 = 24, t_3 = 36 \end{cases}$$

仮定した生存関数

1. 比例ハザード性が成立している場合
2. 特定時点までは差がなく, それ以降に差が開く場合
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)
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. 生存曲線がクロスする場合

1. 比例ハザード性が成立している場合

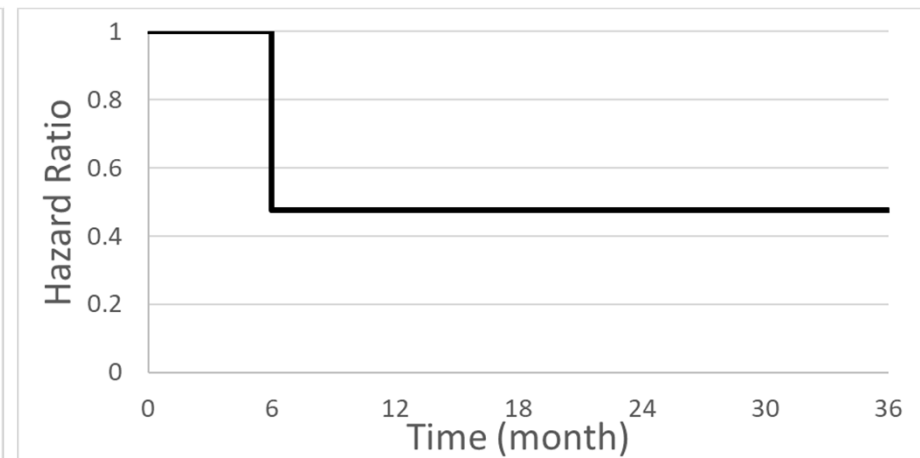
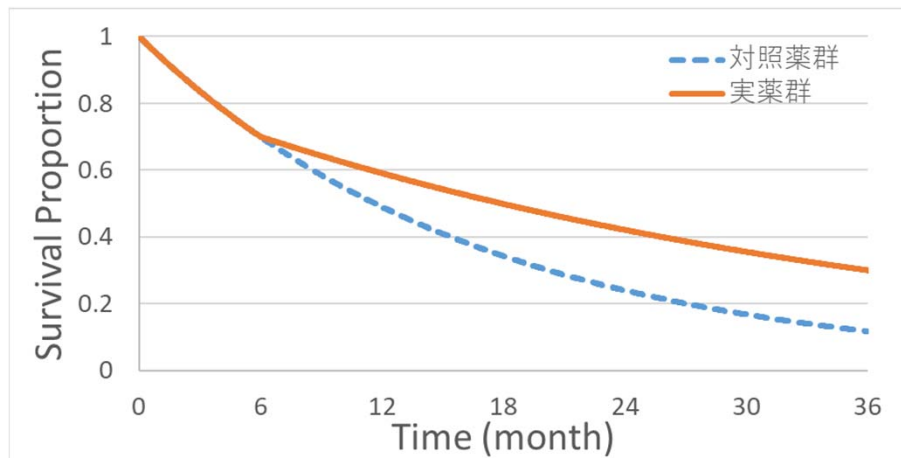
- 対照薬群: (6m, 70%), (36m, 11.8%)
- 実薬群: (6m, 78%), (36m, 22.5%)



Time (month)	生存率 (対照薬群)	生存率 (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%

2. 特定時点までは差がなく、それ以降に差が開く場合

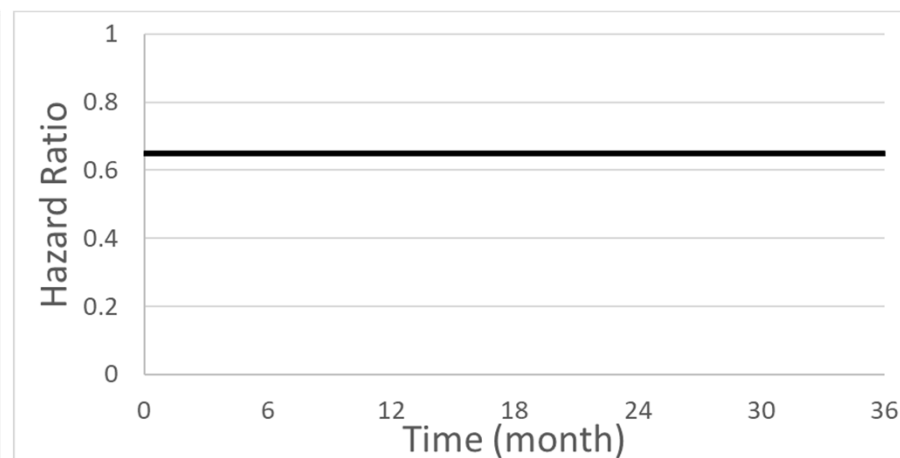
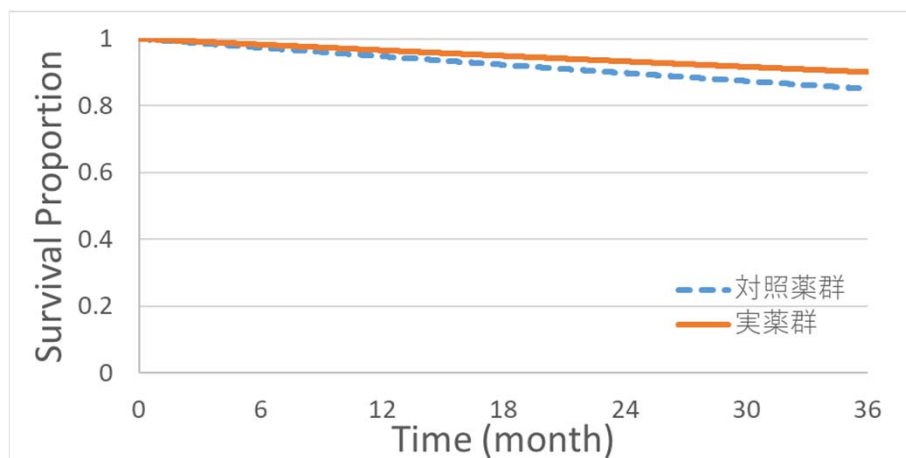
- 対照薬群：(6m, 70%), (36m, 11.8%)
- 実薬群：(6m, 70%), (36m, 30%)



Time (month)	生存率 (対照薬群)	生存率 (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%

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)

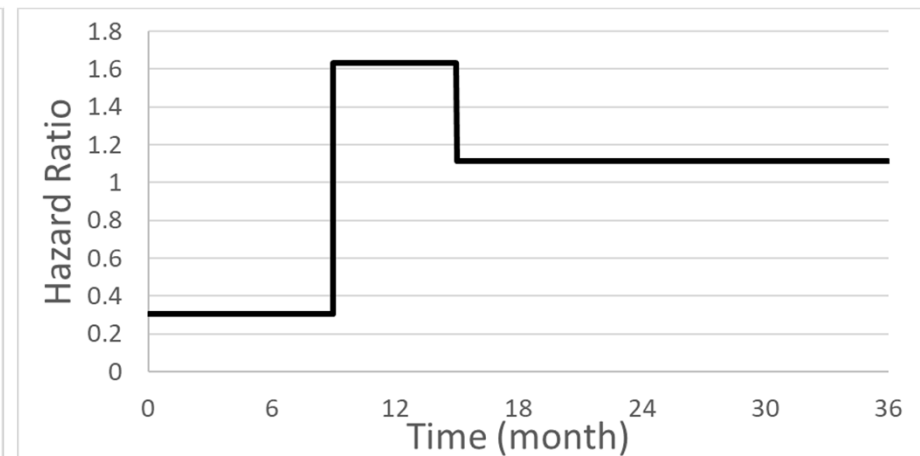
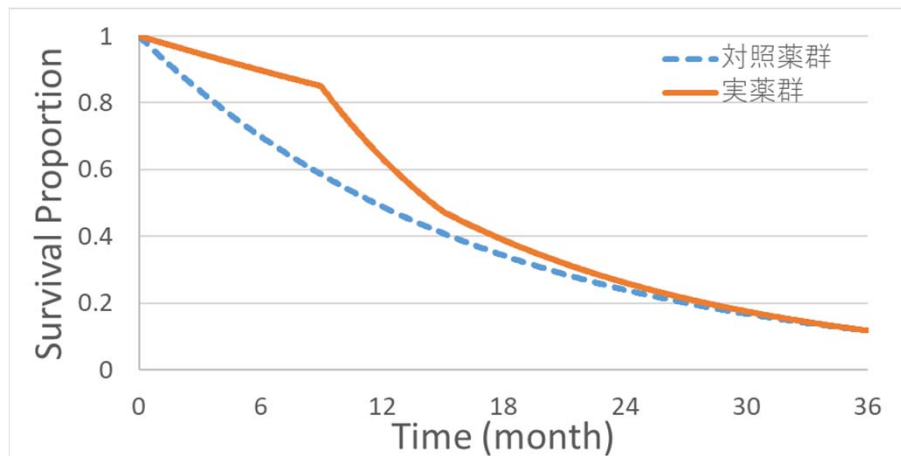
- 対照薬群:(36m, 85%)
- 実薬群:(36m, 90%)



Time (month)	生存率 (対照薬群)	生存率 (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%

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

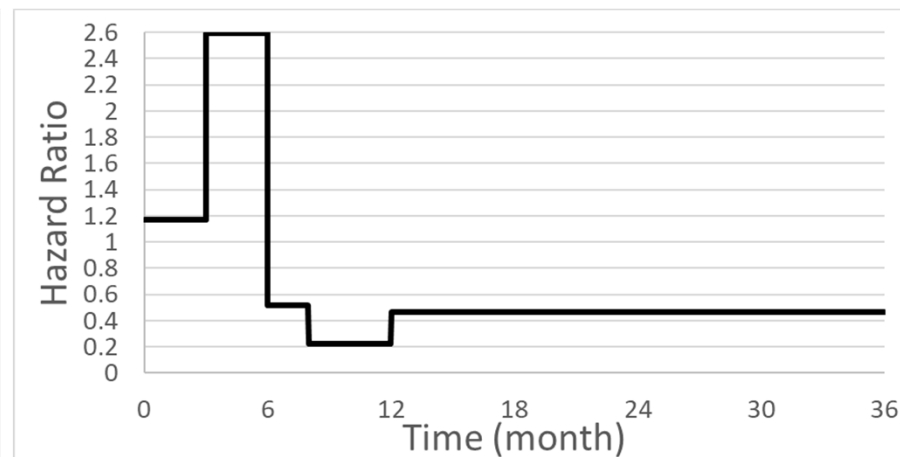
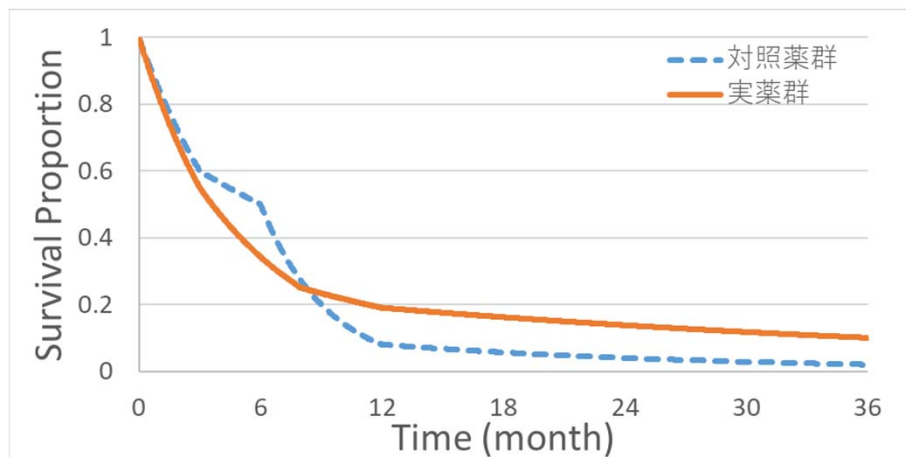
- 対照薬群:(36m, 11.8%)
- 実薬群:(9m, 85%), (15m, 45%), (36m, 11.8%)



Time (month)	生存率 (対照薬群)	生存率 (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%

5. 生存曲線がクロスする場合

- 対照薬群 : (3m, 60%), (6m, 50%), (12m, 8%), (36m, 2%)
- 実薬群 : (3m, 55%), (8m, 25%), (12m, 19%), (36m, 10%)



Time (month)	生存率 (対照薬群)	生存率 (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%

τ の設定と検定に用いる生存時間データ

➤ 検定に用いる生存時間データ

- **RMST**: 試験終了時に得られている生存時間データ
(τ 以降のデータは実質使われない)
- **log-rank[all]**: 試験終了時に得られている生存時間データ
(τ 以降のデータも結果に影響する)
- **log-rank[τ]**: 試験終了時に得られている生存時間データ
(τ 以降のデータは用いない)



➤ 有意水準を片側2.5%として検出力で比較

- log-rank[all] と log-rank[τ] のどちらで比較?

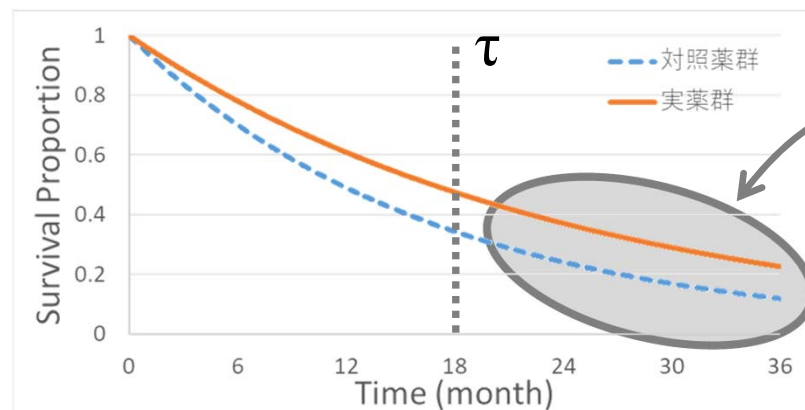
τ の設定と検定に用いる生存時間データ

➤ $\tau = 12$ to 36 by 3

- 登録期間24ヶ月、観察期間12ヶ月という試験期間を考慮

➤ 検定のタイミング

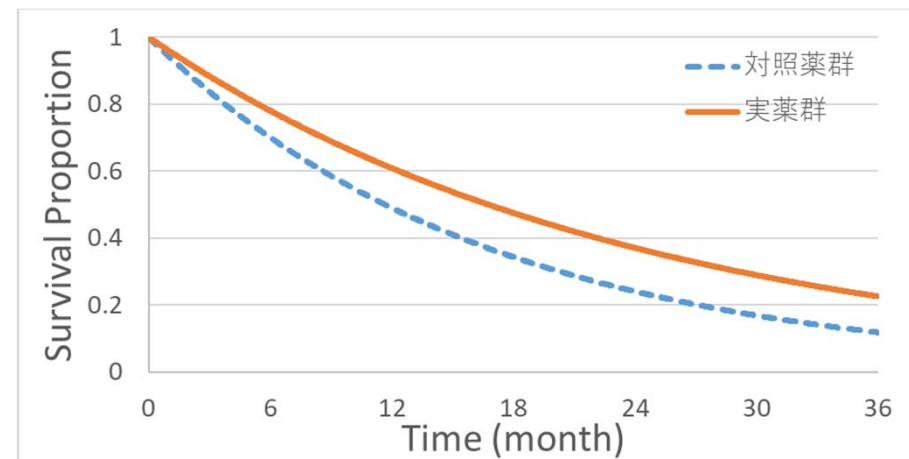
- 試験期間(36ヶ月)終了後とした
- 登録期間、追跡期間と τ との大小関係によっては境界時間 τ まで観察された被験者数が増える
- 検定実施時に τ 以降の生存曲線が得られていても結果には影響しない(RMST及びlog-rank[τ])



3. シミュレーションの結果と考察

1. 比例ハザード性が成立している場合

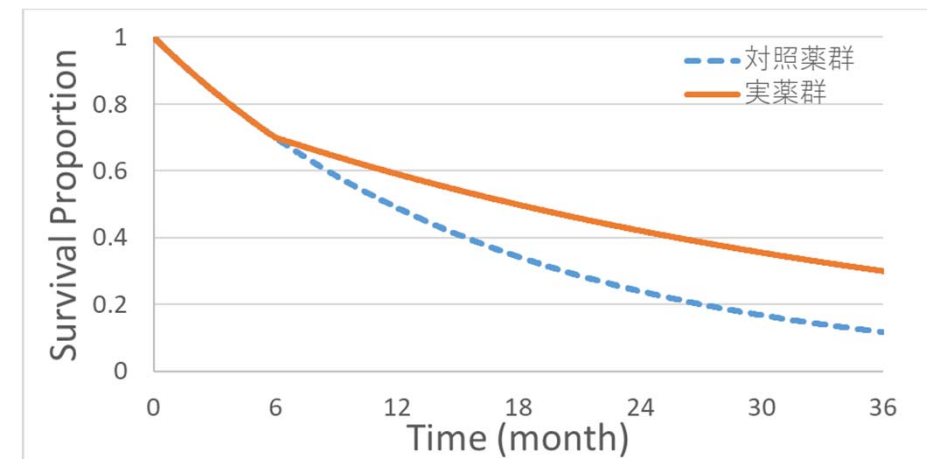
τ	Power			Δ RMST (Mean \pm se)
	RMST	log-rank [τ]	log-rank [all]	
12	68.4	76.0	90.1	0.88 \pm 0.36
15	75.8	82.4		1.25 \pm 0.47
18	81.1	85.6		1.64 \pm 0.58
21	84.7	87.6		2.03 \pm 0.68
24	87.3	88.4		2.43 \pm 0.78
27	88.8	89.4		2.81 \pm 0.88
30	89.6	89.7		3.18 \pm 0.98
33	90.2	90.0		3.54 \pm 1.08
36	90.1	90.1		3.87 \pm 1.19



- 比例ハザード性が成立している下では、境界時間 τ に関わらず、log-rankの検出力が高かった
- τ を大きくすると RMSTと log-rank検定に差はなくなった

2. 特定時点までは差がなく、それ以降に差が開く場合

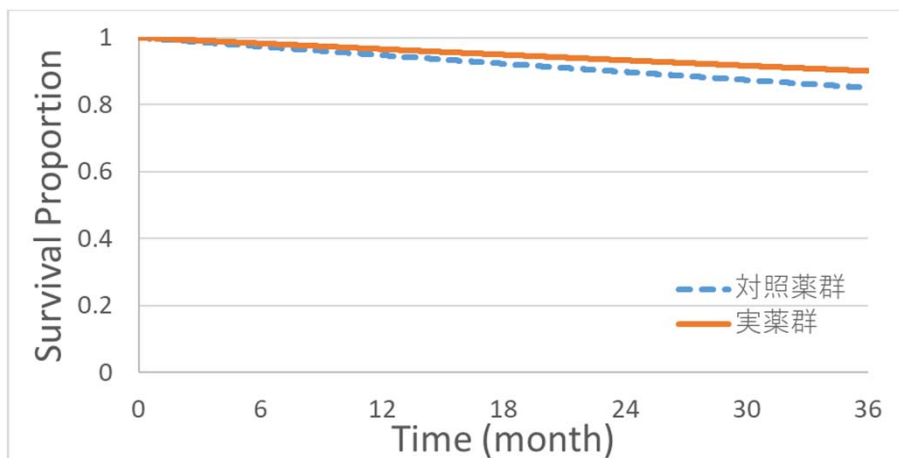
τ	Power			Δ RMST (Mean \pm se)
	RMST	log-rank [τ]	log-rank [all]	
12	14.5	45.3	92.7	0.33 \pm 0.38
15	29.3	66.7		0.68 \pm 0.49
18	45.9	79.4		1.12 \pm 0.60
21	62.6	85.7		1.61 \pm 0.71
24	75.2	89.2		2.14 \pm 0.81
27	83.4	91.1		2.69 \pm 0.92
30	88.5	92.1		3.26 \pm 1.02
33	91.8	92.6		3.82 \pm 1.13
36	93.8	92.7		4.37 \pm 1.24



- 6ヶ月目までは生存割合に差がないため、RMSTでは τ が小さいほど検出力が大きく低下している
- log-rank[τ] と比較しても RMST の検出力は小さかった

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)

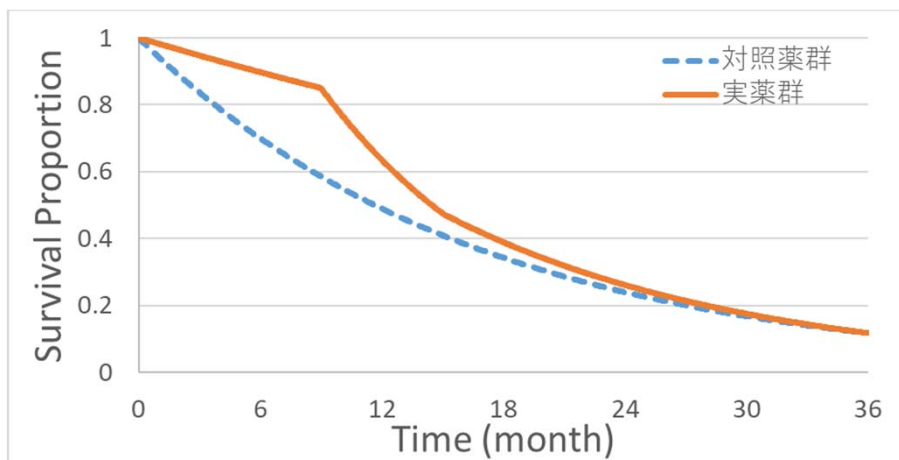
τ	Power			Δ RMST (Mean \pm se)
	RMST	log-rank [τ]	log-rank [all]	
12	13.5	16.6	27.1	0.11 \pm 0.13
15	16.0	19.1		0.17 \pm 0.18
18	18.7	21.0		0.25 \pm 0.23
21	21.1	23.3		0.33 \pm 0.29
24	23.1	25.0		0.43 \pm 0.36
27	24.5	25.8		0.54 \pm 0.43
30	25.6	26.9		0.66 \pm 0.51
33	26.7	27.1		0.80 \pm 0.60
36	27.5	27.1		0.94 \pm 0.69



- 時点 τ でイベントが殆ど発現していない(3-10%)では RMSTでも log-rank検定でも検出力は低かった
- 傾向としてはパターン1と同様であった

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

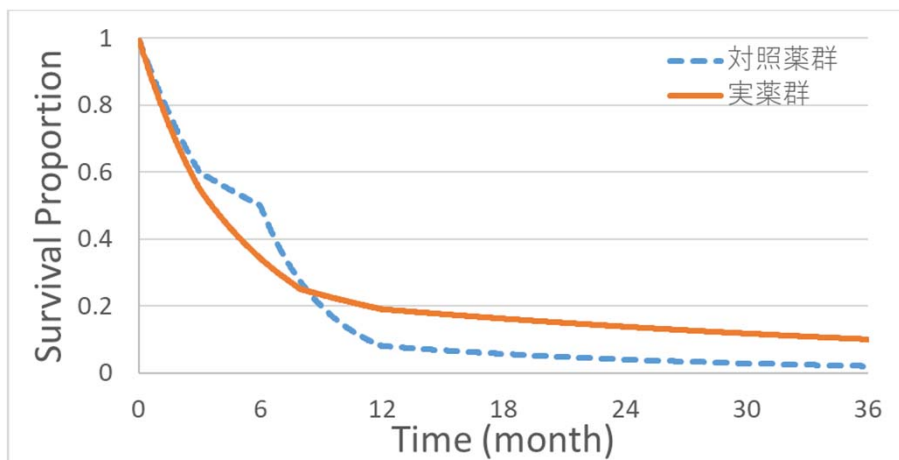
τ	Power			Δ RMST (Mean \pm se)
	RMST	log-rank [τ]	log-rank [all]	
12	100.0	98.3	65.1	1.94 \pm 0.32
15	100.0	83.3		2.24 \pm 0.42
18	99.5	76.5		2.40 \pm 0.52
21	97.9	72.1		2.52 \pm 0.62
24	95.0	68.6		2.60 \pm 0.72
27	90.6	66.7		2.65 \pm 0.81
30	84.3	65.8		2.68 \pm 0.90
33	77.7	65.1		2.69 \pm 0.99
36	70.6	65.1		2.70 \pm 1.08



- log-rank検定では(相対的に)後期のイベントに対して重みが高い為、log-rank[all]では検出力低かった
- RMSTはいずれの τ でもlog-rankより検出力が高く、生存割合に殆ど差のなくなる27ヶ月まで検出力90%を保った

5. 生存曲線がクロスする場合

τ	Power			Δ RMST (Mean \pm se)
	RMST	log-rank [τ]	log-rank [all]	
12	0.2	9.0	14.8	-0.33 \pm 0.36
15	2.3	10.7		-0.01 \pm 0.43
18	8.9	12.0		0.31 \pm 0.49
21	19.2	13.1		0.62 \pm 0.56
24	30.8	13.9		0.92 \pm 0.62
27	42.2	14.4		1.21 \pm 0.68
30	51.6	14.7		1.48 \pm 0.74
33	58.7	14.8		1.74 \pm 0.80
36	63.7	14.8		1.99 \pm 0.86



- Δ RMSTを見ると15ヶ月までは対照薬群の方が良い結果になっている
- 全体的に検出力は低いですが, RMSTでは τ を大きくすることで検出力が増加傾向にある
- 生存曲線がクロスする場合には多角的な検討が必要

4. まとめ

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得られるであろう、比例ハザード性が成立しない生存関数を想定してRMSTの差による検定と log-rank 検定の結果の違いを検出力ベースで確認した
- 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の生存曲線では総じて log-rank 検定の検出力が高かった
- 試験序盤で生存割合に差が付き、後半では差がなくなるようなケースではRMSTの方が検出力が高かった

- 生存関数がクロスするようなケースではRMSTの差を用いることで検出力を確保できる可能性がある
- RMSTと log-rank検定の性能評価という点では、比較の方法にも留意すべきである
 - 検定の実施時期, 検定に用いるデータ範囲
- JPMA TF4 報告書の中ではシミュレーション検討の一環として検出力に加え, P値の挙動も示しているので, ご興味のある方はご参照ください

- 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、意味のある τ 、試験期間を設定することで RMSTを用いた検定でも検出力の高い解析となる可能性がある